Clustering Coefficient Increasing Local Link Switching Algorithm

YihjiaTsai

Department of Computer Science of Information Engineering
Tamkang University
151 Ying-chuan Road, Tamsui, Taipei
TAIWAN
eplusplus@gmail.com http://iAIC.csie.tku.edu.tw/

Abstract: This paper proposed a local link switching algorithm which effectively increases the clustering coefficient of a network while preserving the network node degree distributions. This link switching algorithm is based on local neighborhood information. Link switching algorithm is widely used in producing similar networks with the same degree distribution, that is, it is used in 'sampling' networks from the same network pool. Therefore, the switching pairs of links are selected rather 'globally' from the network. The proposed algorithm focus on increasing an important network characteristic while selecting candidate pairs of links locally. Clustering coefficient characterizes the relative tightness of a network and is a defining network statistics that appears in many 'real-world' network data. Simulation results on three different types of model networks have demonstrate the effectiveness of this algorithm.

Key words: Local link switching, clustering coefficient, complex networks

1 Introduction

Over the past decade the studies of complex networks have yielded many valuable insights into the nature of "real-world" networks [1, 2, 5, 10]. Results from those researches provides more detailed picture in summarizing, categorizing and comparing complex data sets arising naturally from real world. Among those network structural statistics, clustering coefficient is an important concept in summarizing the relationship among neighbors in a network [1, 4].

Networks with high clustering coefficients are often considered to be modeled naturally by small world networks. In this paper, a local link switching algorithm is proposed to produce networks with high clustering coefficients. In the past, link switching technique is used in generating connected networks with prescribed degree sequence [6, 7, 8]. The selection of links to be switched is a random process so the algorithm can effectively, uniformly generate a large number of networks with the same degree sequence. The proposed algorithm works locally with the objective of increasing the network clustering coefficients.

2 Notations

Before introducing the algorithm, basic notations of network are presented in this section. A network $G = \{V, E\}$ consists of two sets, $V = \{1, 2, \cdots, N\}$ is the set of vertices and $E = \{e_{ij} \mid i, j \in V\}$ is the set of edges. The existence of edge e_{ij} in network G is denoted by $e_{ij} = 1$. The number of elements in set V is represented by its absolute value $N = |V|$ and the number of edges for a set of nodes V is denoted by $E_V = |E|$. Let $\beta(i)$ be the set of neighboring nodes of i, $\beta(i) = \{n \mid n \in V, e_{ni} \in E\}$, and the number of elements in set $\beta(i)$ is denoted by $\beta_i = |\beta(i)|$. Node degree d_i of a node i is defined as the number of links node i has, it follows that $d_i = \beta_i$. The intersection of two sets $\beta(i)$ and $\beta(j)$ is represented by $O(i, j) = \beta(i) \cap \beta(j)$ and the number of elements in that set is $O_{i,j} = |O(i, j)| = \beta(i) \cap \beta(j)$. The clustering coefficient of node i is given by the formula $C_i = 2E_{\beta(i)} / \beta_i(\beta_i - 1)$. The clustering coefficient of network G is therefore the average of all clustering coefficient of the vertices in the network $C_V = \sum_{i=1}^{N} C_i / N$.

3 Local Link Switching Algorithm

Link switching is a technique used in many studies of network properties such as generating random uniform graphs with a prescribed degree sequence and random walk behavior in a sequence of Markov chains [6]. Link switching refers to the selection of two links (i_1, i_2), (j_1, j_2), and replace them by the pair of links (i_1, j_1), (i_2, j_2) provided that there is no direct link between i_1 and j_1, i_2 and j_2, as shown in Figure 1. An alternative replacement can be by the pair of links (i_1, j_2) and (i_2, j_1) under the condition that those two direct links do not exist before switching. The selections of those two links are considered randomly and uniformly from the whole network and the main objective of link switching is to create different networks with the same degree characteristics.

In this section, the proposed local link switching is discussed and analyzed with the focus on increasing clustering coefficients. Local link switching refers to the selection of involving nodes is from the neighboring nodes or the next-neighboring nodes, rather than from the whole network.

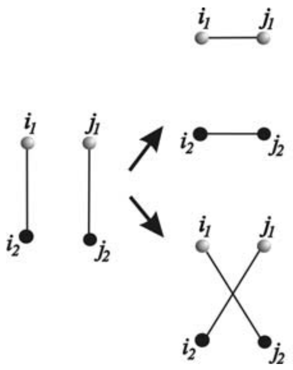


Figure 1 node degree preserving link switching strategy. By switching between a pair of links, the node degree of the involving four nodes remains the same after the link switching operation.

The proposed local link switching strategy begins with a given node, i_1, which can be called the *pivot* node. The first step is to select a *target* node j_1 from among those nodes that have the largest number of common neighbors with pivot node i_1 but without direct link connecting to i_1. After the selection of node j_1, two other candidate nodes are to be determined next. Consider the location of those two candidate nodes in the neighboring nodes of both nodes i_1 and j_1, there are three different scenarios in selecting the two remaining candidate nodes.

I. Both candidate nodes are not in the common neighboring area of nodes i_1 and j_1.
II. One candidate nodes are in the common neighboring area of nodes i_1 and j_1.
III. Both candidate nodes are in the common neighboring area of nodes i_1 and j_1.

As shown in Figure 2 (a), candidate nodes i_2 and j_2 are neighbors of nodes i_1 and j_1 respectively. However, both candidate nodes are not in the common neighboring area of i_1 and j_1.

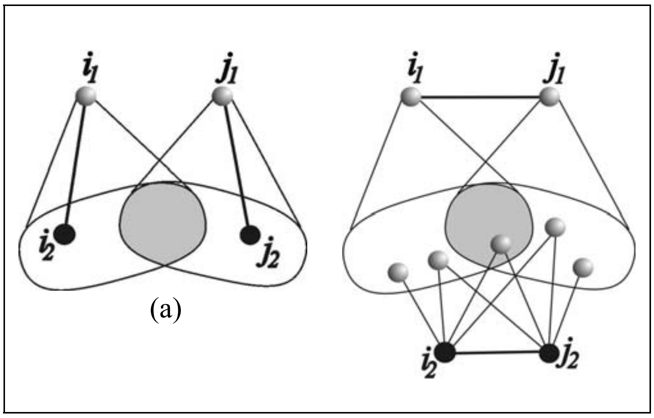


Figure 2 (a) nodes i_1 and j_1 locate in their neighboring nodes i_2 and j_2 respectively and swapping connecting links (i_1, i_2), (j_1, j_2) with links (i_1, j_1) and (i_2, j_2). (b) after swapping, candidate nodes are no longer belongs to the neighboring are of both i_1 and j_1.

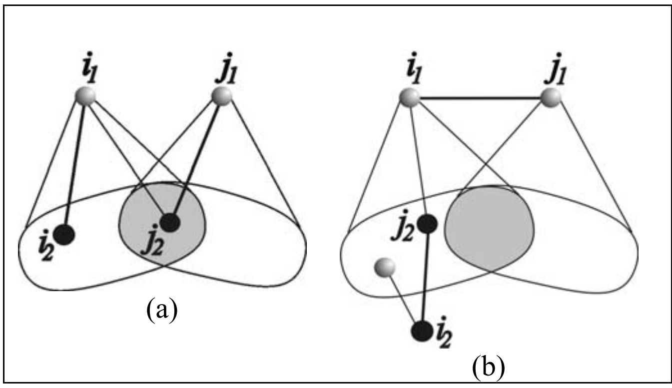


Figure 3 (a) one candidate node, in this case, node j_2, is selected from the common neighboring are of nodes i_1 and j_1 while the other candidate node i_2 is a neighboring node of node i_1 only. (b) after link switching, node i_2 is not a neighboring node of either node i_1 or j_1 while node j_2 becomes a neighbor of i_1 only.

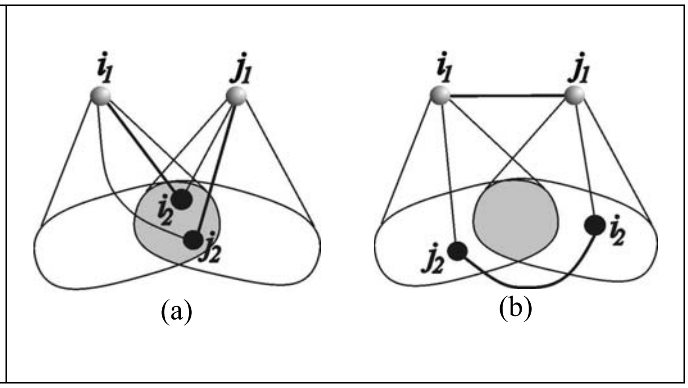


Figure 4 (a) both candidate nodes are located in the common neighbors of nodes i_1 and j_1. (b) after link switching, both candidate nodes are excluded from the common neighbors while remaining as a neighbor of either node i_1 or j_1.

The second scenario is that one candidate node is selected from the common neighbors of both pivot node i_1 and target node j_1, as illustrated in Figure 3(a). This case has two variations, either i_2 or j_2 can be the one in the common neighbor area, however, both cases can regard as one in the analysis. In Figure 4(a) is the third scenario, where both candidate nodes are selected from the common neighbors of both i_1 and j_1. After link switching operation, both nodes remain in the neighbor of pivot node i_1 and target node j_1.

Now we proceed to analyze the change of clustering coefficients under those three scenarios. The strategy is to focus on those four involving nodes and their neighboring areas, as clustering coefficients in those are affected under link switching operation. As illustrated in Figure 5(a), nodes i_1 and j_1 have a common neighboring area colored in grey. A node m_1 is used to show both links connecting to nodes i_1 and j_1. Suppose the node degree of i_1 is d_{i_1}, the clustering coefficient of node i_1 is

$$C_{i_1} = \frac{\sum_{k \in \beta_{i_1}} |\beta(k) \cap \beta(i_1)|}{d_{i_1}(d_{i_1}-1)/2}.$$

The switching of a link connecting i_1 and i_2 to form a new link between i_1 and j_1, as illustrated in Figure 5(b), will add to the numerator by an amount of $|\beta(i_1) \cap \beta(j_1)|$, which is the number of nodes in the grey area. This operation will also subtracted an amount from the numerator, which is $|\beta(i_1) \cap \beta(i_2)|$.

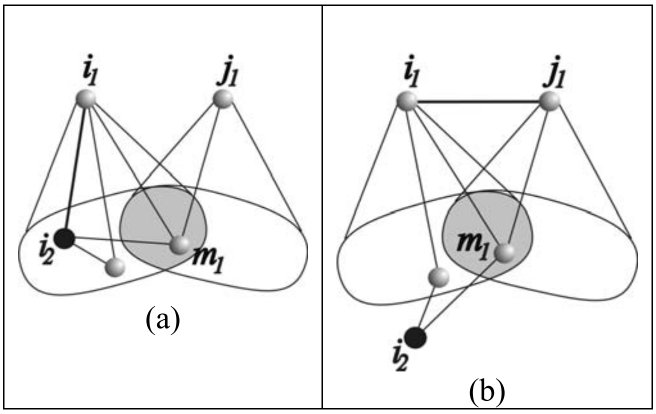


Figure 5 (a) two separated nodes i_1 and j_1 with a common neighboring area, colored in grey. (b) the switching of a link between i_1 and i_2 to will increase the degree of node i_1 by 1 and the number of links between neighboring nodes of i_1 by an amount of the number of nodes in the common area.

The degree of node i_1 remains the same after the switching operation. Therefore, the resulting clustering coefficient after link addition is

$$C'_{i_1} = \frac{\sum_{k\in\beta_{i_1}} |\beta(k)\cap\beta(i_1)| + |\beta(j_1)\cap\beta(i_1)| - |\beta(i_2)\cap\beta(i_1)|}{d_{i_1}(d_{i_1}-1)/2}$$

Therefore, the resulting change of clustering coefficient in node i_1 is

$$\triangle C_{i_1} = \frac{|\beta(j_1)\cap\beta(i_1)| - |\beta(i_2)\cap\beta(i_1)|}{d_{i_1}(d_{i_1}-1)/2}$$

Which can be written as $\triangle C_{i_1} = \dfrac{O_{i_1,j_1} - O_{i_1,i_2}}{d_{i_1}(d_{i_1}-1)/2}$

For those nodes in the common neighboring area, their clustering coefficient is increased by a fixed amount. For node m_1 with degree m_1 in the common area, its change of clustering coefficient from link addition is

$$\triangle C_{m_1} = \frac{1}{d_{m_1}(d_{m_1}-1)/2}$$

Thus for the whole common area, the change of clustering coefficient will be

$$\triangle C_{\beta(i_1)\cap\beta(j_1)} = \sum_{m_1\in\beta(i_1)\cap\beta(j_1)} \frac{1}{d_{m_1}(d_{m_1}-1)/2}$$

Based on similar analysis, for the three scenarios shown in Figure 2(a), 3(a), 4(a), we can calculate the change of clustering coefficient for those three cases and summarize the results in Table 1.

Table 1 change of clustering coefficients for the three scenarios

	Case I.	Case II.	Case III.
$\triangle C_{i_1}$	$\frac{O_{i_1,j_1}-O_{i_1,i_2}}{d_{i_1}(d_{i_1}-1)/2}$	$\frac{O_{i_1,j_1}-1-O_{i_1,i_2}}{d_{i_1}(d_{i_1}-1)/2}$	$\frac{O_{i_1,j_1}-2-O_{i_1,i_2}}{d_{i_1}(d_{i_1}-1)/2}$
$\triangle C_{j_1}$	$\frac{O_{i_1,j_1}-O_{j_1,j_2}}{d_{j_1}(d_{j_1}+1)/2}$	$\frac{O_{i_1,j_1}-1-O_{j_1,j_2}}{d_{j_1}(d_{j_1}+1)/2}$	$\frac{O_{i_1,j_1}-2-O_{j_1,j_2}}{d_{j_1}(d_{j_1}+1)/2}$
$\triangle C_{i_2}$	$\frac{O_{i_2,j_2}-O_{i_1,i_2}}{d_{i_2}(d_{i_2}-1)/2}$	$\frac{O_{i_2,j_2}-1-O_{i_1,i_2}}{d_{i_2}(d_{i_2}-1)/2}$	$\frac{O_{i_2,j_2}-2-O_{i_1,i_2}}{d_{i_2}(d_{i_2}-1)/2}$
$\triangle C_{j_2}$	$\frac{O_{i_2,j_2}-O_{j_1,j_2}}{d_{j_2}(d_{j_2}+1)/2}$	$\frac{O_{i_2,j_2}-1-O_{j_1,j_2}}{d_{j_2}(d_{j_2}+1)/2}$	$\frac{O_{i_2,j_2}-2-O_{j_1,j_2}}{d_{j_2}(d_{j_2}+1)/2}$
$\triangle C_{O_{i_1,j_1}}$	$\sum_{m_1\in O_{i_1,j_1}} \frac{1}{d_{m_1}(d_{m_1}+1)/2}$	$\sum_{m_1\in O_{i_1,j_1},\ m_1\neq j_2} \frac{1}{d_{m_1}(d_{m_1}+1)/2}$	$\sum_{m_1\in O_{i_1,j_1},\ m_1\neq i_2,j_2} \frac{1}{d_{m_1}(d_{m_1}+1)/2}$
$\triangle C_{O_{i_2,j_2}}$	$\sum_{m_1\in O_{i_2,j_2}} \frac{1}{d_{m_1}(d_{m_1}+1)/2}$	$\sum_{m_1\in O_{i_2,j_2},\ m_1\neq j_1} \frac{1}{d_{m_1}(d_{m_1}+1)/2}$	$\sum_{m_1\in O_{i_2,j_2},\ m_1\neq i_1,j_1} \frac{1}{d_{m_1}(d_{m_1}+1)/2}$
$\triangle C_{O_{i_1,i_2}}$	$\sum_{m_1\in O_{i_1,i_2}} \frac{-1}{d_{m_1}(d_{m_1}+1)/2}$	$\sum_{m_1\in O_{i_1,i_2}} \frac{-1}{d_{m_1}(d_{m_1}+1)/2}$	$\sum_{m_1\in O_{i_1,i_2},\ m_1\neq j_1} \frac{-1}{d_{m_1}(d_{m_1}+1)/2}$
$\triangle C_{O_{j_1,j_2}}$	$\sum_{m_1\in O_{j_1,j_2}} \frac{-1}{d_{m_1}(d_{m_1}+1)/2}$	$\sum_{m_1\in O_{j_1,j_2}} \frac{-1}{d_{m_1}(d_{m_1}+1)/2}$	$\sum_{m_1\in O_{j_1,j_2},\ m_1\neq i_1} \frac{-1}{d_{m_1}(d_{m_1}+1)/2}$

4 Model networks

To investigate the effectiveness of the proposed algorithm, we use three different kinds of network models. The first model is two dimensional grid networks and the second one is random networks while the third is the small world networks. The clustering coefficient of two dimensional grid networks is zero for every node. Random networks are known to have low clustering coefficient for low link probability. Small world networks have a higher clustering coefficient as compared with random networks.

Two dimensional grid networks are a special kind of regular networks where each node has the same amount of neighboring nodes except those nodes locating at the border of the grid. Because there is no direct link connecting the next neighboring nodes, therefore, the clustering coefficients of every node in the two dimensional grid networks is zero.

Random networks are a network model that has been studied for over 40 years [3]. Before the introduction of small world network models, and scale-free models, for large and seemingly irregular network data, random networks are a natural candidate. For a fixed sized random network, the existence of any link is determined by a probability. It is known that such random link creation process results in a normal node degree distribution. Another property of random network is that for low link creating probability, the resulting network has a rather low clustering coefficient.

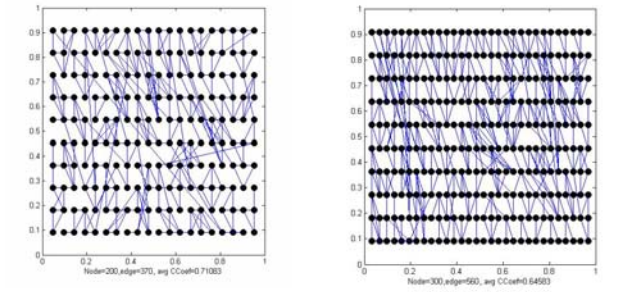
As compared with random network models, small world networks are introduced rather recently [11]. Small world refers to the fact that in a friendship network, one's friends tend to be friends of each other. In more concrete terms, the clustering coefficient of a small world network is relatively high as compared with random network. A small world network is created combining a regular network and a low probability random network. The underlying regular network gives small world network its high clustering coefficients.

Two network sizes are used in the study the effect of local link switching operations. The first network size is 200 nodes and the second one contains 300 nodes. In a fixed size grid network, the number of links is fixed. A 200 nodes (10×20) grid network contains 370 links whereas a 300 (10×30) grid network has 560 links. Based on these figures, both random networks and small work networks are selected to contain similar number of edges with the corresponding grid network.

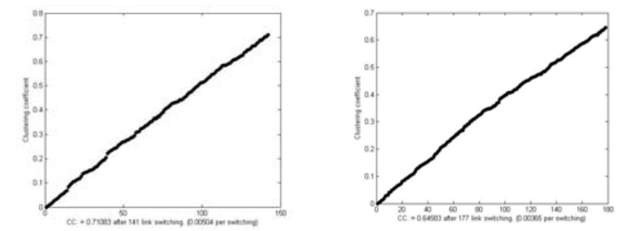
At the beginning of switching operation, a node is chosen randomly to serve as the pivot node, and then the other three candidate nodes are selected based on local information. The second candidate node is selected from the pool of next neighboring nodes without direct link connection with the first pivot node. The selection of the third candidate node is based on the number of common neighboring nodes with the first pivot node. This third candidate node has direct connection with the pivot node while the number of common neighboring nodes with the pivot node is kept at a minimum or zero. The forth candidate nodes is selected by the second candidate with the condition that it has no direct connection with the third candidate node. If failure in selecting any of the four candidate nodes, the iteration is abandoned for the pivot node and another pivot node is selected to begin the next iteration.

After selection all those four candidate nodes, the change of clustering coefficients can be calculated from the formula listed in Table 1. If the resulting summation of change in clustering coefficients is positive, then the link switching will actually take place. Next iteration proceeds with the new network and another pivot node is selected randomly from the entire network.

For each network size, a total of 500 randomly selected pivot nodes are used. However, among those 500 iterations, there are many failed attempts for local link switching operations. The simulation results are summarized in Table 2 which in Figure 6, 7, 8, 9, 10, and 11 are results of different network models of different network size.



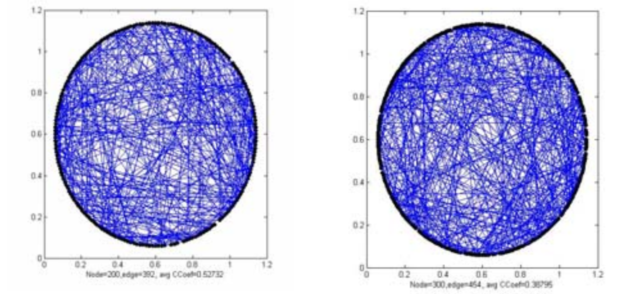
(a) (b)



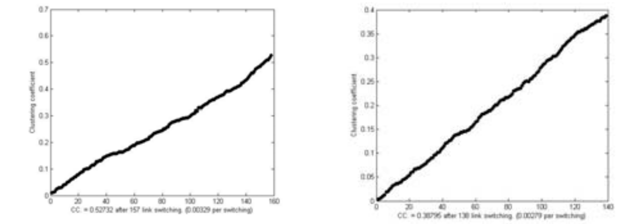
(a) (b)

Figure 6 (a) grid network of size 10×20 after 141 local link switching operations. (b) grid network of size 10×30 after 177 local link switching operations.

Figure 7 (a) change of clustering coefficients in a 10×20 grid network after 141 local link switching. (b) change of clustering coefficients in a 10×30 grid network over 177 local link switching.



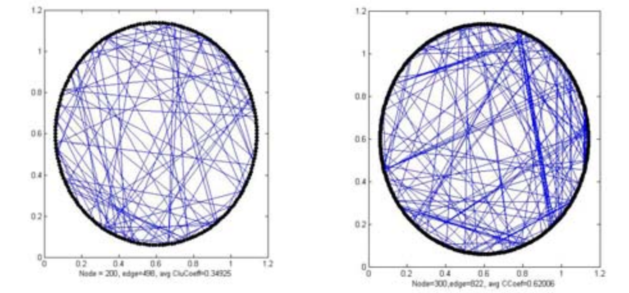
(a) (b)



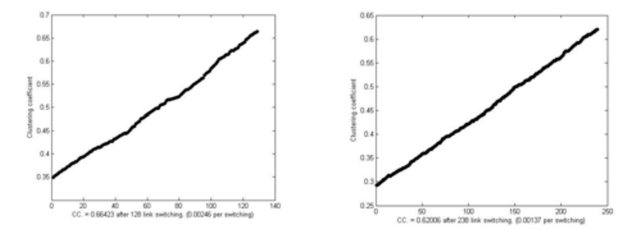
(a) (b)

Figure 8 (a) after 157 local link switching operations, a 200 nodes, 392 edges random network with connection probability 0.02, has a clustering coefficient of 0.52732. (b) Random network with 300 nodes, 560 edges, connection probability 0.01, after 138 local link switching operations, resulted in a network clustering coefficient of 0.38795.

Figure 9 (a) the increase of clustering coefficients in a 200 nodes random network during 157 local link switching. (b) change of clustering coefficients in a 300 nodes random network during the 138 local link switching operations.



(a) (b)



(a) (b)

Figure 10 (a) small world network with 200 nodes and 498 links grid network after 128 local link switching. (b) 300 nodes small world network with 822 links after 238 local link switching.

Figure 11 (a) the increase of clustering coefficients in a 200 nodes random network during 157 local link switching. (b) change of clustering coefficients in a 300 nodes random network during the 138 local link switching operations.

Table 2 results of local link switching on three different network models.

	Grid		Random		Small World	
nodes	200	300	200	300	200	300
edges	370	560	392	454	498	822
cc. orig	0	0	0.01038	0.00344	0.34925	0.29283
cc. after	0.71083	0.64583	0.52732	0.38795	0.66423	0.62006
steps	141	177	157	138	128	238
Incr. per. step	0.00504	0.00365	0.00329	0.00279	0.00246	0.00137
success ratio	0.282	0.354	0.314	0.276	0.256	0.476

5 Simulation results

Table 2 summarizes the results of three network modes. The probability of locating a successful local switching operation for all three kinds of networks is similar, ranging from 25.6% to 35.4%. However, for small world network with a larger number of nodes and links, the probability is a bit higher, up to around 47.6%. For grid network, the increment of clustering coefficient per switching step is higher than the other two networks. As shown in Figure 6(a) and (b), the resulting networks contain many local clusters, which cause the higher clustering coefficients. For random networks, the resulting clustering coefficients are the lowest among the three network models. As compared with grid networks, we can conclude that it is hard to exploit the local structure in random networks with local connecting probability through the proposed local switching algorithm. For small world networks, the unit of increase per switching step is the smallest among the three network models.

Conclusion

Local link strategy can increase the clustering coefficient of a given network. The proposed algorithm works effectively for plane grid networks. This algorithm can produce similar network with the same degree distribution yet different clustering coefficients. One future application of the proposed algorithm is to apply the resulting networks as alternative network models with high clustering coefficients other than the popular small world models.

References

[1] Barrat, A., M. Barthelemy, R. Pastor-Satorras and A. Vespignani, "*The architecture of complex weighted networks,*" PNAS vol. 101 no. 11 3747-3752, March 16, 2004.

[2] Bassett, D.S., E. Bullmore, *"Small-world brain networks,"* Neuroscientist vol. 12 no. 6, pp.512–523, 2006.

[3] Erdős, P. and A. Rényi "*On Random Graphs"* in Publ. Math. Debrecen 6, pp. 290–297, 1959.

[4] Fagiolo, G., *"Clustering in complex directed networks,"* Phys. Rev. E 76, 026107, 2007.

[5] Ferrer-i-Cancho, R. and R.V. Sole, *"The Small World of Human Languag,."* Proceedings of the Royal Society London B 268, 2261-2266, 2001.

[6] Jerrum, M. and A. Sinclair, *"Fast uniform generation of regular graphs,"* Theoretical Computer Science, vol. 73 , iss. 1, pp. 91 -100, June 1990.

[7] McKay, B.D., "*Subgraphs of random graphs with specified degrees,"* Congressus Numerantium, 33, pp.213-223, 1981.

[8] McKay, B.D. and N.C. Wormald, "*Uniform Generation of Random Regular Graphs of Moderate Degree,"* Journal of Algorithms 11, 52-67, 1990.

[9] Raine, D.J. and V. Norris. "*Network Structure of Metabolic Pathway,."* Journal of Biological Physics and Chemistry 1, 89-94, 2001.

[10] Wagner, A. and D.A. Fell. *"The Small World Inside Large Metabolic Networks,"* Proceedings of Royal Society London B 268, 1803-1810, 2001.

[11] Watts, D., and S. Strogatz. "*Collective dynamics of small-world networks*," Nature 393, pp. 440-442, 1998.