2011 International Symposium on Nonlinear Theory and its Applications
NOLTA2011, Kobe, Japan, September 4-7, 2011

Controlling Clustering Coefficient of Graphs by Means of 2-Switch Method

Tatsuya Fukami† and Norikazu Takahashi‡

†Graduate School of Information Science and Electrical Engineering, Kyushu University
744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan.
‡Faculty of Information Science and Electrical Engineering, Kyushu University
744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan.
Email: fukami@kairo.csce.kyushu-u.ac.jp, norikazu@inf.kyushu-u.ac.jp

Abstract—A clustering coefficient control algorithm for simple connected undirected graphs is studied. This algorithm is based on the 2-switch which is a well-known graph transformation that preserves the degree of each node. We first derive an explicit formula for the amount of change in the clustering coefficient caused by a single 2-switch. We next show through computer simulations that the clustering coefficient of scale-free graphs can be controlled over a wide range by using this algorithm.

1. Introduction

It is well known that most of large complex networks in the real world have three common properties: 1) The average path length is short; 2) The degree distribution follows the power law; 3) The clustering coefficient is high. Watts-Strogatz (WS) model [1] and Barabási-Albert (BA) model [2] are widely known as two important models for complex networks, but these do not have all of the above properties. The WS model has the first and third properties but not the second one. On the other hand, the BA model has the first and second properties but not the third one. Hence there have been many attempts to construct network models having all of the three properties [3–6].

Among those attempts, we focus our attention in this paper on the 2-switch based approach [7,8]. The 2-switch is a well-known graph transformation that preserves the degree of each node [9]. Therefore, by applying a sequence of 2-switches to a graph generated by the BA model, we may obtain a graph having all of the above-mentioned properties. Furthermore, since the amount of change in the clustering coefficient caused by a single 2-switch is very small, the clustering coefficient may be finely controlled.

In this paper, we first derive an explicit formula for the amount of change in the clustering coefficient caused by a single 2-switch. We next show experimentally that the clustering coefficient of scale-free graphs can be controlled over a wide range by using an algorithm based on 2-switches, while the average path length is kept short.

2. Notations and Definitions

Throughout this paper, we consider only simple connected undirected graph $G = (V(G), E(G))$ where $V(G) = \{1, 2, \ldots, n\}$ is the set of nodes and $E(G) = \{e_1, e_2, \ldots, e_m\}$ is the set of links. Each member of $E(G)$ is an unordered pair of distinct nodes and denoted by $e_k = \{i, j\}$. For a node i of G, the set of all nodes j such that $\{i, j\} \in E(G)$ is called the neighborhood of the node i and denoted by $N_i(G)$. If G has three nodes i, j and k such that $\{\{i, j\}, \{j, k\}, \{k, i\}\} \subseteq E(G)$ then we say that G has a triangle with nodes i, j and k. When we add a link $\{i, j\}$ to $G = (V(G), E(G))$ such that $\{i, j\} \notin E(G)$, we express this operation as $E(G)+\{i, j\}$. Similarly, when we remove a link $\{i, j\}$ from $G = (V(G), E(G))$, we express this operation as $E(G) - \{i, j\}$. The degree vector of a graph G is defined as

$$K(G) = (k_1(G), k_2(G), \ldots, k_n(G))$$

where $k_i(G)$ is the degree of node i. The clustering coefficient [1] of a graph G is defined as

$$C(G) = \frac{1}{n} \sum_{i \in V(G)} C_i(G)$$

where $C_i(G)$ is the clustering coefficient of node i which is defined as

$$C_i(G) = \begin{cases} \frac{t_i(G)}{k_i(G)(k_i(G)-1)/2}, & \text{if } k_i(G) \geq 2 \\ 0, & \text{if } k_i(G) = 0, 1 \end{cases}$$

where $t_i(G)$ represents the number of unordered pairs of nodes $\{j, k\}$ such that $\{\{i, j\}, \{j, k\}, \{k, i\}\} \subseteq E(G)$, that is, the number of triangles containing node i.

3. Control of Clustering Coefficient

3.1. 2-Switch Method

Let us begin with the following lemma.

Lemma 1 *Let $G = (V(G), E(G))$ be a graph having four nodes i, j, k, l such that $\{i, j\}, \{k, l\} \in E(G)$ and $\{i, l\}, \{j, k\} \notin E(G)$. Let H be the graph obtained from G by removing links $\{i, j\}$ and $\{k, l\}$ and adding links $\{i, l\}$ and $\{j, k\}$. Then G and H have the same degree vector.*

It is clear from Fig. 1 that Lemma 1 holds true. The transformation from one graph G into another graph H in Lemma 1 is called the 2-switch in Reference [9]. So we

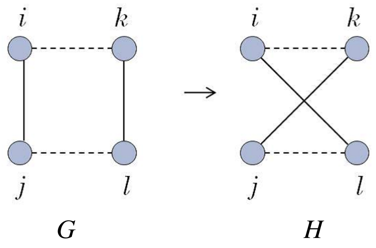

Figure 1: 2-switch

also use this terminology in the following. It is also shown in Reference [9] that a sequence of 2-switches can transform G into H if and only if $K(G) = K(H)$.

The amount of change in the clustering coefficient caused by a 2-switch is given in the following theorem.

Theorem 1 *Let G and H be the same as in Lemma 1. Then the difference of the clustering coefficient between G and H is given by*

$$\begin{aligned} &C(H) - C(G) \\ &= \frac{1}{n}\Bigg(\sum_{a \in N_{il}(G)} \frac{1}{k_a(G)(k_a(G)-1)/2} \\ &\quad + \sum_{b \in N_{jk}(G)} \frac{1}{k_b(G)(k_b(G)-1)/2} \\ &\quad - \sum_{c \in N_{ij}(G)} \frac{1}{k_c(G)(k_c(G)-1)/2} \\ &\quad - \sum_{d \in N_{kl}(G)} \frac{1}{k_d(G)(k_d(G)-1)/2} \\ &\quad + \frac{|N_{il}(G)| - |N_{ij}(G)|}{k_i(G)(k_i(G)-1)/2} + \frac{|N_{kj}(G)| - |N_{ij}(G)|}{k_j(G)(k_j(G)-1)/2} \\ &\quad + \frac{|N_{kj}(G)| - |N_{kl}(G)|}{k_k(G)(k_k(G)-1)/2} + \frac{|N_{il}(G)| - |N_{kl}(G)|}{k_l(G)(k_l(G)-1)/2} \Bigg) \quad (1) \end{aligned}$$

where $N_{ij}(G) = (N_i(G) \cap N_j(G)) \cap (V(G) - \{i, j, k, l\})$ *and* $N_{jk}(G)$, $N_{kl}(G)$, $N_{il}(G)$ *are defined similarly.*

Proof: By the definition of the clustering coefficient and Lemma 1, we have

$$\begin{aligned} C(H) - C(G) &= \frac{1}{n}\sum_{a=1}^{n}(C_a(H) - C_a(G)) \\ &= \frac{1}{n}\sum_{a=1}^{n}\frac{t_a(H) - t_a(G)}{k_a(G)(k_a(G)-1)/2} \\ &= \frac{1}{n}\Bigg(\sum_{a \in V(G)-\{i,j,k,l\}}\frac{t_a(H) - t_a(G)}{k_a(G)(k_a(G)-1)/2} \\ &\quad + \sum_{a \in \{i,j,k,l\}}\frac{t_a(H) - t_a(G)}{k_a(G)(k_a(G)-1)/2}\Bigg) \quad (2) \end{aligned}$$

We first investigate the first term of (2) in more detail. Let a be any node other than i, j, k and l. Then we easily observe that the following statements hold true.

1. If $a \in N_{ij}(G)$ then G has a triangle with nodes a, i and j but H does not have this triangle.

2. If $a \in N_{kl}(G)$ then G has a triangle with nodes a, k and l but H does not have this triangle.

3. If $a \in N_{il}(G)$ then H has a triangle with nodes a, i and l but G does not have this triangle.

4. If $a \in N_{jk}(G)$ then H has a triangle with nodes a, j and k but G does not have this triangle.

5. If G has a triangle with nodes a, b and c such that $\{b, c\} \neq \{i, j\}, \{k, l\}, \{i, l\}, \{j, k\}$ then H also has this triangle because none of three links $\{a, b\}$, $\{b, c\}$ and $\{c, a\}$ is removed by the 2-switch. Conversely, if H has a triangle with nodes a, b and c such that $\{b, c\} \neq \{i, j\}, \{k, l\}, \{i, l\}, \{j, k\}$ then G also has this triangle.

From these observations, we have

$$\begin{aligned} &t_a(H) - t_a(G) \\ &\qquad = \mathbf{I}_{N_{il}(G)}(a) + \mathbf{I}_{N_{jk}(G)}(a) - \mathbf{I}_{N_{ij}(G)}(a) - \mathbf{I}_{N_{kl}(G)}(a) \end{aligned}$$

where $\mathbf{I}_A(a)$ is the indicator function of a subset A of $V(G)$ defined by

$$\mathbf{I}_A(a) = \begin{cases} 1, & \text{if } a \in A \\ 0, & \text{if } a \notin A \end{cases}$$

Therefore, the first term of the right-hand side of (2) can be transformed as follows.

$$\begin{aligned} &\sum_{a \in V(G)-\{i,j,k,l\}} \frac{t_a(H) - t_a(G)}{k_a(G)(k_a(G)-1)/2} \\ &= \sum_{a \in V(G)-\{i,j,k,l\}} \frac{1}{k_a(G)(k_a(G)-1)/2} \\ &\quad \times(\mathbf{I}_{N_{il}(G)}(a) + \mathbf{I}_{N_{jk}(G)}(a) - \mathbf{I}_{N_{ij}(G)}(a) - \mathbf{I}_{N_{kl}(G)}(a)) \\ &= \sum_{a \in N_{il}(G)} \frac{1}{k_a(G)(k_a(G)-1)/2} \\ &\quad + \sum_{a \in N_{jk}(G)} \frac{1}{k_a(G)(k_a(G)-1)/2} \\ &\quad - \sum_{a \in N_{ij}(G)} \frac{1}{k_a(G)(k_a(G)-1)/2} \\ &\quad - \sum_{a \in N_{kl}(G)} \frac{1}{k_a(G)(k_a(G)-1)/2} \end{aligned}$$

Let us next consider the second term of (2). In the following, we deal only with the case where $a = i$ because the remaining cases are similar. Since neither G nor H contains the triangle with nodes i, j and l, all triangles in G and H containing node i are divided into three disjoint classes: 1) Node j is contained; 2) Node l is contained; 3) Neither node j nor node l is contained. For these classes of triangles, the following statements hold true.

1. G has a triangle with nodes i, j and c if and only if $c \in N_{ij}(G)$, while H does not have such triangles because $\{i, j\} \notin E(H)$.

2. H has a triangle with nodes i, l and c if and only if $c \in N_{il}(G)$, while G does not have such triangles because $\{i, l\} \notin E(G)$.

3. If G has a triangle with nodes i, b and c such that $j \notin \{b, c\}$ and $l \notin \{b, c\}$ then H also has this triangle because none of three links $\{i, b\}, \{b, c\}$ and $\{c, i\}$ is removed by the 2-switch. Conversely, if H has a triangle with nodes i, b and c such that $j \notin \{b, c\}$ and $l \notin \{b, c\}$ then G also has this triangle.

From these observations, we have

$$t_i(H) - t_i(G) = |N_{il}| - |N_{ij}|$$

which completes the proof for the case where $a = i$. □

3.2. Algorithm for Controlling Clustering Coefficient

By applying a sequence of 2-switches to a graph, we can finely control the clustering coefficient while keeping the degree vector unchanged. To be more specific, if we repeat two operations: 1) selecting four nodes i, j, k, l such that the right-hand side of (1) is positive (negative, resp.) and 2) applying a 2-switch to the selected four nodes then the clustering coefficient increases (decreases, resp.) gradually. In addition, by using Theorem 1, it is easy to check whether there exists a 2-switch that can increase (or decrease) the clustering coefficient of a graph.

The idea of using 2-switches to control the clustering coefficient of graphs is not new (see, for example, [7] and [8]). However, to the best of the authors' knowledge, Theorem 1 is the first to provide an explicit formula for the amount of change in the clustering coefficient caused by a 2-switch.

In order to confirm the effectiveness of Theorem 1, we show that graphs such that not only they have the scale-free property but also their clustering coefficients are as high (or low) as possible can be generated by using 2-switches. A simple algorithm for constructing such a graph is shown in Fig. 2 where G, the input to the algorithm, is assumed to be a graph generated by the BA model and $D_{(i,j,k,l)}(G)$ represents the right-hand side of (1). It is important to note that this algorithm terminates if and only if there is no 2-switch to increase the clustering coefficient of G_t.

We first applied our algorithm to small graphs with 10 nodes generated by the BA model. Some of the results are shown in Fig. 3. Although it is difficult to perceive the difference between the initial and final graphs, the clustering coefficient is certainly increased by the algorithm to about 0.8 in all cases. We next applied our algorithm to larger graphs with sizes ranging from 50 to 250 nodes. The results are summarized in Fig. 4 where the horizontal axis represents the number of nodes, the vertical axis represents the clustering coefficient, and Δm is the number of links connecting a new node to the existing nodes in the BA model.

Set G_0 to G
Initialize t to 0
Initialize (i, j, k, l) to $(1, 2, 3, 4)$
while $(i, j, k, l) \neq (n-2, n-1, n, n+1)$ **do**
  if $\{i, j\}, \{k, l\} \in E(G_t)$ **and** $\{i, l\}, \{j, k\} \notin E(G_t)$ **then**
    if $D_{(i,j,k,l)}(G_t) > 0$ **then**
      Apply 2-switch to G_t and let the obtained graph be denoted by G_{t+1}
      Add 1 to t
      Reset (i, j, k, l) to $(1, 2, 3, 4)$
    else
      Update (i, j, k, l)
    end if
  end if
end while
return G_t and $C(G_t)$

Figure 2: Algorithm for constructing a graph that has the same degree vector as G and a high clustering coefficient.

For each value of $n \in \{50, 60, \ldots, 250\}$ and $\Delta m \in \{2, 3, 4\}$, the average of ten results is plotted in the graph. We see from Fig. 4 that the clustering coefficient of the final graph is about 0.8 in all cases, while the clustering coefficient of the initial graph is between 0.07 and 0.28. Since the amount of change in the clustering coefficient caused by a single 2-switch is very small, this result means that a sequence of 2-switches can control the clustering coefficient of scale-free graphs over a wide range.

Finally, we investigated the effect of our algorithm on the average path length. The results are shown in Fig. 5 where the horizontal axis represents the number of nodes and the vertical axis represents the average path length. From these results, we see that the average path length is increased by our algorithm but still low.

By using 2-switches, we can also decrease the clustering coefficient while keeping the degree vector unchanged. In order to verify the effectiveness of this approach, we applied an algorithm similar to the one in Fig. 2 to graphs with sizes ranging from 50 to 250 generated by the BA model. We then observed that the clustering coefficient of the generated graph was exactly or nearly zero in all cases.

4. Conclusions

We have studied a 2-switch based clustering coefficient control algorithm for undirected graphs. We first derived an explicit formula for the amount of change in the clustering coefficient caused by a single 2-switch. We next applied the algorithm to graphs generated by the BA model with sizes ranging from 50 to 250 nodes. Experimental results show that the clustering coefficient can be finely controlled over a wide range between about 0 and 0.8, while the average path length is kept short. This means that the 2-switch based algorithm is a useful tool for constructing graphs having

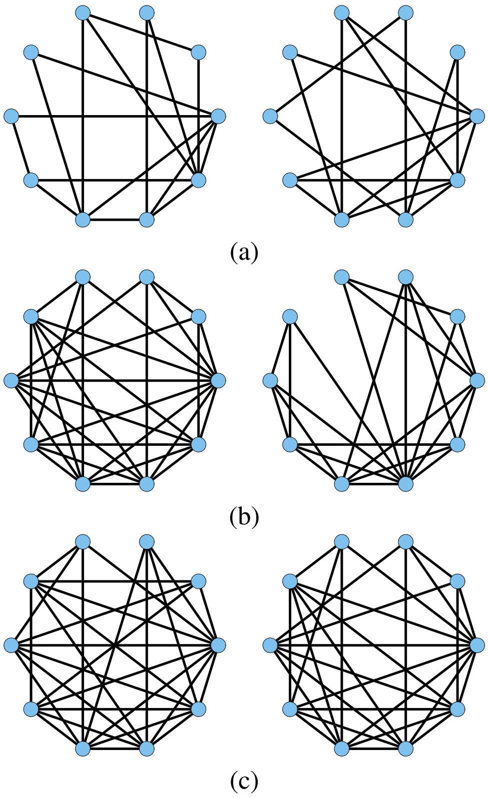

Figure 3: Results of the application of our algorithm to small graphs. The initial and final graphs are shown at left and right, respectively. The clustering coefficients are (a) 0.453333 (left) and 0.793333 (right), (b) 0.618333 (left) and 0.801667 (right), (c) 0.694524 (left) and 0.790714 (right).

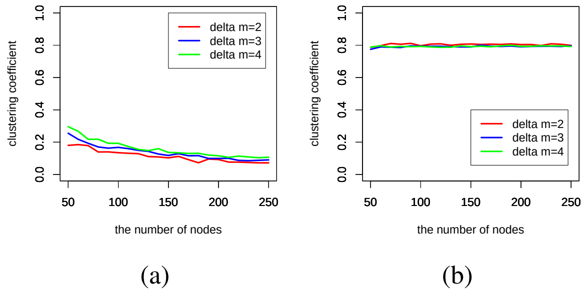

Figure 4: Results of the application of our algorithm to larger graphs. (a) Clustering coefficients of initial graph generated by the BA model. (b) Clustering coefficient of the final graph.

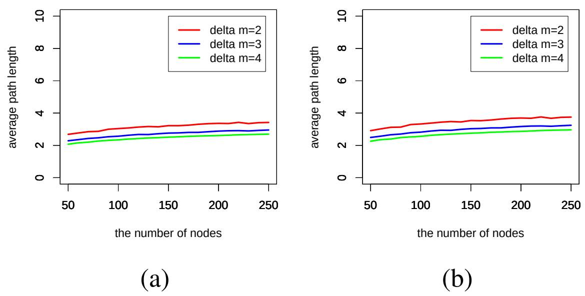

Figure 5: Effect of our algorithm on the average path length. (a) Average path length of the initial graph. (b) Average path length of the final graph.

the three properties mentioned in Section 1.

Acknowledgments

This work was supported by KAKENHI 21560068.

References

[1] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440–442, 1998.

[2] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509–512, 1999.

[3] P. Holme and B. J. Kim, "Growing scale-free networks with tunable clustering," *Physical Review E*, vol. 65, no. 2, 26107, 2002.

[4] K. Klemm and V. M. Eguiluz, "Highly clustered scale-free networks," *Physical Review E*, vol. 65, no. 3, 36123, 2002.

[5] K. Klemm and V. M. Eguiluz, "Growing scale-free networks with small-world behavior," *Physical Review E*, vol. 65, no. 5, 57102, 2002.

[6] W. M. Tam, F. C. M. Lau, and C. K. Tse, "Construction of scale-free networks with adjustable clustering," in *Proceedings of 2008 International Symposium on Nonlinear Theory and its Applications*, pp. 257–260, September 2008.

[7] S. Maslov and K. Sneppen, "Specificity and stability in topology of protein networks," *Science*, vol. 296, pp. 910–913, May 2002.

[8] B. J. Kim, "Performance of networks of artificial neurons: The role of clustering," *Physical Review E*, vol. 69, no. 4, 045101, April 2004.

[9] S. Bereg and H. Ito, "Transforming graphs with the same degree sequence," in *Computational Geometry and Graph Theory*, H. Ito, M. Kano, N. Katoh, and Y. Uno, Eds., vol. 4535 of *Lecture Notes in Computer Science*, pp. 25–32. Springer Berlin / Heidelberg, 2008.