

# Extended Clustering Coefficients of Small-World Networks

Wenjun Xiao<sup>1</sup>, Yong Qin<sup>1,2</sup>, and Behrooz Parhami<sup>3</sup>

<sup>1</sup> Dept. Computer Science, South China University of Technology, Guangzhou 510641, P.R. China

wjxiao@scut.edu.cn

<sup>2</sup> Center of Computer Network and Information, Maoming University, Maoming 525000, P.R. China

<sup>3</sup> Dept. Electrical & Computer Eng., University of California, Santa Barbara, CA 93106-9560, USA

**Abstract.** The clustering coefficient  $C$  of a network, which is a measure of direct connectivity between neighbors of the various nodes, ranges from 0 (for no connectivity) to 1 (for full connectivity). We define extended clustering coefficients  $C(h)$  of a small-world network based on nodes that are at distance  $h$  from a source node, thus generalizing distance-1 neighborhoods employed in computing the ordinary clustering coefficient  $C = C(1)$ . Based on known results about the distance distribution  $P_\delta(h)$  in a network, that is, the probability that a randomly chosen pair of vertices have distance  $h$ , we derive and experimentally validate the law  $P_\delta(h)C(h) \leq c \log N/N$ , where  $c$  is a small constant that seldom exceeds 1. This result is significant because it shows that the product  $P_\delta(h)C(h)$  is upper-bounded by a value that is considerably smaller than the product of maximum values for  $P_\delta(h)$  and  $C(h)$ .

## 1 Introduction

Complex systems with many components and associated interactions abound in nature, prevail within society, and define many human artifacts. The interconnection or interaction structure of such systems are typically neither random (amenable to probabilistic analysis) nor regular (mathematically tractable), rendering the systematic study of their properties a rather challenging undertaking. Interactions in such systems can be modeled by networks/graphs composed of vertices/nodes and undirected or directed edges/links. A network or graph  $G = (V, E)$  has a set  $V$  of  $N$  vertices or nodes and a set  $E$  of  $M$  edges or links, where each edge is defined by a pair of vertices (ordered pair, for directed graphs).

Two models of actual complex networks have been studied extensively [1,2,3,4]: the small-world model and the scale-free one. Our focus in this paper is on small-world networks that feature localized clusters connected by occasional long-range links, leading to an average distance between vertices that grows logarithmically with the network size  $N$ .

Watts and Strogatz [2] studied mechanisms via which a regular network can be transformed into a small-world network, with little or no change in the vertex-degree distribution, and quantified the parameters that characterize the resulting structures. One feature shared by small-world networks is that their clustering coefficients are fairly high compared with random networks[2]. Clustering coefficient is defined as follows. Let a vertex  $v$  of  $G$  have  $k(v)$  neighbors; that is,  $v$  has degree  $k(v)$ . These  $k(v)$  neighbors can potentially be connected via  $k(v)(k(v) - 1)/2$  edges. The fraction of this maximum possible number of edges that actually exist between neighbors of  $v$  is its clustering coefficient  $C_v$ ; the average of clustering coefficients over all  $v \in V$  is the clustering coefficient  $C$  of the network  $G$ . A network with  $C$  close to 1 may consist of highly connected clusters or cliques, perhaps with sparser connections between the local clusters.

## 2 Extended Clustering Coefficients

We extend the clustering coefficient of Watts and Strogatz [2] in the following way. Define the  $h$ -neighbors of a vertex  $v$  as vertices of  $G$  that are at distance  $h$  (measured in number of hops) from  $v$ . Assume that  $v$  has  $k_h(v)$  such  $h$ -neighbors, where  $k_1(v)$  is the same as  $k(v)$  defined earlier (see Section 1). Then there can be at most  $k_h(v)(k_h(v) - 1)/2$  edges connecting  $h$ -neighbors of vertex  $v$ . The fraction  $C_v(h)$  of allowable edges that actually exist between  $h$ -neighbors  $v$  is the  $h$ -clustering coefficient of  $v$ . We assume that  $C_v(h) = 1$  when  $k_h(v) = 1$ , which also covers the special case  $h = 0$ . The average of  $C_v(h)$  over all  $v \in G$  is the  $h$ -clustering coefficient  $C(h)$  of  $G$ . The 1-clustering coefficient  $C(1)$  is the clustering coefficient  $C$  as defined in Section 1.

Thus, while the definition of clustering coefficient is based on the immediate neighborhood of vertices, extended clustering coefficient relates to a wider neighborhood defined by the distance parameter  $h$ . Using experimental data from a wide variety of actual complex networks, along with a deterministic model of small-world networks that we have developed, we seek to relate  $C(h)$  and the distance distribution  $P_\delta(h)$  of a network, defined as the probability that a randomly chosen pair of vertices are at distance  $h$  from each other.

Note that all distances referred to in this paper are shortest distances. However, in view of the results of Kim et al. [5], distances obtained from a routing algorithm with localized decisions are not fundamentally different from shortest distances in complex networks. Thus, our results are expected to remain valid when this latter definition of distance is used in lieu of shortest distance.

## 3 Experimental Observations

For an  $N$ -vertex network with  $M$  edges, we have  $P_\delta(0) = 1/N$  and  $P_\delta(1) = 2M/N^2 > 1/N$ . Beyond  $h = 1$ , however, a precise expression for the value of  $P_\delta(h)$  cannot be supplied, except in the case of certain regular networks. However, for many networks (small-world or otherwise), the value of  $P_\delta(h)$  rises with  $h$  until it reaches a maximum value and then declines as the distance  $h$  gets

closer to the network diameter  $D$ . This is confirmed experimentally for several complex networks of practical interest in Fig1b, Fig2b.

For extended clustering coefficients, the trend begins by a decrease in clustering, from  $C(0) = 1$  to  $C(1) = C$ , and is then followed by further reductions. This is owing to the fact that as  $h$  increases, the number  $q_h$  of nodes at distance  $h$  from a given node increases, and such nodes are apt to belong to several cliques; hence, the presence of many edges between them is quite unlikely. As  $h$  approaches  $D$ , however, a different effect may take hold. Consider, for example, one extreme case where each node in the network is at distance  $D$  from exactly one node (it has a single diametrically opposite node). This leads to  $C(D) = 1$ . In this same situation,  $C(D - 1)$  is likely to be large as well, given the common presence of multiple diametral paths to the same opposite vertex. Note that the preceding argument suggests that  $C(h)$  can be large when  $h$  approaches  $D$ ; it does not imply that  $C(h)$  must be large in this case. Figures 1c and 2c confirm these trends.

Given the opposing trends of  $P_\delta(h)$  (up, then down) and  $C(h)$  (down, then possibly up), one might be led to believe that the product  $P_\delta(h)C(h)$  has an upper bound. Based on the evidence presented in Fig1a and Fig2a, we conjecture that this is in fact the case. That is, for a constant  $c$  in the vicinity of and seldom exceeding 1, we have:

$$P_\delta(h)C(h) \leq c \log N / N \quad (1)$$

In the special case of  $h = 1$ , equation (1) implies  $P_\delta(1)C(1) \approx \log N / N$ . We have  $P_\delta(1) = 2M/N^2 \approx \log N / N$  for small-world networks. This is consistent with  $C(1) = C$  being large for such networks.

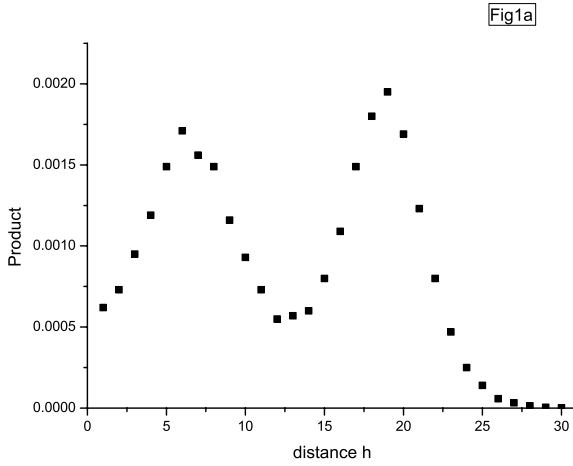
## 4 Model-Based Validation

We now present additional evidence on the validity of equation (1), using a model of deterministic small-world networks that we have developed [6]. In fact, it was through this model that we first became aware of the trend represented in equation (1) and experimentally confirmed in Fig1 and Fig2. A review of our deterministic model, which is based on Cayley graphs [7], has been provided in the supporting information, where we also show that the model yields the clustering coefficient:

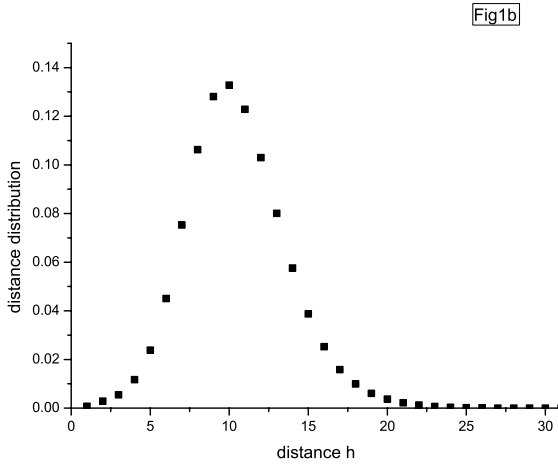
$$C = \frac{at(at - 1)}{(at + t - l)(at + t - l - 1)} \quad (2)$$

In this model,  $t = \log_2 N$  and  $a = (2^l - 1)/t$  is a free tuning parameter that is related to the interconnection density, thereby affecting the value of  $C$ . Note that for very large networks ( $N, t \rightarrow +\infty$ ),  $C$  tends to  $a^2/(a + 1)^2$  when  $a$  is a constant. By suitably choosing  $a$ , we can obtain different clustering coefficients, while maintaining a small vertex degree equal to  $at + t - l = (a + 1)\log_2 N - 1$ .

Unlike actual networks for which the computation of  $C(h)$  is extremely difficult, our deterministic model is amenable to mathematical analysis that yields



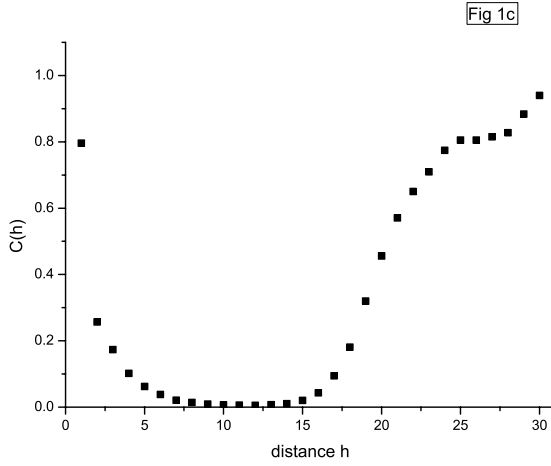
**Fig. 1a.** The plot of  $\text{Product}(P_\delta(h)C(h))$  versus  $h$  in the maximum component  $\Delta$  of NCSTRL graph [9], with 6396 vertices and diameter of 31



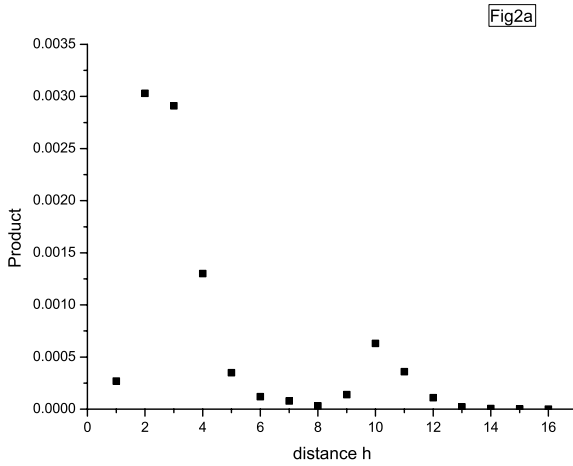
**Fig. 1b.** The plot of distance distribution  $(P_\delta(h))$  versus  $h$  in  $\Delta$

an approximate closed-form expression for the extended clustering coefficients. In our deterministic model, the number  $m$  of adjacent vertex pairs among the  $h$ -neighbors of any vertex is given by the expression:

$$m = (2^l - 1)(2^{l-1} - 1) \binom{t-l}{h-1} \quad (3)$$



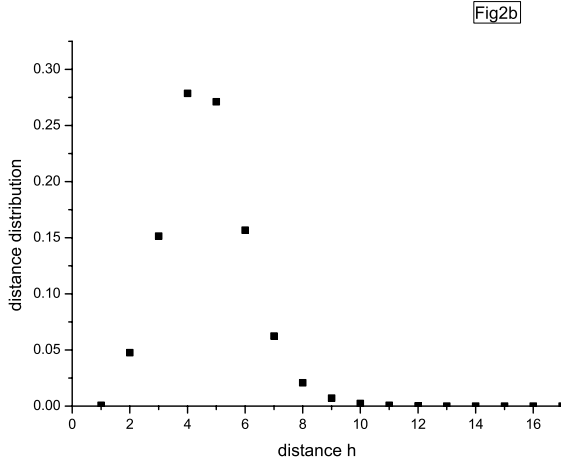
**Fig. 1c.** The plot of  $C(h)$  versus  $h$  in  $\Delta$



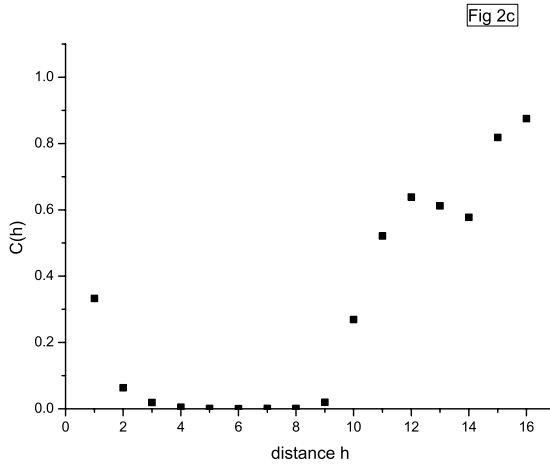
**Fig. 2a.** The plot of  $\text{Product}(P_\delta(h)C(h))$  versus  $h$  in the maximum component  $\Delta_1$  of Linux graph [10] with 5285 vertices and diameter of 17

On the other hand, the number  $k_h(v)$  of  $h$ -neighbors of a vertex  $v$  is bounded as:

$$(2^l - 1) \binom{t-l}{h-1} \leq k_h(v) \leq (2^l - 1) \binom{t-l}{h-1} + \binom{t-l}{h} \quad (4)$$



**Fig. 2b.** The plot of distance distribution( $P_\delta(h)$ ) versus  $h$  in  $\Delta 1$



**Fig. 2c.** The plot of  $C(h)$  versus  $h$  in  $\Delta 1$

Given that the extended clustering coefficient  $C(h)$  is proportional to  $m/(k_h(v))^2$ , we readily find:

$$C(h) \approx 1/\left(\frac{t-l}{h-1}\right) \quad (5)$$

In a companion paper[8], we have derived the distance distribution for small-world networks:

$$P_\delta(h) \approx \binom{\log N}{h} / N \quad (6)$$

Here, we have  $\log N \approx D$ . Because the diameter of our deterministic network model is  $D = t - l + 1$ , we conclude:

$$P_\delta(h) \approx \binom{t-l-1}{h} / N = (t-l+1) \binom{t-l}{h-1} / (hN) \quad (7)$$

Equations (5) and (7) lead to:

$$P_\delta(h)C(h) \leq c \log N / N \quad (8)$$

Equation (8) confirms our hypothesis in equation (1), thereby supplementing the previously supplied experimental evidence of its validity.

## 5 Conclusion

We have shown that extended clustering coefficients are generalizations of ordinary clustering coefficient and are governed by laws that are also generalizations of those pertaining to the latter. We have presented experimental and analytical evidence that the inequality  $P_\delta(h)C(h) \leq c \log N / N$  holds for small-world networks. This result is significant because it shows that the product  $P_\delta(h)C(h)$  is upper-bounded by a value that is considerably smaller than the product of maximum values for  $P_\delta(h)$  and  $C(h)$ . Thus, extended clustering coefficients offer new insights into the structure of small-world networks and open up further avenues for exploration of their properties. Additionally, different shapes for the variations of  $C(h)$  and  $P_\delta(h)C(h)$ , exemplified by Fig1 and Fig2, can be used to categorize small-world networks in order to facilitate their study.

## Acknowledgments

The authors thank M. E. J. Newman for providing the NCSTRL data used in Fig. 1. Research of W. Xiao is supported by the Natural Science Foundation of China and Guangdong Province.

## References

1. Barabasi, A.-L., & Albert, R. (1999) *Science* 286, 509-512
2. Watts, D. J., & Strogatz, S. H. (1998) *Nature* 393, 440-442.
3. Albert, R., & Barabasi, A.-L. (2002) *Rev. Mod. Phys.* 74, 47-97.
4. Newman, M. E. J. (2003) *SIAM Rev.* 45, 167-256.
5. Kim, B. J., Yoon, C. N., Han, S. K., & Jeong, H. (2002) *Phys. Rev.* E65, 027103.
6. Xiao, W. J., & Parhami, B. (2006) *Info. Proc. Lett.* 97, 115-117.
7. Biggs, N. (1993) *Algebraic graph theory* (Cambridge Univ. Press).
8. Xiao, W.J., & Parhami, B. (2005) On conditions for scale-freedom in complex networks, submitted for publication.
9. Newman, M. E. J. (2001) *Phys. Rev. E* 64, 016131.
10. Myrnes, C. R. (2003) *Phys. Rev. E* 68, 046116.