A COMPARATIVE ANALYSIS ON MODIFICATION OF SOCIAL NETWORKS OVER LINKS WITH MECHANISM COMBINATIONS

by

Merve Ünal

B.S., Industrial Engineering, Yildiz Technical University, 2016

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Industrial Engineering
Boğaziçi University
2021

ACKNOWLEDGEMENTS

Firstly, I would like to express my gratitude to my thesis supervisor, *Assoc. Prof. Gönenç Yücel*, for his support and useful comments throughout my thesis process. Whenever he thought that I was in need of help, he consistently directed me to find my way. He encouraged me in many subjects during my thesis process. Even when I had a private problem, he always supported me to freeze my registration and find my path, and motivated me to accomplish.

I would like to thank my jury committee, *Assoc. Prof. Aybek Korugan* and *Assist. Prof. Mert Edalı* for participating my thesis defense and valuable contributions.

I would like to thank my friends, *Zehra* and *Pelin* for their support, especially in format issues.

Last but not least, I also would like to express special thanks to my beloved family; especially my mother, *Nimet*; my father, *Ergün*; my sister, *Aslı*; my grandmother, *Melahat* and *Lokum* for believing me and their endless support. As a family, we have experienced several ups and downs in the past two years. When I did not have enough motivation to continue, they always encouraged me with their unconditional love and support to not to give up. I would not have completed my thesis without them.

ABSTRACT

A COMPARATIVE ANALYSIS ON MODIFICATION OF SOCIAL NETWORKS OVER LINKS WITH MECHANISM COMBINATIONS

In this study, the problem of preserving the characteristics of social networks changing dynamically is investigated. Distinct mechanism combinations have been implemented on the networks selected. Mechanism combinations providing to be preserved its properties as the small world network modifies are discovered. In total, 36 experiments are carried out considering two network classes, six link addition and three link deletion mechanisms. In the experiments, the structural properties of the network such as the average shortest path, average clustering coefficient and degree distribution of the snapshots of each network are analyzed and compared with the results of initial networks. It is seen that experiments related to some mechanism combinations implemented on the small world network yield similar results to the properties of initial small world network. In addition, dynamic social networks with different characteristics are created by making changes over the links in the experiments carried out. Moreover, mechanism combinations that fragment or randomize the structure of the network have been revealed in the experiments.

ÖZET

MEKANİZMA KOMBİNASYONLARI İLE LİNKLER ÜZERİNDEN SOSYAL AĞLARIN MODİFİKASYONU ÜZERİNE KARŞILAŞTIRMALI BİR ANALİZ

Bu çalışmada sosyal ağların dinamik şekilde değişiminin sağlanarak özelliklerinin korunabilmesi problemi incelendi. Buna ilişkin olarak çeşitli mekanizma kombinasyonları seçilen ağlara uygulandı. Küçük dünya ağının değiştikçe özelliklerinin korunmasını sağlayan mekanizma kombinasyonları belirlenmiştir. 2 farklı ağ sınıfı, 6 link ekleme ve 3 link çıkarma mekanizması göz önüne alınarak toplamda 36 farklı deney yapılmıştır. Yapılan deneylerde her ağın anlık görüntüsünün ortalama en kısa yol, ortalama kümeleme katsayısı ve derece dağılımı gibi ağın yapısal özellikleri analiz edilmiştir ve ilk ağın sonuçlarıyla karşılaştırılmıştır. Küçük dünya ağı üzerine uygulanan bazı mekanizma kombinasyonlarına ilişkin deneylerin başlangıçta yaratılan küçük dünya ağının özellikleriyle benzerlik gösterdiği görüldü. Ek olarak, yapılan deneylerde farklı karakterlere sahip dinamik sosyal ağlar da oluşturuldu. Aynı zamanda yapılan deneylerde ağın yapısını parçalayan veya rastsallaştıran mekanizma kombinasyonları da ortaya çıkarılmıştır.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS/ABBREVIATIONS

ACC	Average Clustering Coefficient
ASP	Average Shortest Path
D	Degree
HB1	Hybrid 1
HB2	Hybrid 2
LCD	The Least Common Degree
PA	Preferential Attachment
R	Random
RN	Random Network
SWN	Small World Network
TR1	Triadic Closure 1
TR2	Triadic Closure 2
THR	Threshold Parameter

1. INTRODUCTION

Social networks are indispensable parts of people's lives. Thanks to social networks, people can share their opinions and connect with people having similar ideas. In general, a social network is a structure which shows social actors and their interactions with each other. For instance, students in a class constitute a social network by interacting with each other. Information propagation can be provided through the interactions established among students. As another example, people can connect with one another if they know each other in some way via online social networking sites such as Facebook and Linkedin, which are commonly used in every single person's life.

Graphs are used to analyze social networks mathematically. A graph is a representation of a certain number of objects and interactions between them. Basically, a graph $G = (V, E)$, where the vertex set V represents nodes, and the edge set E symbolizes link groups among nodes [1]. Graphs can be classified into two subcategories, which are directed and undirected graphs. The former has a specific direction from one node to another. Specifically, if node A interacts with node B, node B does not have to connect with node A (as can be shown in Figure 1.1). One of the online social networking sites, Instagram can be an example of this kind of graphs.

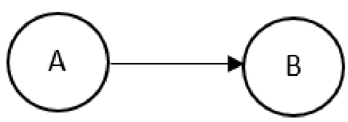

Figure 1.1. Basic Directed Graph Representation.

However, the latter has two-way of interaction. If there is a link between node A and node B, they are adjacent [2] (as can be shown in Figure 1.2). Facebook is an instance of a social network belonging to the undirected graph category. A graph can also be used to represent networks in different fields such as computer science [3, 4], biology [5] and chemistry [6] other than social networks.

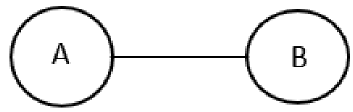

Figure 1.2. Basic Undirected Graph Representation.

In terms of temporal dimension, two different network types are used to explain social networks, which are static networks and dynamic networks. In static networks, edges and nodes are constant over time. In other words, there is no change or evolution. Most studies focus on static networks or a snapshot at a single point of time by ignoring temporal dimension [7]. However, it is far from real networks and neglects modification of networks. Static networks are not realistic to explain real networks, which are expected to evolve. In dynamic networks evolving over time, interactions among people always change. Individuals can establish a new connection or end their existing interaction for some reason.

Links between nodes represent relationships, which can change over time based on node features such as the degree of a node. Similarly, new nodes can be added to the network or existing ones can be removed from the network. Therefore, it is important to analyze how the structure of a social network changes and what these relationships could be affected from. The initial network classes chosen at t_0 and modification mechanisms have distinctive effects on the structure of the network modified. After the properties of networks modified are evaluated, analyzing how they change based on network modification mechanisms can be meaningful. There are few studies on how a link is formed and removed on dynamic networks. Therefore, the way people connect with each other over time and structural effects of dynamic network modification on global network measures are an emerging area to research [8], which brings some important questions with it: How does a network evolve? How do the global measures of network classes change according to network modification mechanisms over time?

In this study, how the characteristics of an undirected small world network created at time $t = 0$ are kept the same with link addition/deletion mechanisms will be investigated. Especially, which link addition and deletion mechanisms keep network structure the same will be studied. Since real networks have specific network features, such as small world structure [9] and power law scaling [10], the modification combinations that convert a random network into a network having these features will also be explained. Since a great number of experiments are carried out with various link addition and deletion mechanisms, mechanisms that convert the network into random network and fragment the network will be revealed. Lastly, since the networks modified in different snapshots are analyzed, how they deviate from the initial networks and network properties change will be examined. In addition to degree distribution, average shortest path and average clustering coefficient are measures that will be used to compare modified networks with initial networks.

The next chapter covers the literature regarding network science, main social network measures, network classes and dynamic social network evolution. The third chapter includes research objectives. In the fourth chapter, assumptions and parameters of models chosen to construct networks are explained. Network evolution mechanisms including link addition and link deletion are clarified by using pseudo codes. In the next chapter, 36 experiments have been reported. All experiments have been completed via Python. Some of 36 experiments do not necessitate deeper research. After the ones which require detailed analysis are selected, their detailed experiments and comparisons of global measures will be analyzed. The next chapter, discussion, includes the summary of experiment results. In the last one, it is concluded with evaluations, and the future research, which need to be conducted.

2. LITERATURE REVIEW

2.1. The measures of networks

Distinct network types from biological networks to power grids can be represented as graphs consisting of nodes and edges [11,12]. In order to compare the characteristics of networks to one another, two different critical metrics are used, which are the average shortest path and average clustering coefficient [13].

2.1.1. Average Shortest Path (ASP)

In a network graph G, the average shortest path is defined as the average number of steps between any two pairs of vertices. In order to calculate the average shortest path of the network, the distance between nodes is calculated at first. The average shortest path of a network is calculated considering all nodes whose degrees are not equal to zero. The average shortest path of the graph G is calculated by Equation 2.1;

$$ASP(G) = \frac{1}{N(N-1)} \sum_{x \neq y} d(x,y) \tag{2.1}$$

where x and y are two nodes in a network G and N is the number of nodes in the network G. For $x \neq y$, minimum number of links from x to y, $d(x,y)$ is calculated.

Degree is a metric belonging to nodes in networks, which means the number of links that a node has. In some cases, the degree of a node is zero, which shows that the node does not have any link in that network and is named as an isolated node. Since there is no path from this isolated node to the rest of the nodes, it is not taken into account to calculate the average shortest path of the network. Additionally, some networks can split into subnetworks, which are also known as components or subgraphs. In this case, there are more than one average shortest path calculation for each clustered component.

A giant component is a component, which refers to the biggest subgraph containing the maximum number of nodes in complex networks [14]. If an edge is created between two different components, then these two components combine into one component [15]. In dynamic networks, graphs containing a certain number of links and nodes mostly change since the structures of those networks alter over time.

2.1.2. Average Clustering Coefficient (ACC)

The other important metric shows how nodes are inclined to be clustered. The clustering coefficient is related to the edges connecting the neighbors of a node. The more these edges are, the higher the clustering coefficient is in a network. Before calculating the average clustering coefficient of a network, the same measure should be calculated for a node as shown in Equation 2.2;

$$CC(x) = \frac{2k_x}{d_x(d_x - 1)} \tag{2.2}$$

where k_x represents the number of links between the neighbors of node x, d_x shows the number of neighbors that node x has and N is the number of nodes in the network G. After getting the measure for each node, the average clustering coefficient of the network G can be calculated as seen in Equation 2.3;

$$ACC(G) = \frac{1}{N} \sum_{x=1}^{N} CC(x) \tag{2.3}$$

where N is the number of nodes.

2.2. Network Topology

Network topology can be expressed as a schematic description of a network, which changes depending on network classes. When the clustering coefficient of a network equals zero, it is seen that there are no edges connecting the neighbors of any nodes in the network. This is named a star network topology.

When this ratio is one, neighbors of all nodes have connected with each other in the network. This type of topology is known as fully connected mesh [16]. Different network classes have different properties in terms of the shortest path, clustering coefficient, and degree distribution. Random and regular networks are extreme in terms of their structural properties. While random networks have fully random characteristics, regular networks are fully ordered. In the next part, after explaining these two networks comprehensively, other network classes, such as small world networks and scale free networks, corresponding to real network characteristics will be also mentioned.

2.2.1. Random Network

Consider that $G(N, p)$ is a random graph model including set of N nodes and edges between nodes linked with probability p [17]. Basically, when p converges to zero, the network is sparsely connected and when p equals zero, there are no edges between nodes. Namely, all nodes in the network become isolated. The largest component size is one and all nodes are independent of one another. As p increases, the network starts to get connected and p is equal to one, the network has been fully connected [18].

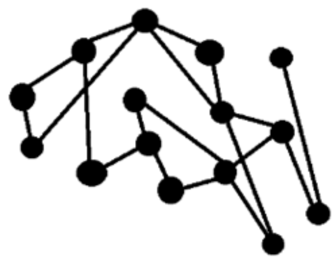

Figure 2.1. A Random Network.

The algorithm regarding random network generation has been proposed by Erdős-Rényi. To construct a random network, a node pair from the node set is selected and a random number (between 0 and 1) is generated. If the generated random number is higher than p, the link is created between this node pair. Otherwise, they remain disconnected. This process is repeated until all possible node pairs are processed. For a network with N nodes, there should be $\frac{N\times(N-1)}{2}$ possible node pairs in total.

The expected number of links in the network can be calculated as $p\times$ $C(N,2)$ [11]. A node can reach the other ones in a few steps, which implies that random networks have a low average shortest path. On the other hand, random networks tend to have a low level of transitivity, which means that the probability of neighbors of a node being connected to one another is low. Additionally, the degree distribution (distribution related to the number of neighbors that each node has over the whole network) of a random network is bell shaped and two tails in each side.

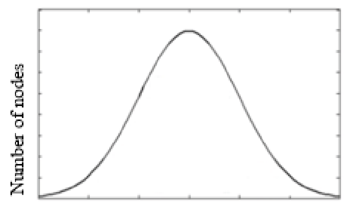

Figure 2.2. Degree Distribution of a Random Network.

When 10 real networks studied by Barabási and Albert [11], are observed in terms of clustering coefficient, it is seen that this metric is not dependent on the network size. However, the average clustering coefficient of a random network declines as the network size increases. Therefore, random networks cannot represent most of the real networks in terms of clustering coefficient [11].

2.2.2. Regular Network

Another network class is a regular network whose model type is regular lattice that all nodes are arranged to raise the level of clustering coefficient. The nodes aligning in a circle have connected with the k nearest nodes. In other words, every node in the network has a certain number of connections with either side [19]. Thus, the fact the number of triangles goes up, makes the clustering coefficient increase. In regular networks, nodes have the same number of neighbors.

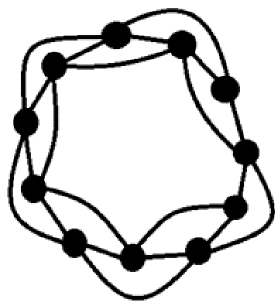

Figure 2.3. A Regular Network.

2.2.3. Small World Network

In a social network, when two people having links with each other are observed, it is seen that they mostly have a mutual acquaintance. The first study on the shrinkage of Planet Earth is brought forward by Karinthy in 1929 [20]. In another research, any two people may have interacted with each other through at most two intermediate links when the network size is approximately 1000 [21]. After that, Milgram made an experiment in order to explore how many intermediate acquaintance links are needed before randomly selected two people are connected [9, 22]. He has selected a certain number of people to send packages from Omaha, Nebraska and Wichita, Kansas to Boston. If people living in starting points know the target person, they have sent packages directly to that person. If not, they have sent packages to the person who is likely to get to know the target person living in Boston. The process is repeated until the packet arrives at the final recipient. According to the experiment done, the median number of steps from the first recipients to the last one is between five and six. It demonstrates that the distances from one node to another are shorter than what is thought, which brings a term known as small world.

A random network and a regular network constitute a small world network whose algorithm is developed by Watts and Strogatz [13]. A random rewiring procedure is used to create a small world network. Consider a ring lattice with N nodes and k edges per node. At first, each node has k number of neighbors.

According to this procedure, each edge is rewired with probability p in a random way. In this sense, some edges between nodes are created randomly and the other ones are formed in the light of regular network. Thanks to the rewired edges formed randomly, the average shortest path is relatively short compared to a regular network. Additionally, the clustering coefficient of small world network is quite high thanks to the lattice concept compare to random network.

As seen in Figure 2.4, when the rewiring probability p converges to zero, it is an utterly regular lattice since there are no rewired edges. As p approaches to one, all pairs of nodes start to create edges in a random way, which brings a random network. In a network whose average shortest path is L and average clustering coefficient is C, the rewiring probability corresponding to $L \geq L_{random}$ and $C \gg C_{random}$ is determined, and a small world network is generated [13].

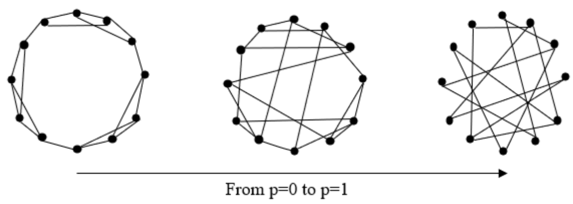

Figure 2.4. Regular to Random Network.

$N \gg k \gg \ln N \gg 1$ is needed, where $k \gg \ln N$ ensures that a random graph will be connected [23]. As p approaches zero and one, the average shortest path and the average clustering coefficient are calculated by Equation 2.4 and Equation 2.5;

$$L \sim \frac{N}{2k} \gg 1 \qquad and \qquad C \sim \frac{3}{4} \qquad as \qquad p \to 0 \tag{2.4}$$

$$L \approx L_{random} \sim \frac{lnN}{lnk} \qquad and \qquad C \approx C_{random} \sim \frac{k}{N} \ll 1 \qquad as \qquad p \to 1 \tag{2.5}$$

where N is the network size and k is the number of neighbors each node has [13].

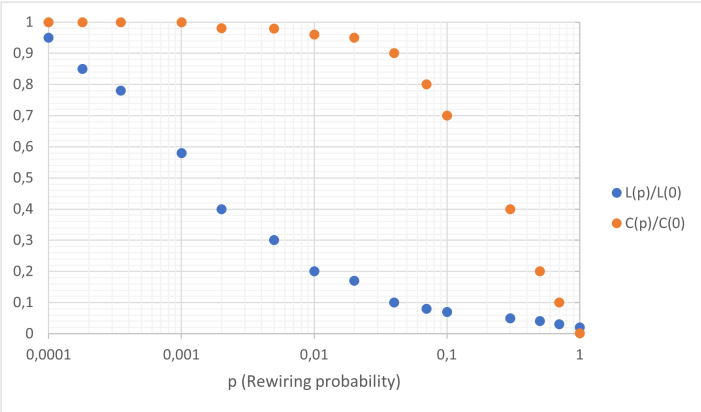

Figure 2.5. L and C Characteristics.

According to Figure 2.5, as the rewiring probability increases, average shortest path length ratio quickly goes down while the clustering coefficient ratio tends to remain high at first. When the rewiring probability p increases if the ratio of $\frac{C(p)}{C(0)}$ is almost constant, it is seen that the small world network behavior is indeterminable. In a small world network, the average shortest path length grows with the logarithm of N in the network [13], which is seen in Equation 2.6;

$$L \propto \log N \tag{2.6}$$

where N is the number of nodes.

A small world network is homogenous in terms of degree distribution [24]. In this respect, a random network and a small world network are similar in degree distribution.

2.2.4. Scale Free Network

One of the complex networks is a scale free network, whose degree distribution is a power law form. The main mechanism underlying that kind of network is the preferential attachment. In this mechanism, people having higher degrees tend to link more, which characterizes hub structure [11].

In this network, a few nodes have significantly high degrees. In terms of the degree distribution, the scale free network is inhomogeneous in nature because of the reasons mentioned above.

When a scale free network is compared to a random network in terms of average shortest path, it is indicated that average shortest path is smaller in a scale free network than in a random network due to uprise of hubs when parameters like degree and network size are kept the same.

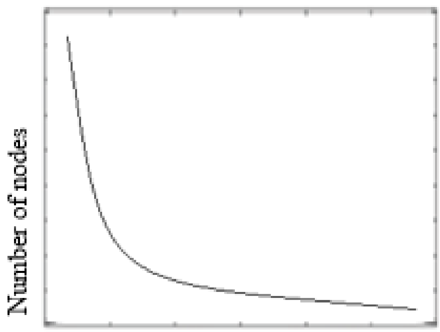

Figure 2.6. Degree Distribution of a Scale Free Network.

The average shortest path of the scale free network can be calculated as seen in Equation 2.7 [25];

$$L \sim \frac{ln(N)}{ln \ln(N)} \tag{2.7}$$

where N is the number of nodes. The clustering coefficient is higher in a scale free network compared to a random network [26]. As the network size increases, it begins to decrease [27]. Its clustering coefficient grows as seen in Equation 2.8;

$$C \sim N^{\frac{-3}{4}} \tag{2.8}$$

where N is the number of nodes.

2.3. Dynamic Network Evolution

Dynamic networks alter their network topology over time unlike static networks. They can be obtained by addition/deletion nodes and/or edges.

According to Albert & Barabási [27], most real networks such as World Wide Web have a scale free property. Whereas the aim of the models of the other network classes such as a random network and a small world network is to generate a graph, the purpose of the model of the scale free network is to capture critical mechanisms behind network evolution. These mechanisms are growth and preferential attachment. In the scope of growth, a new node with a specific degree is added at every single time point. In the preferential attachment mechanism, the probability of creating links is increased with the degree of the node. $\pi(D_i)$ represents the probability that a new node will be linked to node i, which is calculated by Equation 2.9;

$$\pi(D_i) = \frac{D_i}{\sum_j D_j} \tag{2.9}$$

where D_i shows the degree of node i and $\sum_j D_j$ expresses the total number of degrees in the network [27].

Bianconi *et al.* has proposed a model [27, 28] that a fitness parameter and the parameters in Albert & Barabási [27] are considered. The fitness parameter is defined once for each node and represents the ability to make a connection of a node. The value of that parameter does not change while the network evolves. A new node connects to the other node with a probability that depends on those parameters. Both Bianconi Barabási model and Barabási Albert model use preferential attachment structure. What differentiates Bianconi Barabási is the fitness distribution. If an existing node has a low degree and high fitness value, its probability of being selected to be linked is higher than that of the Barabási Albert model. It is indicated that nodes having high fitness values tend to attract the more.

Nguyen *et al.* [29] has generated a general framework that includes link formation rules such as reciprocity, transitivity, cycle, common-in neighbor and common-out neighbor. The framework, which provides a model regarding link formation behaviors, is defined in a dynamic directed social network at the node level and instance level. An instance represents a subgraph consisting of nodes and their edges. Two different behaviors, which are rule usage and rule confidence, are defined for each link formation rule. Rule usage is related to the number of links that a node or an instance has based on each rule. Rule confidence is basically regarding proportion of reciprocal links. These link formation rules are measured on one of the datasets belonging to dynamic directed social networks, Epinions (http://www.trustlet.org/wiki/Extended Epinions dataset). In this data set, nodes having at least 20 in-links and 20 out-links are selected to analyze. The results indicate that common out neighbors, common in neighbor and transitivity are the most frequently used rules for rule usage behavior of both an instance and a node. However, reciprocity is the rule having higher value in terms of rule confidence.

Newman [30] analyzed how scientific collaboration network grows over time. Two main networks are chosen from different fields, Biology and Physics. When any two scientists have published papers together, it is assumed that they have connected to each other. The results show that the probability of collaboration between two scientists selected from these two fields increases with the number of common acquaintances they have and the number of past collaborators.

Stattner *et al.* [8] proposed a method to understand how important the structure and characteristics of a dynamic network are during the propagation process. The process works as follows. A susceptible node become infected with a certain probability. Then, one of the evolution strategies is carried out to evolve the initial network. The strategies cover only link formation rules, which are random, preferential attachment, triadic closure and global connection. Then, the infected node can recover with a probability and does not return to susceptible one anymore.

There are four initial networks considered at time $t = 0$, two of which are obtained by Erdos-Renyi and Barabasi Albert models, GEN1 and GEN2 respectively. The others, HS and PL, are real and synthetic networks representing a small world network and a scale free network. Additionally, the evolution speed is considered based on the number of links generated at every iteration and set at 10, 50, 100, 150. Network properties are investigated after the network evolution. Disease spreading on DynSpread, which is a 2D simulation tool, indicates how different spreading emerges according to link formation rules. When the evolution speed is low, GEN1 and PL networks do not generate an epidemic in any of strategies. When the speed is 50, they can only generate an epidemic with the preferential attachment strategy. Also, it is seen that GEN2 network generates epidemic with all the rules mentioned. Preferential attachment is the strategy that gives the highest score in epidemic dissemination compared to the other rules.

Toivonen *et al.* [31] compares dynamic network evolution models for social networks to understand how they evolve. In these models, an algorithm runs until the structural properties of the network do not change anymore. Dynamic network evolution models consist of three network evolution models, which are DEB, MVS and KOSKK. In DEB model proposed by Davidsen *et al.* [32], if the number of neighbors of a node i chosen randomly is less than two, another random node is selected and they are connected. If not, two neighbors of a node i are selected randomly and connected to each other. To delete links, a random node is selected, and all links of that node are deleted with a certain probability. The second model proposed by Marsili *et al.* [33] is related to triadic closure mechanism. Basically, a random node is connected to another one randomly with a certain probability. Additionally, second level neighbors of a node i are selected. Then one of them is connected to the node i with a probability. In link deletion phase, a link is selected randomly and deleted with a predefined probability. The last one, KOSKK model, developed by Kumpula *et al.* [34] uses triadic closure mechanism as in the previous one. The difference between MVS and KOSKK is that KOSKK considers an interaction strength. A node i is selected randomly as the first node. One (e.g. node k) of the second level neighbors of node i is selected.

If node i and node k are connected to each other, the strength of the links along the path from node i to node k is increased by δ. If not, a link is created between them with a certain probability and the strength of the links along that path is increased by δ. Additionally, node i is connected to another one randomly with probability p_r. If degree of node i is zero, p_r equals to one. New links are mostly generated through strong links. Its link deletion mechanism is the same as that of DEB model. The models mentioned are implemented on two real networks, email and lastfm, whose degree distribution decay exponentially and slowly over time, respectively. When the results are observed, DEB and KOSKK models produce the best match to real networks in terms of clustering and the number of cliques. However, none of dynamic network evolution models explain all structural properties of those networks completely.

Stattner *et al.* [35] generates a network evolution mechanism that involves more than one rule in the algorithm rather than a single link addition rule. A telecommunication network including communication activities of users for 20 days is analyzed to understand underlying evolution mechanisms. The degree distribution of the network at time $t = 0$ is similar to power law distribution. The evolving network after 20 days has still scale free property. To understand if the preferential attachment is only mechanism explaining that network, a hybrid mechanism is generated. The hybrid mechanism including random, preferential attachment and triadic closure rules works as follows. Node i and node j are selected from the network and their connection probabilities are calculated. Basically, degree of node i is divided by sum of degrees of each node in the network to calculate p_i. p_j is also calculated in the same way. They are compared and the higher probability is selected. If the threshold predefined is less than or equal to calculated probability, the algorithm selects preferential attachment. If node i and node j have more than or equal to one common neighbor, the triadic closure rule is chosen to be worked. Otherwise, the random mechanism is activated. When the results are considered, it is seen that 20-30% of new links are from preferential attachment mechanism, 30-50% of the new links are from triadic closure and the rest of them are from the random mechanism. It is concluded the network is not evolved by only preferential attachment although the scale free structure is maintained.

Davidsen *et al.* [32] has proposed a model including both link addition and deletion mechanisms to understand how a small world behavior of a network is evolved. The model consists of two mechanisms, which are transitive linking and change of nodes. In the first mechanism, after selecting a random node in the network, its two neighbors are picked in a random way and they are introduced to one another if they have not already met. If the node selected at first does not have at least two neighbors, it is connected to another node selected randomly. In the second one, a deletion algorithm is evolved, which deletes a node, and all links of that node with probability p. Then, it is modified by a new node with one acquaintance selected randomly. The number of network size is constant during network evolution. When p equals to 0.025, the clustering coefficient of evolving network is 0.63. The degree distribution of that network is power law. The clustering coefficient of a random network with the same parameters is 0.0021. When the result is observed, it is seen that this network has high clustering coefficient thanks to the first mechanism, which demonstrates the existence of small world behavior. For the small values of probability p, the evolving network turns into the scale free network and the degree distribution of that network has power law property. However, as p increases, the degree distribution of the network is exponential.

Wang *et al.* [36] has proposed a model to understand how computer viruses spread through e-mail messages. In that model, a directed e-mail network is evolved over time by adding and deleting links. In this network, an e-mail address book is represented as a node and the relevant e-mail addresses in that book are represented as links. If the defined link addition threshold is less than the number of e-mails that are sent from node i to node j, the e-mail address of node j is seen in the e-mail address book of node i. If not, there is no change regarding link addition. When comes to link deletion from the network, the number of e-mails that are sent from node i to node j is compared to the link deletion threshold and if the calculated number is less than deletion threshold, relevant e-mail addresses are deleted. When the network size is 1000, the number of e-mails that are sent from node i to other nodes is determined as 20.

According to the analysis, the statistical properties of the network have changed based on different parameters determined. For instance, when the link addition threshold is increased, it is seen that the average number of links decreases. Similarly, the average number of links decreases as the link deletion threshold is risen. Additionally, the evolving network corresponds to small world characteristics as the network size drops.

Jin *et al.* [37] proposed two models to observe community evolution considering dynamic features of social networks. In the first model, the probability of creating a link between two nodes increases as the number of mutual nodes rises. There are also three other components taken into account, which are a maximum limit of the number of acquaintances that a node has, deletion strength probability considering the time difference from the last time they meet to today and having constant number of nodes. When the first model is implemented on an empty network with 250 nodes and without any links, it is indicated that there is high clustering coefficient calculated approximately as 0.45 in such social networks. When a random graph is created with the same parameters, the clustering coefficient is nearly 0.02. Therefore, it is seen that the first model provides a significant increase in clustering coefficient. In the second model a parameter rather than the strength probability is assigned to delete a link. The first and second model is quite similar in link addition rule, maximum degree limit rule and having constant number of nodes. When the second model is carried out on the same empty network as in the previous model, its clustering coefficient is calculated as 0.53 when each node reaches the maximum limit of degree. As seen above, these two models create similar results in clustering coefficient. Additionally, when the decay parameter is selected high, the evolving network converges to a random network.

There are several studies related to dynamic social networks. However, in the literature, there are very few studies related to how the characteristics of small world networks are preserved over time, therefore there is an open point that needs to be studied. In the next chapter, differences of this study from other studies will be compared and objectives will be specified.

3. RESEARCH OBJECTIVES

There is a variety of models regarding dynamic network evolution. Some models only focus on link and/or node addition. In that case, the network tends to grow permanently. Therefore, it is not close enough to real networks because they can grow and shrink. There are also some models that both link addition and link deletion mechanisms are considered. Wang *et al.* [36] has analyzed the change in network structure for a directed e-mail network according to model including the mechanisms of link addition and link deletion. However, there are also real networks, consisting of undirected links, such as a Facebook network and a cooperation network.

There are also several studies on undirected networks. Toivonen *et al.* [31] has implemented three models on two real networks to understand how they are evolved by network evolution models. In other words, these models are implemented to two real networks. However, the degree distributions of these chosen real networks have a tendency to decay exponentially. These models have been analyzed on networks similar to scale free networks.

Davidsen *et al.* [32] has analyzed how a small world behavior can arise dynamically. In addition to link addition and deletion, nodes have been removed and added at different time points in their model. They have analyzed how the network structure is influenced by the change of the probability of adding/removing nodes. However, in some networks, node sets may not change. Consider a network where nodes represent deputies and links represent interactions between them. After the elections, deputies are not expected to change during a particular time. Instead, changes in interactions among them are expected. As a second example, in a university network, the number of students does not change unless there are incoming and graduating students, but the links among students always change. As another example, for a colleague network where nodes represent employees and edges represent the interactions between them, employees are not expected to leave the network unless they leave their jobs.

Similarly, a person is not expected to be added to the network unless that person is newly recruited. Since interacting with each other is more likely than addition/deletion of people, it can be thought that interactions can be made with the same people. Therefore, in this study, node addition/deletion in the network will be neglected. The number of nodes is kept constant, assuming that the number of node addition/deletion is proportionally much less than the number of link addition/deletion. All changes regarding network evolution will be made over links.

Jin *et al.* [37] has developed models that the only attribute changing is links (but not nodes) to understand how the social network evolves. The models are implemented over an empty network including solely nodes (without links) and the change of the structural properties of the network has been investigated. At first, deletion process does not occur. After the network reaches a certain size, it has been started. However, link addition and deletion mechanisms are not expected to run separately in a network. For instance, changes in interactions among students in a university network, link additions and link deletions, occur in similar periods. Therefore, in this study, link addition and deletion mechanisms will be run at the same time. Although the purposes of studies of Davidsen *et al.* [32] and Jin *et al.* [37] are similar to this study, the working mechanism of the models in this study and research objectives are differentiated from them.

The aim of this study is to determine relevant link addition and link deletion mechanism combinations that keep the structure of an undirected small world network the same. Additionally, the mechanism combinations that make the network structure change (e.g. from a small world network to a lattice or from a random network to a network with high average clustering coefficient and positively skewed degree distribution) will be analyzed. Moreover, mechanisms deteriorating the network structure (e.g. converge to a random network) will be uncovered. In this study, random networks and small world networks are the initial network classes that have been focused on. The results will be compared with the results of Davidsen *et al.* [32] and Jin *et al.* [37] in terms of average clustering coefficient and degree distribution.

4. METHODOLOGY

In this study, after a random network or a small world network is constructed, its characteristics such as average shortest path, the average clustering coefficient and the degree distribution are calculated. Then, 36 experiments different network modification combinations are carried out. After these characteristics of each experiment are calculated, they are compared to those of the initial network. For each experiment, replications have been done and the results of them are averaged. Since modified networks of some experiments converge to random network or have been fragmented, they are not interesting and therefore, ignored to discuss comprehensively. Experiments whose modified networks start to have small world network characteristics or scale free network characteristics are interesting to analyze and so will be discussed.

There are 18 network modification combinations containing six link addition and three link deletion mechanisms. 36 experiments including two network classes and 18 network modification mechanisms are analyzed in different snapshots where a fraction of the total links have been changed, and consequently how each snapshot changes is observed. In each snapshot, a certain fraction of total links is modified. For instance, the network at $t_{0.25}$ is the snapshot of the network after quarter of its links are modified according to one of the network modification combinations. To compare evolving networks with the initial network, average shortest path and average clustering coefficient are calculated in addition to the degree distribution. Additionally, sub-experiments are carried out with different parameters to observe if the small world network property has still maintained for those experiments. Different number of replications are carried out, which depends on the result of the experiments. At first, six replications are implemented for each experiment to understand the behavior of experiments. If modified networks remain a random network or converge to a random network, the relevant experiments are not meaningful enough. Due to the low variance of the replication results of these experiments and computationally expensive, the number of replications of these experiments is not increased.

If the small world property is discovered, the evolving network does not converge to a random network or the structure of the network is not fragmented, 25 replications have been carried out. Network construction and modification processes are implemented in Python by using networkx [38] and dynetx packages. The former is a package for complex networks used in static network models. The latter, which is the add-on of networkx, is also a package for complex networks and is used in dynamic network models [39].

In the study, all networks that are studied are undirected networks. Furthermore, equal link strength is considered, which shows that it is not a weighted network but an unweighted network. Nodes in the network do not change during network modification process. Rather than nodes, links change over time. Links between nodes can be deleted or added. Additionally, the number of total links in the network does not change at the end of the network modification process, because the same number of links are added and deleted in each replication.

4.1. Network Construction based on network classes

The network size, also known as the number of nodes in the network, is determined as 2000. In order to analyze the influence of different network sizes on the average shortest and the average clustering coefficient, the network size of some experiments is chosen as 3000 other than 2000.

Random networks and small world networks are the initial networks to be constructed. Random networks are created by the Erdős Rėnyi algorithm in which the network size and the probability of link creation between nodes are considered [17]. The probability of link creation specifies the number of links in the network. The expected number of links is calculated as $\frac{N\times(N-1)}{2} \times p$. The link creation probability represented as p and the network size represented as N are chosen as 0.02 and 2000, respectively.

Another network class is a small world network constructed by the Watts-Strogatz algorithm whose parameters are the network size, average number of neighbors of nodes, and rewiring probability [13]. Basically, a small world network has both regular network and random network properties. Thanks to rewiring probability chosen, some edges between nodes are recreated randomly over a regular network. Because the small world network has regular network properties, its average clustering coefficient is high (lower than that of a regular network). Additionally, since small world network has random network properties, its average shortest path is low (higher than that of a random network). If the rewiring probability represented as p equals 0, then the network converges to a regular network whose average shortest path is quite high. When the rewiring probability equals one, a random network is generated, and its clustering coefficient significantly decreases. In this study, the rewiring probability is chosen as 0.1 in order to decrease the clustering coefficient as little as possible and drop the average shortest path as much as possible. The average number of neighbors of a node, represented as k is chosen as 40 for most of the experiments (30 for some to understand the influence of k on characteristics of the modified networks). For hybrid mechanisms, threshold values enable one of the alternative mechanisms to be selected. In some sub-experiments, different threshold values (the value that make one of random and triadic closure mechanisms to be chosen) are used such as 0.1, 0.2, 0.3, 0.4, and 0.5 to understand the influence of change in threshold values on the characteristics of the network.

Before any network is created, a seed for network construction is determined in addition to the parameters mentioned above, which makes the specific network constructed find at any time. For each network class and its different parameter set, 25 different networks are constructed considering different network construction seeds. Then, they are averaged and used as reference values representing the initial networks.

For each network created at time $t = 0$, the main structural properties such as the degree distribution, the average shortest path and the average clustering coefficient are calculated.

4.2. Network modification

After a network is created, it is started to be modified. The network modification is carried out over links. There are six link addition mechanisms that are implemented, which are *random*, *preferential attachment*, *triadic closure 1*, *triadic closure 2*, *hybrid 1* and *hybrid 2*. In addition to this, three link deletion mechanisms, which are *random*, *degree* and *the least common degree*, are considered. Six link addition mechanisms and three link deletion mechanisms constitute 18 network modification combinations. Basically, some links among nodes are created according to one of the link addition mechanisms and some links are deleted according to one of the link deletion mechanisms. In order to create a link between nodes or delete a link from the network, the first node is selected randomly in all mechanisms. The second node is chosen according to the network modification mechanisms. Then, a link between them is formed or deleted.

A network is modified as much as a certain fraction of the number of total links. This fraction is selected as 0.25, 0.5, 0.75, 1, 2, 3, 4, and 5. Thus, eight snapshots are created. For example, the network at $t_{2.00}$ is the snapshot of the network changed as much as two times the number of total links. The average shortest path, the average clustering coefficient and the degree distribution of each snapshot are calculated and compared to those of the initial network.

The number of snapshots changes depending on the result of experiments.Namely, it has not been increased much if the network converges to random network immediately. If the network does not reach a steady state without converging to random network, it has been increased. At most eight different snapshots are generated for each experiment. Additionally, a network modification seed is chosen for each replication in addition to the seeds for network construction. Thus, replications with different seeds are provided for each experiment. The iteration logic of network modification is explained with the help of the pseudo code in Figure 4.1.

Set a seed for network construction

Construct the network n with its parameters

Calculate the average shortest path, the average clustering coefficient and the degree distribution

Compute the total number of links L

Set $i \Leftarrow [0.25, 0.5, 0.75, 1, 2, 3, 4, 5]$

Set $k \Leftarrow 0$

Set a seed for network modification

while $length(i) > k$ **do**

  for t in range(1, $i[k] \times L$) **do**

    Add a link according to the link addition mechanism a ;

    Delete a link according to the link deletion mechanism d ;

  end for

  Calculate the average shortest path, the average clustering coefficient and the degree distribution

  Reconstruct the initial network

  $k = k + 1$

end while

Figure 4.1. Network Modification Mechanism.

After a seed to construct a network is set, the network n which can be a random network or a small world network is created with relevant parameters. If the network n to be created is a random network, the network size N and the probability p of link creation between nodes are set. If it is a small world network, the network size N, average number of neighbors k of nodes, and rewiring probability p are determined.

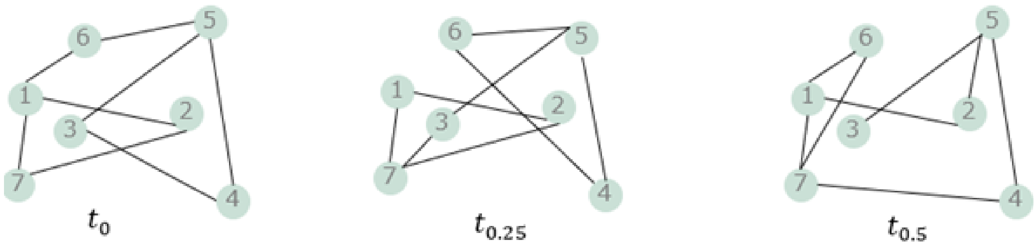

Figure 4.2. An Example of Network Modification Mechanism.

In order to explain the network modification mechanism more clearly, an example (in Figure 4.2) can be given. As seen, there are three network snapshots at t_0, $t_{0.25}$ and $t_{0.5}$. The initial network (the network at t_0) has eight edges. This network is constructed after the network construction seed and relevant network parameters are set. Then, the network is modified over links considering a network modification seed. The number of links that will be changed is calculated with the multiplication of the total number of links (eight in this example) and the coefficient expressing the specific time point. For the network at $t_{0.25}$, the number of links changed is two. Namely, two links will be added and deleted between node pairs. The same process is repeated for the other snapshots. For instance, in the network at $t_{0.5}$, the number of links changed is four. In order to compare the modified networks with the initial network, the modified network is reconstructed at the end of each snapshot and the modification process is repeated.

4.2.1. Link addition mechanisms

The network modification mechanism is run t times as explained in Figure 4.1. For each t, the link addition mechanism a is selected and implemented to add links.

4.2.1.1. Random (R). In this mechanism type, the first node and the second node are chosen randomly. In the pseudocode explained in Figure 4.3, N_x represents neighbors of the node x and PN expresses the potential node set that will be linked.

Select node x randomly
Set $N_x \Leftarrow$ List neighbors of the node x
Set $PN \Leftarrow$ Remove N_x and the node x from the total node set
Select node y from PN randomly
Add the link $[x, y]$

Figure 4.3. Pseudocode of Randomly Link Addition.

4.2.1.2. Preferential Attachment (PA). After selecting the first node randomly, the second one is chosen proportional to the square of the degree each potential node has. PN_d represents the square of the degree that each potential node has. $P_{normalized}$ represents a ratio of the cumulative sum of PN_d to the maximum value of the cumulative sum of PN_d. After calculating $P_{normalized}$ of each potential node, it is compared to a random number generated.

To explain the working mechanism of preferential attachment, Figure 4.5 is demonstrated. After the first node (node X in this example) is selected, the degrees of potential nodes are found. It is assumed that node X has three potential nodes whose degrees are two, three and four, respectively. According to the mechanism rule, squares of degrees are calculated. Then, each square degree is cumulatively sum by row. After that, the normalized probability is calculated. To calculate this probability cumulative sum value of each potential node is divided by maximum value of cumulative sum (4/29, 13/29, 29/29 in this example, respectively). Lastly, a random number is generated and the node belonging to relevant range is selected as a second node. For instance, when it is assumed that it is 0.45, the node having four degrees is selected as the second node since it is between 0.44 and 1.

```
Select node x randomly
Set N_x ⇐ Neighbors of node x
Find PN by removing N_x and the node x from node set
Sort PN based on degrees and list degrees PN_degrees
List nodes PN_nodes according to sorted degrees of PN
Set PN_d ⇐ List square of PN_degrees
for i < len(PN_d)  do
  PN_cumsum[i] = ∑_{k=0}^{i} PN_d[k]
end for
Set PN_normalized ⇐ PN_cumsum/The maximum value of PN_cumsum
Generate a random number RN
Compute length of PN_normalized named as len
for k in range(0, len) do
  if RN < PN_normalized[k] then
    Break
  else
    k = k + 1
  end if
end for
Find the node y from PN_nodes[k]
Add the link [x, y]
```

Figure 4.4. Pseudocode of Preferential Attachment.

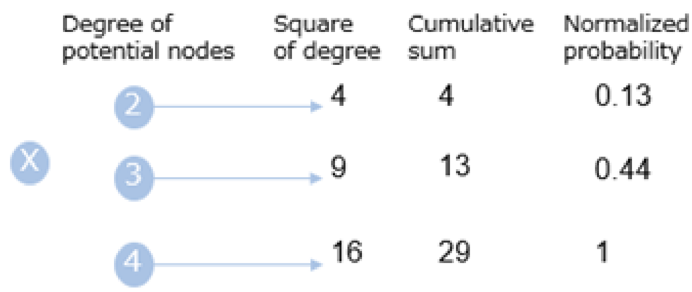

Figure 4.5. An Example of Preferential Attachment Link Addition Mechanism.

4.2.1.3. Triadic Closure 1 (TR1). After the first node is chosen randomly from the network, the second node is selected based on triadic closure 1. In this mechanism, total potential nodes contain union of second level neighbors of the first node. N_{XY} demonstrates second level neighbors of node x. Distinct potential node list is created by excluding repetitive nodes after determining total potential nodes TPN_Y. Then, the second node is selected from that list.

To clarify the working mechanism of this link addition mechanism, an example is demonstrated in Figure 4.6. According to this mechanism, when it is assumed that node 1 is the first node selected, the neighbors of the first node are listed at first (node 2, node 3, node 8 and node 4 in this example). Then, their neighbors (or the second level neighbors of the first node) are found and they are assigned to a potential list (node 5, node 5, node 7, node 4 and node 8). In this list, a node is found only once (node 5, node 7, node 4 and node 8). Then, they are checked if they have already linked with the first node. (Node 4 and Node 8 have already linked with the first node.) The ones which are not linked with the first node, are assigned to unique potential list. In this example, the potential node list has two nodes, node 5 and node 7. One of them is randomly selected from that list.

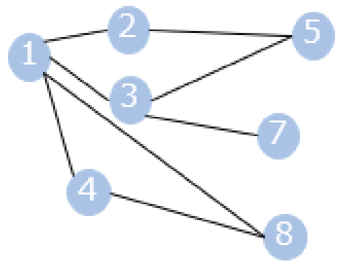

Figure 4.6. An Example of Triadic Closure 1 Link Addition Mechanism.

Select node x randomly

Set $N_x \Leftarrow$ Neighbors of node x

$TPN_Y = [\,]$

for each $k \in N_x$ **do**

  Set $N_{X_Y} \Leftarrow$ Neighbors of k

  Set $PN_Y \Leftarrow$ Remove the node x from N_{X_Y}

  Set $TPN_Y \Leftarrow PN_Y \cup TPN_Y$

end for

Set $UPN_Y \Leftarrow$ Find relevant unique potential node list from TPN_Y

for i in UPN_Y **do**

  if one of neighbors of i = Node x **then**

    $UPN_Y = UPN_Y - i$

  end if

end for

Select the node y from UPN_Y randomly

Add the link $[x, y]$

Figure 4.7. Pseudocode of Triadic Closure 1.

4.2.1.4. Triadic Closure 2 (TR2). After selecting the first node in a random way, the second one is determined according to the triadic closure 2 mechanism. What differentiates from triadic closure 1 is that repetitive nodes are not excluded after finding the sum of second level neighbors of the first node. In this way, the more repeated nodes there are, the more likely they are to be selected as a second node. In other words, repetitive nodes have higher probability to be selected as a second node. If the neighbors of the first node have common neighbors, they are shown in the potential list to be selected as a second node repeatedly. If the neighbors of the first node have common neighbors that have not already linked with the first node, they are shown in the potential list in a repetitive way. Therefore, their probability to be selected is higher than the nodes found in the list once.

As mentioned, according to the example in Figure 4.6, node 5 and node 7 are the nodes in the potential list for triadic closure 1. However, in the triadic closure 2, the potential list consists of node 5, node 5 and node 7. Namely, a node can be found more than once in this mechanism different from triadic closure 1.

4.2.1.5. Hybrid 1 (HB1). This link addition mechanism is a combination of random and triadic closure 1. There is a threshold level that is determined in advance. According to the threshold level and random number generated, one of the alternatives is chosen. In most of the experiments, the threshold level, THR is selected as 0.4. If THR is greater than or equal to random number, the random link addition mechanism is activated. Otherwise, the triadic closure 1 mechanism is selected for link addition process. Different threshold levels are also set in an experiment to compare the results with one another.

4.2.1.6. Hybrid 2 (HB2). Hybrid 2 is the other hybrid link addition mechanism, which combines random and triadic closure 2 mechanisms.

```
Select node x randomly
Set N_x ⇐ Neighbors of node x
TPN_Y = [ ]
for each k ∈ N_x do
   Set N_{X_Y} ⇐ Neighbors of k
   Set PN_Y ⇐ Remove the node x from N_{X_Y}
   Set TPN_Y ⇐ PN_Y ∪ TPN_Y
end for
for i in TPN_Y do
   if one of neighbors of i  = Node x then
      TPN_Y = TPN_Y − i
   end if
end for
Select the node y from TPN_Y randomly
Add the link [x, y]
```

Figure 4.8. Pseudocode of Triadic Closure 2.

```
Set threshold THR between 0 and 1
Generate a random number RN
if THR ≥ RN then
   Choose Random link addition mechanism
else
   Choose Triadic Closure 1 link addition mechanism
end if
```

Figure 4.9. Pseudocode of Hybrid 1.

Set threshold THR between 0 and 1
Generate a random number RN
if $THR \geq RN$ **then**
  Choose Random link addition mechanism
else
  Choose Triadic Closure 2 link addition mechanism
end if

Figure 4.10. Pseudocode of Hybrid 2.

4.2.2. Link deletion mechanisms

The network modification process is carried out t times as explained in Figure 4.1. For each t, the link deletion mechanism d is selected and implemented to delete links between nodes selected.

4.2.2.1. Random (R). After finding neighbors of the first node, one of them is selected randomly. Then, the link between the first node and the second node is deleted from the network.

Select the node x randomly
Set $N_x \Leftarrow$ Neighbors of node x
Select the node y from N_x randomly
Remove the link $[x, y]$

Figure 4.11. Pseudocode of Random link deletion.

4.2.2.2. Degree (D). The first node is selected randomly. A link is eliminated from the network based on the degrees of neighbors of the first node. The node having the least neighbors is selected as a second one.

Select the node x randomly

Set $N_x \Leftarrow$ Neighbors of node x

$Degreelist =$ [ ]

for each $k \in N_x$ **do**

  Set $N_{degreelist} \Leftarrow$ List $[(k, degree\ of\ k)]$

  $Degreelist \Leftarrow$ Append $N_{degreelist}$ to $Degreelist$

end for

Sort $Degreelist$ according to $degree\ of\ k$ in an ascending way

Set $y \Leftarrow$ Find k from the first index of $Degreelist$

Remove the link $[x, y]$

Figure 4.12. Pseudocode of Degree.

4.2.2.3. The Least common degree (LCD). As mentioned in previous mechanisms, the first node is chosen in a random way. In this mechanism, the neighbors of the first node and second level neighbors of the first node are found. Then, neighbors of the first node and its neighbors of each neighbor of the first node are intersected and common degree is calculated. This process is repeated for each neighbor of the first node. The results are appended to a list. The node having the least common degree is selected as a second node. The link between the first node and the second node is deleted from the network. CN_X represents nodes that the neighbors of the first node and its neighbors of each neighbor of the first node have in common.

To explain this mechanism more clearly, an example in Figure 4.14 is demonstrated. It is assumed that node 1 is the first node selected at first. Then, the neighbors of the first node (node 2, node 4, node 5 and node 6) are found. For each neighbor of node 1, its neighbors are listed. For instance, neighbors of node 2 is node 3, node 9, node 4, node 5 except for node 1. The neighbors of node 2 and node 1 are intersected and common degree is found. The same process is repeated for each neighbor of node 1 and the node having the least common degree is selected as a second node.

Select the node x randomly

Set $N_x \Leftarrow$ Neighbors of node x

$Commondegreelist =\ [\,]$

for each $k \in N_x$ **do**

  Set $N_{X_Y} \Leftarrow$ Find the neighbors of k

  Set $CN_X \Leftarrow\ N_{X_Y} \cap\ N_x$

  Set $CCN_X \Leftarrow$ List $[(k, length\ of\ CN_X)]$

  $Commondegreelist \Leftarrow$ Append CCN_X to $Commondegreelist$

end for

Sort $Commondegreelist$ according to $length\ of\ CN_X$ in an ascending way

Set $y \Leftarrow$ Find k from the first index of $Commondegreelist$

Remove the link $[x, y]$

Figure 4.13. Pseudocode of the Least Common Degree.

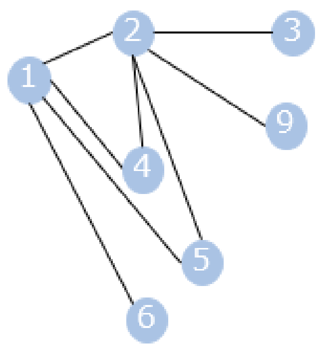

Figure 4.14. An Example of the Least Common Degree Link Deletion Mechanism.

5. EXPERIMENTATION

Reference values should be determined to compare experiments and understand whether the class of the modified network has changed over time. Therefore, 25 seeds for network construction are selected. Then, a random network or a small world network is created for each seed. The fundamental metrics, which are the average shortest path (ASP) and the average clustering coefficient (ACC), are calculated and averaged. Namely, for each parameter set explained in Table 5.1 and Table 5.2, networks are constructed with 25 different seeds and then the results are averaged. These averaged metric values are considered as reference values, shown as a red dashed line for the random network and black dashed line for the small world network in the experiments.

Table 5.1. Minimum, Maximum and Average Values of ASP.

Network class	Parameters	Across replications		
		MIN - ASP	MAX - ASP	AVG - ASP
Random Network	N=2000 p=0.02	2,4131	2,4280	2,4210
Small World Network	N=2000 K=40	2,8193	2,8304	2,8242
Small World Network	N=3000 K=30	3,2397	3,2590	3,2468
Small World Network	N=3000 K=40	2,9390	2,9529	2,9441

Table 5.2. Minimum, Maximum and Average Values of ACC.

Network class	Parameters	Across replications		
		MIN - ASP	MAX - ASP	AVG - ASP
Random Network	N=2000 p=0.02	0,0193	0,0204	0,0200
Small World Network	N=2000 K=40	0,5318	0,5393	0,5356
Small World Network	N=3000 K=30	0,5265	0,5349	0,5302
Small World Network	N=3000 K=40	0,5318	0,5399	0,5346

Other than average shortest path and average clustering coefficient, the degree distributions of the initial networks at t_0 are also calculated and demonstrated as follows.

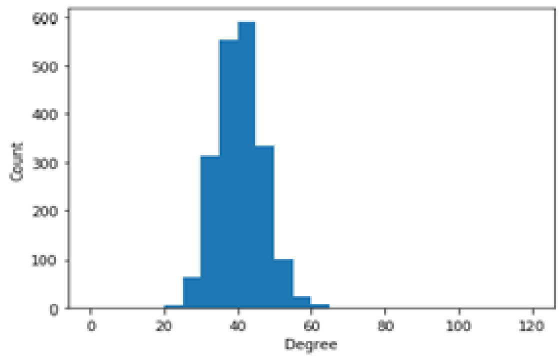

Figure 5.1. The Degree Distribution of a Sample Random Network at t_0.

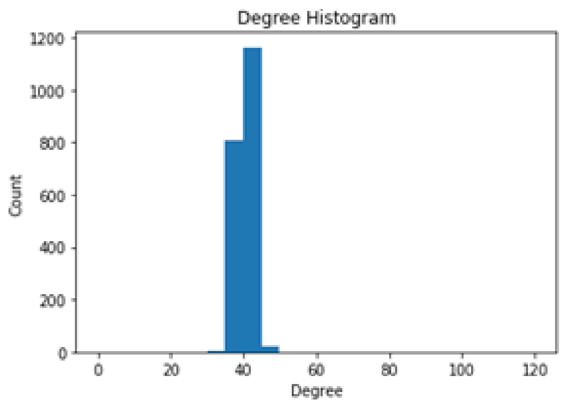

Figure 5.2. The Degree Distribution of a Sample Small World Network at t_0.

Real networks have specific network characteristics, such as small world structure [9] and large hub scaling [10]. A random network does not usually meet characteristics of real networks. Therefore, the experiments that remain a random network and converge to a random network will not be discussed in detail.

Considering two network classes, six link addition mechanisms and three network deletion mechanisms, 36 experiments are carried out. 22 of the experiments are ignored from detailed analysis and discussions because of the reasons mentioned above.

14 experiments have been selected to be analyzed comprehensively. The results of experiments will be analyzed under different subsections for each network class.

After all replications are carried out for each experiment, the average shortest path and average clustering coefficient of these replications are calculated and averaged.

5.1. Experiments on Random Networks

As seen in Table 5.3, 18 experiments are conducted on random networks. Link addition and link deletion mechanisms are abbreviated as + and -, respectively. The value in the brackets shows the number of replications that is carried out for each experiment.

Table 5.3. Experiments on the Random Network.

Exp. No	**Network modification mechanisms**	**Exp. No**	**Network modification mechanisms**
1	R+ R- (6)	10	TR2+ R- (6)
2	R+ D- (6)	11	TR2+ D- (6)
3	**R+ LCD- (25)**	12	**TR2+ LCD- (25)**
4	**PA+ R- (25)**	13	HB1+ R- (6)
5	PA+ D- (6)	14	HB1+ D- (6)
6	**PA+ LCD- (25)**	15	**HB1+ LCD- (25)**
7	TR1+ R- (6)	16	HB2+ R- (6)
8	TR1+ D- (6)	17	HB2+ D- (6)
9	**TR1+ LCD- (25)**	18	**HB2+ LCD- (25)**

At first, six replications are carried out for each experiment. For the experiments which are not discussed comprehensively, the number of replications is not increased because it is computationally expensive and the replication results have low variance. After experiments giving promising results are selected, the number of replications is increased to 25.

In this part, seven experiments demonstrated as bold in Table 5.3 will be discussed.

The reason why 11 experiments will not be discussed is that they remain a random network or the networks have been fragmented during network modification. As an example, the result of TR1+ R- experiment is demonstrated below. This experiment is carried out until the network at $t_{1.00}$ is found. As seen in Figure 5.3 and Figure 5.4, the average clustering coefficient and the average shortest path of the modified network are similar to those of the random network. For each network construction seed demonstrated below, two replications are carried out and averaged (Six replications in total). It is seen that the variance of results based on seeds is considerably low. Therefore, the number of replications is not increased for those experiments. In the experiments that degree link deletion mechanism is used, random networks have been fragmented and still have random network characteristics.

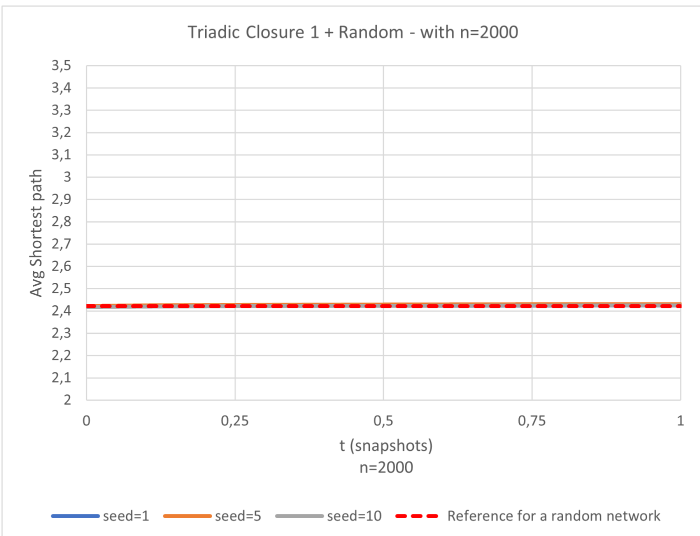

Figure 5.3. Average Shortest Path of the Random Network Modified by TR1+ R-.

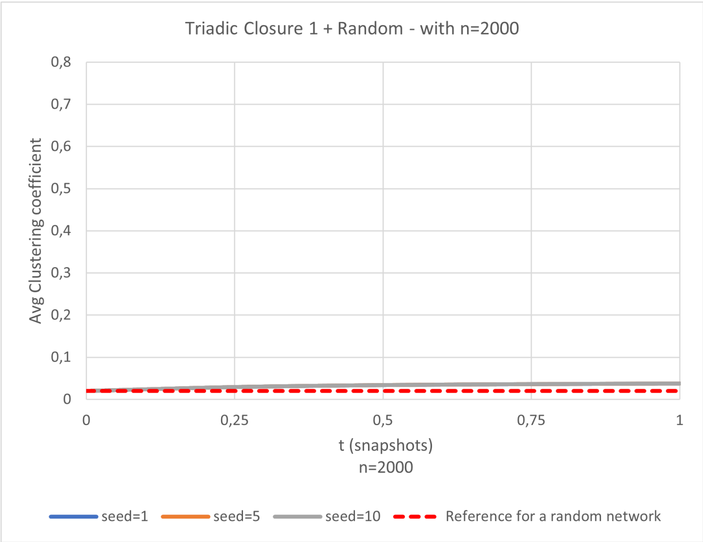

Figure 5.4. Average Clustering Coefficient of the Random Network Modified by TR1+ R-.

To compare networks modified by mechanism combinations to networks generated using Watts and Strogatz algorithm [13] with different rewiring probabilities and observe whether the final network has small world network characteristics, several networks with different rewiring probabilities (from 0.1 to 1 with the increase of 0.1) are generated. The average shortest path and the average clustering coefficient of them are demonstrated in Figure 5.5 and Figure 5.6. For each rewiring probability of the network considered, 25 replications are carried out and the results are averaged. The metric results whose rewiring probability is between 0 and 0.1 (with the increase of 0.01) are demonstrated in Table 5.4.

Table 5.4. Metric Results With Rewiring Probabilities Between 0 and 0.1.

Rewiring probability	Average clustering coefficient			Average shortest path		
	N=2000 k=40	N=3000 k=40	N=3000 k=30	N=2000 k=40	N=3000 k=40	N=3000 k=30
0	0.73	0.73	0.72	25.49	37.99	50.48
0.01	0.71	0.71	0.70	3.77	4.10	4.75
0.02	0.69	0.69	0.68	3.41	3.65	4.17
0.03	0.67	0.67	0.66	3.22	3.42	3.88
0.04	0.65	0.65	0.64	3.10	3.29	3.69
0.05	0.63	0.63	0.62	3.02	3.19	3.56
0.06	0.61	0.61	0.60	2.96	3.12	3.47
0.07	0.59	0.59	0.58	2.91	3.06	3.40
0.08	0.57	0.57	0.57	2.88	3.02	3.34
0.09	0.55	0.55	0.55	2.85	2.98	3.29
0.1	0.54	0.53	0.53	2.82	2.94	3.25

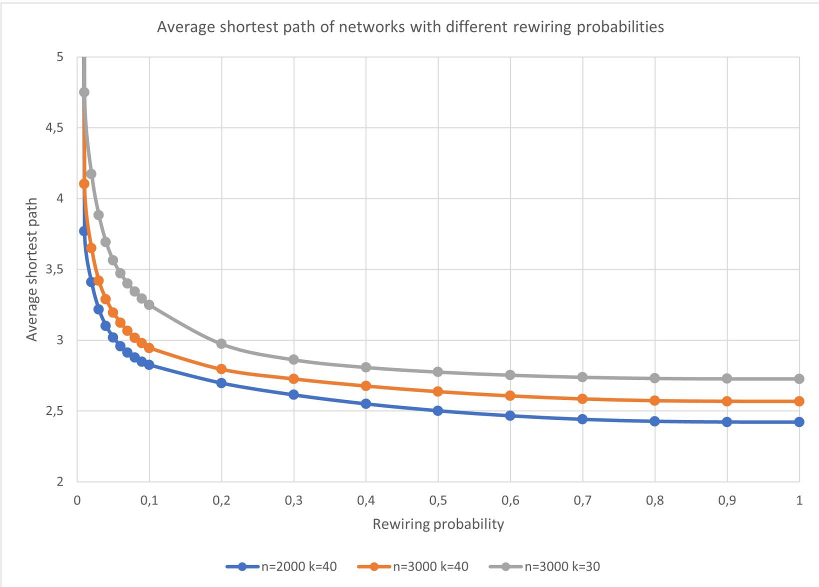

Figure 5.5. Average Shortest Path of Small World Network With Different Rewiring Probabilities.

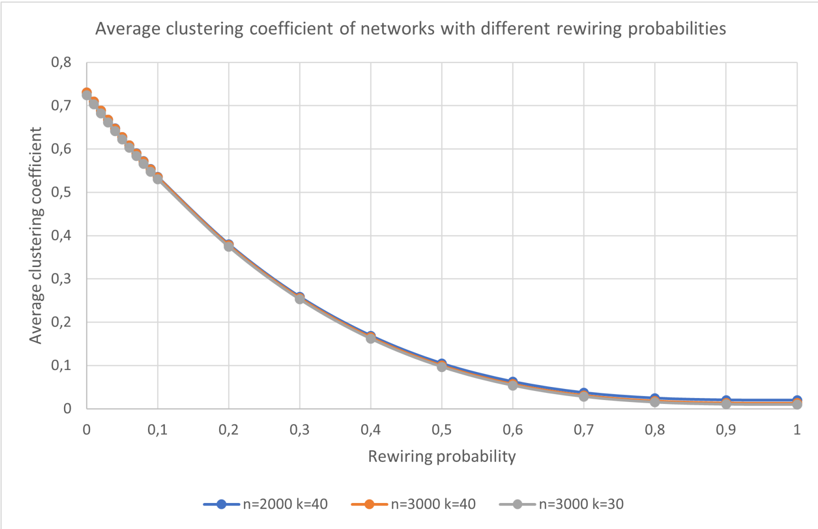

Figure 5.6. Average Clustering Coefficient of Small World Network With Different Rewiring Probabilities.

5.1.1. The R+ LCD- experiment

In this experiment, 25 replications have been carried out, then the metric results, the average shortest path and the average clustering coefficient, are averaged for each snapshot. Although the initial network class and the link addition mechanism are random, the clustering coefficient of this experiment is relatively high compared to the result of the random network, which demonstrates the importance of the least common degree mechanism. According to this rule, a link is deleted based on the number of mutual neighbors as mentioned in 4.2.2.3, which leads to an increase in the average clustering coefficient. Thus, the structure of the evolving network starts to have small world network characteristics.

Eight snapshots of evolving networks are analyzed for this experiment. The evolving network at $t_{3.00}$ has hardly any isolated nodes. In this snapshot, it is seen that there is a minor change in the average shortest path as seen in Figure 5.7 compared to the initial network. However, the average clustering coefficient, as shown in Figure 5.8, has a moderate increase over time.

The peak value, the number of nodes having 40 degrees, has decreased in the degree distribution of the network at $t_{3.00}$ and tails have been long compared to the one in the initial network. The degree distribution of the modified network is bell-shaped curve without skewness.

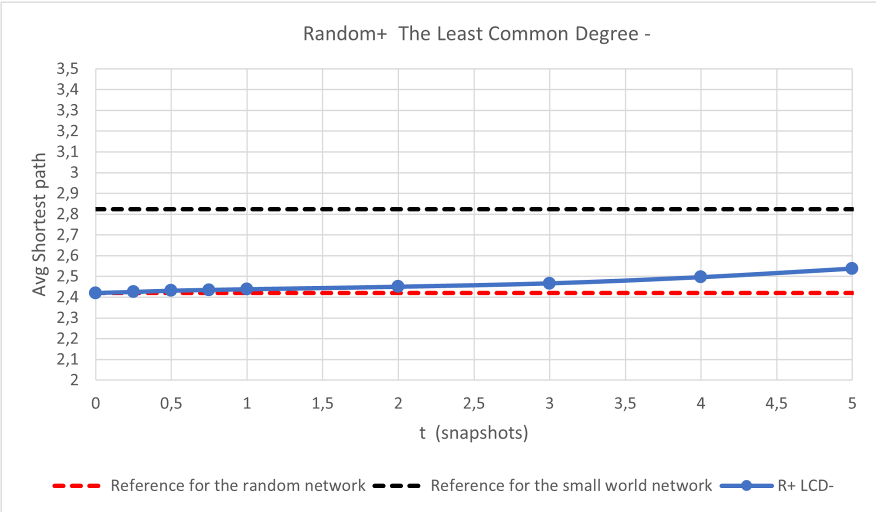

Figure 5.7. The Average Shortest Path of the Random Network Modified by R+ LCD-.

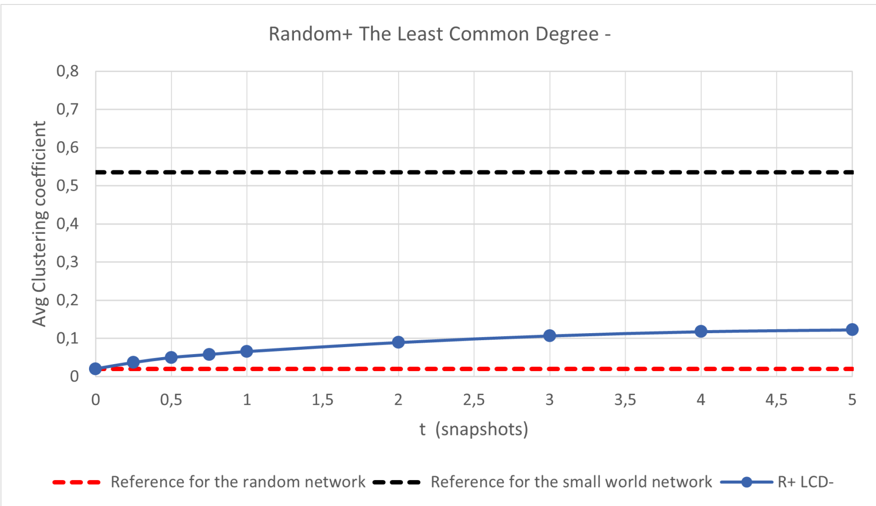

Figure 5.8. The Average Clustering Coefficient of the Random Network Modified by R+ LCD-.

In order to understand if the network at $t_{3.00}$ converges to the small world network, the average clustering coefficient and average shortest path of that network are compared to that of networks created with different rewiring probabilities in Figure 5.5 and Figure 5.6. It is seen that the modified network is similar to the network whose rewiring probability is almost 50%. Therefore, it can be concluded that the modified network is similar to the one having less random characteristics than the initial network at t_0 thanks to this mechanism. Additionally, the variances of the experiment results are quite low. The maximum variances of the average shortest path and the average clustering coefficient are 0.00012 and 0.00003 considering all snapshots, respectively.

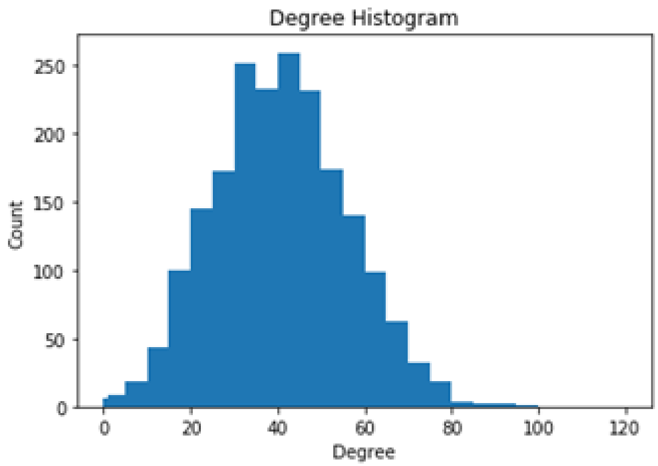

Figure 5.9. The Degree Distribution of the Network at $t_{3.00}$ Modified by R+ LCD-.

Since the average clustering coefficient does not reach a steady state at $t_{3.00}$, the network is modified until the snapshot at $t_{5.00}$ is found. The degree distribution of that network has been slightly flattened, which is seen in Figure 5.10. The degree distribution is also found by changing the scale of the vertical axis to check if the structure of the degree distribution has changed visually and it is seen that the flat degree distribution is still observed. As seen in Figure 5.10, the number of nodes having different degrees is almost the same.

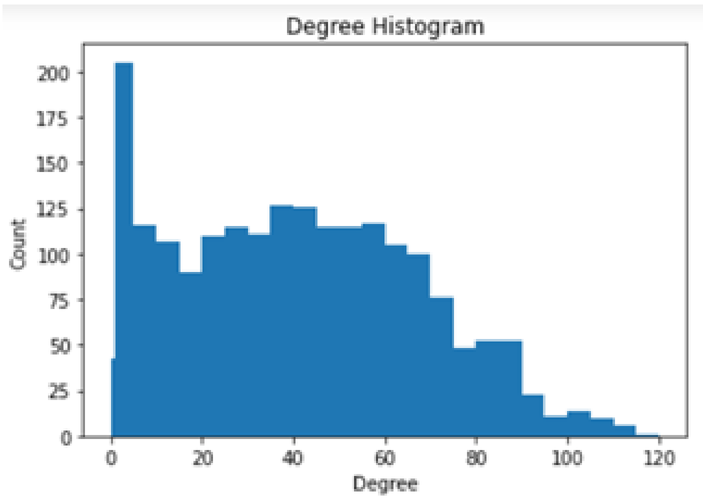

Figure 5.10. The Degree Distribution of the Network at $t_{5.00}$ Modified by R+ LCD-.

5.1.2. The PA+ R- experiment

Eight snapshots are analyzed after 25 replications are completed. In the preferential attachment mechanism, after the first node is selected randomly, the second one is chosen proportional to the square of the degree each potential node has. Removing a link is completed in a random way.

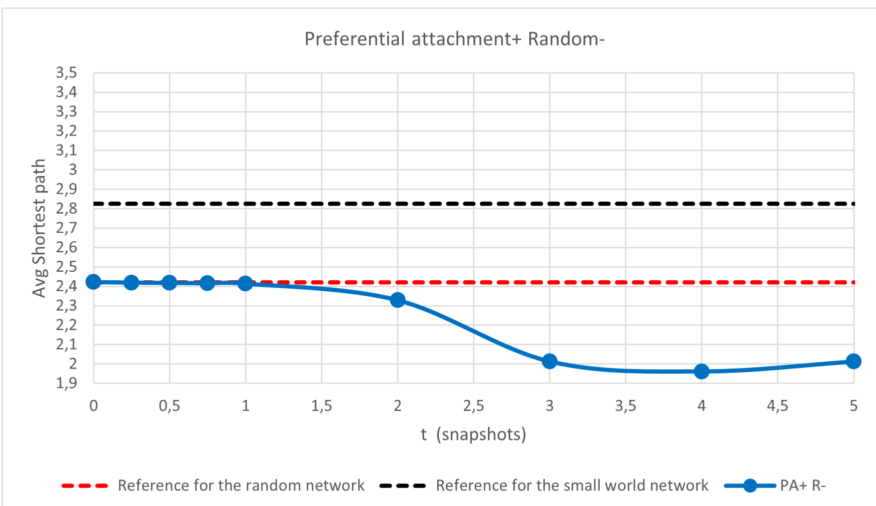

Figure 5.11. The Average Shortest Path of the Random Network Modified by PA+ R-.

As seen, the average shortest path considerably decreases, and the average clustering coefficient remarkably increases after the network at $t_{2.00}$ is observed. To understand this considerable change, nodes are ordered based on their degrees in a descending way and the first five nodes and the first 20 nodes are chosen for each snapshot, separately. It is checked whether the first five nodes of each snapshot are connected to the first 20 nodes. It is seen that top five nodes are linked with two of the 20 nodes on average in the evolving network at $t_{1.00}$.

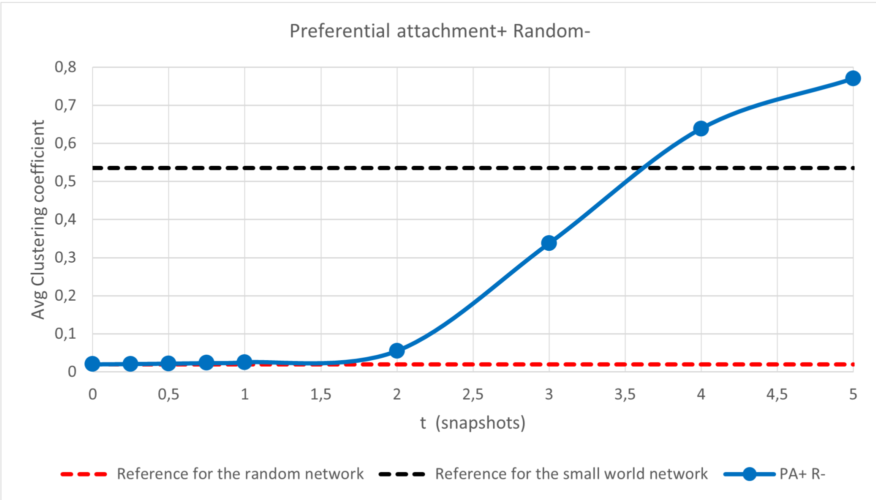

Figure 5.12. The Average Clustering Coefficient of the Random Network Modified by PA+ R-.

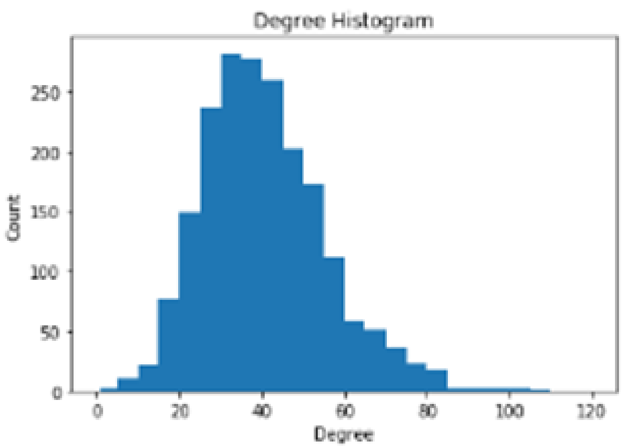

Figure 5.13. The Degree Distribution of the Network at $t_{1.00}$ Modified by PA+ R-.

Top five nodes of the evolving network at $t_{2.00}$ are linked with six of 20 nodes on average. After that, the first five nodes interact with most of 20 nodes. In the network at $t_{3.00}$, these five nodes have interacted with 18 nodes and each other. Nodes have high probability of linking with the ones having high degrees due to the preferential attachment mechanism, which creates the presence of hubs. Thus, a node can reach another node in fewer steps, which provides a decrease in the average shortest path. As explained, the first five nodes, which are the ones having very high degrees have linked with each other as the number of snapshots proceeds. It contributes to the increase in average clustering coefficient. As an additional analysis to understand huge increase in average clustering coefficient and decrease in average shortest path, 10 nodes whose degrees are not considerably high but middle level (between 15 and 55) are determined. For the networks at $t_{1.00}$, $t_{2.00}$, $t_{3.00}$, $t_{4.00}$, $t_{5.00}$, the neighbors of these 10 nodes and the degrees of their neighbors are analyzed. For the network at $t_{1.00}$, the maximum degree of neighbors of 10 nodes is 122 and there are two neighbors whose degrees are over 100. The maximum degree of neighbors of 10 nodes is 634 and 30 neighbors whose degrees are over 100 in the network $t_{2.00}$. Since the network at $t_{3.00}$, 30 neighbors of 10 nodes have over 1000 degrees and their maximum degree is 1725. In the network at $t_{4.00}$, 80 neighbors of 10 nodes have over 1000 degrees and the results of the network at $t_{5.00}$ is similar to those of its previous network. It is concluded that nodes having middle level degrees have tendency to link with central nodes as the number of snapshots proceeds. It explains the changes in average shortest path and average clustering coefficient. Additionally, the maximum variances of the average shortest path and the average clustering coefficient are 0.04 and 0.003 respectively when all snapshot results are considered.

At the end of the analysis, approximately 40 nodes are isolated due to these mechanisms. When the degree distribution of the final network is observed, it is seen that the peak value has decreased compared to the random network. The distribution is positively skewed as seen in Figure 5.14 and there are few nodes having high degrees.

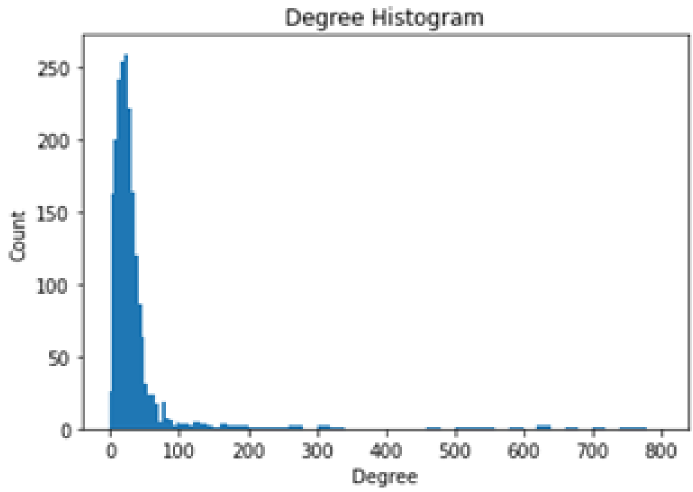

Figure 5.14. The Degree Distribution of the Network at $t_{5.00}$ Modified by PA+ R-.

5.1.3. The PA+ LCD- experiment

Eight snapshots are analyzed to identify how the fundamental structural properties have changed after 25 replications are completed.

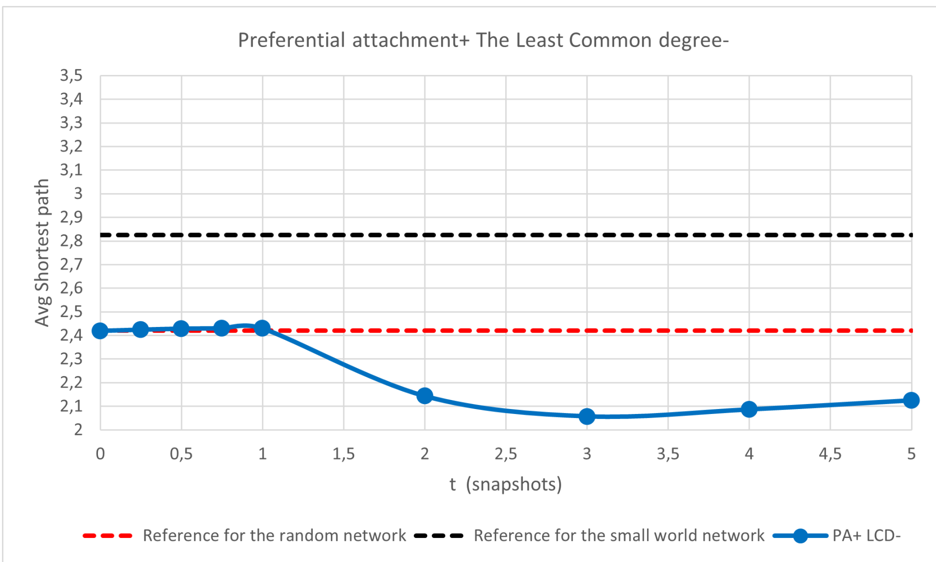

Figure 5.15. The Average Shortest Path of the Random Network Modified by PA+ LCD-.

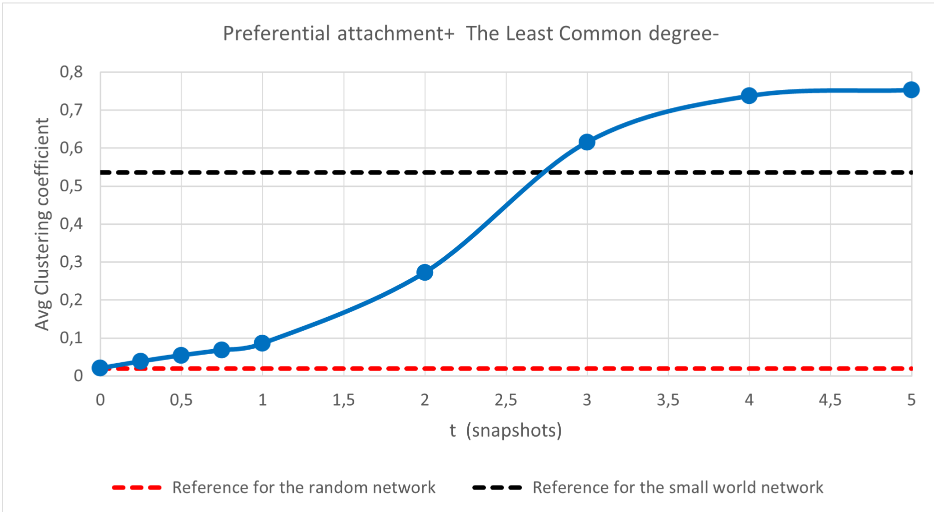

Figure 5.16. The Average Clustering Coefficient of the Random Network Modified by PA+ LCD-.

As seen in Figure 5.15 and Figure 5.16, these measures have considerably changed especially after the evolving network at $t_{2.00}$. To identify the reason underlying this change, the same study done in the previous experiment is carried out. Namely, all nodes are sorted in descending order based on their degrees. Then the first five nodes and 20 nodes are filtered for each snapshot. It is analyzed whether the first five nodes are connected to 20 nodes over time.

It is seen that top five nodes link with five of 20 nodes on average in the evolving network at $t_{1.00}$. The number of isolated nodes is approximately four. Its degree distribution is seen in Figure 5.17.

The top five nodes have interacted with 18 of the 20 nodes and the top three nodes in the network at $t_{2.00}$. The degree distribution of that network (as seen in Figure 5.18) starts being skewed to the right. There are nearly 100 isolated nodes.

The top five nodes are connected to almost each of 20 nodes in the network at $t_{3.00}$. Thus, each node can reach the other one in a few steps, which decreases the average shortest path. Its degree distribution is seen in Figure 5.19.

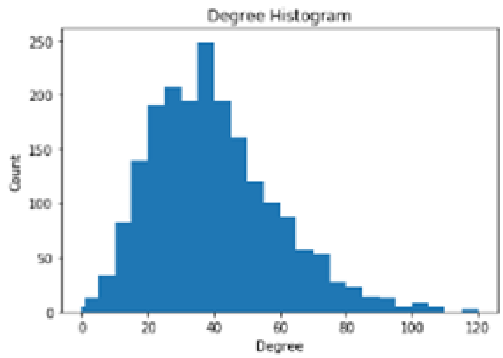

Figure 5.17. The Degree Distribution of the Network at $t_{1.00}$ Modified by PA+ LCD-.

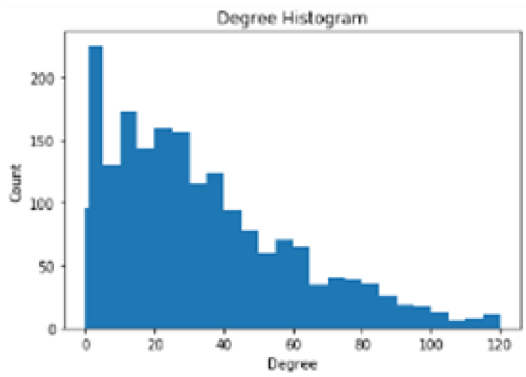

Figure 5.18. The Degree Distribution of the Network at $t_{2.00}$ Modified by PA+ LCD-.

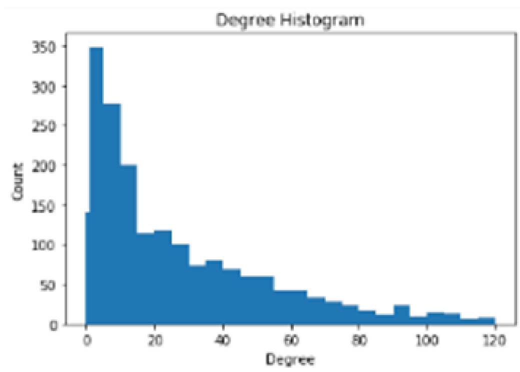

Figure 5.19. The Degree Distribution of the Network at $t_{3.00}$ Modified by PA+ LCD-.

All nodes selected are interacted with each other in the evolving network at $t_{5.00}$. Also, 10 nodes having middle level degrees (between 15 and 55) are selected and the degrees of their neighbors are analyzed to understand their tendencies to be linked over time. In the network at $t_{1.00}$, 10 nodes selected have 283 neighbors in total. There are seven neighbors whose degrees are over 100 and the maximum degree is 174. In the network at $t_{2.00}$, 85 neighbors of 10 nodes have at least 100 degrees and 11 neighbors have more than 1000 degrees. The number of neighbors whose degrees are more than 100 and 1000 are 143 and 37 in the network $t_{3.00}$. Then, each of the neighbors of 10 nodes has at least 100 nodes in the network $t_{5.00}$.

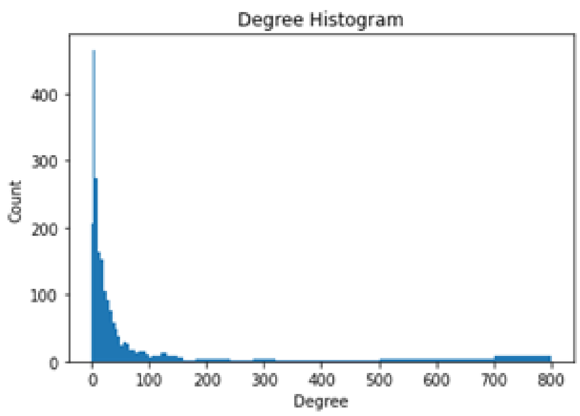

Figure 5.20. The Degree Distribution of the Network at $t_{5.00}$ Modified by PA+ LCD-.

The nodes having higher degrees tend to have interacted with one another and the nodes whose degrees are not high tend to link with central nodes, which makes average clustering coefficient increase and average shortest path decrease. The degree distribution of the network at $t_{5.00}$ is skewed right. Compared to the degree distribution of the initial network at t_0 shown in Figure 5.1, the peak value has decreased. Additionally, considering all snapshots, the maximum variances of the average shortest path and the average clustering coefficient are 0.008 and 0.003, respectively.

At the end of the analysis, the network has almost 200 isolated nodes. Due to the preferential attachment mechanism, a node with low degrees tends to link less, which causes the increase in the isolated nodes.

Additionally, the least common degree mechanism deletes links between nodes having the least number of common neighbors. It is concluded that the average shortest path and average clustering coefficient of the network has tendency to decrease dramatically and increase considerably over snapshots, respectively since nodes with very high degrees tend to link with each other, and nodes with medium level degrees have tendency to link with more central nodes. Therefore, the degree distribution of the network is positively skewed over time.

5.1.4. The TR1+ LCD- experiment

The triadic closure mechanism tends to create links between the first node selected and its second level neighbors, which provides an increase in the clustering coefficient. The second node is selected from a list that each potential node is found once. For instance, if node x has two neighbors (node 1 and node 2), the neighbors of node 1 and node 2 are found. It is assumed that node 1 has two neighbors (node 3 and node 4) and node 2 has a neighbor, node 4. Assuming that node x is not interacted with node 3 and node 4, it is seen that node 3 and node 4 are found only once in the potential list. The second node is selected from that list. Similarly, the link deletion mechanism deletes links between nodes having the least common neighbors, which also makes the clustering coefficient increase. It is seen that the triadic closure and the least common degree mechanisms provide an increase in the average clustering coefficient.

The average shortest path of the network at $t_{2.00}$ remains nearly the same as that of the random network. Then, this measure starts declining as seen in Figure 5.21. The decrease can be caused by both link addition and link deletion mechanism directing nodes to link with nodes having common neighbors.

As seen in Figure 5.22, the average clustering coefficient has a tendency to increase until the network at $t_{2.00}$ and then reach a steady state. Since the network at $t_{4.00}$ has started to lose small world network characteristics in terms of degree distribution shown in Figure 5.24, the network modification process is interrupted.

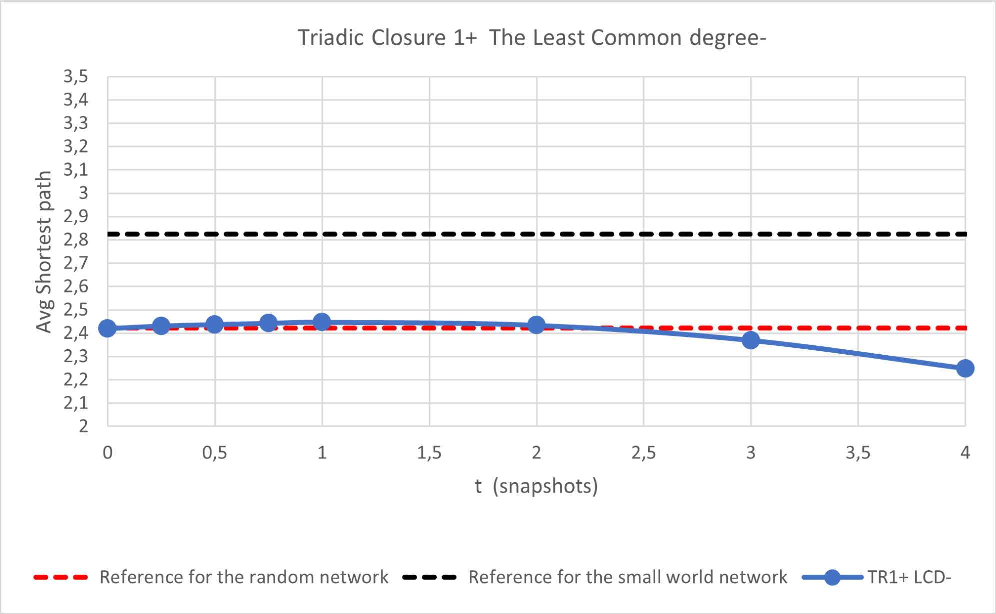

Figure 5.21. The Average Shortest Path of the Random Network Modified by TR1+ LCD-.

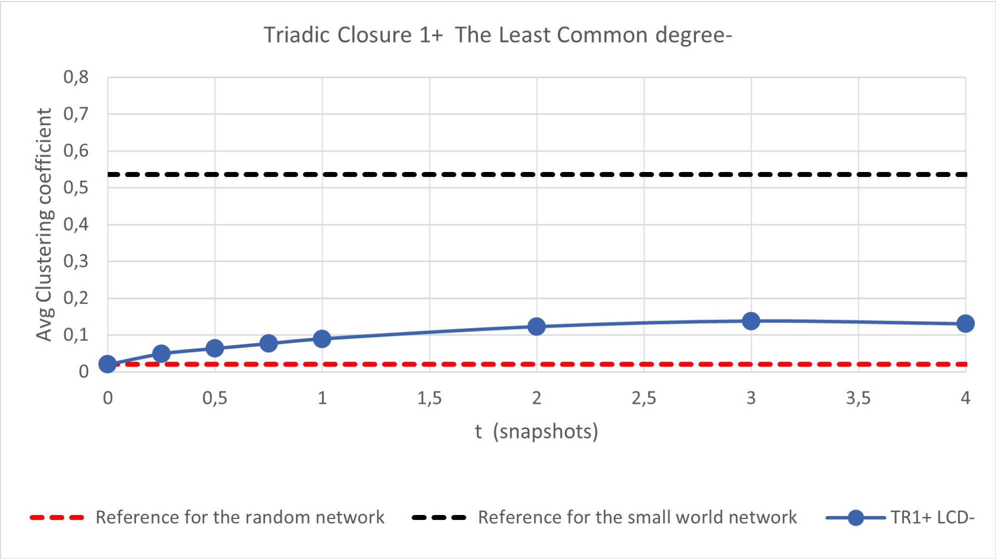

Figure 5.22. The Average Clustering Coefficient of the Random Network Modified by TR1+ LCD-.

In order to identify if the evolving network at $t_{2.00}$ converges to small world network, the average clustering coefficient of that network is compared to that of networks constructed with different rewiring probabilities in Figure 5.6. It is seen that the network at $t_{2.00}$ is similar to the network whose rewiring probability is between 40% and 50%. Compared to the initial network, the tails in the degree distribution of the network at $t_{2.00}$ have been long and the peak has gone down. It seems symmetrically distributed and hardly ever skewness shown in Figure 5.23.

The network at $t_{2.00}$ has more than 50 isolated nodes. As mentioned, since the average clustering coefficient of the network at $t_{2.00}$ does not reach steady state, the structural properties of the evolving network at $t_{3.00}$ and $t_{4.00}$ are also evaluated. The degree distribution of the network at $t_{4.00}$ is flattened, as seen in Figure 5.24. When the vertical axis is rescaled to observe if the distribution is a bell-shaped curve, it is seen the flat degree distribution does not change. When all results of the snapshots are considered, maximum variances of the average shortest path and the average clustering coefficient are 0.00007 and 0.00002.

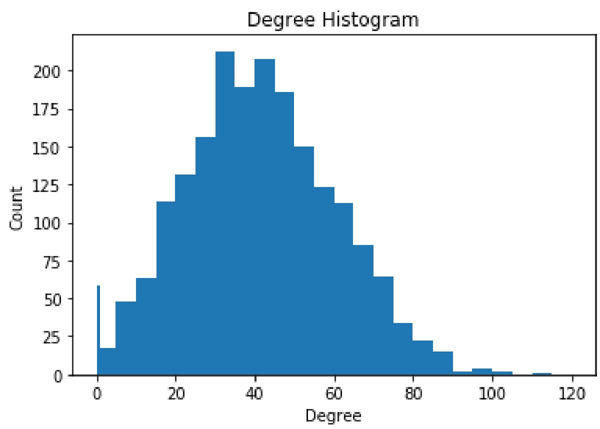

Figure 5.23. The Degree Distribution of the Network at $t_{2.00}$ Modified by TR1+ LCD-.

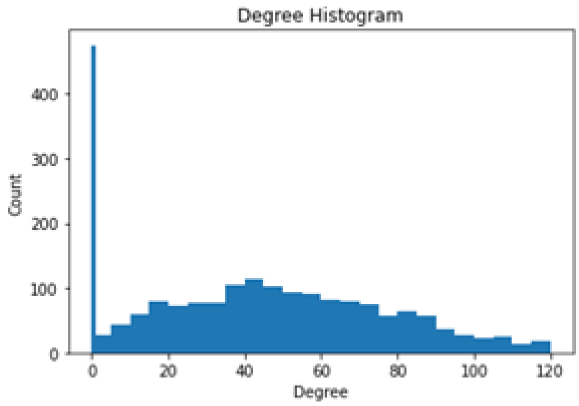

Figure 5.24. The Degree Distribution of the Network at $t_{4.00}$ Modified by TR1+ LCD-.

5.1.5. The TR2+ LCD- experiment

In the triadic closure 2 mechanism, after the second level neighbors of the first node are found, they are assigned to a list. The second node is selected from that list that each potential node can be found more than once. For instance, if a node x has two neighbors (node 1 and node 2), the neighbors of node 1 and node 2 are found. It is assumed that node 1 has 2 neighbors (node 3 and node 4) and node 2 has a neighbor, node 4. Assuming that node x is not interacted with node 3 and node 4, it is seen that the potential list consists of node 3, node 4 and node 4. The second node is selected from that list.

As seen in Figure 5.25, the average shortest path does not change much until the evolving network at $t_{2.00}$. After that, this metric considerably decreases compared to the initial network, which may be caused by both link addition and link deletion mechanism directing nodes to link with nodes having common neighbors.

The degree distribution of the network at $t_{2.00}$ has slightly started to skew to the right. It starts to have a long tail to the right, therefore it is a positively skewed distribution. Its average clustering coefficient has a tendency to increase.

Several networks with different rewiring probabilities are created as seen in Figure 5.6. It is seen that the network modified at $t_{2.00}$ is similar to the network whose rewiring probability is nearly 0.4. It is concluded that the network at $t_{2.00}$ shows similarity with the one having less random characteristics than the initial network at t_0 thanks to this mechanism.

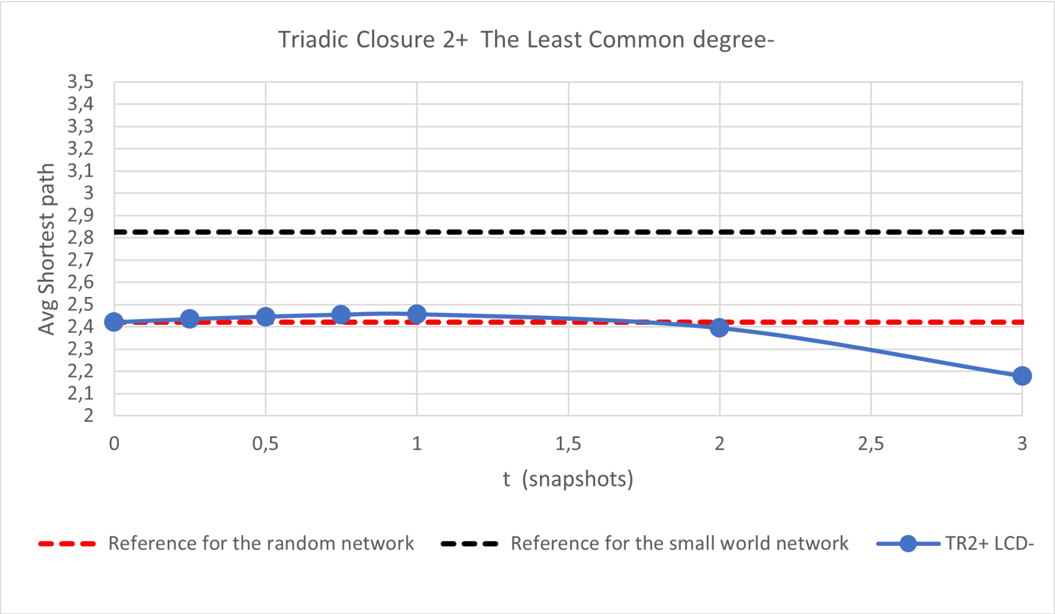

Figure 5.25. The Average Shortest Path of the Random Network Modified by TR2+ LCD-.

The clustering coefficient of the network at $t_{2.00}$ does not reach a steady state but continues increasing as seen in Figure 5.26. Therefore, the number of snapshots evaluated is increased. It is seen that average clustering coefficient has reached a steady state. The degree distribution of the network at $t_{3.00}$ is flattened as seen in Figure 5.28. It is seen that the number of nodes having different degrees is almost the same. Addionally, maximum variances are 0.0003 for the average shortest path and 0.00002 for the average clustering coefficient over snaphots.

In conclusion, this network modification combination provides a certain amount of change on structural properties of a random network. The evolving network at $t_{2.00}$ has small world network characteristics.

However, it is seen that those characteristics have been changed and flat degree distribution has emerged when the evolving network at $t_{3.00}$ is observed. By excluding the isolated nodes on the graph, the degree distribution is rescaled to observe if there is any change in the shape. It is seen that there is almost no change, and the flat degree distribution is still observed.

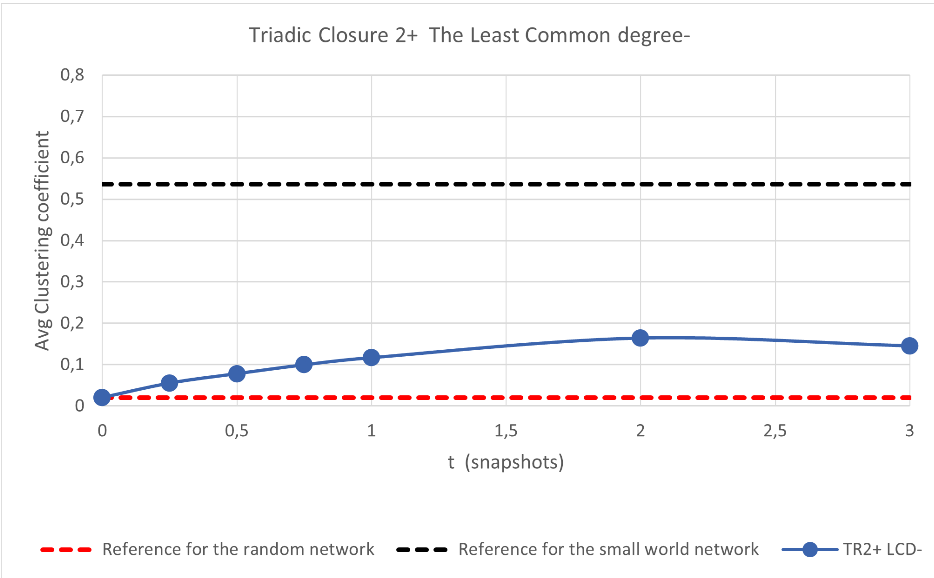

Figure 5.26. The Average Clustering Coefficient of the Random Network Modified by TR2+ LCD-.

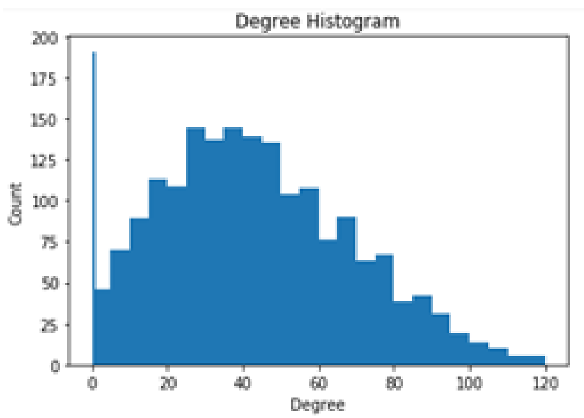

Figure 5.27. The Degree Distribution of the Network at $t_{2.00}$ Modified by TR2+ LCD-.

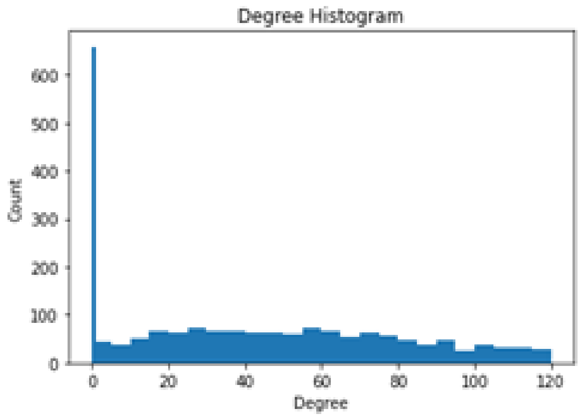

Figure 5.28. The Degree Distribution of the Network at $t_{3.00}$ Modified by TR2+ LCD-.

5.1.6. The HB1+ LCD- experiment

Hybrid 1 mechanism consists of random and triadic closure 1 link addition mechanisms. After the threshold level determined, random number is generated and one of the mechanism combinations is selected. The threshold level is 0.4 for all replications done. Namely, the probabilities that triadic closure 1 and random are selected for link addition are 60% and 40%, respectively. As a result of 25 replications carried out, it is seen that the average clustering coefficient has increased.

The average shortest path (as seen in Figure 5.29) of the evolving network at $t_{2.00}$ is almost the same as that of the initial network. In that snapshot, the average clustering coefficient does not reach a steady state but continues increasing. Therefore, the number of snapshots is increased until that metric of the network reaches the steady state. From t_0 to $t_{5.00}$, during network modification process, the average shortest path does not change much. The average clustering coefficient has a tendency to increase at first and then has reached a steady state as seen in Figure 5.30.

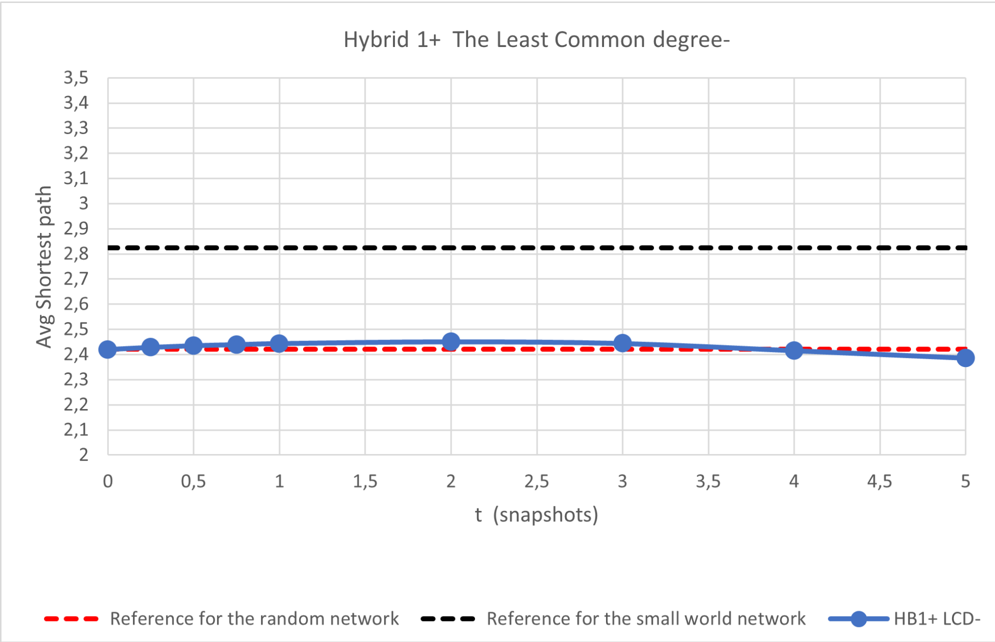

Figure 5.29. The Average Shortest Path of the Random Network Modified by HB1+ LCD-.

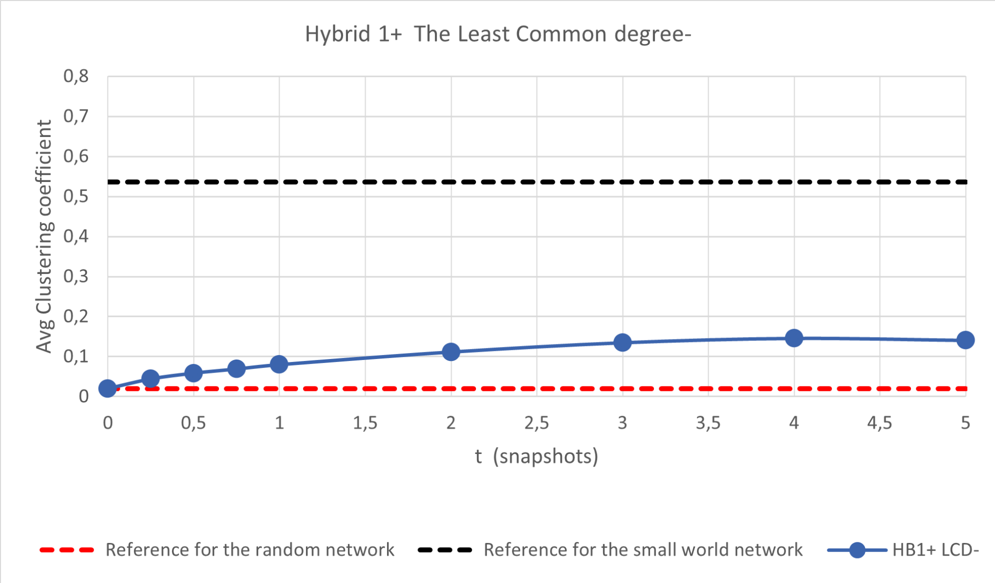

Figure 5.30. The Average Clustering Coefficient of the Random Network Modified by HB1+ LCD-.

The evolving network at $t_{2.00}$ has 20 isolated nodes. Its degree distribution has two tails symmetrically. The mode and mean value are equal, and the graph is bell-curved shape. The modified network at $t_{2.00}$ has small world properties because there is a moderate increase in the average clustering coefficient and almost no change in the average shortest path compared to those characteristics of the initial random network. Therefore, networks with different rewiring probabilities are generated as demonstrated in Figure 5.6. The clustering coefficient of the network at $t_{2.00}$ is similar to that of the network whose rewiring probability is 0.5. However, related rewiring probability is very high in order that the modified network is a small world network.

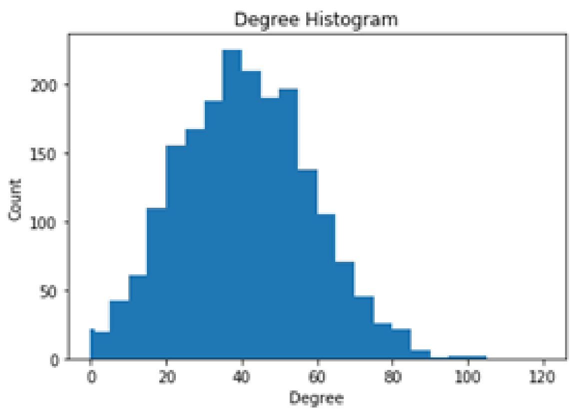

Figure 5.31. The Degree Distribution of the Network at $t_{2.00}$ Modified by HB1+ LCD-.

Since the average clustering coefficient of the network until $t_{2.00}$ tends to increase, this experiment continues until this measure reaches a steady state. The degree distribution of the network at $t_{3.00}$ starts changing and is skewed to the right as seen in Figure 5.32. However, its average clustering coefficient still increases.

As seen in Figure 5.30, the average clustering coefficient of the network at $t_{5.00}$ remains unchanged. However, its degree distribution is flattened. After rescaling of vertical axis, it is observed it is still similar to flat degree distribution. Additionally, maximum variances are 0.0001 for the average shortest path and 0.00001 for the average clustering coefficient when all snapshots are considered.

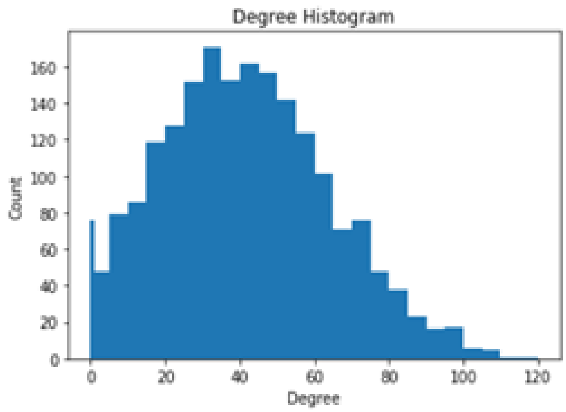

Figure 5.32. The Degree Distribution of the Network at $t_{3.00}$ Modified by HB1+ LCD-.

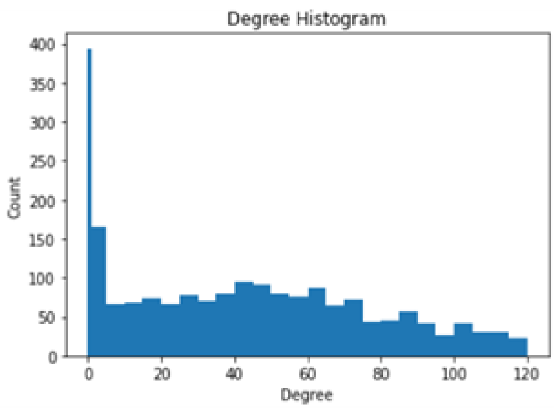

Figure 5.33. The Degree Distribution of the Network at $t_{5.00}$ Modified by HB1+ LCD-.

As a result, the evolving network at $t_{5.00}$ starts to lose small world network characteristics in terms of the degree distribution. Although the average shortest path and the average clustering coefficient of that network do not change much compared to those of the network at $t_{2.00}$, its degree distribution is flattened.

5.1.7. The HB2+ LCD- experiment

When the hybrid 2 link addition, a combination of random and triadic closure 2 link addition mechanisms, and the least common degree link deletion mechanisms are implemented on the random network, it is seen that the average shortest path remains almost the same for nearly all snapshots. The threshold value is selected as 0.4 for all 25 replications. The probabilities that triadic closure 2 and random link addition mechanisms are selected are 60% and 40%, respectively.

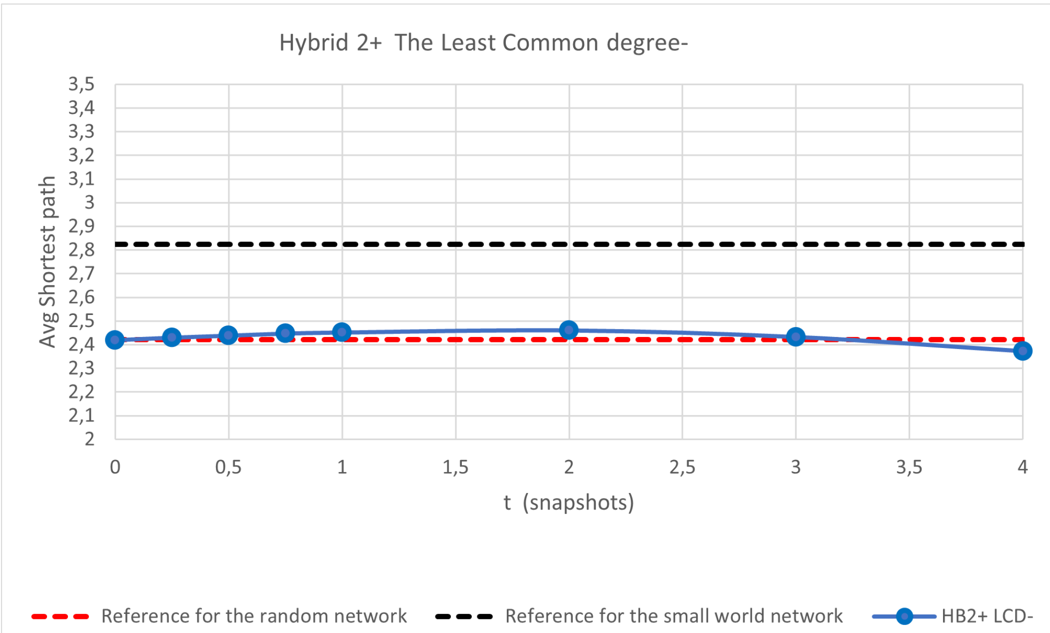

Figure 5.34. The Average Shortest Path of the Random Network Modified by HB2+ LCD-.

That network at $t_{2.00}$ has small world network properties since its average clustering coefficient of is similar to that of the network whose rewiring probability is almost 0.4. A few networks with different rewiring probabilities are generated as shown in Figure 5.6. Compared to the initial network, there is almost no change in the average shortest path.

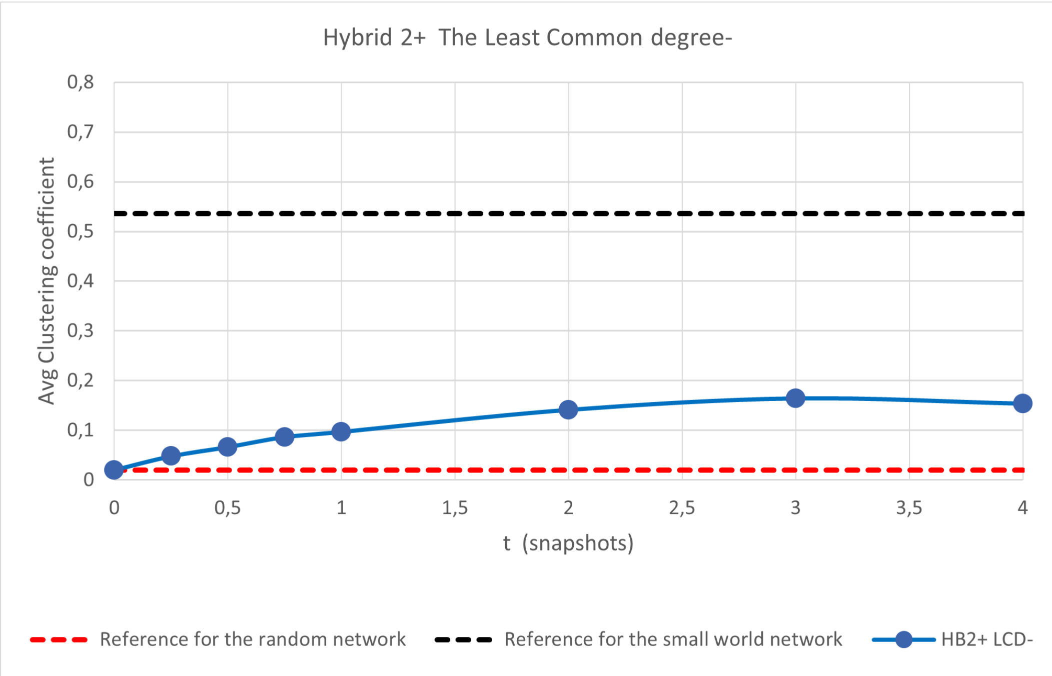

Figure 5.35. The Average Clustering Coefficient of the Random Network Modified by HB2+ LCD-.

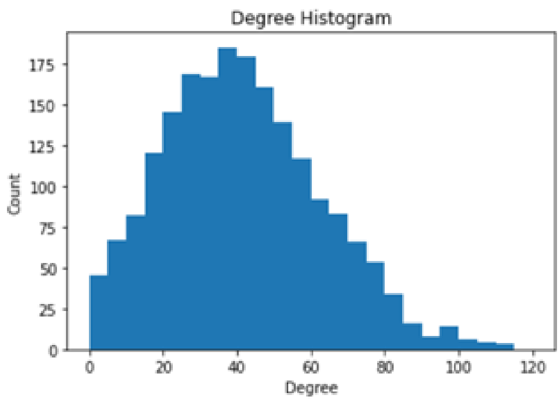

Figure 5.36. The Degree Distribution of the Network at $t_{2.00}$ Modified by HB2+ LCD-.

As seen in Figure 5.35, the average clustering coefficient of the network at $t_{2.00}$ does not reach a steady state. Therefore, the number of snapshots is increased until the clustering coefficient of the network has a stable level.

That measure of the network at $t_{4.00}$ remains unchanged, but its degree distribution is flattened as demonstrated in Figure 5.37. Then, the size of the vertical axis is changed, and it is seen the flat degree distribution is remained. It is observed that the initial random network has changed its characteristics. The final network reaches a steady state in terms of the average shortest path and average clustering coefficient with a flat degree distribution. Additionally, maximum variances of the average shortest path and the average clustering coefficient are 0.0003 and 0.00005 when all snaphot results are considered, respectively.

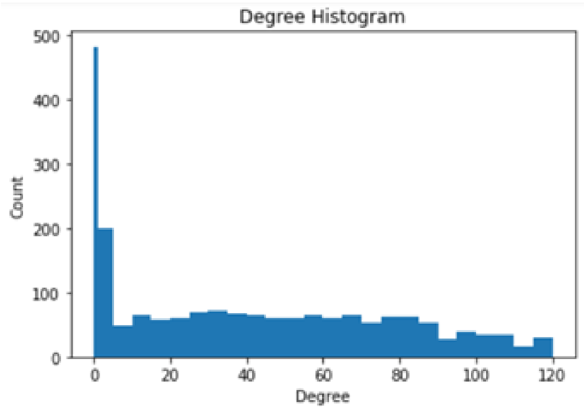

Figure 5.37. The Degree Distribution of the Network at $t_{4.00}$ Modified by HB2+ LCD-.

5.2. Experiments on Small World Networks

18 experiments are carried out on small world networks. Table 5.5 shows the related experiments, and the number of replications in brackets. Link addition and link deletion mechanisms are abbreviated as + and -, respectively. The modified networks converging to random network will not be discussed further. The other experiments marked as bold will be explained comprehensively.

As an example of experiments which will not be discussed in detail, the experiment result of HB1+ R- is demonstrated below.

As shown in Figure 5.38 and Figure 5.39, the average shortest path and average clustering coefficient of the evolving network at $t_{2.00}$ converge to those measures of the random network. Since the modified network converges to random network, it will not be analyzed in detail.

Table 5.5. Experiments on the Small World Network

Exp. No	**Network modification mechanisms**	**Exp. No**	**Network modification mechanisms**
1	R+ R- (6)	10	TR2+ R- (6)
2	R+ D- (6)	11	TR2+ D- (6)
3	**R+ LCD- (2 sub-experiments with 50 replications)**	12	**TR2+ LCD- (25)**
4	**PA+ R- (25)**	13	HB1+ R- (6)
5	PA+ D- (6)	14	HB1+ D- (6)
6	**PA+ LCD- (25)**	15	**HB1+ LCD- (25 replications for each of seven sub-experiments)**
7	TR1+ R- (6)	16	HB2+ R- (6)
8	TR1+ D- (6)	17	HB2+ D- (6)
9	**TR1+ LCD- (25)**	18	**HB2+ LCD- (25)**

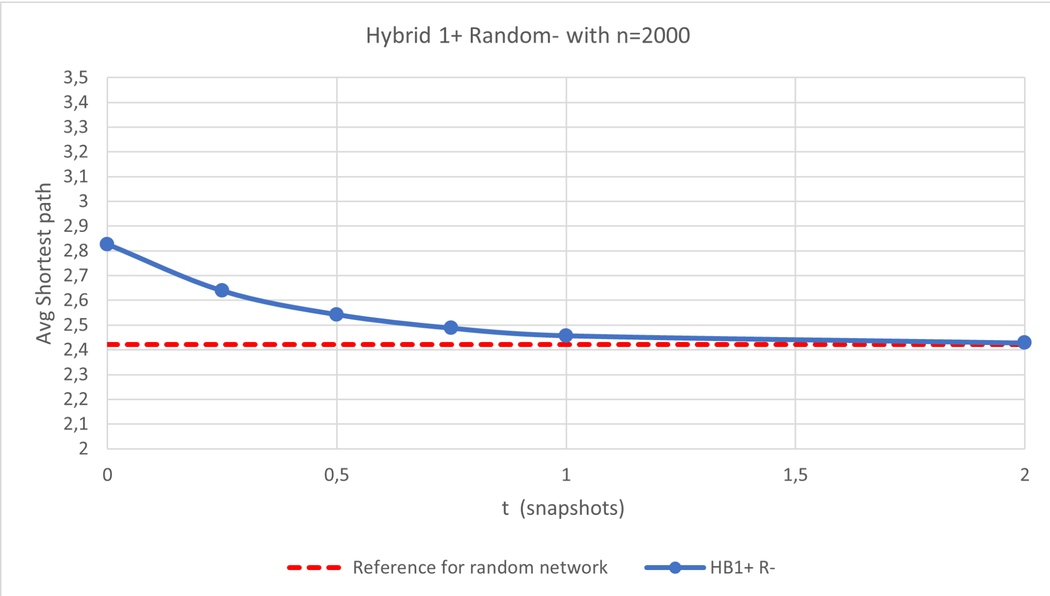

Figure 5.38. The Average Shortest Path of the Small World Network Modified by HB1+ R-.

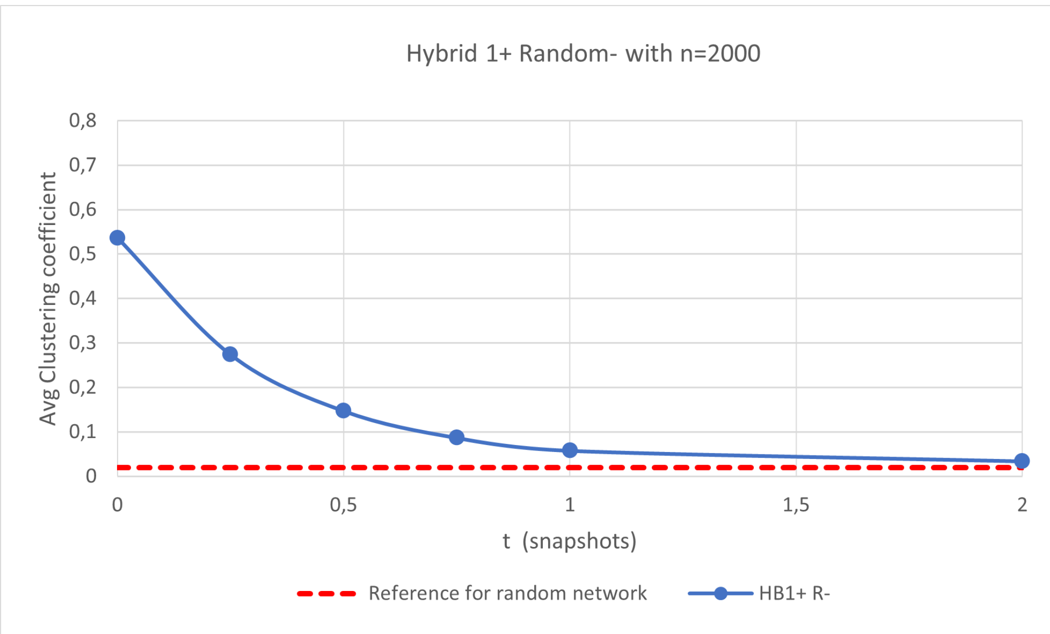

Figure 5.39. The Average Clustering Coefficient of the Small World Network Modified by HB1+ R-.

5.2.1. The R+ LCD- experiment

There are two sub-experiments that different parameters are taken into account as shown in Table 5.6. In total, 50 replications are carried out within the scope of these sub-experiments. The network is proceeded to be modified until the average shortest path and the average clustering coefficient have a steady state. At first, the first sub-experiment is carried out. The results demonstrate that the small world network characteristics are preserved. Then, the second sub-experiment is conducted in order to check if there is a special case for those parameters.

Table 5.6. Parameters of Two Sub-experiments Modified by R+ CD- in the SMW.

Parameters	**First sub-experiment**	**Second sub-experiment**
Number of nodes	2000	3000
Average number of edges	40	30
Rewiring probability	0.1	0.1

For the first sub-experiment, the average shortest path and the average clustering coefficient are seen in Figure 5.40 and Figure 5.41.

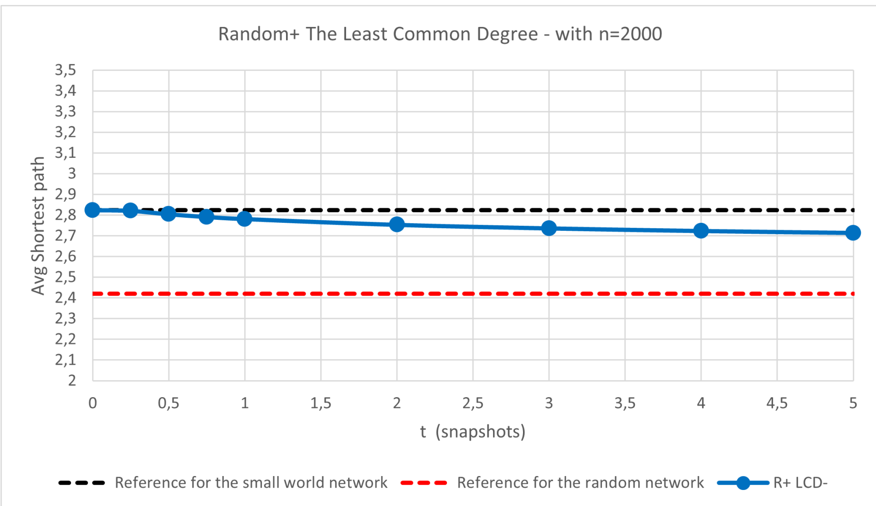

Figure 5.40. The Average Shortest Path of the Small World Network Modified by R+ LCD- with N=2000.

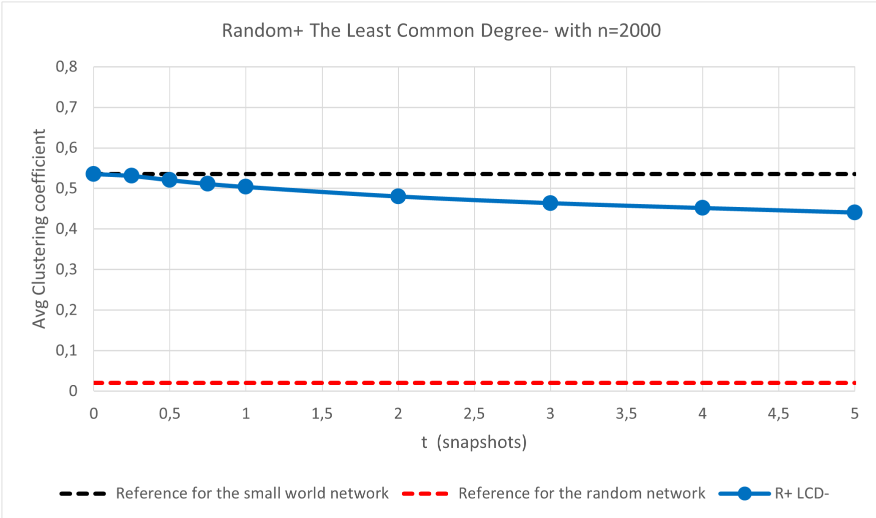

Figure 5.41. The Average Clustering Coefficient of the Small World Network Modified by R+ LCD- with N=2000.

The evolving network at $t_{5.00}$ has a decrease in the average shortest path and average clustering coefficient compared to those measures of the initial network. However, this network has still small world network properties since these metrics do not decrease dramatically. To identify how that network changes, several small world networks are generated with different rewiring probabilities as exhibited in Table 5.4 and compared to those measures. It is deduced that the network at $t_{5.00}$ is similar to the network whose rewiring probability is almost 15% in terms of the clustering coefficient and the average shortest path. Maksimum variances are 0.00001 for average shortest path and 0.00006 for average clustering coefficient when all snaphot results are taken into account.

In the second sub-experiment, the network size and the number of neighbors have been changed. After 25 replications done, the averaged results of these metrics are demonstrated below.

Similarly, the average shortest path and the average clustering coefficient of the network at $t_{5.00}$ decrease compared to those of the initial network.

Therefore, these values are compared to the values of networks generated with different rewiring probabilities in Figure 5.6. It is seen that the network modified is similar to the network whose rewiring probability is 20%.

Compared to the degree distribution of the initial network shown in Figure 5.2, the degree distribution in the network at $t_{5.00}$ is spread in wider range, tails have been long and the peak value has decreased for both sub-experiments. They are symmetrically distributed and bell-shaped curve. As an example, the degree distribution of the first experiment is demonstrated in Figure 5.44. The evolving network at $t_{5.00}$ does not have any isolated nodes for both of the sub-experiments. The parameters used in the second sub-experiment will be also used in the HB1+ LCD- experiment shown in 5.2.6. Comparisons between them will be mentioned in that part.

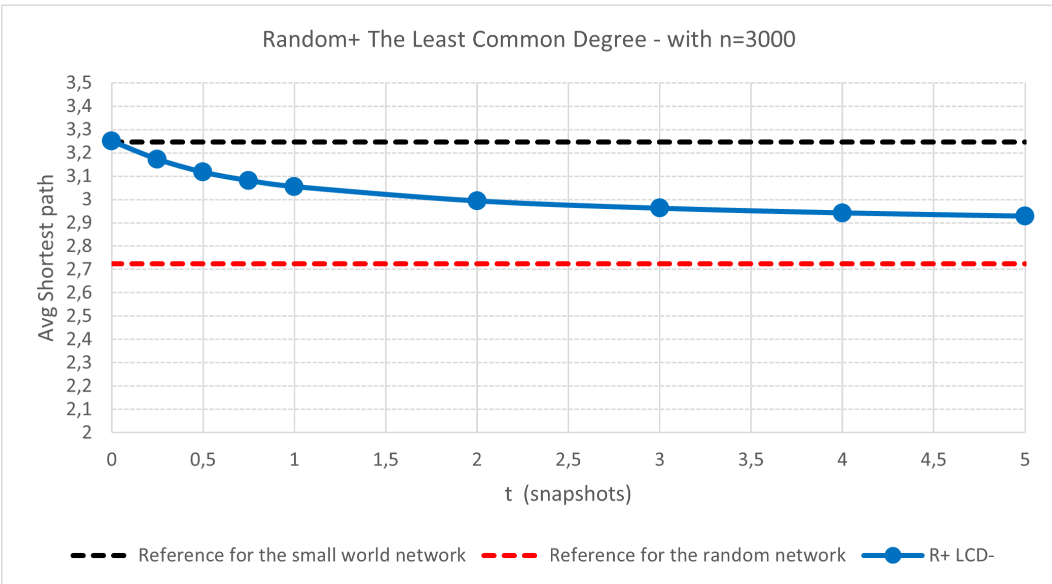

Figure 5.42. The Average Shortest Path of the Small World Network Modified by R+ LCD- with N=3000.

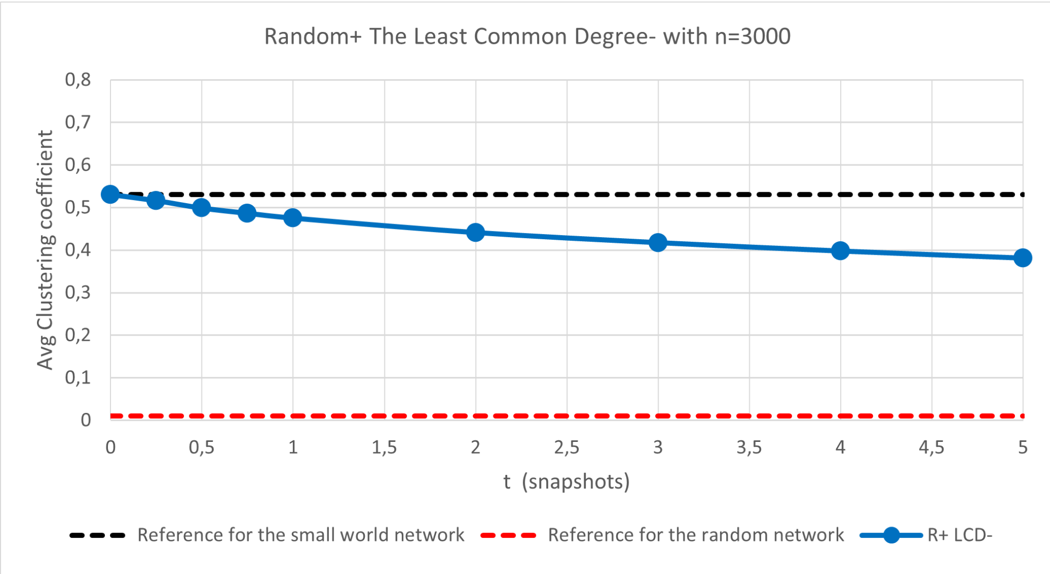

Figure 5.43. The Average Clustering Coefficient of the Small World Network Modified by R+ LCD- with N=3000.

As a result, it is concluded that the least common degree is an extremely important mechanism to maintain the structure of the small world network. Because, when random link addition and random link deletion mechanisms are implemented on the small world network, it converges to the random network rapidly. However, the structure of the small world network is sustained when the link deletion rule is replaced with the least common degree.

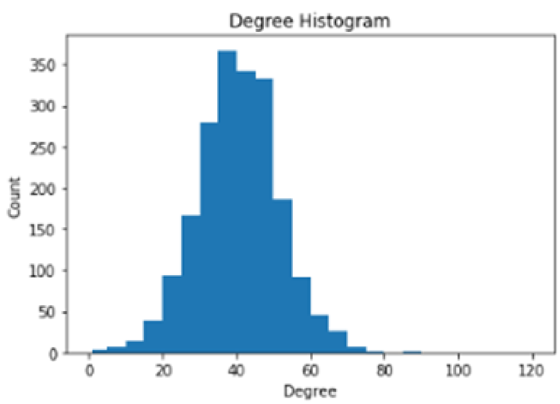

Figure 5.44. The Degree Distribution of the Network at $t_{5.00}$ Modified by R+ LCD-.

5.2.2. The PA+ R- experiment

After 25 replications are carried out, the average shortest path and the average clustering coefficient are demonstrated in Figure 5.45 and Figure 5.46.

The network at $t_{2.00}$ converts into the random network since the average shortest path and the average clustering coefficient converge to those of the random network. This experiment and the PA+ R- experiment carried out on the random network (in the part 5.1.2) follow similar pattern after the network at $t_{2.00}$ is reached. There are two different analyses carried out to understand dramatic increase in average clustering coefficient and remarkable decrease in average shortest path. The first one is on how the nodes having very high degrees (top five) connect to each other and top 20 nodes as the number of links changed by these mechanisms increases. It is seen that from the network $t_{3.00}$, top five nodes have connected to 18 of 20 nodes. Also, these top five nodes are linked with each other in that snapshot. This explains the changes in average shortest path and average clustering coefficient of the network thanks to nodes having very high degrees. The other analysis is to understand how the nodes having middle level degrees contribute to the changes in these measures. Therefore, 10 nodes are selected and analyzed whether their neighbors tend to central nodes over time. In the network at $t_{1.00}$, there is no neighbor whose degree is more than 100 (the maximum degree is 88). In the network at $t_{2.00}$, there are 37 neighbors whose degrees are over 100 but no neighbor having 1000 degrees. Then, the network at $t_{3.00}$ has started to be clustered because the neighbors of these nodes have started to link with central nodes. The number of neighbors whose degrees are more than 100 and 1000 are 90 and 19, respectively. These degrees have increased to 138 and 64 in the network at $t_{4.00}$. Lastly, there are 131 and 92 neighbors whose degrees are more than 100 and 1000 in the network at $t_{5.00}$. These nodes have interacted with more central ones over time, which contributes to the increase in average clustering coefficient and the decrease in average shortest path.

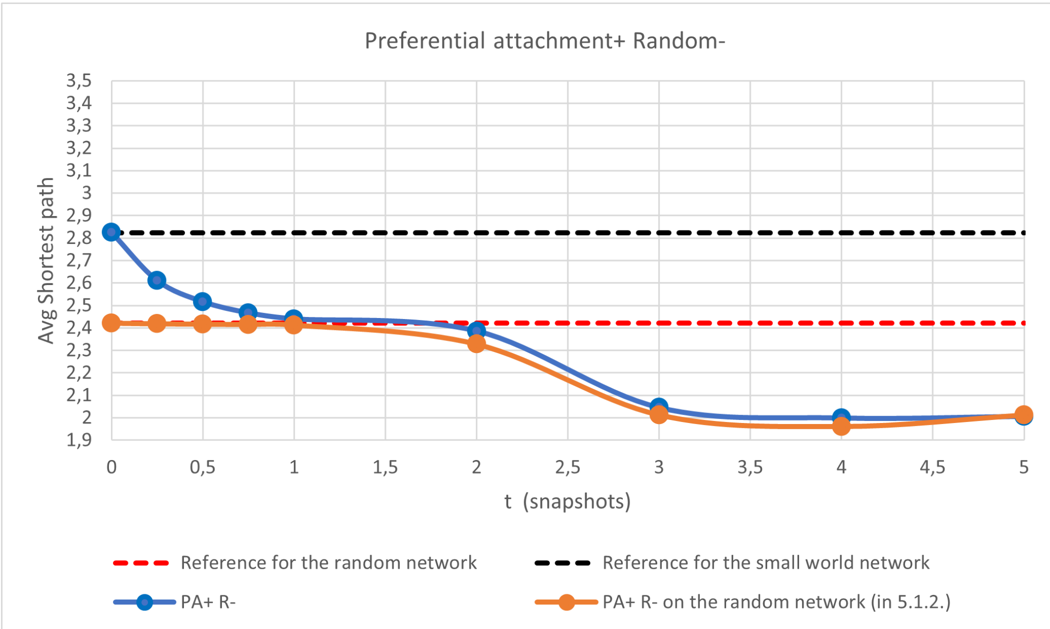

Figure 5.45. The Average Shortest Path of the Small World Network Modified by PA+ R-.

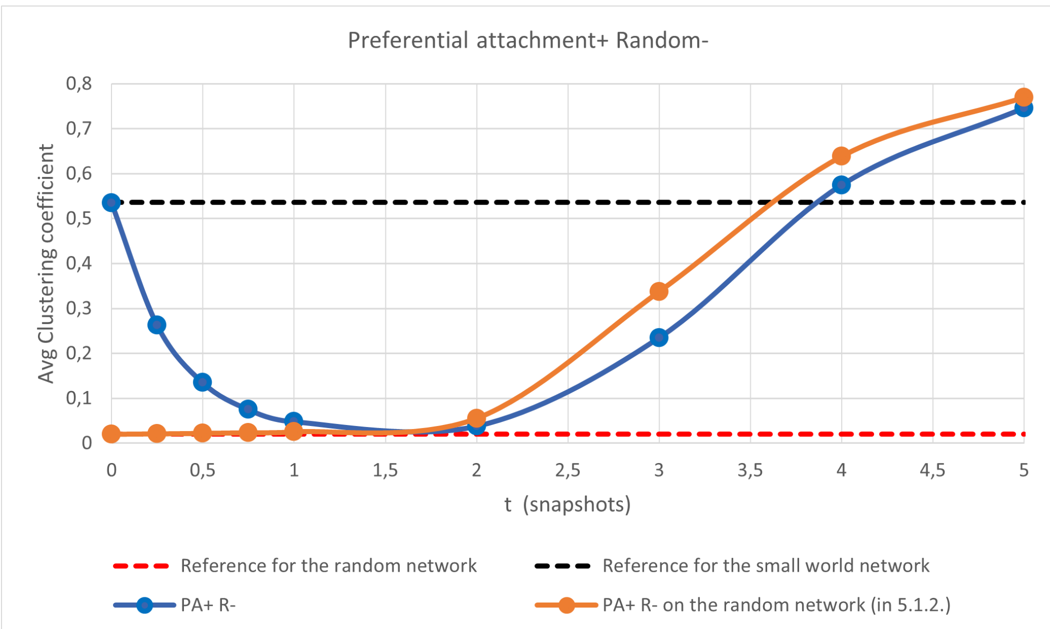

Figure 5.46. The Average Clustering Coefficient of the Small World Network Modified by PA+ R-.

The degree distribution of the network at $t_{3.00}$ (as seen in Figure 5.47) is right skewed and the peak has decreased compared to the initial network.

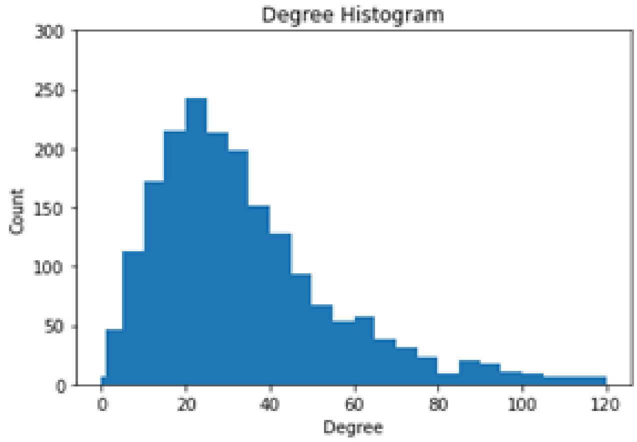

Figure 5.47. The Degree Distribution of the Network at $t_{3.00}$ Modified by PA+ R-.

While the degree distribution of the initial small world network is symmetrically distributed, the degree distribution of the network at $t_{5.00}$ is skewed to the right due to preferential attachment link addition mechanism. At the end of the analysis, the number of isolated nodes is 25 on average and there are no disconnected components. Since the degree distribution of the network has similar pattern and conducting more snapshots is computationally expensive, additional snapshots will not be carried out although the average clustering coefficient does not reach a steady state in Figure 5.46. Additionally, maximum variances of the average shortest path and average clustering coefficient are 0.01 and 0.06 considering all snapshots, respectively.

As seen, the experiments of PA+ R- for both random network and small world network have similar results related to average shortest path and average clustering coefficient. What differentiates the second one is that the small world network converges to random network at first. The evolving network at $t_{2.00}$ converges to a random network in terms of the average shortest path and the average clustering coefficient. After that, the behavior of the network is similar to the one explained in 5.1.2. The reason is that nodes with high degrees have interacted with each other and the ones with middle level degrees have tendency to link with more central nodes.

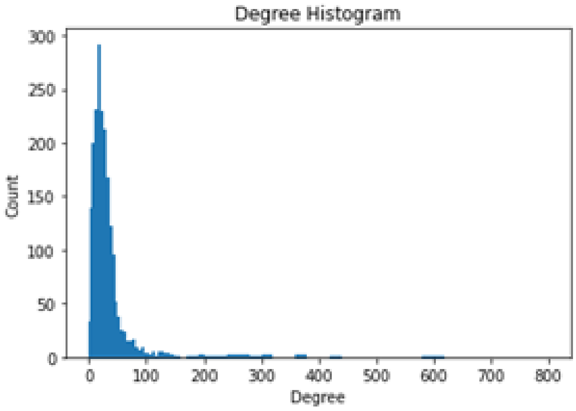

Figure 5.48. The Degree Distribution of the Network at $t_{5.00}$ Modified by PA+ R-.

5.2.3. The PA+ LCD- experiment

In total, 25 replications are carried out. The experiment is performed until the evolving network at $t_{5.00}$ is reached.

The average shortest path of the network at $t_{2.00}$ is almost the same as that of the initial small world network. In the network at $t_{2.00}$, the average clustering coefficient does not change much as seen in Figure 5.50.

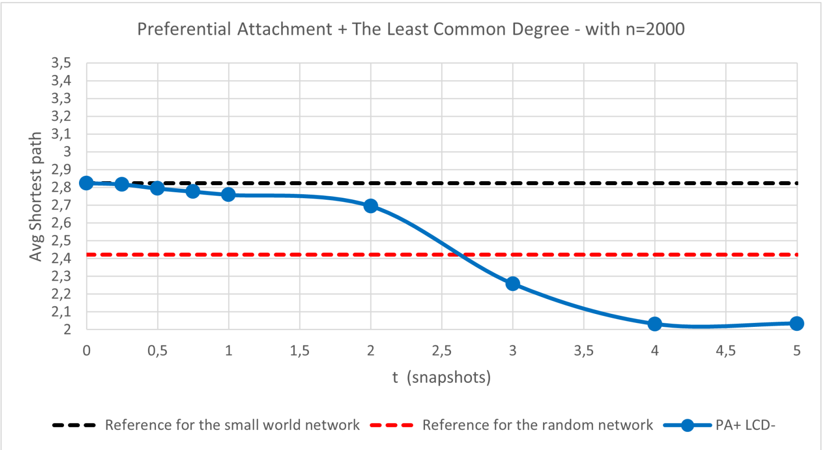

Figure 5.49. The Average Shortest Path of the Small World Network Modified by PA+ LCD-.

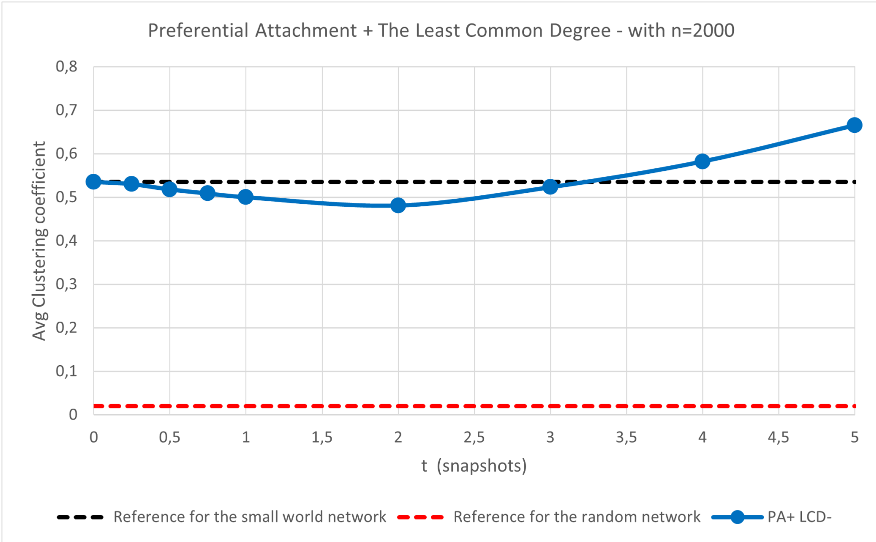

Figure 5.50. The Average Clustering Coefficient of the Small World Network Modified by PA+ LCD-.

Additionally, it is seen that the degree distribution of the evolving network at $t_{2.00}$ is symmetric and bell-shaped curve. Compared to the initial network, the peak value has dropped,the tails in each side have been long and its degree distribution is spread in wider range. Additionally, the network at $t_{2.00}$ does not have any isolated nodes.

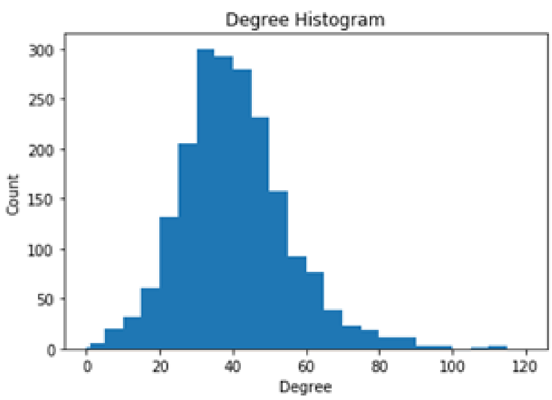

Figure 5.51. The Degree Distribution of the Network at $t_{2.00}$ Modified by PA+ LCD-.

The fundamental measures have remarkably changed especially after the network at $t_{2.00}$.

To understand this considerable change in the average shortest path and the average clustering coefficient, nodes are ordered in a descending way based on their degrees. Then, the first five and 20 nodes are selected. It is checked whether the five nodes are connected to the 20 nodes over time. In the network at $t_{2.00}$, it is seen that top five nodes are connected to four of 20 nodes. Then, it is seen that top five nodes of the network at $t_{3.00}$ have interacted with approximately half of 20 nodes. By the nature of preferential attachment link addition mechanism, nodes have tendency to link with the ones having higher nodes. Therefore, the degree distribution starts to be skewed to the right as the number of links changed by PA+ LCD- increases. These five nodes have linked with all 20 nodes in the network at $t_{5.00}$ and its degree distribution is positively skewed. Also, there are over 200 isolated nodes.

As an additional analysis to understand the increase in average clustering coefficient and decrease in average shortest path, 10 nodes whose degrees are between 15 and 55 are selected. It is checked if the neighbors of these nodes have linked with central nodes over time. 10 nodes have 408 neighbors in total in the network at $t_{1.00}$ and there is not any neighbor whose degree is more than 100 (The maximum degree is 62). In the network at $t_{2.00}$, there is one neighbor whose degree is over 100. Then, the number of neighbors whose degrees are more than 100 increases to 46. Also, 5 neighbors have more than 1000 degrees in the network $t_{3.00}$. The number of neighbors whose degrees are over 100 and 1000 are 123 and 40, respectively for the network at $t_{4.00}$. Lastly, in the network at $t_{5.00}$, the number of neighbors whose degrees are more than 100 and 1000 are 205 and 60, respectively. It shows that these nodes tend to link with central nodes over time, which explains the increase in average clustering coefficient. Since nodes can reach the other ones in a few steps, this analysis also explains the decrease in average shortest path. Additionally, maximum variances of the average shortest path and the average clustering coefficient are 0.002 and 0.001, respectively when all snapshots are considered.

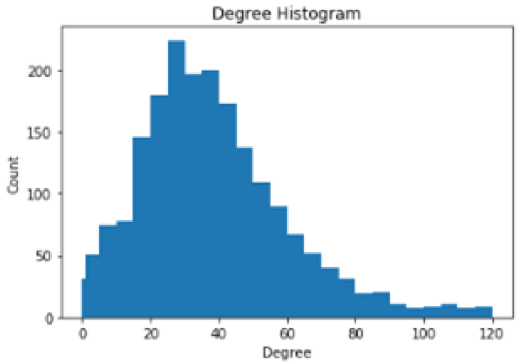

Figure 5.52. The Degree Distribution of the Network at $t_{3.00}$ Modified by PA+ LCD-.

The class of the network can change in different snapshots. For instance, it is seen that the structural properties of the network at $t_{2.00}$ show similarity with the small world network. After that, the class of the network has changed. Thanks to the link addition mechanism, nodes tend to connect to the ones having high degrees and nodes having middle level degrees tend to link with central nodes, contributes to increase average clustering coefficient. The link deletion mechanism also supports this effect. Because deleting a link via the least common degree mechanism pushes the network to be clustered more. Also, nodes can reach each other more quickly, which decreases average shortest path of the network. The network at $t_{5.00}$ has over 200 isolated nodes as seen in Figure 5.53.

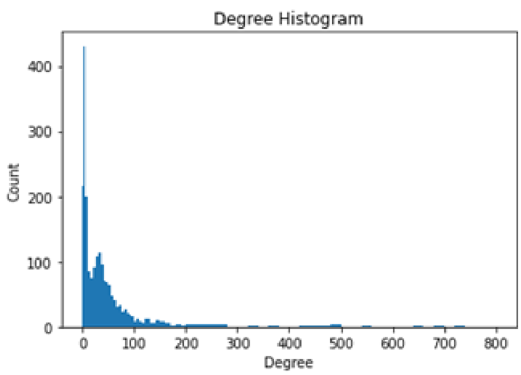

Figure 5.53. The Degree Distribution of the Network at $t_{5.00}$ Modified by PA+ LCD-.

5.2.4. The TR1+ LCD- experiment

The experiment that 25 replications are carried out is ended when the average shortest path and average clustering coefficient reach a steady state. It is seen that small world network properties have been preserved.

As seen from the Figure 5.54, the average shortest path of the network at $t_{5.00}$ is 2.93 on average. The literature review demonstrates that it grows the logarithm of the network size in the small world network, which corresponds to $\log 2000 = 3.30$ [13]. Since the value calculated is less than 3.30, the class of the network is still the small world network.

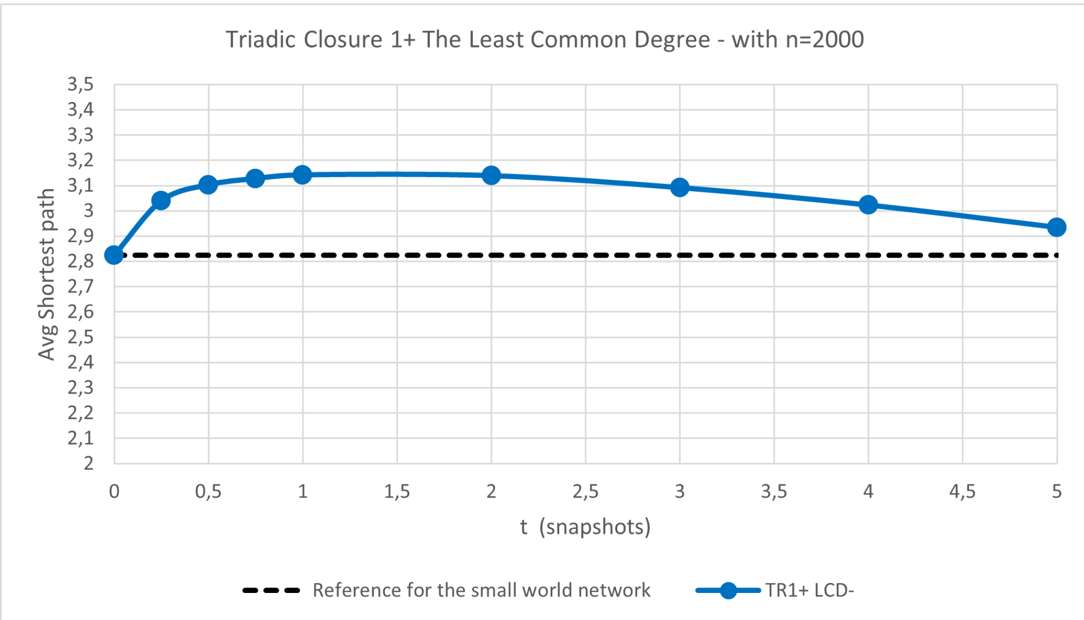

Figure 5.54. The Average Shortest Path of the Small World Network Modified by TR1+ LCD- with N=2000.

The average clustering coefficient of the network at $t_{5.00}$ remains nearly the same as that of the initial network, which is seen in Figure 5.55.

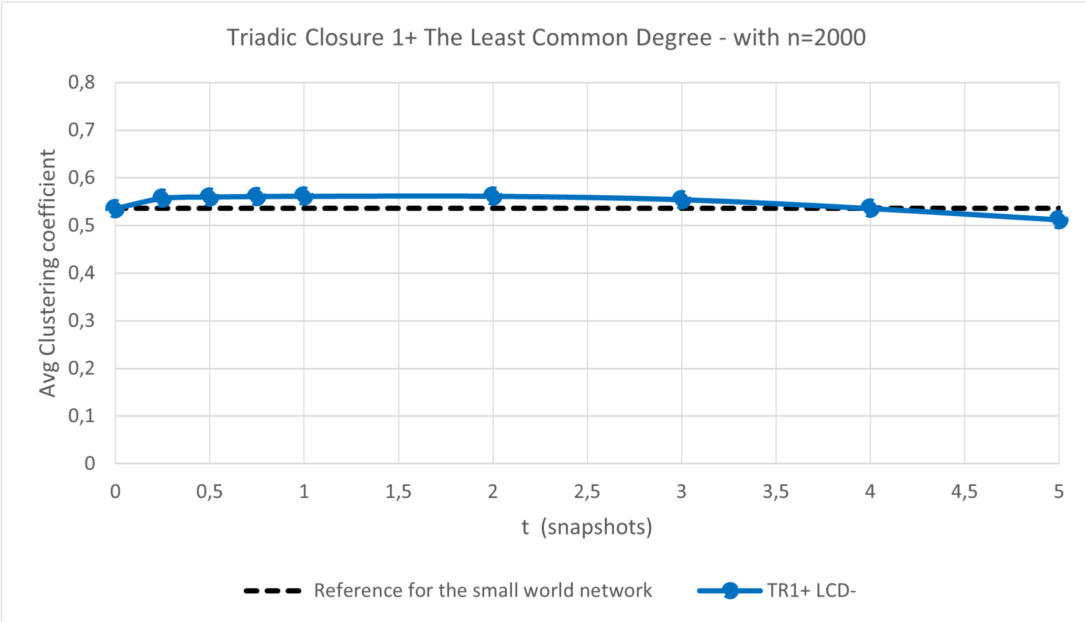

Figure 5.55. The Average Clustering Coefficient of the Small World Network Modified by TR1+ LCD- with N=2000.

At the end of the analysis, the results show that the network class does not change. The networks with different rewiring probabilities are generated, which are exhibited in Table 5.4 and compared the results of the network at $t_{5.00}$. The average shortest path of the network at $t_{5.00}$ is almost the same as the value in the small world network whose rewiring probability is between 6% and 7%. Similarly, its average clustering coefficient is close to that of the small world network whose rewiring probability is approximately 10% (namely the initial network). It is concluded that the modified network has still small world network characteristics. The degree distribution the network at $t_{1.00}$ is symmetrically distributed with two tails in each side. Compared to the initial network, the peak has decreased, the tails have been long and the degree values are spread out in wider range. It is observed that there are no isolated nodes. In other words, each node is connected to the others in some way.

The mean value of the degree distribution of the network at $t_{5.00}$ is almost the same as that of initial network. It has longer tails than that of the network at $t_{1.00}$. The peak value has decreased a bit more.

Additionally, maximum variances of the average shortest path and the average clustering coefficient are 0.002 and 0.0006 over time, respectively.

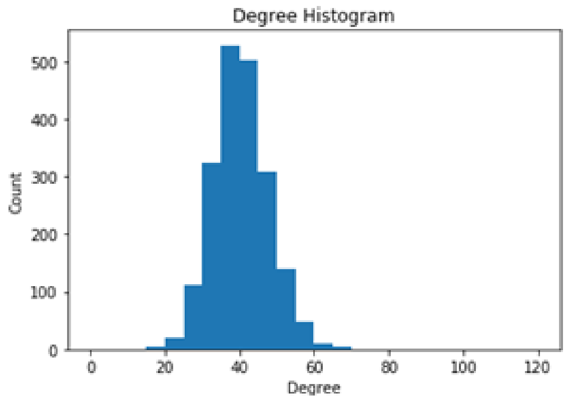

Figure 5.56. The Degree Distribution of the Network at $t_{1.00}$ Modified by TR1+ LCD-.

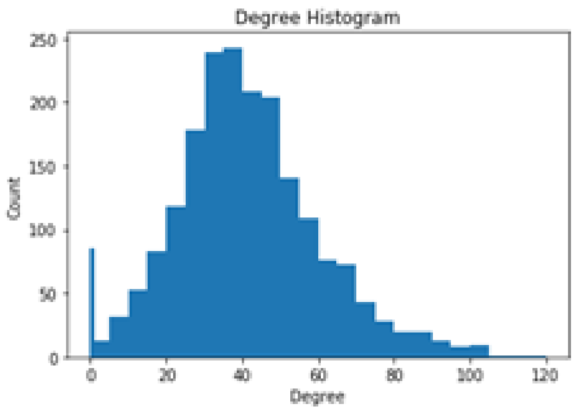

Figure 5.57. The Degree Distribution of the Network at $t_{5.00}$ Modified by TR1+ LCD-.

5.2.5. The TR2+ LCD- experiment

In this experiment, it is seen that there is a considerably high difference between the average shortest path of the network at $t_{0.25}$ and the average shortest path of the network at $t_{0.50}$. The experiment is carried out until the network at $t_{2.00}$ is found. Although its average shortest path does not reach a steady state, there will be no additional experiments done for further snapshots.

Because it is computationally expensive and the network at $t_{2.00}$.has been fragmented to subgraphs. Since the average shortest path of this experiment has dramatically increased, the scale of y axis in Figure 5.58 is increased.

In theory, the average shortest path grows with $n/2k$ and the average clustering coefficient is 3/4 for the lattice network [13]. Therefore, the network class is the lattice network if its average shortest path and its average clustering coefficient of the evolving network converge to 25 and 0.75, respectively.

As shown in Figure 5.58 and Figure 5.59 , the average shortest path of the network at $t_{0.50}$ increases to 15 on average. Its average clustering coefficient is nearly 0.70, which indicates that the class of the network is similar to the lattice network. In this snapshot, there are no isolated nodes or components. Its degree distribution is bell curve shaped, and the tails are not so long. The nodes have degrees between 20 and 60. The range of the degree distribution in the network at $t_{0.50}$ is slightly wider than that of initial network There is hardly ever skewness compared to the initial network but there are less nodes having mean degrees than the initial network.

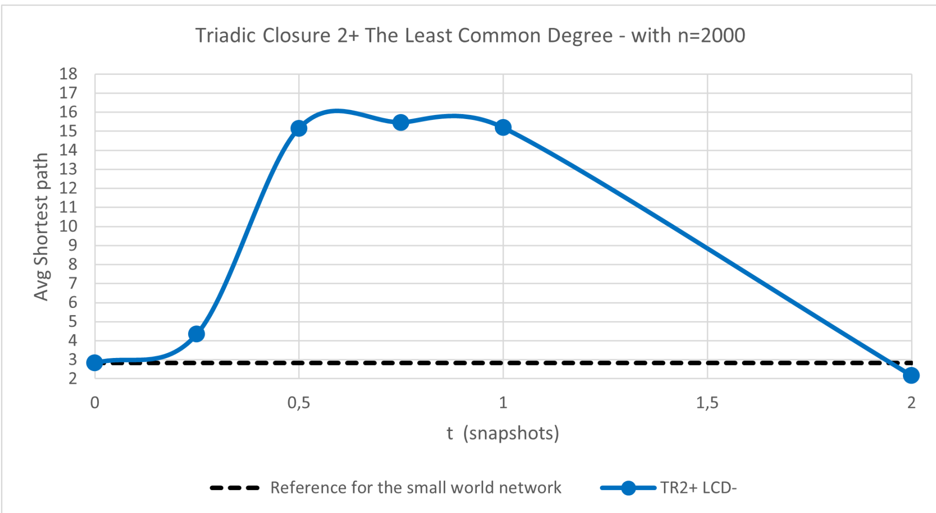

Figure 5.58. The Average Shortest Path of the Small World Network Modified by TR2+ LCD-.

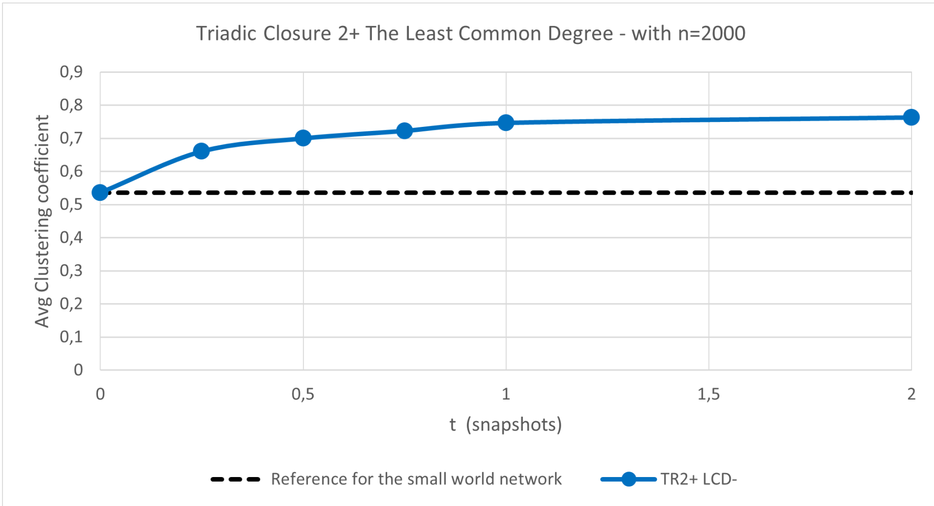

Figure 5.59. The Average Clustering Coefficient of the Small World Network Modified by TR2+ LCD-.

The results of the structural properties of the network at $t_{1.00}$ are more or less the same as those of the network at $t_{0.50}$. In this snapshot, there are approximately five isolated nodes.

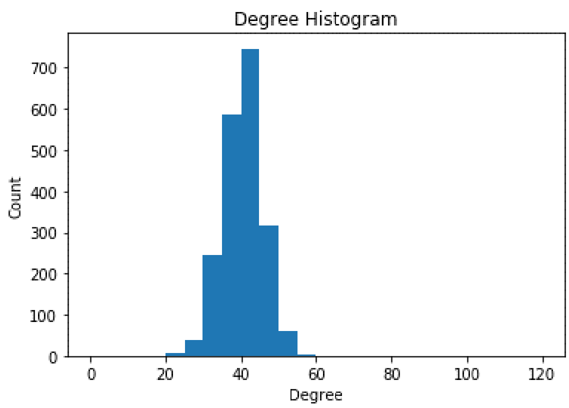

Figure 5.60. The Degree Distribution of the Network at $t_{0.50}$ Modified by TR2+ LCD-.

However, the network at $t_{2.00}$ is fragmented. The reason of the dramatic decrease in the average shortest path is the least common degree mechanism. In this network, some links connect different sub-graphs to each other.

Due to the fact that those links are deleted based on the common degree mechanism, the network breaks into different components. Therefore, each component has different average shortest paths. There are approximately 150 isolated nodes, and ten different sub-graphs, which demonstrates that the network is splitted. Its degree distribution has longer tails than those of previous snapshots. It is slightly skewed to the left, but still close to symmetrically distributed as seen in Figure 5.62. The reason in the increase the number of nodes having degrees between 1 and 20 can be the sub-graphs. Since they are clustered in themselves, the number of nodes with low degrees increases. Additionally, maximum variances of the average shortest path and the average clustering coefficient are 2.12 and 0.00006 over snapshots, respectively.

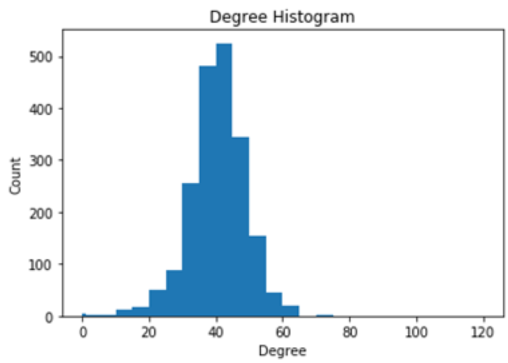

Figure 5.61. The Degree Distribution of the Network at $t_{1.00}$ Modified by TR2+ LCD-.

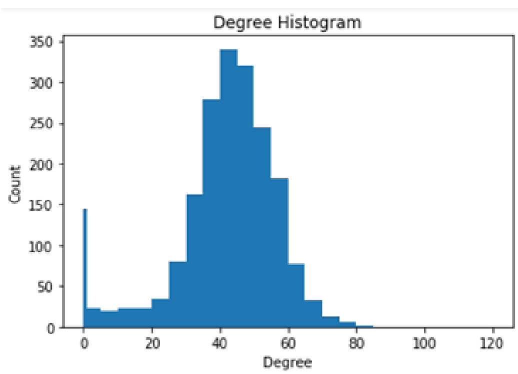

Figure 5.62. The Degree Distribution of the Network at $t_{2.00}$ Modified by TR2+ LCD-.

5.2.6. The HB1+ LCD- experiment

In this experiment, seven sub-experiments are carried out by using different parameters as shown in Table 5.7. Some of them have been carried out to identify the influence of the threshold value determined in hybrid 1 link addition mechanism on the structural properties of the evolving network. Threshold parameters are chosen as 0.1, 0.2, 0.3, 0.4, and 0.5. Other than the threshold parameter, some sub-experiments having different number of nodes have been completed, as well. Additionally, the ones with different k values have been also conducted to identify the influence on the structure of the evolving network. In total, 175 replications are done. Each sub-experiment is performed until the network at $t_{5.00}$ is found.

Table 5.7. The Parameter Values of the Sub-experiments.

Parameters	Sub-exp. 1	Sub-exp. 2	Sub-exp. 3	Sub-exp. 4	Sub-exp. 5	Sub-exp. 6	Sub-exp. 7
The number of nodes (N)	2000	2000	2000	2000	2000	3000	3000
The number of edges (k)	40	40	40	40	40	30	40
The rewiring probability (p)	0.1	0.1	0.1	0.1	0.1	0.1	0.1
The threshold value (THR)	0.1	0.2	0.3	0.4	0.5	0.4	0.4

For the first five sub-experiments, the average shortest path of the network in the sub-experiment 1 is the highest for each snapshot and the average shortest paths of these five sub-experiments are higher than that of the initial network. The sub-experiments whose the percentage of using random link addition mechanism are relatively higher than others (e.g. sub-experiment 4 and sub-experiment 5) yield more similar results to the initial network over time.

Although the average shortest paths of all sub-experiments increase at first, then they converge to that of the initial network as the number of links changed increases as shown in Figure 5.63.

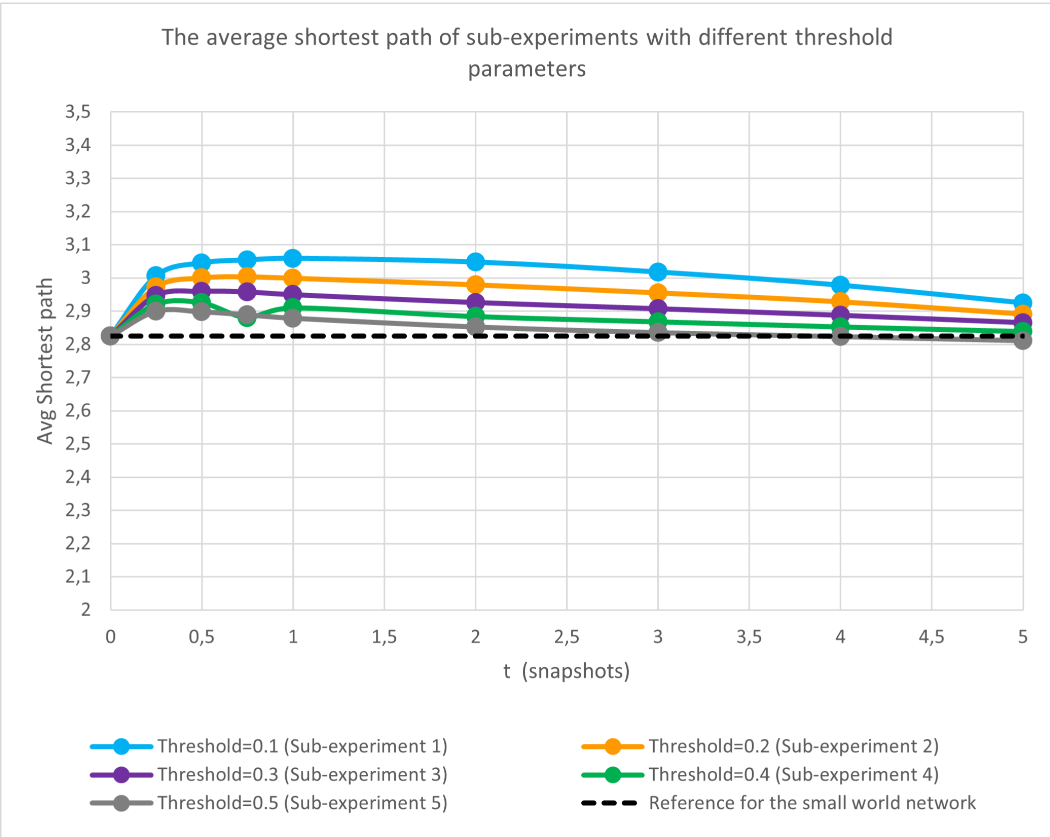

Figure 5.63. The Average Shortest Path of the Small World Network Modified by HB1+ LCD- With Different Threshold Parameters.

As seen in Figure 5.64, the average clustering coefficients of the sub-experiments are so close to each other that they are almost in parallel with the line of the average clustering coefficient of the initial network. When the number of links changed by HB1+ LCD- considerably increases (e.g. five times total links), average clustering coefficients of all sub-experiments have slightly decreased compared to the initial network. In that snapshot, the sub-experiment 1 (the average clustering coefficient of 0.51) is the closest to that of the initial network.

For those sub-experiments, the degree distributions of the evolving networks have bell shaped curve. There is hardly ever skewness to any side. It is symmetrically distributed and two tails in each side. Compared to the initial network, the peak has decreased and degree values are spread in wider range. As an example, the degree distribution of the sub-experiment 4 is demonstrated in Figure 5.65. The number of isolated nodes increases as the threshold parameter decreases. When the threshold parameters are selected as 0.1, 0.2, 0.3, 0.4 and 0.5, the isolated nodes of the network are approximately 80, 40, 30, 20, 10, respectively.

In conclusion, the threshold parameter has relatively less influence on sustaining the structural properties of the small world network than the link addition and the link deletion mechanisms. The network modified by HB1+ LCD- is still small world network characteristics. The sub-experiment having the highest clustering coefficient is the one with the lowest threshold value in the network at $t_{5.00}$.

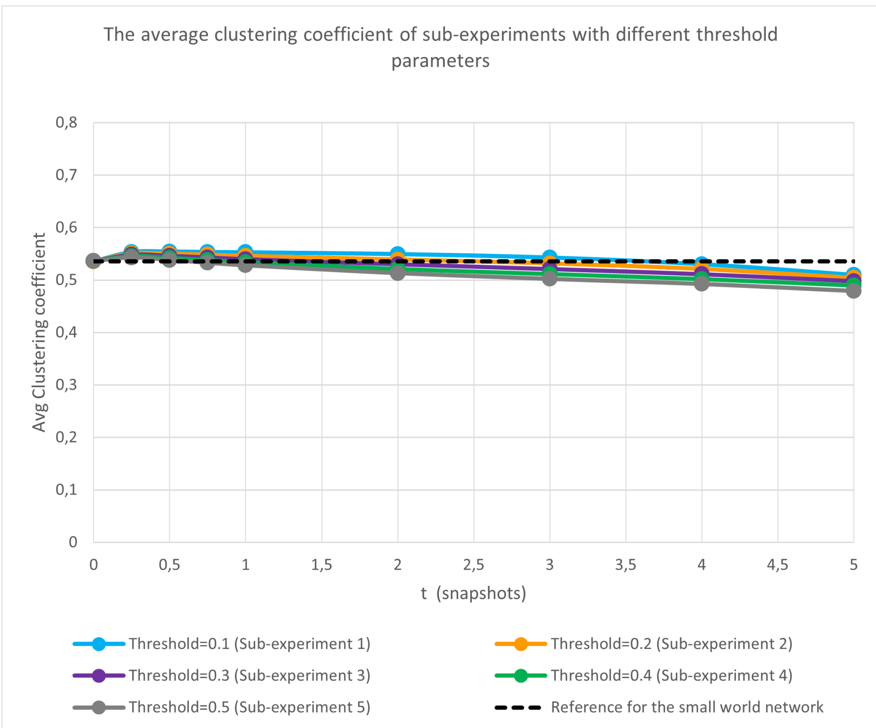

Figure 5.64. The Average Clustering Coefficient of the Small World Network Modified by HB1+ LCD- With Different Threshold Parameters.

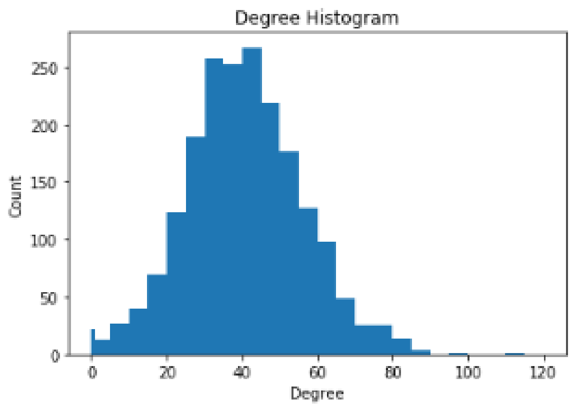

Figure 5.65. The Degree Distribution of the Network at $t_{5.00}$ Modified by HB1+ LCD- (N=2000, k=40, THR=0.4).

Alternatively, the threshold parameter and the number of the links that a node has are kept constant and the network size is chosen as 2000 and 3000. The dashed lines in Figure 5.66 represent the values of the initial small world networks. The comparison between the sub-experiment 4 and the sub-experiment 7 is explained below.

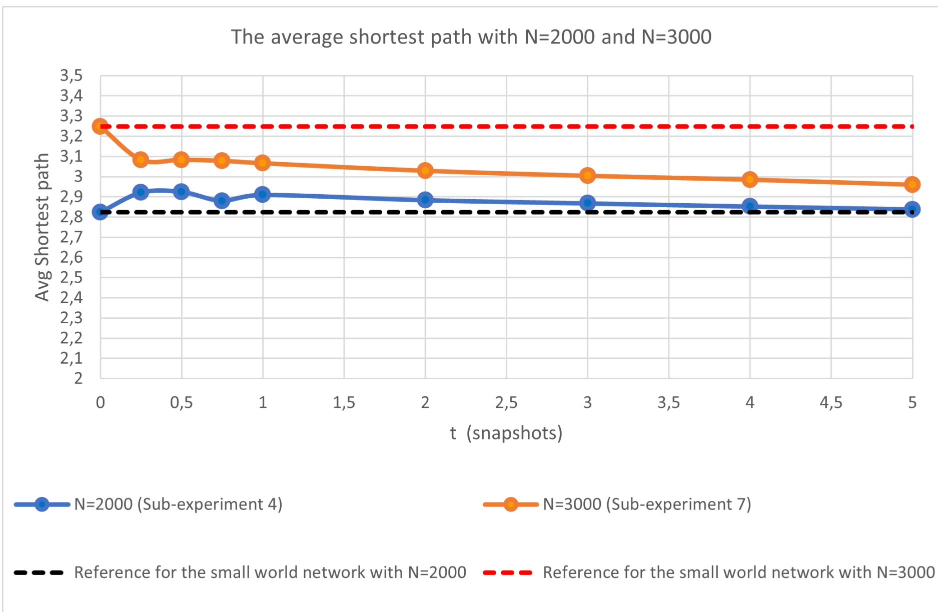

Figure 5.66. The Average Shortest Path of the Small World Network Modified by HB1+ LCD- With Different Network Size.

As seen in Figure 5.66, the average shortest path of the network in the sub-experiment 4 fluctuates at first and then remains steady as the number of links changed with HB1+ LCD- increases. However, the average shortest path of the sub-experiment 7 has gradually decreased. It is expected to see higher average shortest path in the sub-experiment 7 than the other because this metric grows with the logarithm of the number of nodes N for a small world network.

As seen in Figure 5.67, the change in the network size does not have a remarkable influence on the average clustering coefficient. When the sub-experiment 7 is compared to the sub-experiment 4, it is seen that their clustering coefficients remain almost the same when the number of snapshots increases. It demonstrates that HB1+ CD- experiment on the small world network can produce meaningful results for higher number of nodes. The degree distributions of the sub-experiment 7 and the sub-experiment 4 are similar to each other (the sub-experiment 4 is shown in Figure 5.65). When the number of nodes are 2000 and 3000, the average number of isolated nodes are 20 and 40 on average, respectively.

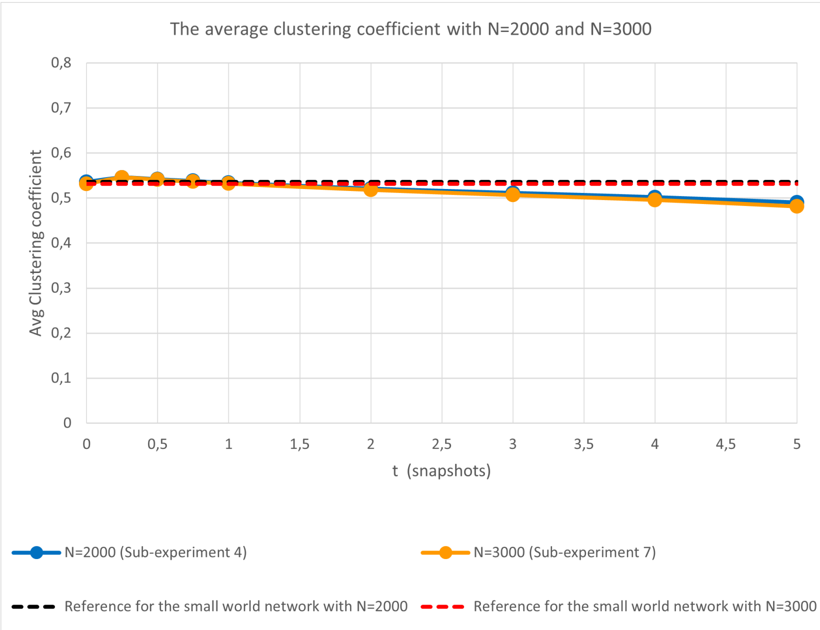

Figure 5.67. The Average Clustering Coefficient of the Small World Network Modified by HB1+ LCD- With Different Network Size.

There are other sub-experiments done considering different number of neighbors of a node, 30 and 40. Structural properties of the sub-experiment 6 and the sub-experiment 7 are compared to each other and exhibited in Figure 5.68 and Figure 5.69.

When the structural properties of these two sub-experiments are compared to the networks generated with different rewiring probabilities, it is seen that the sub-experiment 6 is similar to the network whose rewiring probability is between 0.1 and 0.2. The other one shows similarity with the small world network whose rewiring probability is almost 0.1.

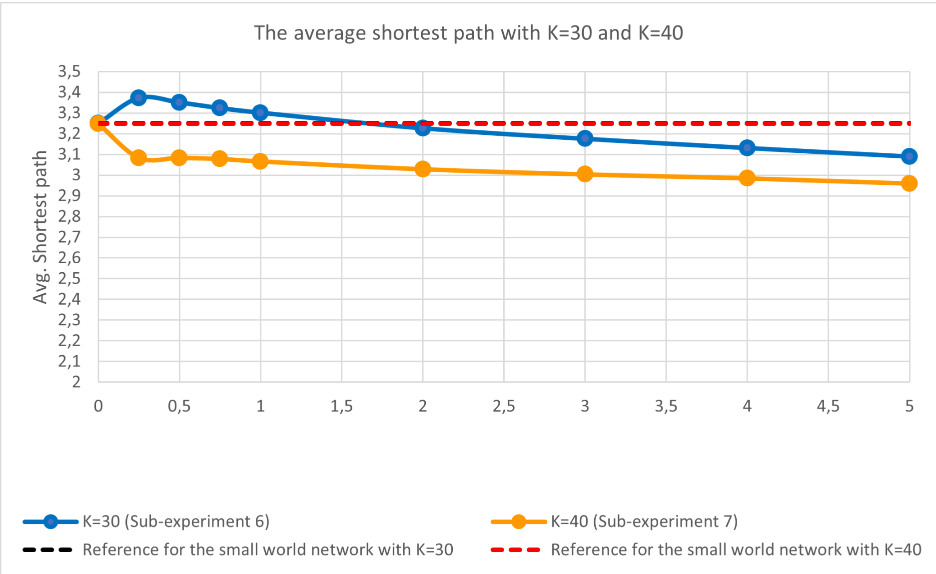

Figure 5.68. The Average Shortest Path of the Small World Network Modified by HB1+ LCD- With Different Number of Neighbors.

As the number of links changed by HB1+ LCD- mechanism increases, the average clustering coefficient of the experiment having higher k (The sub-experiment 7) declines less as expected. Because total links and average degrees of nodes in the network at t_0 are higher in the sub-experiment 7 than the sub-experiment 6. The degree distributions of these networks at $t_{5.00}$ have bell curve shaped and two tails in each side almost symmetrically.

Compared to initial network, tails have been long, peaks have decreased and the degree range has become wider. As an example, the degree distribution of the sub-experiment 7 is shown in Figure 5.70.

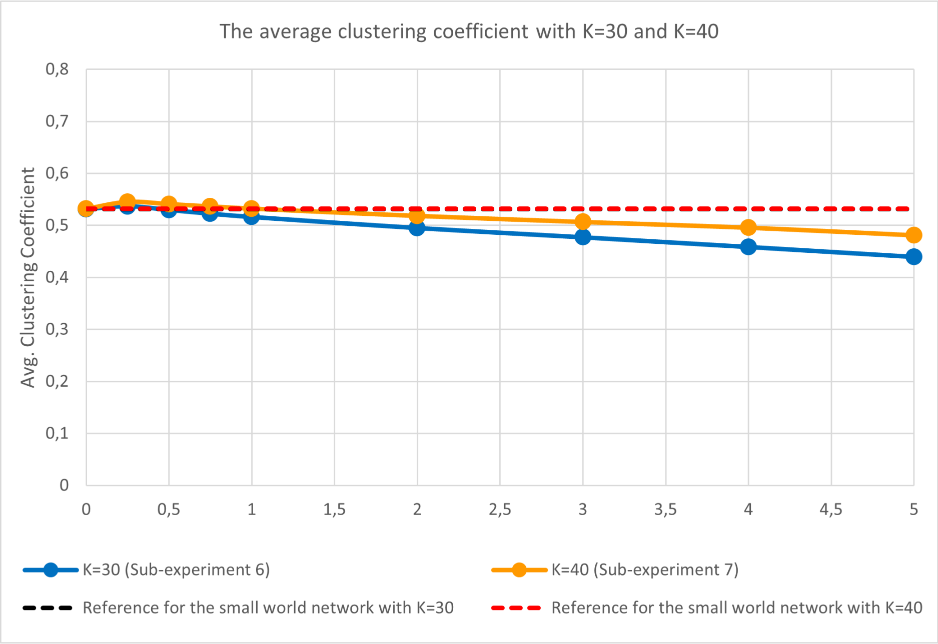

Figure 5.69. The Average Clustering Coefficient of the Small World Network Modified by HB1+ LCD- With Different Number of Neighbors.

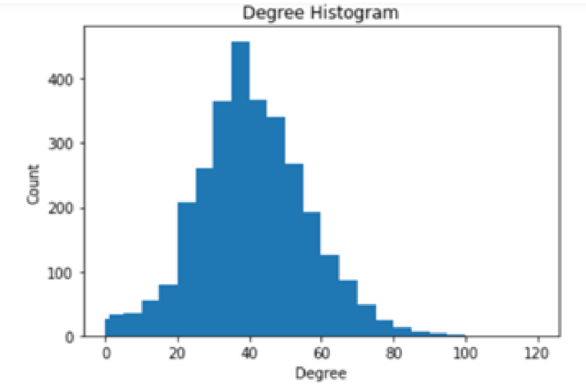

Figure 5.70. The Degree Distribution of the Sub-experiment 7.

Also, the sub-experiment 6 is compared to the experiment R+ LCD- shown in 5.2.1 to understand how link addition mechanisms influences the average shortest path and the average clustering coefficient over time. In these experiments, the network size is 3000 and the number of neighbors that a node has is 30.

As seen from Figure 5.71 and Figure 5.72, HB1 link addition mechanism gives more promising result than the other in terms of average shortest path and average clustering coefficient. The average clustering coefficient and average shortest path of HB1+ CD- provides closer results to that of the initial small world network in each snapshot. The variances of sub-experiment results are quite low.
For instance, the maximum variances of the sub-experiment 4 are 0.001 for the average shortest path and 0.0001 for the average clustering coefficient when all snapshots are considered.

Overall, the HB1+ CD- link modification mechanism is quite important when it is implemented on the small world network. Because the networks modified by HB1+ LCD- mechanism still have small world network characteristics as seen from all sub-experiments done.

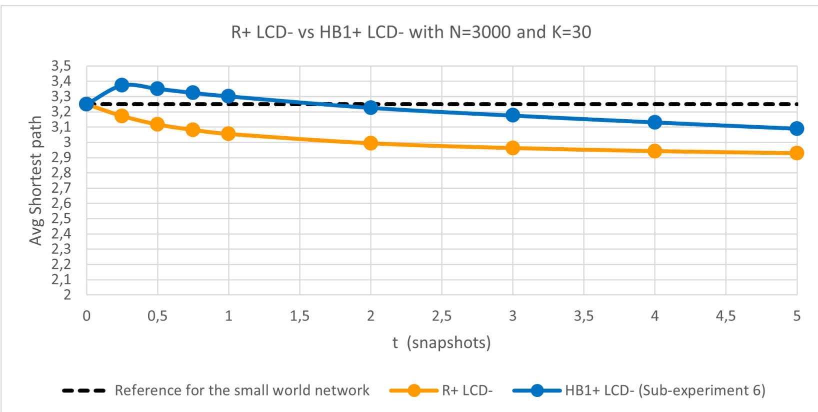

Figure 5.71. The Average Shortest Path of R+ LCD- vs HB1+ LCD- With N=3000 and k=30.

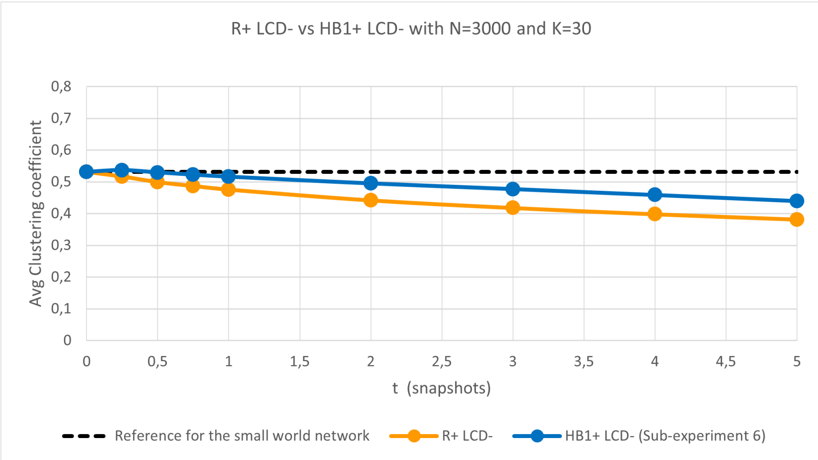

Figure 5.72. The Average Clustering Coefficient of R+ LCD- vs HB1+ LCD- With N=3000 and k=30.

5.2.7. The HB2+ LCD- experiment

For this experiment, it is seen that there is a rapid increase in the average shortest path of the network at $t_{0.5}$ compared to that of the small world network. After that, it reaches a steady state as shown in Figure 5.73.

The average clustering coefficient tends to increase as the number of snapshots proceeds. In theory, the average shortest path grows with $n/2k$, which corresponds to 25 for this experiment, and the average clustering coefficient is 3/4 for the lattice network [13]. Since there is no significant increase in the average shortest path, it is considered that the final network is not a lattice network. However, it has been clustered more over time as seen in Figure 5.74. The structural properties of the final network are compared to those of the networks generated with different rewiring probabilities, which are shown in Table 5.4. It is concluded that the final network is similar to the small world network whose rewiring probability is almost 0.02.

Compared to the initial network as seen in Figure 5.2, the peak has dropped in the degree distribution of the network at $t_{2.00}$. It is nearly symmetrically distributed without skewness. Additionally, the number of isolated nodes is nearly 10.

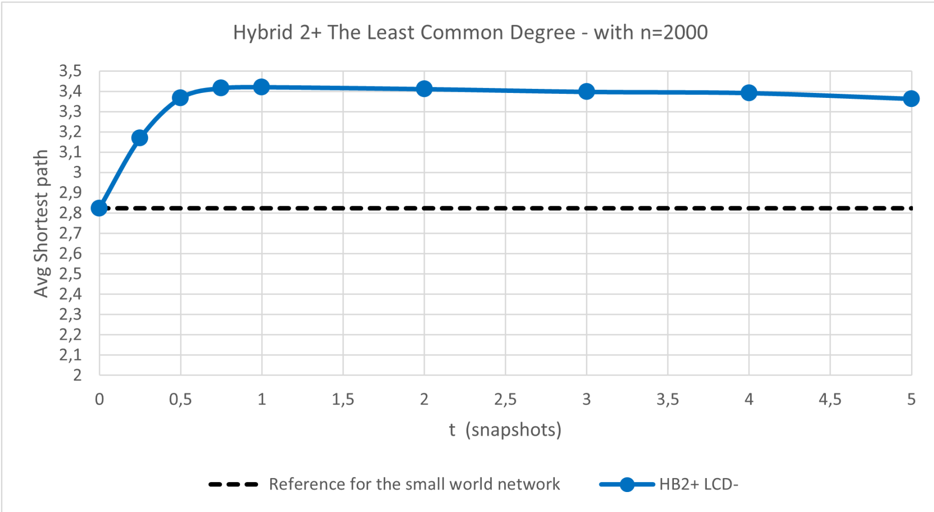

Figure 5.73. The Average Shortest Path of the Small World Network Modified by HB2+ LCD-.

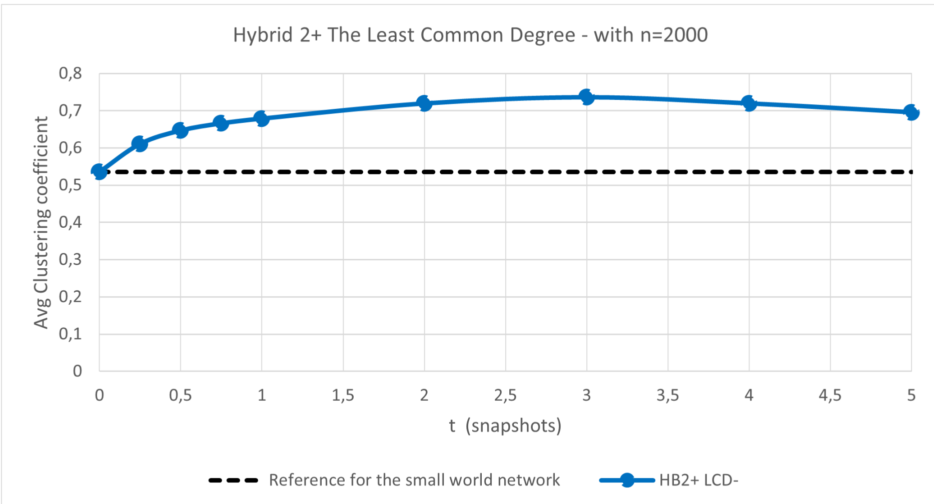

Figure 5.74. The Average Clustering Coefficient of the Small World Network Modified by HB2+ LCD-.

The number of nodes having low degrees begins to increase in the evolving network at $t_{3.00}$. The number of isolated nodes is approximately 80.

The degree distribution of the network at $t_{5.00}$ has two peak points. The left tail is slightly longer than the right tail and there are over 200 isolated nodes.

Additionally, maximum variances over snapshots are 0.001 for the average shortest path and 0.0001 for the average clustering coefficient.

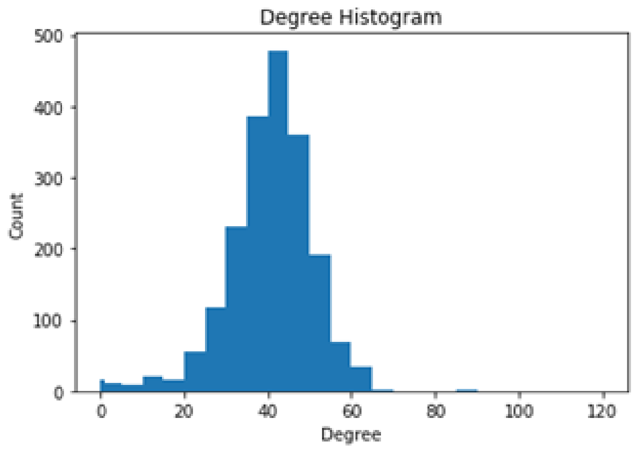

Figure 5.75. The Degree Distribution of the Network at $t_{2.00}$ Modified by HB2+ LCD-.

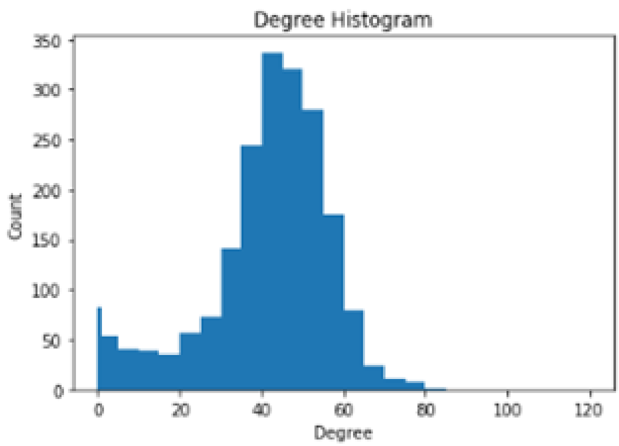

Figure 5.76. The Degree Distribution of the Network at $t_{3.00}$ Modified by HB2+ LCD-.

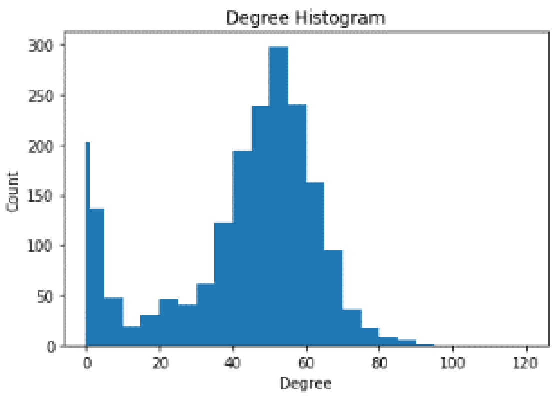

Figure 5.77. The Degree Distribution of the Network at $t_{5.00}$ Modified by HB2+ LCD-.

The summary of the experiments can be demonstrated as follows for each network class in terms of average shortest path and average clustering coefficient. Table 5.8 and Table 5.9 are experiment results on random networks with $N = 2000$ and $p = 0.02$. Table 5.10 and Table 5.11 are experiment results on small world networks with $N = 2000$, $k = 40$ and $p = 0.1$.

Table 5.8. ASP of Experiments on Random Networks.

Random network - ASP	R+	PA+	TR1+	TR2+	HB1+	HB2+
R-	Remained almost the same	Decreased	Remained almost the same	Remained almost the same	Remained almost the same	Remained almost the same
D-	Remained almost the same	Remained almost the same	Remained almost the same	Remained almost the same	Remained almost the same	Remained almost the same
LCD-	Slightly increased	Decreased	Slightly decreased	Slightly decreased	Remained almost the same	Remained almost the same

Table 5.9. ACC of Experiments on Random Networks.

Random network - ACC	R+	PA+	TR1+	TR2+	HB1+	HB2+
R-	Remained almost the same	Increased dramatically	Remained almost the same	Remained almost the same	Remained almost the same	Remained almost the same
D-	Remained almost the same	Remained almost the same	Remained almost the same	Remained almost the same	Remained almost the same	Remained almost the same
LCD-	Increased	Increased dramatically	Increased	Increased	Increased	Increased

Table 5.10. ASP of Experiments on Small World Networks.

Small world network - ASP	R+	PA+	TR1+	TR2+	HB1+	HB2+
R-	Decreased	Decreased	Decreased	Decreased	Decreased	Decreased
D-	Decreased	Decreased	Decreased	Decreased	Decreased	Decreased
LCD-	Slightly decreased	Decreased	Increased at first then decreased gradually	Increased considerably then decreased since the network is fragmented	Remained almost the same	Increased

Table 5.11. ACC of Experiments on Small World Networks.

Small world network - ACC	R+	PA+	TR1+	TR2+	HB1+	HB2+
R-	Decreased	Decreased at first then increased significantly	Decreased	Decreased	Decreased	Decreased
D-	Decreased	Decreased	Decreased	Decreased	Decreased	Decreased
LCD-	Decreased	Decreased at first then increased significantly	Remained almost the same	Increased	Remained almost the same	Increased

6. DISCUSSION

When (R+ LCD-), (TR1+ LCD-), (TR2+ LCD-), (HB1+ LCD-) and (HB2+ LCD-) mechanisms have been implemented on random networks, it is seen that the modified networks have converged first to networks with small world characteristics. When the number of links that are changed during network modification process is more than three times the total links, it is observed that their degree distributions tend to be flattened. All five experiments have the same link deletion mechanism, LCD, and their link addition mechanisms are different from each other. As the number of links changed by these mechanisms increases, degree distributions of all five experiments have converted into flat degree distribution. Therefore, it can be concluded that the influence of link addition mechanism decreases on the structure of degree distribution as the number of snapshots proceeds.

When the initial network is a small world network, it is observed that there are network modification combinations that small world network characteristics are preserved. For instance, the average shortest path and the average clustering coefficient of the network modified by (R+ LCD-) mechanism combination have decreased slightly compared to the initial small world network generated at t_0. When the network size is 2000 and the number of neighbors that a node has is 40, (The parameters selected are similar to the ones in literature studies. The network size is selected as 1000 and average degree is 20 in the study of Wang *et al.* [36], the network size and maximum degree are 250 and 5 respectively in the study of Jin *et al.* [37]) the network at $t_{5.00}$ is similar to the one whose rewiring probability is 15%. Its degree distribution has two tails and one side mirrors the other one without skewness. In that sense, it is similar to that of the initial network. The difference between them is that the values are spread in a wider range in the network at $t_{5.00}$ than the initial one. This experiment demonstrates the importance of the link deletion mechanism. Because R+ R- mechanism combination implemented on a small world network converts the network into a random network.

However, when the random link deletion mechanism is replaced with the least common degree mechanism, it is observed that the characteristics of small world network have been preserved. It demonstrates that the least common degree link deletion mechanism has a meaningful influence on sustaining the small world network structure.

Another mechanism combination that changes the network dynamically by preserving small world network characteristics is (TR1+ LCD-). It also provides the network to reach a steady state in terms of average clustering coefficient. By combining (TR1+ LCD-) and (R+ LCD-) mechanism combinations, another link modification combination, (HB1+ LCD-) is generated that also provides a small world network maintaining its properties. In this experiment, some conclusions are reached by performing controlled experiments (sub-experiments) with different parameters. The parameters that are changed are the threshold parameter, the network size, and the number of neighbors that a node has. In the five sub-experiments that the threshold parameter is changed, their average clustering coefficients are close to each other and almost on the line of that of initial small world network. When the links that will be changed are five times the total edges, the average clustering coefficient of each sub-experiment has slightly decreased and the network with the least threshold value (sub-experiment 1 with the threshold parameters THR=0.1) is more similar to the initial network in terms of average clustering coefficient. Namely, as the fraction of links that are changed by this mechanism increases, it is observed that the experiment with low threshold parameters is similar to the small world network generated at first in terms of average clustering coefficient. While the threshold value is lower, it is expected that links are changed by the triadic closure 1 mechanism more. Since the average clustering coefficient of the network in the sub-experiment 1 is closer to that of the initial network as the fraction of links that are changed increases, it can be concluded that the triadic closure 1 mechanism is more effective than random mechanism to preserve the properties of the initial network over time. (R+ LCD-) and (HB1+ LCD-) mechanisms are also compared to each other after the network size and average neighbor are changed from 2000 to 3000 and from 40 to 30, respectively.

In this study, (HB1+ LCD-) provides similar result to the initial network in terms of average clustering coefficient and the average shortest path for all snapshots.

Other than the threshold parameter chosen for hybrid 1 link addition mechanism, sub-experiments with (HB1+ LCD-) mechanism combination implemented on small world networks are also conducted on networks of different sizes. According to the results of the experiments that the network sizes are taken as 2000 and 3000, it can be concluded that average clustering coefficient has been influenced relatively less than from the number of nodes. It can be an implication that (HB1+ LCD-) can also be used in larger networks. As expected, the average shortest path has increased proportional to the network size.

Sub-experiments with (HB1+ LCD-) mechanism combination implemented on small world networks have also been conducted for different average degree. As the average degree increases, nodes can reach each other in fewer steps because the number of total links increases due to the increase in average degree. This explains the decrease in the average shortest path. The clustering coefficient is relatively higher because the probability of being clustered of the network increases when average degree increases and the network in the experiment 7 (with k=40) is similar to that of the first network for all snapshots considered.

Additionally, (HB2 + LCD-) mechanism combination implemented on the small world network makes the characteristic of the initial network preserve. When the number of links added and deleted by this mechanism are two times the total links (in the network at $t_{2.00}$), the average shortest path and the average clustering coefficient have reached a steady state. When the number of links added and deleted is five times the total links (in the network at $t_{5.00}$), the average shortest path increases from 2.82 to 3.36, and the average clustering coefficient increases from 0.53 to 0.69. It is seen that the network at $t_{5.00}$ is similar to the small world network whose rewiring probability is 0.02.

Some mechanism combinations implemented on a random network or a small world network change the network properties. (PA+ R-) and (PA+ LCD-) are examples of these combinations. When each of them is implemented on a random network, the average clustering coefficient increases dramatically while the average shortest path drops. One reason underlying is that the nodes with high degrees (top five nodes in the analysis) are linked with each other over time. The second reason is that nodes having medium level degrees have tendency to interact with more central nodes as the number of snapshots proceeds. For each experiment conducted by (PA+ R-) and (PA+ LCD-), since the network modified has few nodes having high degrees, the degree distribution becomes positively skewed over time compared to the degree distribution of the initial network. When the experiments carried out by these combinations are compared to each other, similar results are observed although link deletion mechanisms are different from each other. This reveals that the influence of PA link addition mechanism is higher than R+ and LCD- link deletion mechanisms on the results of the experiments. When the same combinations are implemented on a small world network, the network modified converges to a random network at first. Then it changes in parallel with modification of the random network mentioned before.

Another mechanism combination that changes the structure of the network when it is implemented on a small world network is (TR2 + LCD-). While TR2 adds a link with an increasing rate based on the common neighbors of the neighbors of the first node, LCD deletes a link of the node pair having the least common neighbor. In this experiment, the network structure shifts in different snapshots. When the number of links changed is equal to the total links (in the network at $t_{1.00}$), the network is similar to the lattice network. The link addition mechanism makes the average shortest path increase. Because a node has high probability to be linked with the other node having common neighbors, but the first node has no probability to be linked with any other nodes that are not its second level neighbors. Due to both link addition and link deletion mechanisms, average clustering coefficient has increased. Since the links are deleted according to the least common degree, the deletion of the links connecting the subgraphs to each other causes the network to be fragmented over time.

Therefore, average shortest path decreases dramatically. Since the network is fragmented to the sub-graphs (approximately 10 subgraphs in this study), different number of average shortest paths have been created. Each sub-graph has started to be clustered in itself. It is concluded that (TR2+ LCD-) converts the structure of the small world network into lattice at first compared to initial network. Then, it causes the network to be fragmented.

There are two similar studies done, one of them is Davidsen *et al.* [32] and the other is Jin *et al.* [37]. Davidsen *et al.* has studied on a dynamical model where a network can be modified by local interactions. The mechanisms that are considered are transitive linking (similar to TR1 link addition mechanism in this study but not exactly the same) and the death and birth of nodes. This study demonstrates how the degree distribution changes as the death and birth rate (p) is increased or decreased. It is concluded that the degree distribution of the evolving network having small p values is a power law. Thanks to transitive linking, the evolving network has a high clustering coefficient. Their experiment results are similar to the results of (TR1+ LCD-), (R+ LCD-), (HB1+ LCD-) and (HB2+ LCD-) combinations implemented on small world networks in this study in terms of high clustering coefficient. The degree distributions of these four experiments are bell curve shaped and there is almost no skewness in any side. In that sense, they are similar to the degree distribution of the initial small world network. In the model studied by Davidsen *et al.* [32], the degree distribution is in between the exponential distribution and the degree distribution with a power law tail depending on the death and birth of nodes. The results of (PA+ R-) and (PA+ LCD-) mechanism combinations carried out on random networks and small world networks are similar to those of Davidsen *et al.* in terms of average clustering coefficient. As seen in Figure 6.1, the clustering coefficient of the network whose p value is 0.0025 is 0.63. In this study, the average clustering coefficients of the networks at $t_{3.00}$, $t_{4.00}$ and $t_{5.00}$ modified by the (PA+ R-) and (PA+ LCD-) mechanism combinations are shown in Table 6.1.

Since the number of nodes is 7000 for the study of Davidsen and 2000 for this study, one-to-one comparison cannot be made, but it is deduced that the study of Davidsen *et al.* with small p values, and (PA+ R-) and (PA+ LCD-) experiments of this study are similar in terms of average clustering coefficient.

Table 6.1. The ACC of the Networks by (PA+ R-) and (PA+ LCD-).

Snapshot of the network	PA+ R- on RN	PA+ LCD- on RN	PA+ R- on SWN	PA+ LCD- on SWN
The network at $t_{3.00}$	0.33	0.61	0.23	0.52
The network at $t_{4.00}$	0.63	0.73	0.57	0.58
The network at $t_{5.00}$	0.76	0.75	0.75	0.66

In the study of Davidsen *et al.* [32], it is concluded that the degree distribution tends to power law as p decreases in Figure 6.1. To illustrate, the degree distribution with log-log scale of the network at $t_{5.00}$ in (PA+ LCD-) experiment implemented on a small world network is demonstrated in Figure 6.2. In order to understand if the modified network is similar to exponential or power law, power law package in Python is downloaded and $distribution\_compare('powerlaw', 'exponential')$ function is used. If the ratio measuring whether the network is similar to power law or exponential is higher than zero, the network is similar to the first one in this function. Otherwise, it is similar to the second [40]. Additionally, degree exponent is also calculated. In the literature, when it is between two and three, the network is in scale free regime [11]. In this network, that value is 2.92. Since the ratio is higher than zero and the degree exponent is between two and three, it is deduced that the network at $t_{5.00}$ in the (PA+ LCD-) experiment has power law property. Therefore, it is concluded that the network in the (PA+ LCD-) experiment is also similar to the network studied by Davidsen *et al.* in terms of degree distribution.

The networks at $t_{3.00}$ modified by (PA+ R-) on random networks and small world networks have relatively low average clustering coefficient than those networks modified by (PA+ LCD-) because of the random link deletion mechanism.

The average clustering coefficient of the network at $t_{4.00}$ is similar to that of the study of Davidsen *et al.* [32]. It is also deduced that a small world network behavior with a power law degree distribution can be reached by only making changes on the links, without making changes on the nodes. Another difference from the study of Davidsen *et al.*, in this study, while the small world networks modify, network modification combinations are discovered which makes that network preserve its characteristics (average shortest path, average clustering coefficient and degree distribution).

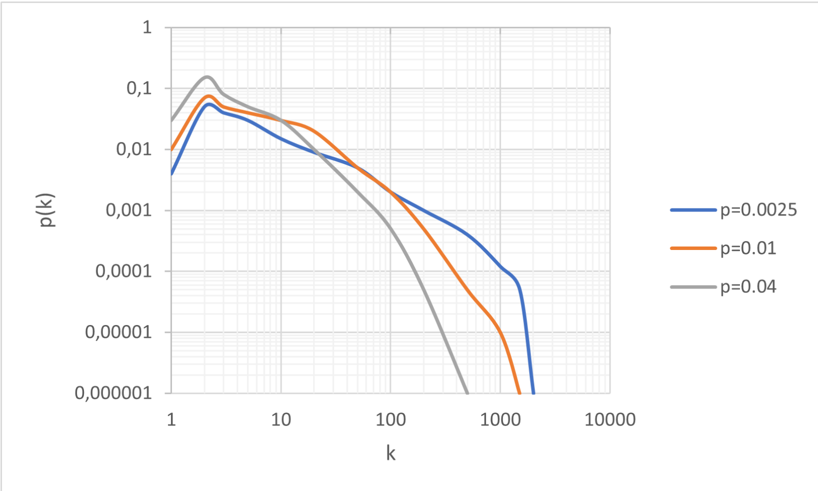

Figure 6.1. The Degree Distribution of Networks Depending on p.

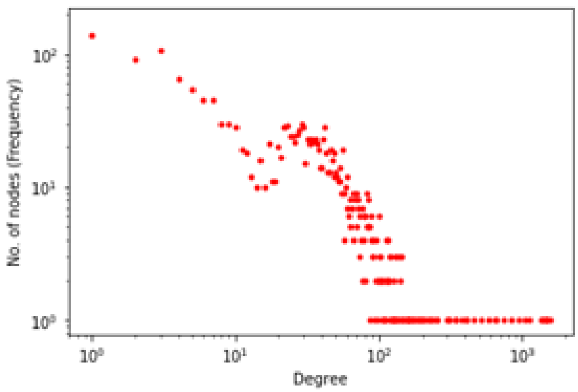

Figure 6.2. The Degree Distribution of the Network at $t_{5.00}$ Modified by PA+ LCD-.

Jin *et al.* [37] has studied on models for evolution of social networks. These models have generated networks having high clustering coefficient (e.g. 0.45 and 0.53).

The average clustering coefficient of the random network with the same parameters is 0.02. Since the maximum link number is determined, the network studied by Jin *et al.* has the degree distribution with a sharp peak but there are no figures related to degree distribution in order to compare with this study. Additionally, average shortest path has not been calculated in their study. In this respect, this study makes a difference since the clustering coefficient, degree distribution, and shortest path metric results will be calculated and compared for each experiment. The link deletion parameter is increased in the model of Jin *et al.* after the network reaches a certain size. However, in this study, link addition and deletion are not carried out separately from each other. On the contrary, every time a link is added, another link is deleted. When the study of Jin *et al.* is compared to this one in terms of clustering coefficient, it is seen that the results of (TR1+ LCD-), (R+ LCD-), (HB1+ LCD-) and (HB2+ LCD-) mechanism combinations show similarity with the model by Jin *et al.*

7. CONCLUSION

In this study, network modification combinations are discovered that makes preserving of characteristics of small world network. It is analyzed whether the structure of the network deteriorates (e.g. convergence to a random network) or how it changes (e.g. convergence to lattice), thanks to the link addition and deletion mechanisms implemented on both the random networks and the small world networks.

As a result of the experiments done, due to the fact that the degree link deletion mechanism tends to fragment the structure of the network, the experiments using degree link deletion mechanism have not ensured network stability. Additionally, in most of the experiments that the random link deletion mechanism is used, this mechanism deteriorates the network regularity and therefore the modified network has had random network characteristics rather than small world network characteristics when the number of links deleted by random increases. 12 experiments are conducted where the random link deletion mechanism is used. Two of them, (PA+ R-) modification combination implemented on random networks and small world networks, give promising results using preferential attachment link addition. The others either remain random networks or converge to random networks.

(R+ LCD-), (TR1+ LCD-), (TR2+ LCD-), (HB1+ LCD-) and (HB2+ LCD-) mechanism combinations implemented on a random network make the degree distribution flatten over time. Compared to the initial random network, the modified networks with these mechanism combinations make average clustering coefficient increase over time. The networks modified by (PA+ R-) and (PA+ LCD-) mechanism combinations have high clustering coefficient and low average shortest path with positively skewed degree distribution. Because nodes with very high degrees have tendency to link with each other and the ones with medium level degrees tend to interact with more central nodes over time.

(TR2+ LCD-) mechanism implemented on a small world network makes the average shortest path increase considerably at first and then cause the network to be fragmented. (R+ LCD-), (TR1+ LCD-), (HB1+ LCD-) and (HB2+ LCD-) mechanism combinations make a small world network modify dynamically by preserving its network characteristics. As seen, the link deletion mechanisms of the experiments with these four mechanism combinations are the same, LCD. When a small world network is modified by random link addition and random link deletion, the modified network is a random network. However, when the link deletion mechanism is replaced with the least common degree, the structure of the initial network is preserved, which shows the importance of LCD.

Unlike the two studies discussed in the previous section, in this study, link addition and deletion mechanism combinations have discovered that makes a small world network modify dynamically by preserving its network characteristics. In addition, relevant mechanism combinations have been discovered that creates networks with high average clustering coefficient and the power law distribution by only modifying the links. Lastly, it has been observed that networks with high clustering coefficients are also created when link addition and deletion processes occur at the same time.

For future studies, several research areas can be mentioned. One of them might be to study on propagation. For instance, a propagation model can be studied in a static network at first. Then, the same model can be studied on dynamic small world networks using (TR1+ LCD-), (R+ LCD-), (HB1+ LCD-) and (HB2+ LCD-) mechanisms, separately. After that, the same propagation model is run on a real network containing small world features. Lastly, they are compared to each other and it can be found in which experiment the propagation is more similar to the propagation of the real network. As a second suggestion, the promising network modification combinations (R+ LCD-, PA+ R-, PA+ LCD-, TR1+ LCD-, TR2+ LCD-, HB1+ LCD- and HB2+ LCD-) can be implemented on several real networks. As a starting point, a snapshot of a real network can be selected. Then, the combinations are implemented on that network.

Lastly, the results based on combinations and the real network itself are compared to each other and it is deduced which combination explains the working mechanism of a real network better. As a third suggestion, the same sub-experiment in 5.2.6 (HB1+ LCD- experiment on small world networks) can be carried out for larger network size (due to computational complexity and high number of experiments, 2 different network sizes have been used in this study, 2000 and 3000) in order to understand if the average clustering coefficients remain almost the same as the number of snapshots proceeds. Thus, the inference related that HB1+ CD- experiment on the small world network can produce meaningful results for larger network size can be strengthened. Lastly, another study can be conducted to identify the relevant mechanism combinations that make the scale-free network change over time by preserving its properties.

REFERENCES

1. Williamson, S. and E. Bender, *Lists, Decisions and Graphs*, University of California, San Diego, 2010.

2. Diestel, R., *Graph Theory*, Springer, Heidelberg, 2000.

3. Kruskal, J., “On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem”, *Proceedings of the American Mathematical Society*, American Mathematical Society, 1956.

4. Prim, R., “Shortest Connection Networks and Some Generalizations”, *Bell System Technical Journal*, Vol. 36, pp. 1389–1401, 1957.

5. Aittokallio, T. and B. Schwikowski, “Graph-based Methods for Analysing Networks in Cell Biology”, *Briefings in Bioinformatics*, Vol. 7, pp. 243–55, 2006.

6. Balaban, A., “Applications of Graph Theory in Chemistry”, *Journal of Chemical Information and Computer Sciences*, Vol. 25, pp. 334–343, 1985.

7. Goldenberg, A., A. Zheng, S. Fienberg and E. Airoldi, “A Survey of Statistical Network Models”, *Foundations and Trends in Machine Learning*, Vol. 2, pp. 129–233, 2009.

8. Stattner, E., M. Collard and N. Vidot, “Diffusion in Dynamic Social Networks: Application in Epidemiology”, *Database and Expert Systems Applications (DEXA) 2011 - 22nd International Conference*, pp. 559–573, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

9. Milgram, S. L., “The Small World Problem”, *Psychology Today*, Vol. 2, pp. 60–67, 1967.

10. Barabasi, A.-L. and R. Albert, “Emergence of Scaling in Random Networks”, *Science*, Vol. 286, pp. 509–12, 1999.

11. Barabási, A.-L. and M. Pósfai, *Network Science*, Cambridge University Press, Cambridge, 2016.

12. Easley, D. and J. Kleinberg, *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*, Cambridge University Press, Cambridge, 2010.

13. Watts, D. and S. Strogatz, “Collective Dynamics of ‘Small-World’ Networks”, *Nature*, Vol. 393, pp. 440–442, 1998.

14. Dong, X., G. Xiong and J. Hou, “Construction of Artificial Grid Systems Based on ACP Approach”, *Service Science, Management, and Engineering, First Edition*, Intelligent System Series, chap. 14, pp. 285–304, Academic Press, Beijing, 2012.

15. Wilensky, U., *NetLogo Models Library Sample Models and Networks*, `https://ccl.northwestern.edu/netlogo/models/GiantComponent`, accessed in October 2020.

16. Lim, F., “A Review-Analysis of Network Topologies for Microenterprises”, *Advanced Science and Technology Letters- CES-CUBE 2016*, Vol. 135, pp. 175–180, SERSC, Quezon, Philippines, 2016.

17. Erdős, P. and A. Rényi, “On Random Graphs I”, *Publicationes Mathematicae Debrecen*, Vol. 6, pp. 290–297, 1959.

18. Hexmoor, H., “Computational Network Science: An Algorithmic Approach”, *Computer Science Reviews and Trends*, Emerging Trends in Computer Science and Applied Computing, pp. 33–43, Morgan Kaufmann, Boston, 2014.

19. Golbeck, J., *Analyzing the Social Web*, Morgan Kaufmann, Waltham, 2013.

20. Karinthy, F., “Láncszemek (Chains)”, *Everything Is Different*, pp. 1–4, 1929.

21. De Sola Pool, I. and M. Kochen, “Contacts and influence”, *Social Networks*, Vol. 1, No. 1, pp. 5–51, 1978.

22. Watts, D., *Six Degrees: The Science of a Connected Age*, W. W. Norton, New York, 2003.

23. Benz, W. and E. Asphaug, “Simulations of Brittle Solids Using Smooth Particle Hydrodynamics”, *Computer Physics Communications*, Vol. 87, pp. 253–265, 1995.

24. Erkol, Ş. and G. Yücel, “Influence Maximization Based on Partial Network Structure Information: A Comparative Analysis on Seed Selection Heuristics”, *International Journal of Modern Physics C*, Vol. 28, p. 1750122, 2017.

25. Bollobás, B. and O. Riordan, “The Diameter of a Scale-Free Random Graph”, *Combinatorica*, Vol. 24, pp. 5–34, 2004.

26. Wang, X. and G. Chen, “Complex Networks: Small-World, Scale-Free, and Beyond”, *IEEE Circuits and Systems Magazine*, Vol. 3, pp. 6–20, 2003.

27. Albert, R. and A.-L. Barabasi, “Statistical Mechanics Of Complex Networks”, *Reviews of Modern Physics*, Vol. 74, pp. 47–97, 2001.

28. Bianconi, G. and A.-L. Barabasi, “Competition and Multiscaling in Evolving Networks”, *Europhysics Letters*, Vol. 54, p. 436, 2001.

29. Nguyen, V.-A., C. Leung and E.-P. Lim, “Modeling Link Formation Behaviors in Dynamic Social Networks”, *Social Computing, Behavioral-Cultural Modeling and Prediction: 4th International Conference*, pp. 349–357, Springer Verlag, College Park, 2011.

30. Newman, M., “Clustering and Preferential Attachment in Growing Networks”,

Physical Review E - Statistical, Nonlinear, and Soft Matter Physics, Vol. 64, pp. 1–13, 2001.

31. Toivonen, R., L. Kovanen, M. Kivelä, J.-P. Onnela, J. Saramäki and K. Kaski, "A Comparative Study of Social Network Models: Network Evolution Models and Nodal Attribute Models", *Social Networks*, Vol. 31, pp. 240–254, 2009.

32. Davidsen, J., H. Ebel and S. Bornholdt, "Emergence of a Small World from Local Interactions: Modeling Acquaintance Networks", *Physical Review Letters*, Vol. 88, pp. 1–4, 2002.

33. Marsili, M., F. Vega-Redondo and F. Slanina, "The Rise and Fall of a Networked Society: A Formal Model", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 101, pp. 1439–1442, 2004.

34. Kumpula, J. M., J.-P. Onnela, J. Saramäki, K. Kaski and J. Kertész, "Emergence of Communities in Weighted Networks", *Physical Review Letters*, Vol. 99, p. 228701, 2007.

35. Stattner, E., "Involvement of Node Attributes in the Link Formation Process into a Telecommunication Network", *Social Network Analysis and Mining*, Vol. 5, p. 64, 2015.

36. Wang, J. and P. De Wilde, "Properties of Evolving E-mail Networks", *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, Vol. 70, p. 066121, 2004.

37. Jin, E., M. Girvan and M. Newman, "Structure of Growing Social Networks", *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, Vol. 64, pp. 461321–461328, 2001.

38. Hagberg, A. A., D. A. Schult and P. J. Swart, "Exploring Network Structure, Dynamics, and Function Using NetworkX", G. Varoquaux, T. Vaught and J. Millman

(Editors), *Proceedings of the 7th Python in Science Conference (SciPy 2008)*, pp. 11–15, Pasadena, CA USA, 2008.

39. Rossetti, G., L. Milli, S. Rinzivillo, A. Sîrbu, D. Pedreschi and F. Giannotti, "NDlib: a Python Library to Model and Analyze Diffusion Processes Over Complex Networks", *International Journal of Data Science and Analytics*, Vol. 5, pp. 61–79, 2018.

40. Alstott, J., E. Bullmore and D. Plenz, "Powerlaw: A Python Package for Analysis of Heavy-Tailed Distributions", *Plos One*, Vol. 9, pp. 1–11, 2014.