# PROPERTIES OF CLUSTERING COEFFICIENT IN RANDOM AGENT NETWORKS

XIAOLONG JIN AND JIMING LIU

*Department of Computer Science,*
*Hong Kong Baptist University,*
*Kowloon Tong, Hong Kong*
*E-mail: {jxl,jiming}@comp.hkbu.edu.hk*

Random agent networks have been used as important contrast roles in the study of topologies of real-world agent networks. As an important property, in this paper, we study the clustering coefficient of random agent networks. Specifically, we give the theoretical average value and lower bound of the clustering coefficient in random agent networks with $n$ agents and $m$ links. As a validation, we also provide some experimental results to support our theoretical results.

*Keywords*: Agent network, Small world, Clustering coefficient

## 1. Introduction

Multi-agent systems have been widely used to problem-solving tasks [2,3,5,6,8]. In order to solve a given task, agents will implicitly or explicitly form an network, where links are the relationships of cooperation or coordination between agents. As an agent usually cooperate or coordinate with some agents, but not all agents, the resulting agent network will not be fully connected. It has been proven that the network topology of a multi-agent system plays an important role in its performance [4]. In the study of topologies in agent networks [1,4], random agent networks are used as contrast roles to other real-world agent networks. Particularly, clustering coefficient is used as an contrast property to differentiate random agent networks and other agent networks [4]. In this paper, we will focus on the clustering coefficient [7] of random agent networks to provide its average value and lower bound.

This paper is organized as follows: Section 2 introduces what is the clustering coefficient of an agent network. In Section 3, we give the theoretical average value of clustering coefficient in random agent networks with $n$ agents and $m$ links. We also give an experimental validation of the theoretical result. Further, in Section 4, we show a lower bound of the clustering coefficient. We propose a conjecture about another lower bound at the end of this section. Section 5 concludes the paper.

## 2. Clustering coefficient

Given an agent network $G =< A, L >$, where $A = \{a_1, a_2, \cdots, a_n\}$, i.e., the set of agents and $L = \{l_1, l_2, \cdots, l_m\}$, i.e., the set of links between agents in $A$. Then, the *clustering coefficient* $C$ of $G$ is defined as follows:

$$C = \frac{1}{n} \sum_{i=1}^{n} c_{a_i} \tag{1}$$

where, $c_{a_i}$ is called the *clustering* of agent $a_i$. Assume $d(a_i)$ is the degree of $a_i$ (i.e., the number of neighboring agents of $a_i$), and $e(a_i)$ is the existing links between agent $a_i$ and its neighbors. Therefore,

$$c_{a_i} = \frac{e(a_i)}{\frac{(d(a_i)+1)\cdot d(a_i)}{2}} = \frac{2 \cdot e(a_i)}{(d(a_i) + 1) \cdot d(a_i)} \tag{2}$$

Clustering coefficient $C$ locally reflects the average connectivity of agents in agent network $G$. In essence, $C$ denotes the possibility of two agents, which have a common neighboring vertex, are neighbors.

## 3. Average value

If an agent network has $n$ agents, then there can exist at most $\frac{n\cdot(n-1)}{2}$ possible links in it. To generate a random agent network with $n$ agents and $m$ links, one can uniformly select $m$ links from $\frac{n\cdot(n-1)}{2}$ possible links at random. In this case, each possible link can be selected with a probability of $\frac{2m}{n\cdot(n-1)}$. If random agent network $G$ has $n$ agents and $m$ links, the average degree of agents in $G$ is $\frac{2m}{n}$. It means an agent is on average connected to $\frac{2m}{n}$ different agents. As a consequence, there can exist at most

$$\frac{(\frac{2m}{n} + 1) \cdot \frac{2m}{n}}{2} \tag{3}$$

possible links between this agent and its neighbors. But, since this agent on average has $\frac{2m}{n}$ neighbors, there are at least $\frac{2m}{n}$ links existing. As to other possible

$$\frac{(\frac{2m}{n} + 1) \cdot \frac{2m}{n}}{2} - \frac{2m}{n} \tag{4}$$

links between this agent and its neighbors, because each link has a probability of $\frac{2m}{n\cdot(n-1)}$ to appear in this agent network, there will on average be

$$[\frac{(\frac{2m}{n}+1)\cdot\frac{2m}{n}}{2}-\frac{2m}{n}]\cdot\frac{2m}{n\cdot(n-1)} \tag{5}$$

links that also appear in agent network G. Hence, the total number of links existing between this agent and its neighbors is:

$$\frac{2m}{n}+[\frac{(\frac{2m}{n}+1)\cdot\frac{2m}{n}}{2}-\frac{2m}{n}]\cdot\frac{2m}{n\cdot(n-1)} \tag{6}$$

Therefore, for any agent $a_i$ in $G$, its clustering $C_{a_i}$ is :

$$\begin{aligned} C_{a_i} &= \frac{\frac{2m}{n}+[\frac{(\frac{2m}{n}+1)\cdot\frac{2m}{n}}{2}-\frac{2m}{n}]\cdot\frac{2m}{n\cdot(n-1)}}{\frac{(2m+n)\cdot m}{n^2}} \\ &= \frac{2n^2\cdot(n-1)+2m\cdot(2m-n)}{n\cdot(n-1)\cdot(2m+n)} \end{aligned} \tag{7}$$

Hence, the clustering coefficient $C$ of agent network $G$ is:

$$C=\frac{1}{n}\cdot\sum_{a_i\in G}C_{a_i}=\frac{2n^2\cdot(n-1)+2m\cdot(2m-n)}{n\cdot(n-1)\cdot(2m+n)} \tag{8}$$

But, we should note that in Eq. 5, it should guarantee that

$$\frac{(\frac{2m}{n}+1)\cdot\frac{2m}{n}}{2}-\frac{2m}{n}\geq 0 \tag{9}$$

i.e., the total number of possible links between agent $a_i$ and its neighbors is larger than or equal to the number of existed links. Simplifying Eq. 9, we get $m\geq\frac{n}{2}$. So, Eq. 8 is applicable under condition of $m\geq\frac{n}{2}$.

Then, what is the average value of the clustering coefficient in the case of $m<\frac{n}{2}$? We can imagine that if $m\ll\frac{n}{2}$, i.e., there are only a few links existing in agent network $G$, the existing links will normally not be connected to the same agent. In this case, the agent connected with a link has clustering 1. The isolated agent has clustering 0. Since there are $2m$ agents with clustering 1, we can readily get the clustering coefficient $C$:

$$C=\frac{2m}{n} \tag{10}$$

Altogether, in a general case, we have the following average value of clustering coefficient in an agent network:

$$C=\begin{cases}\frac{2n^2\cdot(n-1)+2m\cdot(2m-n)}{n\cdot(n-1)\cdot(2m+n)} & if\ m\geq\frac{n}{2}\\ \\ \frac{2m}{n} & if\ m<\frac{n}{2}\end{cases} \tag{11}$$

**Theorem 3.1.** *Let $\{G(n,m)\}$ be a set of all possible random agent networks with $n$ agents and $m$ links, and $C(n,m)$ be the average value of the clustering coefficient in $\{G(n,m)\}$. Then,*

$$C(n,m) = \begin{cases} \frac{2n^2\cdot(n-1)+2m\cdot(2m-n)}{n\cdot(n-1)\cdot(2m+n)} & if\ m \geq \frac{n}{2} \\ \\ \frac{2m}{n} & if\ m < \frac{n}{2} \end{cases} \tag{12}$$

In Fig. 1, we show both the theoretical and experimental result on an example of agent network with 100 agents. We can see from Fig. 1 that the experimental result is very similar to the theoretical result.

In agent network $G(n,m)$, let $k$ denote the average degree of agents. Then we can get:

$$m = \frac{n\cdot k}{2} \tag{13}$$

Substituting the above expression into Eq. 12 in the case of $m \geq \frac{n}{2}$, we get:

$$C(n,k) = \frac{2}{k+1} + \frac{k\cdot(k-1)}{(n-1)\cdot(k+1)} \tag{14}$$

So, we conclude that for a set of random agent networks with $n \gg 1$ and $n \gg k$, their average clustering coefficient value is approximately equal to $\frac{2}{k+1}$.

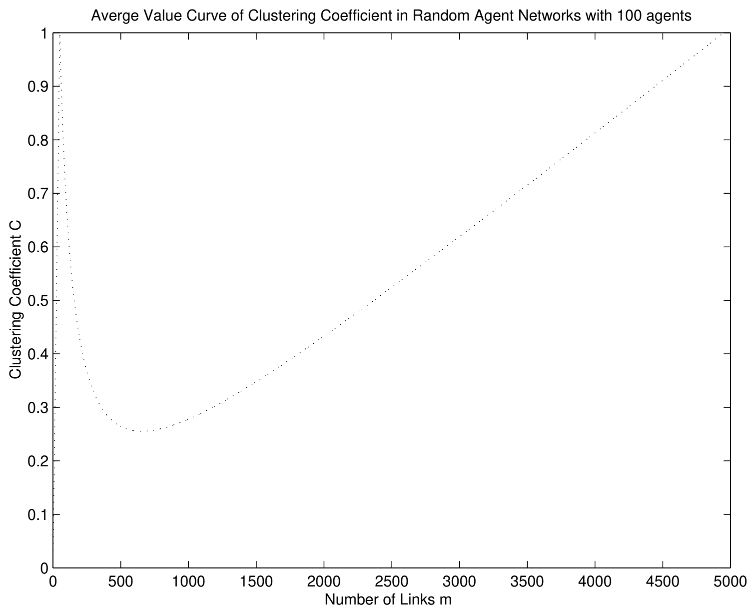
## 4. Lower bounds

In the previous section, we have given the average value of the clustering coefficient in random agent networks. In this section, let us deduce a lower bound of the clustering coefficient in an agent network. Given agent network $G$ has $n$ agents and $m$ links. For $\forall a_i$ $(i \in 1,2,...,n)$, we assume its degree is $d(a_i)$. Then, at most $\frac{(d(a_i)+1)\cdot d(a_i)}{2}$ links can exist between $a_i$ and its neighbors. Further, we assume there actually exist $e(a_i)$ links between $a_i$ and its neighbors. Then, the clustering of $a_i$ is $\frac{2\cdot e(a_i)}{(d(a_i)+1)\cdot d(a_i)}$. Therefore, the clustering coefficient $C$ of $G$ is:

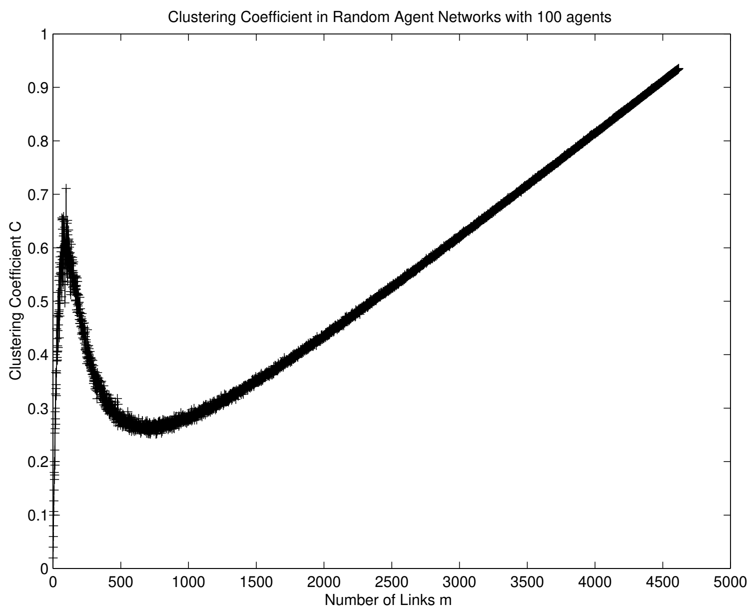
$$C = \frac{1}{n}\cdot\sum_{i=1}^{n}\frac{2\cdot e(a_i)}{(d(a_i)+1)\cdot d(a_i)}$$

In any agent network, it is known that $\forall a_i$, $e(a_i) \geq d(a_i)$. So, we have:

$$C = \frac{1}{n}\cdot\sum_{i=1}^{n}\frac{2\cdot e(a_i)}{(d(a_i)+1)\cdot d(a_i)} \geq \frac{1}{n}\cdot\sum_{i=1}^{n}\frac{2\cdot d(a_i)}{(d(a_i)+1)\cdot d(a_i)}$$




(a)




(b)

Figure 1. (a)The average value curve of the clustering coefficient in agent networks with 100 agents. (b) An experimental result of the clustering coefficient in agent networks also with 100 agents. For each $m \in [1, 4950]$, we generated 100 random agent networks and calculated the average value of the clustering coefficient on these agent networks.

$$= \frac{1}{n} \cdot \sum_{i=1}^{n} \frac{2}{d(a_i)+1} \tag{15}$$

It is also known that in any agent network, there is a relationship between the total number of agent degrees and the number of links, i.e.,

$$\sum_{i=1}^{n} d(a_i) = 2m \tag{16}$$

Then, the problem to find a lower bound of the clustering coefficient $C$ in agent network $G$ becomes an optimization problem:

$$\min \sum_{i=1}^{n} \frac{1}{d(a_i)+1} \quad s.t. \quad \sum_{i=1}^{n} d(a_i) = 2m \tag{17}$$

Let $x_i = d(a_i) + 1$, the above problem can be rewritten as:

$$\min \sum_{i=1}^{n} \frac{1}{x_i} \quad s.t. \quad \sum_{i=1}^{n} x_i = 2m + n \tag{18}$$

Further, let $z = 2m + n$, and $f(x_1, x_2, ..., x_n) = \sum_{i=1}^{n} \frac{1}{x_i}$. From Eq. 18, we have:

$$x_n = z - \sum_{i=1}^{n-1} x_i \tag{19}$$

Substituting $x_n$ in $f(x_1, x_2, ..., x_n)$ with this equation, we get:

$$f(x_1, x_2, ..., x_{n-1}) = \sum_{i=1}^{n-1} \frac{1}{x_i} + \frac{1}{z - \sum_{i=1}^{n-1} x_i} \tag{20}$$

To find the extremal value of the function $f(x_1, x_2, ..., x_{n-1})$, we derivate its partial differential coefficient for $\forall x_i$ $(i = 1, ..., n-1)$:

$$\begin{cases} \frac{\partial f}{\partial x_1} & = \frac{1}{(z-\sum_{i=1}^{n-1} x_i)^2} - \frac{1}{x_1^2} \\ \frac{\partial f}{\partial x_2} & = \frac{1}{(z-\sum_{i=1}^{n-1} x_i)^2} - \frac{1}{x_2^2} \\ \vdots & \vdots \\ \frac{\partial f}{\partial x_{n-1}} & = \frac{1}{(z-\sum_{i=1}^{n-1} x_i)^2} - \frac{1}{x_{n-1}^2} \end{cases} \tag{21}$$

Then let $\frac{\partial f}{\partial x_1} = \frac{\partial f}{\partial x_2} = \cdots = \frac{\partial f}{\partial x_{n-1}} = 0$, we get the following equation system:

$$\begin{cases} \frac{1}{(z-\sum_{i=1}^{n-1} x_i)^2} - \frac{1}{x_1^2} & = 0 \\ \frac{1}{(z-\sum_{i=1}^{n-1} x_i)^2} - \frac{1}{x_2^2} & = 0 \\ \vdots & \vdots \\ \frac{1}{(z-\sum_{i=1}^{n-1} x_i)^2} - \frac{1}{x_{n-1}^2} & = 0 \end{cases} \tag{22}$$

Since, obviously $x_i > 0$ and $z - \sum_{i=1}^{n-1} x_i > 0$, we simplify Eq. 22 and get:

$$\begin{cases} x_1 + \sum_{i=1}^{n-1} x_i & = z \\ x_2 + \sum_{i=1}^{n-1} x_i & = z \\ \vdots & \vdots \\ x_{n-1} + \sum_{i=1}^{n-1} x_i & = z \end{cases} \tag{23}$$

Solving the above Eq. 23, we can easily get $x_1 = x_2 = \cdots = x_{n-1} = \frac{z}{n}$. Substituting them into Eq. 19, we get $x_n = \frac{z}{n}$. So,

$$\sum_{i=1}^{n} \frac{1}{x_n} = n \cdot \frac{n}{z} = \frac{n^2}{2m+n} \tag{24}$$

Therefore,

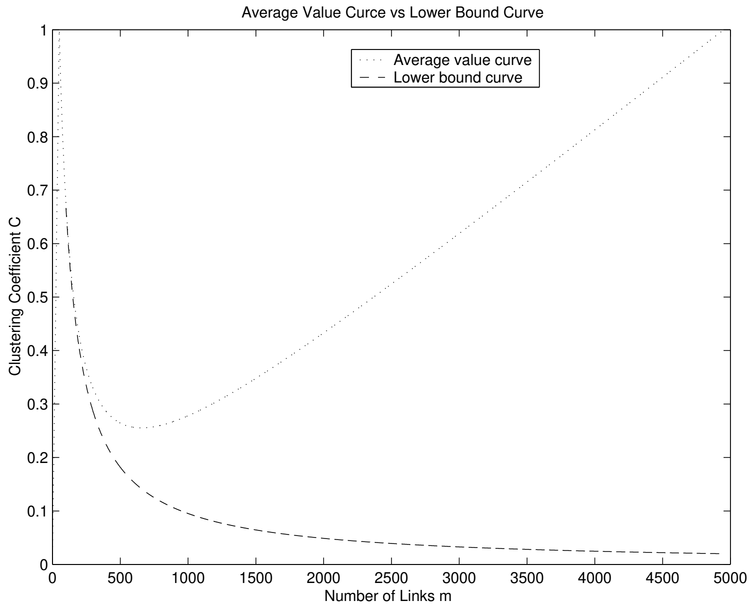
$$C \geq \frac{1}{n} \cdot \sum_{i=1}^{n} \frac{2}{d(a_i)+1} \geq \frac{2}{n} \cdot \frac{n^2}{2m+n} = \frac{2n}{2m+n} \tag{25}$$

i.e., $\frac{2n}{2m+n}$ is a lower bound of the clustering coefficient in an agent network.

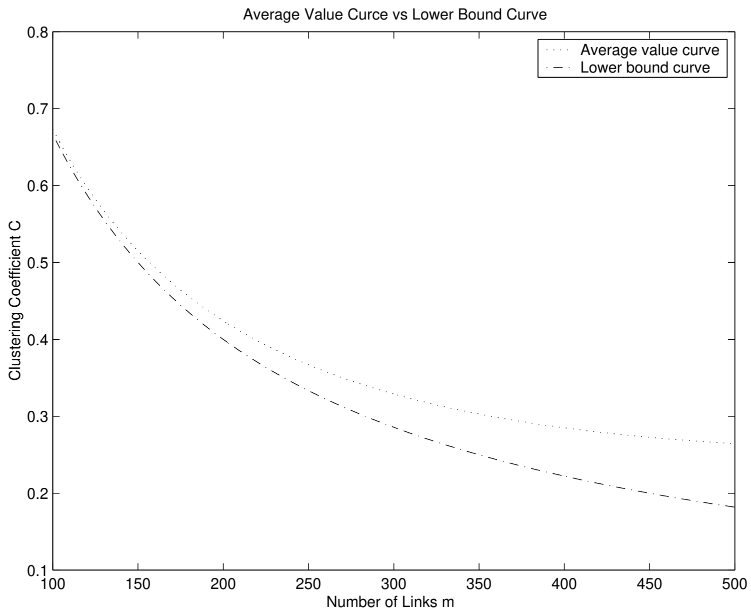
**Theorem 4.1.** *Let $\{G(n,m)\}$ be a set of all possible random agent networks with $n$ agents and $m$ links, and $C_i(n,m)$ be the clustering coefficient of $G_i(n,m) \in \{G(n,m)\}$. Then,*

$$C_i(n,m) \geq \frac{2n}{2m+n} \tag{26}$$

In Fig. 2, we plot the average value curve and lower bound curve of the clustering coefficient in agent networks with 100 agents. From this figure, we can note that the lower bound curve is a monotonously decreasing curve. It is close to the average value curve in the region where $m$ is near $n$. So, in this region, this lower bound is good. But, as the number of links $m$ increases, the lower bound curve gradually gets away from the average value curve. This indicates the lower bound becomes worse. Here, we give a conjecture:




(a)




(b)

Figure 2. (a)The theoretical average value curve and lower bound curve of the clustering coefficient in agent networks with 100 agents. (b)A zoom-in image of the left curves where $m \in [n, 500]$.

**Conjecture 4.1.** *For an agent network $G(n,m)$ with $n$ agents and $m$ links,*

$$f(n,m) = \frac{2m}{n \cdot (n-1)} \tag{27}$$

*is a lower bound of its clustering coefficient.*

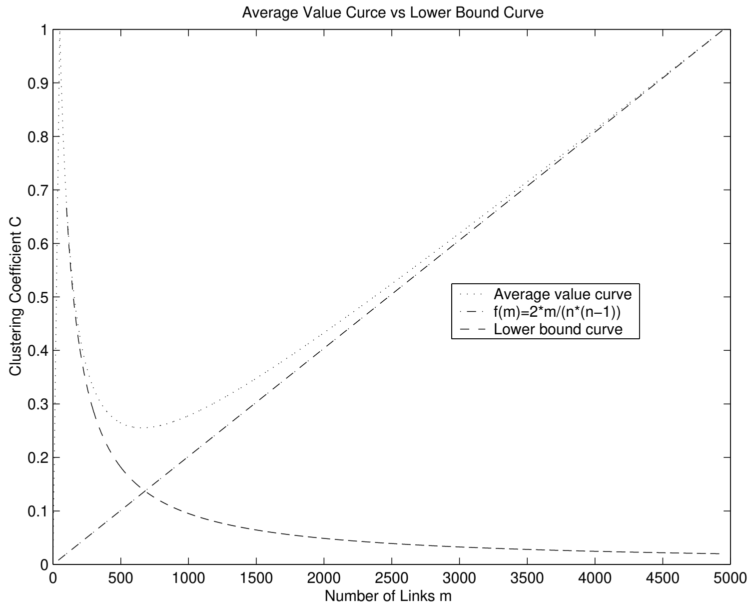



Figure 3. The theoretical average value curve and lower bound curve as well as the conjectured lower bound curve (i.e.,$\frac{2m}{n\cdot(n-1)}$) of the clustering coefficient in agent networks with 100 agents.

In Fig. 3, we plot this conjectured lower bound and compare it with the proven lower bound as well as the average value. We note that as $m \longrightarrow \frac{n\cdot(n-1)}{2}$, $f(n,m)$ curve gets very close to the average value curve. It means $f(n,m)$ becomes a better lower bound as $m \longrightarrow \frac{n\cdot(n-1)}{2}$.

## 5. Conclusions

In this paper, we studied an important property, i.e., *clustering coefficient* of random agent networks. We gave both theoretical and experimental results of the average clustering coefficient in random agent networks with $n$ agents and $m$ links. We further gave a lower bound to the clustering coefficient. But, we noted that, as the number of links increases, this lower bound becomes ineffective. Hence, we proposed a conjecture about another lower bound. Contrary to the previous lower bound, this lower bound (if the conjecture holds) becomes effective gradually as random agent networks approach to be fully connected.

## References

1. J. Delgado. Emergence of social conventions in complex networks. *Artificial Intelligence*, 141:175–181, 2002.
2. K. Hirayama and M. Yokoo. Local search for distributed sat with complex local problems. In *Proceedings of the First International Joint Conference on Autonomous Agent and Multiagent Systems*, pages 1199–1206, Bologna, Italy, 2002.
3. X. Jin and J. Liu. Multiagent sat (massat): Autonomous pattern search in constrained domains. In *Proceedings of the Third International Conference on Intelligent Data Engineering and Automated Learning (IDEAL'02)*, volume 2412 of *LNCS*, pages 318–328, 2002.
4. X. Jin and J. Liu. Agent network topology and complexity. *Submitted to AAMAS 2003*.
5. J. Liu, X. Jin, and J. Han. Distributed problem solving without communication – an examination of computationally hard satisfiability problems. *International Journal of Pattern Recognition and Artificial Intelligence*, 16(8), 2002.
6. J. Liu, H. Jing, and Y. Y. Tang. Multi-agent oriented constraint satisfaction. *Artificial Intelligence*, 136:101–144, 2002.
7. D. J. Watts and S. H. Strogatz. Collective dynamics of small-world networks. *Nature*, 393:440–442, 1998.
8. M. Yokoo and K. Hirayama. Algorithms for distributed constraint satisfaction: A review. *Autonomous Agents and Multi-Agent Systems*, 3(2):198–212, 2000.