IEICE Proceeding Series

On Graphs that Locally Maximize Global Clustering Coefficient

Tetsuro Teraji, Norikazu Takahashi

Vol. 2 pp. 130-133
Publication Date: 2014/03/18
Online ISSN: 2188-5079

On Graphs that Locally Maximize Global Clustering Coefficient

Tetsuro Teraji† and Norikazu Takahashi‡

†Graduate School of Information Science and Electrical Engineering, Kyushu University
744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan
‡Graduate School of Natural Science and Technology, Okayama University
3-1-1 Tsushima-naka, Kita-ku, Okayama 700-8530, Japan
Email: teraji@me.inf.kyushu-u.ac.jp, takahashi@cs.okayama-u.ac.jp

Abstract—The problem of finding graphs that locally maximize the global clustering coefficient (GCC) is considered. We first prove that if a graph is composed of two cliques sharing one vertex then it locally maximizes the GCC. We next prove that if a graph is composed of two cliques connected by a path with an arbitrary length then it locally maximizes the GCC. The first result is the same as the one given in the case of the average clustering coefficient (ACC). On the other hand, the second one does not hold in the case of the ACC.

1. Introduction

Clustering coefficient [1, 2] is one of the most important measures that characterize the structure of complex networks. Roughly speaking, the clustering coefficient of a graph represents the probability that two vertices adjacent to a given vertex are adjacent to each other. It is well known that many networks in the real world have higher clustering coefficients than random graphs. Also, it has been reported that the clustering coefficient is strongly related to the performance of Hopfield neural networks [3], the synchronization of oscillator networks [4], and so on.

There are two kinds of definitions for the clustering coefficient. One is *the average clustering coefficient* introduced by Watts and Strogatz [1], and the other is *the global clustering coefficient* [2]. In order to clarify the fundamental properties of the average clustering coefficient, Koizuka and Takahashi [5] studied the problem of finding graphs that maximize or locally maximize the clustering coefficient. They proved theoretically that if a graph is composed of two or three cliques sharing one vertex then it locally maximizes the clustering coefficient. This is not a surprising result because one can easily imagine that such a graph has a high clustering coefficient. However, this is an important step to deeper understanding of the average clustering coefficient. In fact, their result was recently extended to a more general form [6].

In this paper, we employ the global clustering coefficient for the definition of the clustering coefficient, and study the same problem as Koizuka and Takahashi [5]. The objective of this paper is to see whether the definition of the clustering coefficient affects the result or not. We first prove that if a graph is composed of two cliques sharing one vertex then it locally maximizes the global clustering coefficient. We next prove that if a graph is composed of two cliques connected by a path with an arbitrary length then it locally maximizes the global clustering coefficient. The second result is more important than the first one because not only is it counter-intuitive but also the same statement does not hold for the average clustering coefficient.

2. Two Definitions of Clustering Coefficient

In this paper, by a graph, we mean a simple connected undirected graph $G = (V(G), E(G))$ where $V(G)$ is the set of vertices (nodes) and $E(G)$ is the set of edges (links). The set of all graphs composed of n vertices and m edges is denoted by $\mathcal{G}(n, m)$.

Definition 1 (Average Clustering Coefficient [1])**.** For a given graph $G \in \mathcal{G}(n, m)$, the clustering coefficient of the vertex $i \in V(G)$ is defined by

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

where $d_i(G)$ is the degree of the vertex i and $t_i(G)$ is the number of triangles containing the vertex i, that is,

$$t_i(G) = |\{\{j,k\} \in E(G) \,|\, \{i,j\}, \{i,k\} \in E(G)\}| .$$

The average clustering coefficient (ACC) of the graph $G = (V(G), E(G))$ is then defined by

$$C_A(G) = \frac{1}{n}\sum_{i=1}^{n} C_i(G) .$$

Definition 2 (Global Clustering Coefficient [2])**.** The global clustering coefficient (GCC) of a graph $G \in \mathcal{G}(n, m)$ is defined by

$$C_G(G) = \frac{\sum_{i=1}^{n} t_i(G)}{\sum_{i=1}^{n} d_i(G)(d_i(G)-1)/2} \tag{1}$$

where $t_i(G)$ and $d_i(G)$ are defined as in Definition 1.

Suppose that we randomly select a vertex i with equal probability, and then randomly select, with equal probability, a pair (j, k) of vertices adjacent to i. Then the probability that j and k are adjacent to each other is equal to the

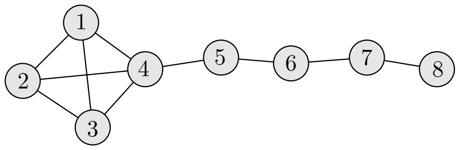

Figure 1: A lollipop graph with 8 vertices and 10 edges.

ACC. Suppose next that we randomly select, with equal probability, a triple (i, j, k) of vertices such that j and k are adjacent to i. Then the probability that j and k are adjacent to each other is equal to the GCC.

In order to see how different these two definitions are, let us consider a lollipop graph $G \in \mathcal{G}(8, 10)$ shown in Fig. 1. The clustering coefficients of 8 vertices are given by $C_1(G) = C_2(G) = C_3(G) = 1$, $C_4(G) = 0.5$ and $C_5(G) = C_6(G) = C_7(G) = C_8(G) = 0$. Hence the ACC of this graph is given by $C_{\mathrm{A}}(G) = \frac{3.5}{8} = \frac{7}{16} = 0.4375$. On the other hand, the GCC is given by $C_{\mathrm{G}}(G) = \frac{3\times4}{3+3+3+6+1+1+1+0} = \frac{2}{3} \approx 0.6667$ which is much greater than the ACC.

3. Global Clustering Coefficient Locally Maximizing Graphs

Clustering coefficient maximizing graphs and locally maximizing graphs are defined as follows.

Definition 3 (Clustering Coefficient Maximizing Graph [5]). If a graph $G \in \mathcal{G}(n, m)$ satisfies $C_{\mathrm{A}}(G) \geq C_{\mathrm{A}}(G')$ ($C_{\mathrm{G}}(G) \geq C_{\mathrm{G}}(G')$, resp.) for all $G' \in \mathcal{G}(n, m)$ then it is called an ACC (a GCC, resp.) maximizing graph in $\mathcal{G}(n, m)$.

Definition 4 (Clustering Coefficient Locally Maximizing Graph [5]). If a graph $G \in \mathcal{G}(n, m)$ satisfies $C_{\mathrm{A}}(G) \geq C_{\mathrm{A}}(G')$ ($C_{\mathrm{G}}(G) \geq C_{\mathrm{G}}(G')$, resp.) for all $G' \in \mathcal{G}(n, m)$ that are obtained from G by rewiring an edge then it is called an ACC (a GCC, resp.) locally maximizing graph in $\mathcal{G}(n, m)$.

In the following, for any $G \in \mathcal{G}(n, m)$, the numerator and the denominator of the right-hand side of (1) are denoted by $T(G)$ and $D(G)$, respectively. That is,

$$\begin{aligned} T(G) &= \sum_{i=1}^{n} t_i(G), \\ D(G) &= \sum_{i=1}^{n} d_i(G)(d_i(G) - 1)/2. \end{aligned}$$

The first result of this paper is as follows.

Theorem 1. If $G \in \mathcal{G}(n, m)$ is composed of two cliques sharing one vertex then it is a GCC locally maximizing graph in $\mathcal{G}(n, m)$.

Proof Let $G = (V(G), E(G))$ be any graph composed of two cliques sharing one vertex (see Fig.2). Then $V(G)$ has a partition $\{V_0, V_1, V_2\}$ such that the following conditions are satisfied.

1. $|V_0| = 1$, $|V_1| = n_1 \geq 1$ and $|V_2| = n_2 \geq 1$.
2. The subgraph of G induced by $V_0 \cup V_1$ and the subgraph of G induced by $V_0 \cup V_2$ are complete.
3. If $i \in V_1$ and $j \in V_2$ then $\{i, j\} \notin E(G)$.

Also, $T(G)$ and $D(G)$ are expressed in terms of n_1 and n_2 as follows:

$$\begin{aligned} T(G) &= (n_1 + 1)\binom{n_1}{2} + (n_2 + 1)\binom{n_2}{2} \\ &= (n_1^3 + n_2^3 - n_1 - n_2)/2, \\ D(G) &= \binom{n-1}{2} + n_1\binom{n_1}{2} + n_2\binom{n_2}{2} \\ &= (n_1^3 + n_2^3 - n_1 - n_2 + 2n_1n_2)/2. \end{aligned}$$

In the following, we assume without loss of generality that $V_0 = \{1\}$, $V_1 = \{2, 3, \ldots, n_1 + 1\}$ and $V_2 = \{n_1 + 2, n_1 + 3, \ldots, n\}$. In order to prove that G is a GCC locally maximizing graph, we have to show that $C_{\mathrm{G}}(G) \geq C_{\mathrm{G}}(G')$ for any graph $G' \in \mathcal{G}(n, m)$ obtained from G by rewiring one edge. There are a number of ways of rewiring one edge in G, but it suffices for us to consider the following four cases: (a) $\{1, 2\}$ is removed and $\{2, n\}$ is added, (b) $\{1, 2\}$ is removed and $\{3, n\}$ is added under the assumption that $n_1 \geq 2$, (c) $\{2, 3\}$ is removed and $\{2, n\}$ is added under the assumption that $n_1 \geq 2$, and (d) $\{2, 3\}$ is removed and $\{4, n\}$ is added under the assumption that $n_1 \geq 3$. We consider only Case (a) due to the space limitation. In this case, since $T(G')$ and $D(G')$ are given by $T(G) - 3(n_1 - 1)$ and $D(G) - (n_1 - 1)$, respectively, we have

$$\begin{aligned} C_{\mathrm{G}}(G) - C_{\mathrm{G}}(G') &= \frac{T(G)}{D(G)} - \frac{T(G) - 3(n_1 - 1)}{D(G) - (n_1 - 1)} \\ &= \frac{(n_1 - 1)(3D(G) - T(G))}{D(G)\{D(G) - (n_1 - 1)\}}. \qquad (2) \end{aligned}$$

The denominator is positive and the numerator is also positive because

$$\begin{aligned} 3D(G) - T(G) &= n_1^3 + n_2^3 - n_1 - n_2 + 3n_1n_2 \\ &= n_1(n_1^2 - 1) + n_2(n_2^2 - 1) + 3n_1n_2 \\ &> 0. \end{aligned}$$

Therefore, the right-hand side of (2) is positive. □

The second result of this paper is as follows.

Theorem 2. If the vertex set $V(G)$ of a graph $G \in \mathcal{G}(n, m)$ has a partition $\{V_0, V_1, V_2\}$ that satisfies the following conditions then G is a GCC locally maximizing graph in $\mathcal{G}(n, m)$.

1. $|V_0| \geq 2$, $|V_1| \geq 1$ and $|V_2| \geq 1$.
2. The subgraph of G induced by V_0 is a path graph with its end vertices α and β.

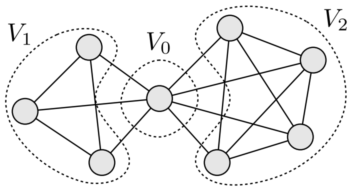

Figure 2: A graph composed of two cliques sharing one vertex.

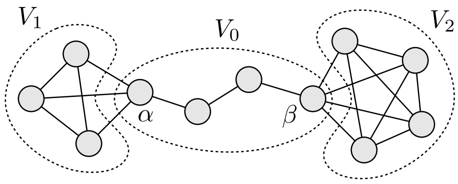

Figure 3: A graph satisfying the conditions in Theorem 2.

3. The subgraph of G induced by $V_1 \cup \{\alpha\}$ is a complete graph. Also, if $i \in V_1$ and $j \notin V_1 \cup \{\alpha\}$ then the edge $\{i, j\}$ does not exist.
4. The subgraph of G induced by $V_2 \cup \{\beta\}$ is a complete graph. Also, if $i \in V_2$ and $j \notin V_2 \cup \{\beta\}$ then the edge $\{i, j\}$ does not exist.

Proof Let $G = (V(G), E(G))$ be any graph satisfying the four conditions (see Fig.3). In the following, we focus our attention on the case where $|V_0| = n_0$, $|V_1| = n_1$ and $|V_2| = n_2$ are sufficiently large due to the space limitation. In addition, we assume without loss of generality that $V_0 = \{1(=\alpha), 2, \dots, n_0(=\beta)\}$, $V_1 = \{n_0+1, n_0+2, \dots, n_0+n_1\}$, and $V_2 = \{n_0+n_1+1, n_0+n_1+2, \dots, n\}$. Then $T(G)$ and $D(G)$ are expressed in terms of n_0, n_1 and n_2 as follows:

$$\begin{aligned} T(G) &= (n_1+1)\binom{n_1}{2} + (n_2+1)\binom{n_2}{2} \\ &= (n_1^3 + n_2^3 - n_1 - n_2)/2\,, \end{aligned} \tag{3}$$

$$\begin{aligned} D(G) &= (n_1+1)\binom{n_1}{2} + (n_2+1)\binom{n_2}{2} + (n_0-2) + n_1 + n_2 \\ &= \{n_1^3 + n_2^3 + 2(n_0-2) + n_1 + n_2\}/2\,. \end{aligned} \tag{4}$$

Let $G' \in \mathcal{G}(n, m)$ be any graph obtained from G by rewiring an edge. Let $\{i_1, i_2\}$ be the edge removed from G and $\{i_3, i_4\}$ be the edge added to G. Then there are five possible cases: (a) $i_1, i_2 \in V_0$, (b) $i_1, i_2 \in V_1$, (c) $i_1 = 1$ and $i_2 \in V_1$, (d) $i_1, i_2 \in V_2$, and (e) $i_1 = n_0$ and $i_2 \in V_2$. In the following, we will show that $C_G(G) \geq C_G(G')$ holds for the first three cases, because Cases (d) and (e) can be analyzed in the same way as Cases (b) and (c), respectively.

(a) It is easily seen that $D(G) + 1 \leq D(G')$ and $T(G) = T(G')$. Therefore $C_G(G) > C_G(G')$ holds in this case.

(b) Let G'' be the graph obtained from G by removing the edge $\{i_1, i_2\}$. Note that $G'' \in \mathcal{G}(n, m-1)$ because n_1 is assumed to be sufficiently large. Since $T(G'')$ and $D(G'')$ are given by $T(G) - 3(n_1-1)$ and $D(G) - 2(n_1-1)$, respectively, we have

$$\begin{aligned} C_G(G) - C_G(G'') &= \frac{T(G)}{D(G)} - \frac{T(G) - 3(n_1-1)}{D(G) - 2(n_1-1)} \\ &= \frac{(n_1-1)(3D(G) - 2T(G))}{D(G)\{D(G) - 2(n_1-1)\}}\,. \end{aligned}$$

Here, the denominator is positive and the numerator is also positive because

$$3D(G) - 2T(G) = \{n_1^3 + n_2^3 + 6(n_0-2) + 5n_1 + 5n_2\}/2 > 0\,.$$

Therefore, $C_G(G) - C_G(G'')$ is positive. Let us next consider the quantity $C_G(G'') - C_G(G')$. Since G' is obtained from G'' by adding the edge $\{i_3, i_4\}$, we immediately see that $D(G') > D(G'')$. So, if $T(G') = T(G'')$, that is, if the number of triangles is not changed by the edge addition, we have $C_G(G) - C_G(G') = C_G(G) - C_G(G'') + C_G(G'') - C_G(G') > 0$. From this fact, we can concentrate our attention on the six cases: (i) $\{i_3, i_4\} = \{i_1, 2\}$, (ii) $i_3 \in V_1 \setminus \{i_1, i_2\}$ and $i_4 = 2$, (iii) $\{i_3, i_4\} = \{1, 3\}$, (iv) $i_3 \in \{2, 3, \dots, n_0 - 3\}$ and $i_4 = i_3 + 2$, (v) $\{i_3, i_4\} = \{n_0 - 2, n_0\}$ and (vi) $i_3 = n_0 - 1$ and $i_4 \in V_2$. In all of these cases, a new triangle appears by the edge addition. Hence we have $T(G') = T(G'') + 3$. On the other hand, the value of $D(G')$ in Case (iv), which is given by $D(G'') + 4$, is smaller than any other case. Therefore, we have

$$\begin{aligned} & C_G(G) - C_G(G') \\ &\quad \geq \frac{T(G)}{D(G)} - \frac{T(G'') + 3}{D(G'') + 4} \\ &\quad = \frac{T(G)}{D(G)} - \frac{T(G) - 3n_1 + 6}{D(G) - 2n_1 + 6} \\ &\quad = \frac{n_1(3D(G) - 2T(G)) - 6(D(G) - T(G))}{D(G)(D(G) - 2n_1 + 6)}\,. \end{aligned}$$

The denominator is positive. Hence it suffices for us to show that the numerator is also positive. By substituting (3) and (4) into the right-hand side, we have

$$\begin{aligned} & n_1(3D(G) - 2T(G)) - 6(D(G) - T(G)) \\ &\quad = n_1(n_1^3 + n_2^3)/2 + (5n_1 - 12)(n_1 + n_2)/2 \\ &\qquad + 3(n_0 - 2)(n_1 - 2)\,, \end{aligned}$$

which is positive because n_1 is sufficiently large.

(c) Let G'' be the graph obtained from G by removing the edge $\{i_1, i_2\}$. Note that $G'' \in \mathcal{G}(n, m-1)$ because n_1 is assumed to be sufficiently large. Since $T(G'')$ and $D(G'')$ are given by

$$\begin{aligned} T(G'') &= T(G) - 3(n_1 - 1)\,, \\ D(G'') &= D(G) - 2n_1 + 1\,, \end{aligned}$$

we have

$$\begin{aligned} & C_G(G) - C_G(G'') \\ = & \frac{T(G)}{D(G)} - \frac{T(G) - 3(n_1 - 1)}{D(G) - 2n_1 + 1} \\ = & \frac{3(n_1 - 1)D(G) - (2n_1 - 1)T(G)}{D(G)\{D(G) - 2n_1 + 1\}}. \end{aligned}$$

Here, the denominator is positive and the numerator can be transformed as

$$\begin{aligned} & 3(n_1 - 1)D(G) - (2n_1 - 1)T(G) \\ = & (n_1 - 2)(n_1^3 + n_2^3)/2 + 3(n_0 - 2)(n_1 - 1) \\ & + (5n_1 - 4)(n_1 + n_2)/2. \end{aligned}$$

Therefore, $C_G(G) - C_G(G'')$ is positive. Let us next consider the quantity $C_G(G'') - C_G(G')$. As in the previous case, we can concentrate our attention on the five cases: (i) $i_3 \in V_1 \setminus \{i_2\}$ and $i_4 = 2$, (ii) $\{i_3, i_4\} = \{1, 3\}$, (iii) $i_3 \in \{2, 3, \ldots, n_0 - 3\}$ and $i_4 = i_3 + 2$, (iv) $\{i_3, i_4\} = \{n_0 - 2, n_0\}$ and (v) $i_3 = n_0 - 1$ and $i_4 \in V_2$. In all of these cases, a new triangle appears by the edge addition. Hence we have $T(G') = T(G'') + 3$. On the other hand, the value of $D(G')$ in Case (iii), which is given by $D(G'') + 4$, is smaller than any other case. Therefore, we have

$$\begin{aligned} & C_G(G) - C_G(G') \\ \geq & \frac{T(G)}{D(G)} - \frac{T(G'') + 3}{D(G'') + 4} \\ = & \frac{T(G)}{D(G)} - \frac{T(G) - 3n_1 + 6}{D(G) - 2n_1 + 5} \\ = & \frac{n_1(3D(G) - 2T(G)) - 6D(G) + 5T(G)}{D(G)(D(G) - 2n_1 + 5)}. \end{aligned}$$

The denominator is positive. Hence it suffices for us to show that the numerator is also positive. By substituting (3) and (4) into the right-hand side, we have

$$\begin{aligned} & n_1(3D(G) - 2T(G)) - 6D(G) + 5T(G) \\ = & (n_1 - 1)(n_1^3 + n_2^3)/2 + 3(n_0 - 2)(n_1 - 2) \\ & + (5n_1 - 11)(n_1 + n_2)/2, \end{aligned}$$

which is positive because n_1 is sufficiently large. □

By letting $|V_2| = 1$ in Theorem 2, we have the following result.

Corollary 1. *If $G \in \mathcal{G}(n, m)$ is a lollipop graph then it is a GCC locally maximizing graph.*

Finally, we will show that the statement of Theorem 2 does not hold if we replace "GCC" with "ACC". As a counterexample, let us consider the graph $G \in \mathcal{G}(6, 8)$ shown in Fig.4. It is a GCC locally maximizing graph because the conditions of Theorem 2 are satisfied by $V_0 = \{1, 2\}$, $V_1 = \{3, 4, 5\}$ and $V_2 = \{6\}$. However, it is not an ACC locally maximizing graph. The ACC of G is calculated as $C_A(G) = \frac{1}{6}\left(\frac{1}{2} + 0 + 1 + 1 + 1 + 0\right) = \frac{7}{12} = 0.583\cdots$ while the ACC of G', which is obtained from G by removing the edge $\{1, 3\}$ and adding the edge $\{1, 6\}$, is given by $C_A(G') = \frac{1}{6}\left(\frac{1}{3} + 1 + 1 + \frac{2}{3} + \frac{2}{3} + 1\right) = \frac{7}{9} = 0.777\cdots$ which is greater than $C_A(G)$.

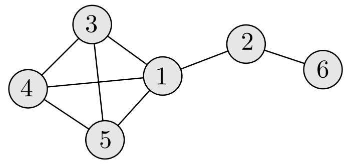

Figure 4: Example of a GCC locally maximizing graph but not an ACC locally maximizing graph.

4. Conclusion

In this paper, we have presented two theoretical results concerning the global clustering coefficient. The second result shows that two definitions of the clustering coefficient differ considerably from the viewpoint of the clustering coefficient locally maximizing graph.

Acknowledgments

This work was supported by KAKENHI 24560076 and 23310104.

References

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

[2] M. E. Newman, S. H. Strogatz and D. J. Watts, "Random graphs with arbitrary degree distributions and their applications," *Physical Review E*, vol.64, 026118, 2001.

[3] B. J. Kim, "Performance of networks of artificial neurons: The role of clustering," *Physical Review E*, vol.69, 045101, 2004.

[4] P. N. McGraw and M. Menzinger, "Clustering and the synchronization of oscillator networks," *Physical Review E*, vol.72, 015101, 2005.

[5] S. Koizuka and N. Takahashi, "Maximum clustering coefficient of graphs with given number of vertices and edges," *Nonlinear Theory and Its Applications, IEICE*, vol.2, no.4, pp.443–457, 2011.

[6] T. Fukami and N. Takahashi, "New classes of clustering coefficient locally maximizing graphs," submitted.