

11-4-2013

Complex Network Analysis

Jian James Ma

Computer Information Systems Department Colorado State University-Pueblo

Daniel Zeng

Chinese Academy of Sciences, Beijing

Richard A. Huff

Computer Information Systems Department Colorado State University-Pueblo

Follow this and additional works at: <https://scholarworks.lib.csusb.edu/jitim>

Recommended Citation

Ma, Jian James; Zeng, Daniel; and Huff, Richard A. (2013) "Complex Network Analysis," *Journal of International Technology and Information Management*: Vol. 22 : Iss. 4 , Article 6.

Available at: <https://scholarworks.lib.csusb.edu/jitim/vol22/iss4/6>

This Article is brought to you for free and open access by CSUSB ScholarWorks. It has been accepted for inclusion in *Journal of International Technology and Information Management* by an authorized editor of CSUSB ScholarWorks. For more information, please contact scholarworks@csusb.edu.

Complex Network Analysis: A Comparison of Two Clustering Coefficient Definitions

Jian James Ma
Computer Information Systems Department
Colorado State University-Pueblo
USA

Daniel Zeng
The Key Lab of Complex Systems and Intelligence Science
Institute of Automation
Chinese Academy of Sciences, Beijing, CHINA
and
Department of Management Information Systems
Eller College of Management
University of Arizona
USA

Richard A Huff
Computer Information Systems Department
Colorado State University-Pueblo
USA

ABSTRACT

The clustering coefficient has two definitions that dominate its use, one by Watts and Strogatz, and the second by Newman. It is critical to identify and report the analytic similarities and differences between the two methods in order to give future researchers guidance on the appropriate method to use in their research. This paper reports on an analytical comparison between the two clustering coefficient definitions. We performed the comparison using analytical derivations to show the mathematical relations between the two definitions, the limits on the analysis, and the impact clustering coefficient values derived from the two definitions.

INTRODUCTION

Quantitative evaluation of software architectures with large sizes is a complicated and often difficult task. Traditional analysis methods and measures utilized in software engineering are no longer suitable to analyze complicated software packages. Many readily available evaluation metrics, such as the number of files in a package, total lines of code, or the number of developers for a given project, are not sufficiently descriptive. Even measures such as complexity (Harrison, Samaraweera, Dobie, & Lewis, 1996; Kang & Bieman, 1999), maintainability (Li & Henry, 1993), and cohesion (Etzkorn, Davis, & Li, 1998), often fail to fully capture the nature of increasingly complex software systems. In order to gain meaningful insights into the structure of software systems, software developers and analysts have begun to search for more powerful, detailed, and informative tools to serve their needs.

The object of our research project, complex network analysis, has received additional attention recently. It is widely adopted in analyzing real world complex systems, including software systems (Albert & Barabasi, 2002; Newman, 2003; Potanin, Noble, Freen, & Biddle, 2005). One of the most useful, and frequently leveraged, topological measures employed when random graph theory is applied in complex network analysis is the clustering coefficient. The clustering coefficient measures the extent to which being a neighbor in a cluster is a transitive property.

The clustering coefficient has two commonly used definitions, one by Watts and Strogatz (1998), and the second by Newman (Newman, 2001; Newman, Watts, & Strogatz, 2002; Newman, 2003). These two definitions share some common considerations and, at the same time, have their own unique viewpoints when used as a measure in clustering circumstances. However, when conducting and reporting their research work, most researchers just adopt one of them based on their own needs, without justifying why they used the method chosen. To the best of our knowledge, we are not aware of any other paper that examined the relationship of these two definitions in an analytic and systematic way. Therefore, it is critical to identify and report the analytic similarities and differences between the two methods in order to give future researchers guidance on the appropriate method to use in their research.

This paper intends to clarify the research situation by reporting on an analytical comparison between the two clustering coefficient definitions developed by Watts and Strogatz (1998) and by Newman (2003), respectively. The comparison was performed by using analytical derivations to show the mathematical relations between the two definitions, showing the limits on the analysis, and leveraging a simulated network example to show the impact of clustering coefficient values derived from those two definitions.

The organization of the paper is as follows. We first introduce some background information in Section 2. In Section 3, we briefly explain the topological metrics that we will be using. Section 4 presents the numeric analysis and the derived formulas of the two clustering coefficient definitions. In Section 5, we introduce several numeric properties of the two definitions, including the impact factors of the clustering values for each definition. Finally, we conclude the paper in Section 6 with a summary of contributions and potential future research opportunities.

BACKGROUND

As an important topological measure in graph theory (Tuttle, 1984), the clustering coefficient measures the extent to which being a neighbor in a cluster is a transitive property (Eggemann & Noble, 2011). The clustering coefficient captures the level of connectivity of a local community within a network. The higher the clustering coefficient, the more connected the community is.

As mentioned earlier, and previously discussed in Ma, Zeng, and Zhao (2010), the clustering coefficient has two commonly used definitions (Watts & Strogatz, 1998; Newman, 2003). Watts and Strogatz (1998) define a clustering coefficient C for any network node i having at least two neighbors (if a node has a clustering coefficient with degree zero or one, it is defined as zero):

$$C_i^{(1)} = \frac{a_i}{k_i(k_i - 1)/2}, \quad (1)$$

where a_i is the number of edges (connections) between the neighbors of node i and k_i is the degree of node i , i.e., the number of edges connected to the node. An equivalent, more graphical formulation is:

$$C_i^{(1)} = \frac{\text{Number of triangles connected to node } i}{\text{Number of connected triples centered on node } i}, \quad (2)$$

A connected triple being a single node connected to an unordered pair of other nodes. The clustering coefficient for the entire network, is then defined as the average

$$C^{(1)} = \frac{1}{N} \sum_{i=1}^N C_i^{(1)}. \quad (3)$$

A second definition of the clustering coefficient, introduced by Newman (Newman, Strogatz, & Watts, 2001; Newman et al., 2002; Newman, 2003), is:

$$C^{(2)} = \frac{3 \times \text{Number of triangles in the network}}{\text{Number of connected triples in the network}}. \quad (4)$$

The constant three being used here to normalize $C^{(2)}$ into the $[0,1]$ range, as each triangle contributes to three connected triples centered on different nodes.

The two definitions are similar in that $C^{(1)}$ calculates the mean of ratios, while $C^{(2)}$ the ratio of means. However, they can give quite different results, as $C^{(1)}$ weights the contributions of low-degree nodes more heavily, while $C^{(2)}$ treat all nodes equally.

TOPOLOGICAL METRICS INTRODUCTION

The networks that we analyze for the purposes of this comparison are un-directed and un-weighted networks. They are simple, meaning no self-loops or multiple edges connecting two nodes are allowed.

N_0 is the number of the nodes in the network, which is commonly referred to as the size of the network. In Table 1, N is the number of nodes whose degrees are greater than 1. N is useful when we calculate the clustering coefficient. On the other hand, N_0 is not directly usable for calculating the clustering coefficient because the nodes with degree 1 do not have any effect on the clustering coefficient value. The variables k_i and a_i have the same meanings as in the previous sections. In order to study the detailed information of an individual node's connections, we define two additional variables, $T_a^{(i)}$ and $T_p^{(i)}$. $T_a^{(i)}$ is the number of triangles around node i . A triangle is a group of three nodes that connect to each other. $T_p^{(i)}$ is the number of triples centered at node i . A triple centered at node i is a group of three nodes that node i is connected to the other two nodes. T_a and T_p are aggregated variables that count the total numbers of

triangles and triples, respectively, in the entire network. $C_{WS}^{(i)}$ is the clustering coefficient in the Watts-Strogatz (1998) definition for node i that is connected to at least two other nodes. Finally C_{WS} and C_{NW} are the dependent variables that are the clustering coefficient in the Watts-Strogatz (1998) definition and Newman (2003) definition, respectively, for the entire network.

Table 1: Symbols of network measures.

Symbol	Measure
N_0	Number of nodes in the network, referred to as the network size
N	Number of nodes in the network whose degrees are greater than 1
k_i	Degree of node i , i.e., the number of edges connected to the node
a_i	Number of edges among the neighbors of node i
$T_a^{(i)}$	Number of triangles around node i . A triangle is a group of three nodes that connect to each other.
$T_p^{(i)}$	Number of triples centered at node i . A triple means a single node connected to two other nodes.
T_a	Total number of triangles in the network
T_p	Total number of triples in the network
$C_{WS}^{(i)}$	Clustering coefficient in the Watts-Strogatz definition for node i with the degree value greater than 1
C_{WS}	Clustering coefficient in the Watts-Strogatz definition for the network
C_{NW}	Clustering coefficient in the Newman definition for the network

NUMERIC ANALYSIS

Using the symbols that we defined in Table 1, we present the formulas to calculate C_{WS} and C_{NW} . Based on equations 1 to 4, we present C_{WS} and C_{NW} 's calculations in the following three equations.

$$C_{WS}^{(i)} = \frac{a_i}{k_i(k_i-1)/2} \quad (5)$$

$$C_{WS} = \frac{1}{N} \sum_{i=1}^N C_{WS}^{(i)} \quad (6)$$

$$C_{NW} = \frac{3 \times T_a}{T_p} \quad (7)$$

Based on the definition, a_i is the number of edges among the neighbors of node i . Since every neighbor is connected to node i by definition, every edge among node i 's neighbors

corresponds to a triangle around node i . On the other hand, every triangle around node i must correspond to an edge connecting a pair of node i 's neighbors. Thus,

$$T_a^{(i)} = a_i \quad (8)$$

$T_p^{(i)}$ is the number of triples centered at node i . A triple centered at node i is a group of three nodes that node i is connected to the other two nodes. Every triple centered at node i corresponds to an unordered pair of node i 's neighbors. Thus, the total number of triples centered at node i is the total number of different combinations of node i 's un-ordered neighbors which is $k_i(k_i - 1)/2$.

$$T_p^{(i)} = k_i(k_i - 1)/2 \quad (9)$$

Implanting equations 8 and 9 into equation 5 leads to a new formula to calculate $C_{WS}^{(i)}$ and then C_{WS} .

$$C_{WS}^{(i)} = \frac{T_a^{(i)}}{T_p^{(i)}} \quad (10)$$

$$C_{WS} = \frac{1}{N} \sum_{i=1}^N \frac{T_a^{(i)}}{T_p^{(i)}} \quad (11)$$

We now consider the formula for C_{NW} . The following two equations are quite straightforward. T_a and T_p are aggregated variables of $T_a^{(i)}$ and $T_p^{(i)}$, respectively. Since every triangle is counted three times when considering each node i , the total number of triangles in the network, T_a , should be the summation of $T_a^{(i)}$ divided by 3. Moreover $T_a^{(i)} = T_p^{(i)} = 0$, if $k_i \leq 1, \forall i = 1, \dots, N_0$. Thus,

$$T_a = \frac{1}{3} \sum_{i=1}^{N_0} T_a^{(i)} = \frac{1}{3} \sum_{i=1}^N T_a^{(i)} \quad (12)$$

$$T_p = \sum_{i=1}^{N_0} T_p^{(i)} = \sum_{i=1}^N T_p^{(i)} \quad (13)$$

We then implant equations 12 and 13 into equation 7, and obtain the formula for C_{NW} as shown in equation 14.

$$C_{NW} = \frac{\sum_{i=1}^N T_a^{(i)}}{\sum_{i=1}^N T_p^{(i)}} \tag{14}$$

We will leverage equations 11 and 14 to compare the two clustering coefficient definitions from now on.

Starting from equation 11, $C_{WS} = \frac{1}{N} \sum_{i=1}^N \frac{T_a^{(i)}}{T_p^{(i)}} = \text{Mean of } \frac{T_a^{(i)}}{T_p^{(i)}}$. Starting from equation 14,

$$C_{NW} = \frac{\frac{1}{N} \sum_{i=1}^N T_a^{(i)}}{\frac{1}{N} \sum_{i=1}^N T_p^{(i)}} = \frac{\text{Mean of } T_a^{(i)}}{\text{Mean of } T_p^{(i)}}.$$

Thus we can say C_{WS} is the mean of the ratio $\frac{T_a^{(i)}}{T_p^{(i)}}$, and C_{NW}

is the ratio of the mean of $T_a^{(i)}$ and the mean of $T_p^{(i)}$.

NUMERIC PROPERTIES

Lower bound

Since $T_a^{(i)} \geq 0, \forall i = 1, 2, \dots, N$, then both C_{WS} and C_{NW} are non-negative. Thus, the minimum of the values for both C_{WS} and C_{NW} variables may be zero. The minimum is zero if and only if $T_a^{(i)} = 0, \forall i = 1, 2, \dots, N$. A formal description is listed below.

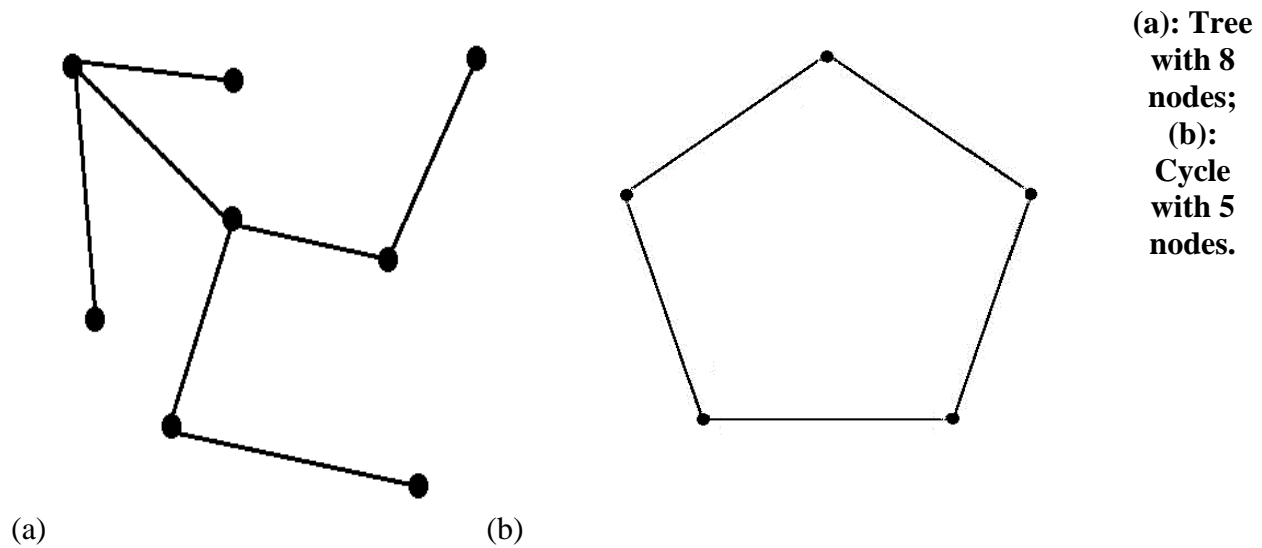
$$C_{WS} \geq 0$$

$$C_{NW} \geq 0$$

$$C_{WS} = C_{NW} = 0 \Leftrightarrow T_a^{(i)} = 0, \forall i = 1, 2, \dots, N$$

In order to satisfy $T_a^{(i)} \geq 0, \forall i = 1, 2, \dots, N$, the network can be a tree where no cycle exists, or a cycle with more than 3 nodes, etc. Figure 1 shows an example for a tree with 8 nodes and a cycle with 5 nodes, respectively. In conclusion, the lower bound of C_{WS} and C_{NW} is met when there does not exist three nodes that are connected to each other in the network.

Figure 1: Examples of both C_{WS} and C_{NW} are 0,



Upper bound

Since $T_a^{(i)} \leq T_p^{(i)}, \forall i = 1, 2, \dots, N$, then both C_{WS} and C_{NW} are less than or equal to 1. Thus the maximum of the values for both C_{WS} and C_{NW} variables may be 1. The maximum is 1 if and only if $T_a^{(i)} = T_p^{(i)}, \forall i = 1, 2, \dots, N$. A formal description is listed below.

$$C_{WS} \leq 1$$

$$C_{NW} \leq 1$$

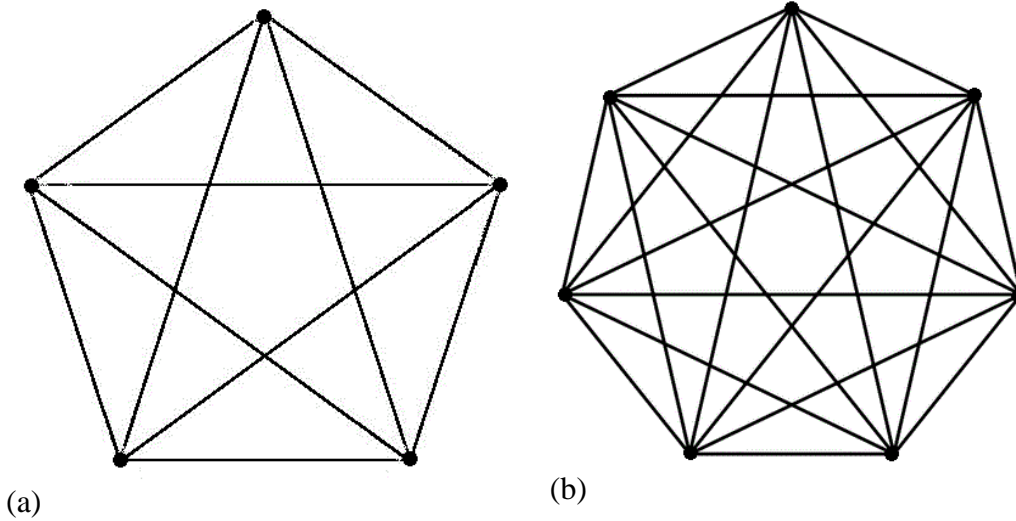
$$C_{WS} = C_{NW} = 1 \Leftrightarrow T_a^{(i)} = T_p^{(i)}, \forall i = 1, 2, \dots, N$$

In order to satisfy $T_a^{(i)} = T_p^{(i)}, \forall i = 1, 2, \dots, N$, the network has to be a complete graph where every node is connected to every other nodes. Thus C_{WS} and C_{NW} values meet the upper bound only when the network is a complete graph. Figure 2 shows two complete graph examples with 5 nodes and 7 nodes, respectively.

Equality

We are interested in exploring the conditions where $C_{WS} = C_{NW}$. If $T_p^{(i)} = T_p^{(j)}, \forall i, j = 1, 2, \dots, N$, then

Figure 2: Examples of both C_{WS} and C_{NW} are 1
(a): Complete graph with 5 nodes; (b): Complete graph with 7 nodes.



$$C_{WS} = \frac{1}{N} \sum_{i=1}^N \frac{T_a^{(i)}}{T_p^{(i)}} = \frac{1}{N \cdot T_p^{(1)}} \sum_{i=1}^N T_a^{(i)}$$

$$C_{NW} = \frac{\sum_{i=1}^N T_a^{(i)}}{\sum_{i=1}^N T_p^{(i)}} = \frac{\sum_{i=1}^N T_a^{(i)}}{\sum_{i=1}^N T_p^{(1)}} = \frac{1}{N \cdot T_p^{(1)}} \sum_{i=1}^N T_a^{(i)} = C_{WS} .$$

Since $T_p^{(i)} = k_i(k_i - 1)/2$, thus $T_p^{(i)} = T_p^{(j)}, \forall i, j = 1, 2, \dots, N$ is equivalent to $k_i = k_j, \forall i, j = 1, 2, \dots, N$. Therefore,

$$C_{WS} = C_{NW} \Leftarrow k_i = k_j, \forall i, j = 1, 2, \dots, N$$

Note $k_i = k_j, \forall i, j = 1, 2, \dots, N$ is the sufficient condition, but not a necessary condition. That is to say, if the degrees of all nodes in a network are the same, then $C_{WS} = C_{NW}$.

Figure 3 shows two graph examples that all nodes in the graph have the same degrees. Figure 3(a) is an octahedron with 6 nodes. Each node has a degree 4 and each node has 4 triangles around it.

$$\text{Thus } C_{WS} = C_{NW} = \frac{4}{4(4-1)/2} = 2/3 .$$

Figure 3: Examples of $C_{WS} = C_{NW}$, (a): Octahedron; (b): Cube.

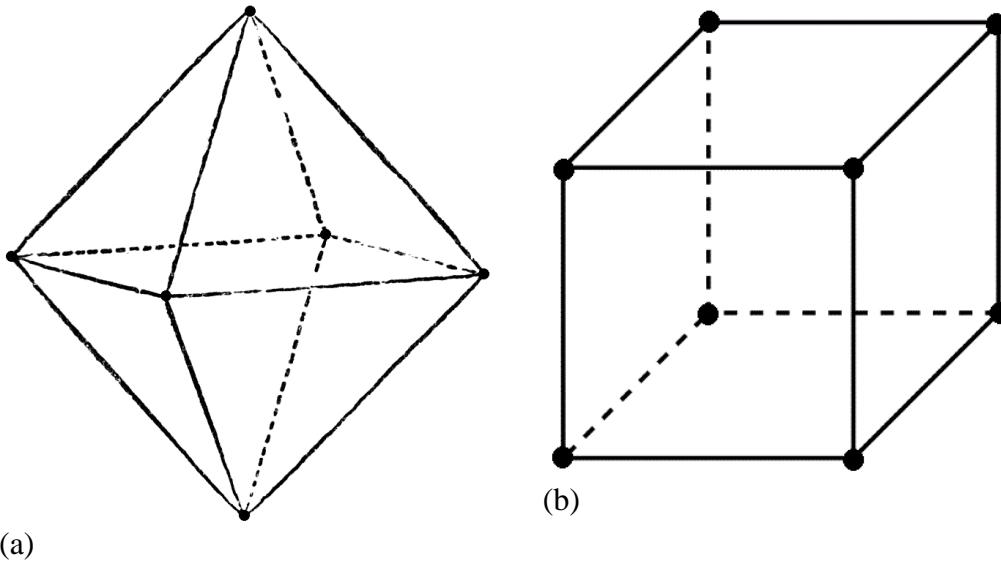
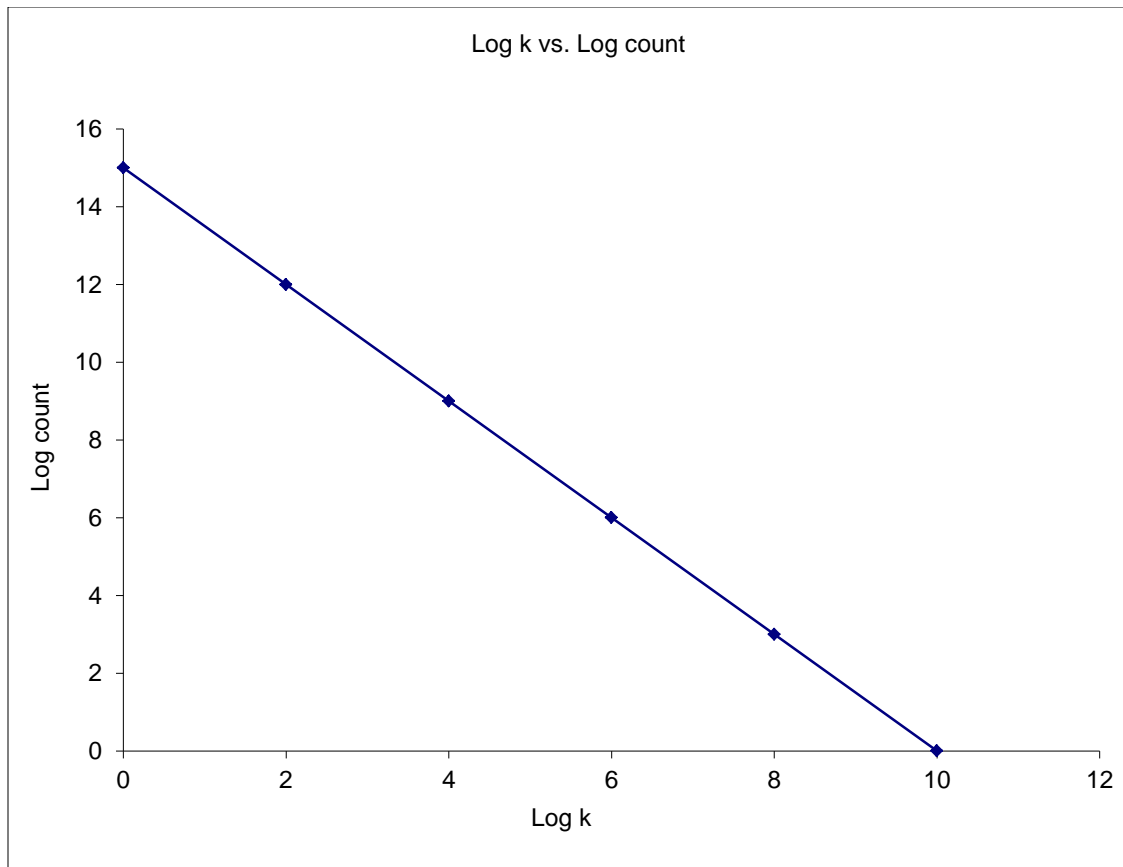


Figure 3(b) is a cube with 8 nodes, each of which has a degree 3. Note each node has 0 triangles around it, which results in $C_{WS} = C_{NW} = 0$. This example not only exhibits the equality between C_{WS} and C_{NW} , but also fits in the lower bound scenario as well.

Figure 4: Log degree k vs. log count p(k).



Impact of nodes with large degrees

Recent research findings show that lots of real world complex networks possess the scale-free properties (Newman & Watts, 1999; Albert & Barabasi, 2000a; Albert & Barabasi, 2000b; Goh, Kahng, & Kim, 2001a; Goh, Kahng, & Kim, 2001b; Cohen, Ben-Avraham, Havlin, 2002; Goh, Oh, Jeong, Kahng, & Kim, 2002; Schwartz, Cohen, Ben-Avraham, Barabasi, & Havlin, 2002; Cohen & Havlin, 2003; Goh, Lee, Kahng, & Kim, 2003; Vazquez, Boguna, Moreno, Pastor-Satorras, & Vespignani, 2003) whose degree distributions follow the power law. Specifically, the logarithmic values of degrees and the logarithmic values of the number of nodes with the same degrees form a decreasing straight line. Intuitively speaking, there are very few nodes with very large degrees, and most nodes have low degrees. Figure 4 presents a simulated example of the scale-free network.

Table 2: Simulated degree distribution.

Log k	Log Count	K	Count	k(k-1)/2 * Count
0	15	1	32,768	0
2	12	4	4,096	24,576
4	9	16	512	61,440
6	6	64	64	129,024
8	3	256	8	261,120
10	0	1,024	1	523,776

We leverage the simulated network above to further compare the two clustering coefficient definitions by Watts-Strogatz (1998) and Newman (2003). Table 2 lists some calculated results derived from the degree distribution. The values of Log k and Log Count form the decreasing straight line presented in Figure 4. From the logarithmic values, we can compute k and count values. The last column, $k(k-1)/2 * \text{count}$, will be used later to analyze C_{NW} .

From the previous section, we know $C_{WS} = \frac{1}{N} \sum_{i=1}^N \frac{T_a^{(i)}}{T_p^{(i)}} = \text{Mean of } \frac{T_a^{(i)}}{T_p^{(i)}}$. The value of C_{WS} is an

average of ratio $\frac{T_a^{(i)}}{T_p^{(i)}}$. And $0 \leq \frac{T_a^{(i)}}{T_p^{(i)}} \leq 1, \forall i = 1, \dots, N$, the values of $\frac{T_a^{(i)}}{T_p^{(i)}}$ are within a limited range

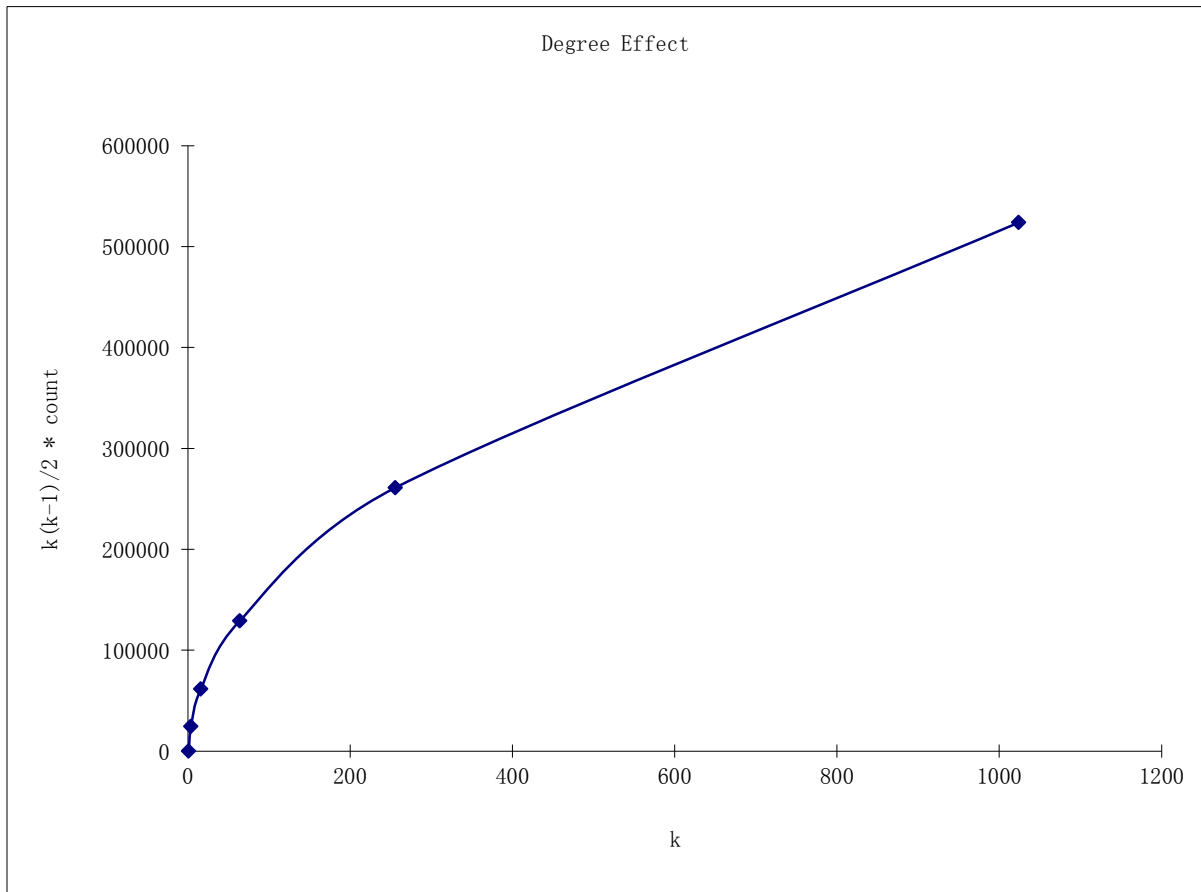
from 0 to 1. If we assume the nodes with the same degrees have the similar a_i values, then the value of C_{WS} largely depends on the majority of nodes who have the same degree values. Specifically, there are 32,768 nodes with the same degree 1. Those nodes do not have any impact on the clustering coefficient based on our definition, so that we can simply ignore those nodes. There are 4,096 nodes with degree 4. Those 4,096 nodes will yield a dominating factor to the overall value of C_{WS} , since C_{WS} is the average of the 4,681 nodes whose degrees are greater than 1. On the other hand, although the one node with the degree 1024 is the most “popular” node of the network, its $C_{WS}^{(i)}$ only accounts for 1/4681 to the overall C_{WS} value. The impact of the popular nodes is smothered by that of the un-popular nodes, which are dominating the node count.

However C_{NW} is totally different from C_{WS} , in that the few popular nodes are the dominating

factors in determining the final value of C_{NW} . Equation 14 states $C_{NW} = \frac{\sum_{i=1}^N T_a^{(i)}}{\sum_{i=1}^N T_p^{(i)}}$. Let us

consider $T_p^{(i)}$, whose value is $k_i(k_i - 1)/2$ (per equation 9). To analyze the impact of the popular and un-popular nodes to the mean of $T_p^{(i)}$, we aggregate the impact of $T_p^{(i)}$ for nodes who have the same degree values. The last column in Table 2 indicates which set of nodes has the dominating factor. Although the un-popular nodes, with degree 4, have the count advantage (count equals 4,096), the overall summation of $T_p^{(i)}$ for the un-popular nodes is only 24,576. On the other hand, the one popular node with degree 1,024 has a huge impact since its $k_i(k_i - 1)/2$ value is 523,776. In the overall $\sum_{i=1}^N T_p^{(i)}$ value, the one popular node has a dominating factor, and the 4,096 unpopular nodes do not play an important role. Figure 5 plots the degree k versus $k(k-1)/2 * \text{count}$.

Figure 5: Degree effect: degree k vs. $k(k-1)/2 * \text{count}$.



In conclusion, the two clustering coefficient definitions, C_{WS} and C_{NW} , have different impact factors. The value of C_{WS} is dominated by the un-popular nodes, which have low degree values, and usually the node count advantage. The few popular nodes with extremely high degrees do not have much impact on C_{WS} . On the other hand, those few popular nodes play the most important role in calculating C_{NW} . The low-degree nodes are not as important as the few extremely popular nodes.

CONCLUSIONS

Network analysis is becoming an important method for studying complex systems, and the clustering coefficient remains one of the most useful measures in examining network characteristics. This paper aims to provide an analytical comparison between two widely adopted clustering coefficient definitions, C_{WS} and C_{NW} , as proposed by Watts and Strogatz, and Newman, respectively. Mathematical derivations were presented to compare the similarities and the differences between those two definitions. Our findings show that the two definitions both depend on $T_a^{(i)}$ and $T_p^{(i)}$, the number of triangles and triples, respectively, around node i . The difference between those two definitions lies in that C_{WS} is the mean of the ratio $T_a^{(i)}$ and $T_p^{(i)}$, and C_{NW} is the ratio of the two means of $T_a^{(i)}$ and $T_p^{(i)}$. We also examined the lower bounds and upper bounds of those two definitions, and the conditions to meet those extreme bounds. Our further analysis shows the impact factors of C_{WS} and C_{NW} values. Using a scale-free simulated network, we found that the extremely popular nodes have little impact on C_{WS} due to the limited number of those popular nodes, whereas those popular nodes are the dominating factors in determining the value of C_{NW} .

Our research findings show detailed properties of the two clustering coefficient definitions, C_{WS} and C_{NW} . It provides researchers more insights when conducting network analysis research. Our findings provide the guidelines on which clustering coefficient definition should be used when analyzing a network. Moreover our results give researchers usable hints when a random network model is needed to explain the topological measures found in real world complex systems. Specifically in software engineering, software practitioners can leverage our research results to analyze complex software products, engineer collaborations, and create product development processes, when complex networks are chosen to conduct the analysis.

FUTURE RESEARCH POSSIBILITIES

Our research examined the condition where $C_{WS} = C_{NW}$ and identified that $k_i = k_j, \forall i, j = 1, 2, \dots, N$ (the degrees of all nodes in a network are the same) is the sufficient condition for the equality to exist. We question if this is also the necessary condition for the equality to exist. This may be an idealized situation that does not exist in actual networks, scale-free or otherwise. Therefore, evaluation of the condition in other contexts presents opportunities for fruitful endeavor.

Another opportunity exists to extend our research to other networks in the examination of the impact of nodes with large degrees on the determination of the clustering coefficient. The

current project only utilized a simulated scale-free network to test our findings. As referenced earlier, many researchers have identified real world scale-free networks. Identification of real world networks with large degree nodes and application of the methodology described herein offers a chance to verify the veracity of our methods.

REFERENCES

- Albert, R., & Barabasi, A. L. (2000a). Dynamics of Complex Systems: Scaling Laws for the Period of Boolean Networks. *Physical Review Letters*, 84, 5660-5663.
- Albert, R., & Barabasi, A. L. (2000b). Topology of Evolving Networks: Local Events and Universality. *Physical Review Letters*, 85, 5234-5237.
- Albert, R., & Barabasi, A. L. (2002). Statistical Mechanics of Complex Networks. *Reviews of Modern Physics*, 74, 47-97.
- Cohen, R., Ben-Avraham, D., & Havlin, S. (2002). Percolation Critical Exponents in Scale-Free Networks. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 66, 36113.
- Cohen, R., & Havlin, S. (2003). Scale-Free Networks Are Ultrasmall. *Physical Review Letters*, 90, 058701.
- Eggemann, N., & Noble, S. D. (2011). The Clustering Coefficient of a Scale-Free Random Graph. *Discrete Applied Mathematics*, 159, 953-965.
- Etzkorn, L.C., Davis, C., & Li, W. (1998). A Practical Look At the Lack of Cohesion in Methods Metric. *Journal of Object-Oriented Programming*, 11, 27-34.
- Goh, K.-I., Kahng, B., & Kim, D. (2001a). Spectra and Eigenvectors of Scale-Free Networks. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 64, 051903.
- Goh, K.-I., Kahng, B., & Kim, D. (2001b). Universal Behavior of Load Distribution in Scale-Free Networks. *Physical Review Letters*, 87, 278701.
- Goh, K.-I., Oh, E., Jeong, H., Kahng, B., & Kim, D. (2002). Classification of Scale-Free Networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 12583-12588.
- Goh, K.-I., Lee, D.-S., Kahng, B., & Kim, D. (2003). Sandpile on Scale-Free Networks. *Physical Review Letters*, 91, 148701.
- Harrison, R., Samaraweera, L. G., Dobie, M. R., & Lewis, P. H. (1996). Comparing Programming Paradigms: An Evaluation of Functional and Object-Oriented Programs. *Software Engineering Journal*, 11, 247-254.
- Kang, B. K., & Bieman, J. M. (1999). A Quantitative Framework for Software Restructuring. *Journal of Software Maintenance: Research and Practice*, 11, 245-284.

- Kim, J.-H., Goh, K.-I., Kahng, B., & Kim, D. (2003). Probabilistic Prediction in Scale-Free Networks: Diameter Changes. *Physical Review Letters*, 91, 058701.
- Li, W., & Henry, S. (2003). Object-Oriented Metrics That Predict Maintainability. *Journal of Systems and Software*, 23, 111-122.
- Ma, J., Zeng, D., & Zhao, H. (2012). Modeling the Growth of Complex Software Function Dependency Networks. *Information Systems Frontiers*, 14, 301-315.
- Newman, M. E. J., & Watts, D.J. (1999). Scaling and Percolation in the Small-World Network Model. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 60, 7332-7342.
- Newman, M. E. J. (2001). Clustering and Preferential Attachment in Growing Networks. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 64, 025102.
- Newman, M. E. J., Strogatz, S.H., & Watts, D.J. (2001). Random Graphs With Arbitrary Degree Distributions and Their Applications. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 64, 026118.
- Newman, M. E. J. (2003). The Structure and Function of Complex Networks. *SIAM Review*, 45, 167-256.
- Newman, M. E. J., Watts, D. J., & Strogatz, S. H. (2002). Random Graph Models of Social Networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 2566-2572.
- Potanin, A., Noble, J., Frean, M., & Biddle, R. (2005). Scale-Free Geometry in OO Programs. *Communications of the ACM*, 48, 99-103.
- Schwartz, N., Cohen, R., Ben-Avraham, D., Barabasi, A. L., & Havlin, S. (2002). Percolation in Directed Scale-Free Networks. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 66, 015104.
- Tutte, W. T. (1984). *Encyclopedia of Mathematics and its Applications, Volume 21: Graph Theory*. Cambridge, U.K.: Cambridge University Press.
- Vazquez, A., Boguna, M., Moreno, Y., Pastor-Satorras, R., & Vespignani, A. (2003). Topology and Correlations in Structured Scale-Free Networks. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, 67, 046111.
- Watts, D. J., & Strogatz, S. H. (1998). Collective Dynamics of "Small-World" Networks. *Nature*, 393, 440-442.
- Watts, D. J. (2003). *Small Worlds: The Dynamics of Networks between Order and Randomness (Princeton Studies in Complexity)*. Princeton, N.J.: Princeton University Press.