Network Science **3** (2): 183–195, 2015. © Cambridge University Press 2015
doi:10.1017/nws.2014.26

A clustering coefficient for complete weighted networks

MICHAEL P. MCASSEY and FETSJE BIJMA
Department of Mathetmatics, VU University Amsterdam, Amsterdam, the Netherlands
(email: m.p.mcassey@auc.nl, f.bijma@vu.nl*)*

Abstract

The clustering coefficient is typically used as a measure of the prevalence of node clusters in a network. Various definitions for this measure have been proposed for the cases of networks having weighted edges which may or not be directed. However, these techniques consistently assume that only a subset of all possible edges is present in the network, whereas there are weighted networks of interest in which all possible edges are present, that is, complete weighted networks. For this situation, the concept of clustering is redefined, and computational techniques are presented for computing an associated clustering coefficient for complete weighted undirected or directed networks. The performance of this new definition is compared with that of current clustering definitions when extended to complete weighted networks.

Keywords: *binary network, clustering coefficient, complete network, directed network, undirected network, weighted network*

Network theory has been developed to model complex systems which involve elements represented as nodes (or vertices) of a network and their mutual connections represented as edges between nodes (Albert & Barabási, 2003; Dorogovtsev & Mendes, 2003; Newman, 2003). Initially, this theory concentrated on networks whose edges are binary (either present or absent) and undirected. While such networks have been sufficient to model many real-world phenomena, there has arisen a need for further complexity to model systems in which heterogeneous strengths of the connections between pairs of nodes must be considered, and systems in which the presence of a connection between node pairs is asymmetric (node i may be connected to node j, but node j is not necessarily connected to node i). These phenomena are modeled respectively by weighted networks and by directed networks. Naturally, these two concepts can be merged to form weighted directed networks (WDNs), in which the edge from node i to node j may be assigned a different weight from that of the edge from node j to node i, and either edge can be absent.

Various structural features of networks have been identified as useful properties which enhance the efficiency of a network in carrying out its essential functionality. Among these is the concept of clustering. This signifies the presence of well-connected neighborhoods of nodes within the network, more than would be found in a random network. The clustering coefficient was developed as a means to measure the degree to which a network manifests this property. The clustering coefficient was first designed with binary undirected networks (BUNs) in mind (Szabó et al., 2004; Watts & Strogatz, 1998). The measure has subsequently been extended to binary directed networks (BDNs) and to weighted undirected networks (WUNs) (Barrat et al., 2004; Li et al., 2009; Onnela et al., 2005; Saramäki et al., 2007; Zhang & Horvath, 2005; Kalna & Higham, 2006), and eventually to WDNs (Fagiolo, 2007).

The clustering coefficient formulations for weighted networks, whether directed or undirected, consistently assume that only a subset of node pairs have an edge between them. See Saramäki et al. (2007) for a comparison among four such definitions in the literature. For a network consisting of N nodes, these definitions may involve the $N \times N$ network adjacency matrix $A = [a_{ij}]$, in which $a_{ij} = 1$ if there is an edge (possibly directed) from node i to node j and $a_{ij} = 0$ otherwise, and the $N \times N$ weight matrix $W = [w_{ij}]$, which provides the weight w_{ij} of the edge from node i to node j. (Typically $a_{ii} = w_{ii} = 0$ for all i, and the weights are normalized so that $0 \leqslant w_{ij} \leqslant 1$ for all pairs i and j. In the directed setting, a_{ij} and w_{ij} may differ from a_{ji} and w_{ji}, respectively.) As with their binary counterparts, the clustering coefficient for an individual node is generally conceived of as some function of the edge weights of all existing triangular paths involving that node and all pairs of adjacent nodes, normalized by some maximum. The clustering coefficient for the network is then the average of its node-wise clustering coefficients. For sparsely connected weighted networks, current definitions of this measure are adequate for describing the prevalence of node clusters in the network.

However, there are many weighted networks which arise in practice for which either directed or undirected edges exist between *every* pair of nodes. Such a network will be referred to as a *complete* network. For example, when the nodes represent some random phenomena and the weighted edges represent the pairwise correlations of these phenomena, it is possible that all correlations are non-zero, although many may be quite small. In such case the network is a complete WUN, with the weights taken as the absolute values of the correlations. This would also be the case if the edge weights are functions of the distances d_{ij} between nodes, as when $w_{ij} = 1/d_{ij}$ normalized by some constant. In another example based on a neuroscience study (McAssey et al., 2013), a network is created in which the nodes are simulated neurons, and each weighted, directed edge from one node to the other represents the expected number of potential synapses in that direction. The displacement between neurons may be more favorable for synaptic connectivity between neurons in one direction as opposed to the other, so the edge weights are not symmetric. Hence this is an example of a complete WDN.

The concept of clustering has not been defined for the case of a complete network, that is, a network consisting of N nodes and $N(N-1)/2$ undirected edges, or $N(N-1)$ directed edges. Clustering does not make sense for complete binary networks, but when the edges are weighted, particularly when most of the edges have small weights, a meaningful definition of clustering is plausible. Applying the currently available definitions for WUNs and WDNs in the complete network context does not produce values for the clustering coefficient of a node which have any practical meaning. Four current definitions for the clustering coefficient of WUNs are described and compared in Saramäki et al. (2007). The definition of Barrat et al. (2004) for the weighted clustering coefficient for node i, when applied to complete WUNs (where the node degree $d_i = N - 1$ and $a_{ij}a_{jk}a_{ik} = 1$ for all i, j, k), always equals $1/(N-2)$

regardless of the weights:

$$C_i^B = \frac{1}{s_i(d_i-1)} \sum_{j,k \neq i} \frac{w_{ij}+w_{ik}}{2} a_{ij} a_{jk} a_{ik} = \frac{\sum_{j \neq i} w_{ij} + \sum_{k \neq i} w_{ik}}{2 s_i (N-2)} = \frac{2 s_i}{2 s_i (N-2)} = \frac{1}{N-2}, \tag{1}$$

where the strength of node i is $s_i = \sum_{j \neq i} w_{ij}$. Note that C_i^B does not involve the weights w_{jk} of the connections among neighbors. The definition given by Onnela et al. (2005) does consider these weights, but the normalization factor preceding the sum does not involve the node strength:

$$C_i^O = \frac{1}{d_i(d_i-1)} \sum_{j \neq i} \sum_{k \neq i,j} (w_{ij} w_{ik} w_{jk})^{1/3}. \tag{2}$$

The definition of Zhang & Horvath (2005) is based entirely on the weights of all triangle edges:

$$C_i^Z = \frac{\sum_{j,k \neq i} w_{ij} w_{jk} w_{ik}}{\sum_{j \neq i} \sum_{k \neq i,j} w_{ij} w_{ki}}. \tag{3}$$

This definition can assign a high clustering coefficient to a node all of whose connections to other nodes have very low weight, a result which is counter-intuitive. The definition of Holme et al. (2007) is very similar to that of Zhang & Horvath (2005) and thus will not be included here. As is demonstrated below, each of these definitions fails to provide an intuitively satisfactory value for the clustering coefficient in the complete WUN context.

To define the clustering coefficient for complete WUNs, it is essential to first decide what an intuitively satisfactory value in this setting should be. Typically, all edge weights are normalized by dividing by some constant, e.g., by the maximum (possible) weight. Hence we assume that each edge weight $w_{ij} \in [0, 1]$ for all i, j. The nearer w_{ij} is to one, the more nodes i and j will be considered "strong" neighbors, while the nearer w_{ij} is to zero the more these nodes will be regarded "weak" neighbors. Those pairs with edge weights in the middle range will be regarded as "moderate" neighbors. It is proposed that the clustering coefficient for any individual node within a complete WUN should have the following characteristics:

C1. The clustering coefficient for node i should be large (close to one) if the set of strong neighbors of i are themselves strong neighbors of each other, and should become smaller as the proportion of its strong neighbors who are themselves weak neighbors increases.

C2. As the weights of the links involving the remaining neighbors of node i increase, the clustering coefficient for node i should also increase proportionately.

C3. The clustering coefficient for node i should be small if it has only weak neighbors, or at most one non-weak neighbor.

To achieve these goals, a new definition is proposed for computing the clustering coefficient of node i in a complete WUN, and then extended to complete WDNs. The idea is to capture the mean cluster prevalence of the network as the scale at which the network is viewed ranges from the zoom-in level (where all edges are visible) to the zoom-out level (where only the strongest edges are visible). First consider the complete WUN case:

1. For $t \in [0,1]$, set $A_t = [\mathbb{1}\{w_{ij} \geqslant t\}]$. This is the adjacency matrix corresponding to the network $\mathcal{N}_t$ formed when an edge is assigned between every pair of nodes having a weight at or above the threshold t. Denote the element in row i and column j of A_t by a_{ij}^t.
2. Let $\gamma_i(t)$ denote the number of triangles formed by consecutive edges with node i at one vertex and any two neighbors of node i as the other two vertices, and let $\Gamma_i(t)$ denote the number of triangles that would be formed with node i at one vertex if *every* pair of neighbors of node i were also neighbors of each other, i.e.,

$$\gamma_i(t) = \sum_{j \neq i} \sum_{k \neq i,j} a_{ij}^t a_{jk}^t a_{ik}^t = [A_t^3]_{ii} \quad \text{and} \quad \Gamma_i(t) = \sum_{j \neq i} \sum_{k \neq i,j} a_{ij}^t a_{ik}^t = [A_t O A_t]_{ii}. \quad (4)$$

   Here $O = \mathbf{1} \cdot \mathbf{1}' - I$, that is, a matrix consisting of zeros on the diagonal and ones in all other positions, I is the $N \times N$ identity matrix, $\mathbf{1}$ is a vector of length N consisting of ones in every position, and $\mathbf{1}'$ is its transpose. The clustering coefficient $C_i(t)$ for node i corresponding to $\mathcal{N}_t$ is then defined as the ratio of these two quantities (which is the established clustering coefficient for a node in a BUN), i.e.,

$$C_i(t) = \frac{\gamma_i(t)}{\Gamma_i(t)} = \frac{[A_t^3]_{ii}}{[A_t O A_t]_{ii}}, \quad (5)$$

   provided $[A_t O A_t]_{ii} \neq 0$. Otherwise set $C_i(t) = 0$.
3. The clustering coefficient C_i for node i corresponding to the complete WUN is then the average of $C_i(t)$ overall t in $[0,1]$:

$$C_i = \int_0^1 C_i(t)\,\mathrm{d}t. \quad (6)$$

   Since $C_i(t)$ is in practice a step function which changes value at the finitely many points at which t equals one of the edge weights in $\mathcal{N}$, the integral decomposes into a finite sum.
4. The clustering coefficient C for network $\mathcal{N}$ is, as usual, the average clustering coefficient over all nodes: $C = N^{-1} \sum_{i=1}^{N} C_i$.

This definition can then be extended to the setting in which the edges are directed. Consider a complete WDN $\mathcal{N}$ in which w_{ij} and w_{ji} are not necessarily equal. Again assume $0 < w_{ij} \leqslant 1$ for all i, j. The desired characteristics C1–C3 for the clustering coefficient in a complete WUN still apply in the complete WDN context, but with the understanding that a neighboring node may be a strong neighbor with respect to one direction, but a weak neighbor with respect to the other. In this scenario, there are eight directed triangles corresponding to each triplet of nodes, based on the eight different combinations of orientations of the three directed edges comprising a triangle (as described in Fagiolo (2007)). Among these eight directed triangles, two are *cyclic* triangles, i.e., triangles in which all three edges have the same cyclic orientation. The clustering coefficient may be defined in terms of all eight triangles, or only in terms of the two cyclic triangles. Based on the characteristics C1–C3, the clustering coefficient at node i should be close to one if, in every directed triangle (or cyclic triangle) for which the two edges involving node i have large weight, the remaining edge also has large weight, and should be lower if most of these remaining edges have small weight. Moreover, nodes which have no more than one directed edge of high weight associated with them should have a low clustering coefficient.

Fagiolo (2007) defines the clustering coefficient for a WDN as

$$C_i^F = \frac{[W^{[1/3]} + (W')^{[1/3]}]^3_{ii}}{2[d_i^{tot}(d_i^{tot} - 1) - 2d_i^{\leftrightarrow}]}, \tag{7}$$

where $W^{[1/3]} = [w_{ij}^{1/3}]$, d_i^{tot} is the total degree and $d_i^{\leftrightarrow}$ is the number of bilateral edges between node i and its neighbors. In a complete WDN, $d_i^{tot} = 2(N-1)$ and $d_i^{\leftrightarrow} = N - 1$. When only cyclic triangles are considered, Fagiolo (2007) defines the clustering coefficient at node i as

$$C_i^{Fc} = \frac{(W^{[1/3]})^3_{ii}}{d_i^{in} d_i^{out} - d_i^{\leftrightarrow}}, \tag{8}$$

where the in- and out-degrees d_i^{in} and d_i^{out} both equal $N-1$ in a complete WDN. In both definitions, the normalization factor does not involve the actual strength of a node, but only its maximum possible strength if all weights equal one, that is, its degree. In a complete WDN, this results in deflated clustering coefficients which cannot fulfill characteristics C1–C3. As the size of the network increases, this deflation can become very serious.

In contrast, the proposed definition of the clustering coefficient for a complete WDN does fulfill these characteristics, independent of network size. As with the proposed complete WUN definition, the matrix A_t corresponding to network $\mathscr{N}_t$ for threshold t in $[0, 1]$ is derived from the asymmetric weight matrix W. The clustering coefficient for node i in $\mathscr{N}_t$ in this setting essentially involves replacing A_t with $(A_t + A_t')/2$ in Equation (5), where A_t' is the transpose of A_t. Dividing by two ensures that the WDN definition will reduce to the WUN version definition (5) when $A_t = A_t'$. Let

$$\gamma_i^D(t) = \sum_{j \neq i} \sum_{k \neq i,j} \left(\frac{a_{ij}^t + a_{ji}^t}{2} \right) \left(\frac{a_{jk}^t + a_{kj}^t}{2} \right) \left(\frac{a_{ik}^t + a_{ki}^t}{2} \right) = \frac{[(A_t + A_t')^3]_{ii}}{8}$$

and

$$\Gamma_i^D(t) = \sum_{j \neq i} \sum_{k \neq i,j} \left(\frac{a_{ij}^t + a_{ji}^t}{2} \right) \left(\frac{a_{ik}^t + a_{ki}^t}{2} \right) = \frac{[(A_t + A_t')O(A_t + A_t')]_{ii}}{4},$$

so that the clustering coefficient $C_i^D(t)$ for node i in $\mathscr{N}_t$ is:

$$C_i^D(t) = \frac{\gamma_i^D(t)}{\Gamma_i^D(t)} = \frac{[(A_t + A_t')^3]_{ii}}{2[(A_t + A_t')O(A_t + A_t')]_{ii}}, \tag{9}$$

provided the denominator is positive (and equals zero otherwise). Hence, the clustering coefficient for node i in $\mathscr{N}_t$ is the ratio of the number of directed triangles involving node i to the maximum number of possible directed triangles involving node i. Then, analogous to its WUN counterpart, the clustering coefficient for node i in the complete WDN $\mathscr{N}$ is $C_i^D = \int_0^1 C_i^D(t)\, \mathrm{d}t$, and the mean clustering coefficient for $\mathscr{N}$ is $C^D = N^{-1} \sum_{i=1}^N C_i^D$.

If, however, one prefers to base the clustering coefficient only on the two cyclic triangles having node i as a vertex from among the eight triangles, the clustering

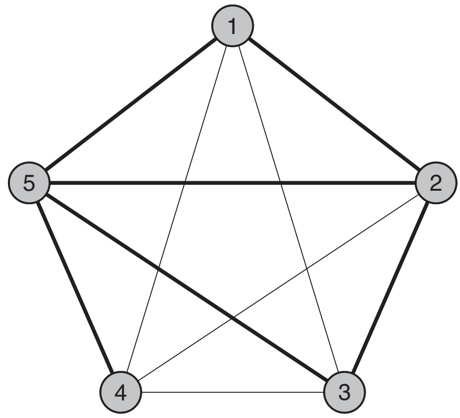

Fig. 1. A complete weighted undirected network consisting of five nodes, with line thickness corresponding to edge weights (thick edge = 0.9, thin edge = 0.1).

coefficient in a complete WDN based on cycles is defined as

$$C_i^c(t) = \frac{\sum_{j\neq i}\sum_{k\neq i,j} a_{ij}^t a_{jk}^t a_{ki}^t}{\sum_{j\neq i}\sum_{k\neq i,j} a_{ij}^t a_{ki}^t} = \frac{[A_t^3]_{ii}}{[A_t O A_t]_{ii}}. \tag{10}$$

Then the clustering coefficient for node i in the WDN $\mathscr{N}$ based on cyclic triangles is $C_i^c = \int_0^1 C_i^c(t)\ \mathrm{d}t$, and the corresponding mean clustering coefficient is $C^c = N^{-1}\sum_{i=1}^N C_i^c$.

To illustrate the proposed definitions, first consider the WUN depicted in Figure 1. The thick edges represent edges whose weights equal 0.9, while the thin edges represent edges whose weights are 0.1. The value of $C_i(t)$ for each of the five nodes as t ranges from 0 to 1 is shown in Figure 2. As can be observed, $C_i(t)$ is a step function. However, this function is not necessarily monotonic: Figure 3 shows the non-monotonic plot of $C_i(t)$ for a single node in a WUN consisting of 100 nodes with edge weights drawn from a uniform distribution on the interval $(0, 1)$. In each of these examples, the clustering coefficient for the corresponding node is simply the area beneath the curve.

Under the proposed definition, the clustering coefficient C_i for any node is a continuous function of the N edge weights in the network, as is the mean clustering coefficient C. To illustrate, suppose the strong edges of the WUN shown in Figure 1 each have value x in $[0.5, 1]$ while the weak edges each have value $1 - x$. As shown in Figure 4, as x varies from 0.5 to 1, the clustering coefficients for all five nodes, and the mean clustering coefficient, vary continuously from a common value of 0.5 to their respective values in the BUN obtained when $x = 1$. Using the same illustration, the mean clustering coefficient using the proposed definition and the three definitions of Onnela et al., Zhang et al., and Barrat et al. as x varies from 0.5 to 1 are compared in Figure 5. As x approaches one it is clear that the different definitions produce very different values.

Consider again the complete WUN shown in Figure 1. Note that node 1 has only two strong neighbors (nodes 2 and 5). These two strong neighbors are themselves strong neighbors, so the clustering coefficient C_1 for node 1 should be close to one, with some room for improvement if the strong edge weights increase above 0.9. The

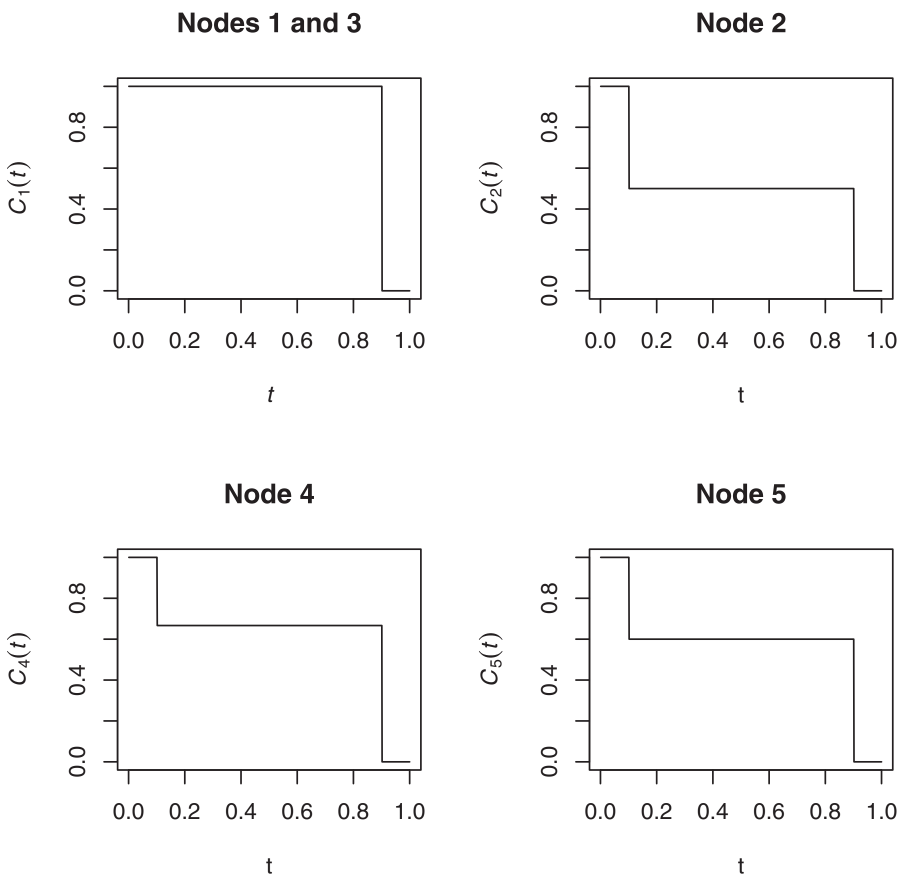

Fig. 2. Evolution of $C_i(t)$ for the five nodes in the WUN depicted in Figure 1 as t ranges from 0 to 1.

two strong neighbors are themselves strong neighbors of the two weak neighbors (nodes 3 and 4), but these weak neighbors are weak neighbors of each other, which diminishes their impact on the value of C_1. Based on definition (6), $C_1 = 0.901$, a value that makes sense in this context. However, using the weighted clustering coefficient definition given in definition (2), C_1^O is only 0.418, due to the greater influence of the two weak edges, while that given in definition (3) is a more sensible $C_1^Z = 0.832$. These values are shown in Table 1. Node 3 has the same configuration as node 1, and hence the same clustering coefficient regardless of the definition selected.

Next, note that node 2 has three strong neighbors (nodes 1, 3, and 5). Among these strong neighbors, there are three edges, two of which are strong. Hence C_2 should be about 2/3, and using definition (6) the value of C_2 is indeed 0.634, as shown in Table 1. But using the definition (2), C_2^O is only 0.514. Meanwhile, node 5 has four strong neighbors, among which there are two strong edges out of the six present. Hence C_5 should be about 1/3, and indeed using definition (6) one obtains $C_5 = 0.368$. But using definition (2) one computes a much higher value of $C_5^O = 0.588$, the largest clustering coefficient among the five nodes based on this definition. Lastly, node 4 has only one strong neighbor, so it cannot participate in

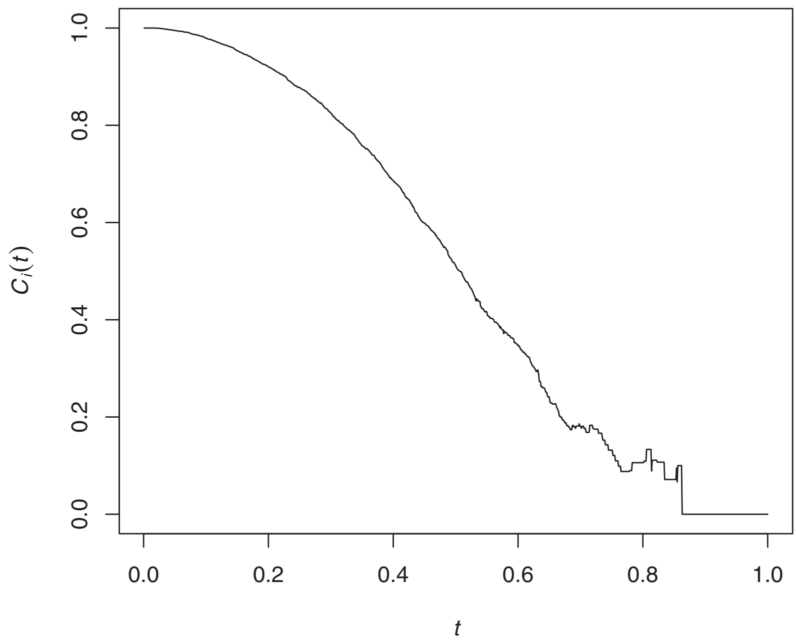

Fig. 3. Evolution of $C_i(t)$ as t ranges from 0 to 1 for one node in a WUN consisting of 100 nodes with edge weights drawn from a uniform distribution on $(0, 1)$.

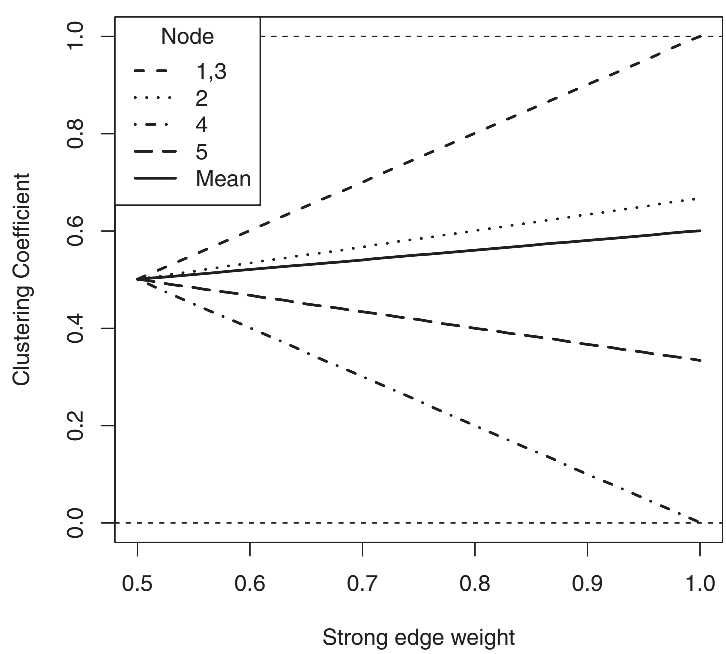

Fig. 4. Continuity of the nodewise and mean clustering coefficients as the common weight of the strong edges in the WUN shown in Figure 1 varies from 0.5 to 1 while the common weight of the weak edges varies from 0.5 to 0.

Table 1. *Clustering coefficients for each of the five nodes in the complete WUN depicted in Figure 1, and for the network itself, based on the proposed definition and two definitions in the current literature.*

Node	Proposed	Onnela, et al.	Zhang, et al.
1	0.901	0.418	0.832
2	0.634	0.514	0.607
3	0.901	0.418	0.832
4	0.101	0.302	0.873
5	0.368	0.588	0.367
Mean	0.581	0.448	0.702

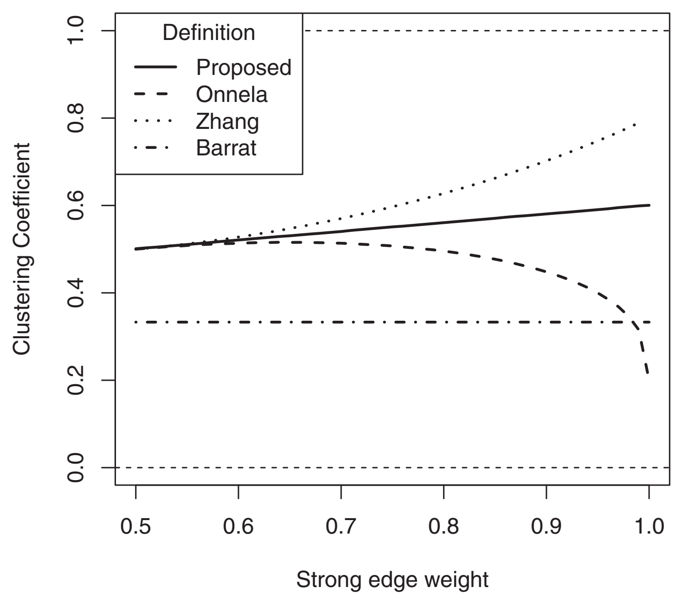

Fig. 5. The mean clustering coefficients based on four different definitions as the common weight of the strong edges in the WUN shown in Figure 1 varies from 0.5 to 1 while the common weight of the weak edges varies from 0.5 to 0.

any strong triangles and thus C_4 should be quite small. Indeed, $C_4 = 0.101$ based on definition (6), while C_4^O is 0.302 using definition (2). Note also that $C_4^Z = 0.873$ using definition (3), which is the one significant difference between that definition and the one proposed here, and is quite the opposite of the desired value expressed in characteristic C3. This is because the normalization factor in definition (3) is based only on the weights of the edges between a node and its neighbors, so that the magnitude of the clustering coefficient does not depend on these weights, but only on the weights of the edges between the neighbors.

Consequently, the proposed definition for the clustering coefficient of individual nodes in a complete WUN corresponds quite well with the desired characteristics C1–C3 identified as desirable for the complete WUN context, while alternative definitions do not.

Moving to the complete WDN context, consider now the complete WDN consisting of five nodes shown in Figure 6. In this figure, a thick arrow indicates a weight

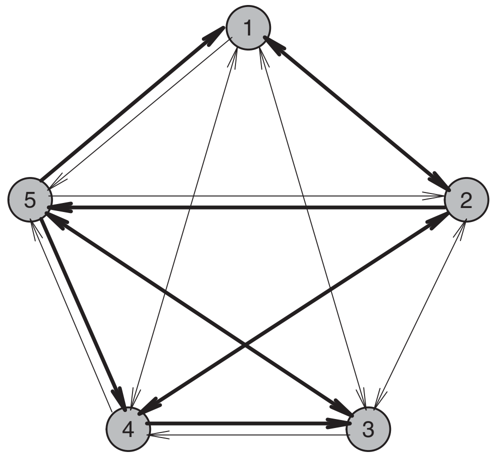

Fig. 6. A complete weighted directed network consisting of five nodes, with arrow thickness corresponding to edge weights (thick edge = 0.9, thin edge = 0.1) and arrow direction corresponding to edge direction. A double-arrow signifies the same weight in each direction.

of 0.9 for that directed edge, and a thin arrow indicates a weight of 0.1. A double-arrow signifies the same weight in both directions. Table 3 provides the clustering coefficient for each node based on all triangles or on cyclic triangles, using both the proposed definitions (9) and (10) and those of Fagiolo (7) and (8). Note that node 1 has two strong neighbors, nodes 2 and 5, but node 5 is only strong with respect to one direction. Among the eight directed triangles involving these three nodes, four of them involve the three strong edges beginning or ending at node 1. Out of these four, two also involve the strong edge from node 2 to node 5. This is shown in Table 2. Hence the clustering coefficient C_1^D for node 1 should be about 0.5. Using the proposed definition (9), $C_1^D = 0.501$. Meanwhile, definition (7) gives $C_1^F = 0.086$, a considerably lower value. If only the two cyclic paths are to be considered (the first and last rows of Table 2), note that nodes 2 and 5 are strong neighbors of node 1 for only one of these, and that nodes 2 and 5 are themselves strong neighbors in this triangle (first row of the table). Then the clustering coefficient based on cyclic paths should be close to one, and indeed the proposed definition (10) yields $C_1^c = 0.901$. This value would approach one as the edge weights are increased. Meanwhile, the corresponding definition (8) produces a much lower value of $C_1^{Fc} = 0.313$. Observe that node 3 has the same topology as node 1, and thus the same clustering coefficient under either definition.

Now consider node 5 in Figure 6. This node has four strong neighbors, although for three of them the strength lies in only one direction. Among the eight triangles involving nodes 5, 1 and 2, two of them include the strong directed edges $5 \to 1$ and $2 \to 5$, and both of these triangles include a strong directed edge between nodes 1 and 2. Continuing in this manner through all neighbor pairs, one can identify 18 triangles in which node 5 has a strong directed edge with both neighbors in the pair, and among these 18 triangles there are six for which the directed edge between the neighbors is also strong. Hence C_5^D should be about $1/3$ according to the proposed definition, and indeed $C_5^D = 0.368$, as shown in Table 3. Meanwhile, definition (7) gives a much lower value of $C_5^F = 0.118$. Also, 5 of the 18 triangles are cyclic

Table 2. *Summary of the eight directed triangles in Figure 6 involving nodes 1, 2 and 5, indicating for which triangles nodes 2 and 5 are both strong neighbors of node 1, and among these, for which triangles nodes 2 and 5 are also strong neighbors, thereby forming a strong triangle.*

Directed triangle	Strong neighbors?	Strong triangle?
$1 \to 2 \quad 2 \to 5 \quad 5 \to 1$	Yes	Yes
$1 \to 2 \quad 2 \to 5 \quad 1 \to 5$		
$1 \to 2 \quad 5 \to 2 \quad 5 \to 1$	Yes	
$1 \to 2 \quad 5 \to 2 \quad 1 \to 5$		
$2 \to 1 \quad 2 \to 5 \quad 5 \to 1$	Yes	Yes
$2 \to 1 \quad 2 \to 5 \quad 1 \to 5$		
$2 \to 1 \quad 5 \to 2 \quad 5 \to 1$	Yes	
$2 \to 1 \quad 5 \to 2 \quad 1 \to 5$		

Table 3. *Clustering coefficients for each of the five nodes in the complete WDN depicted in Figure 6, and for the network itself, when all eight triangles among each node triplet are considered, and when only the two cyclic triangles are considered, based on the proposed definitions and the definitions given in Fagiolo (2007).*

	All triangles		Cyclic triangles	
Node	Proposed	Fagiolo	Proposed	Fagiolo
1	0.501	0.086	0.901	0.313
2	0.301	0.107	0.501	0.398
3	0.501	0.086	0.901	0.313
4	0.421	0.103	0.634	0.371
5	0.368	0.118	0.581	0.418
Mean	0.418	0.100	0.704	0.363

paths, and three of these consist of three strong edges, so the clustering coefficient based only on cyclic paths should be about 3/5. Using definition (10) for cyclic paths, $C_5^c = 0.581$, in agreement with this reasoning. But using definition (8) one has $C_5^{Fc} = 0.418$, which is somewhat lower.

Following the same procedure, one can estimate values favorable to characteristics C1–C3 for the clustering coefficient of the remaining two nodes in Figure 6, and apply the proposed definition to obtain values which match these estimates closely, whether all eight triangles per node triplet are considered or only the two cyclic triangles per triplet are considered (see Table 3). In each case, applying the definitions (7) and (8) of Fagiolo to complete WDNs, leads to clustering coefficients which are considerably smaller and which do not satisfy characteristics C1–C3. This is mostly due to the inflated normalization factor, as discussed above.

The definition proposed here provides intuitively sensible values for the clustering coefficient of a node in a complete weighted network, both when the edges are undirected and when they are directed. Moreover, this definition also produces reasonable values when a network is not complete, and reduces to the usual definitions of the clustering coefficient in BUNs and BDNs that are not complete. Consequently, it could in fact be considered a global definition of the clustering coefficient for networks in general. Furthermore, the methodology employed here to construct a valid and useful clustering coefficient for complete weighted networks can be generalized easily to the construction of other common network metrics, such as the minimum path length or the small-world coefficient, for weighted networks in general. For each $t \in [0, 1]$, form a binary network and compute the metric of interest $\mu(t)$ for the binary network. Then integrate $\mu(t)$ over $[0, 1]$ to obtain the desired metric μ for the weighted network.

The given definition allows researchers who model phenomena using complete weighted networks to discuss the prevalence of tightly-clustered neighborhoods in these networks, which could not be done using methods currently found in the literature. For example, a potential field of application is the modeling of brain networks. Although complete WUNs are studied with increasing frequency in this field, in many studies the complete WUNs are converted into BUNs by using some arbitrary threshold on the edge weights, thereby sacrificing much useful information (see, e.g., the review studies in Tijms et al. (2013) and van den Heuvel & Fornito (2014)). The researchers then proceed to investigate network properties using the BUNs. But with the definition proposed here, the neuroscientists can conduct their investigation using the complete WUNs, and thereby retain the valuable information carried by the edge weights and arrive at conclusions that can be properly substantiated. As a result, this approach opens the door to liberate researchers to investigate network properties of weighted networks to the same extent that they investigate the same properties for binary networks, without having to wrest those weighted networks from their natural context.

References

Albert, R., & Barabási, A. (2003). Statistical mechanics of complex networks. *Reviews of Modern Physics*, **74**, 47–97.

Barrat, A., Barthélemy, M., Pastor-Satorras, R., & Vespignani, A. (2004). The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences*, **101**(11), 3747–3752.

Dorogovtsev, S., & Mendes, J. (2003). *Evolution of networks: From biological nets to the internet and WWW*. Oxford: Oxford University Press.

Fagiolo, G. (2007). Clustering in complex directed networks. *Physical Review E*, **76**, (026107), 1–8.

Holme, P., Park, S. M., Kim, B. J., & Edling, C. (2007). Korean university life in a network perspective: Dynamics of a large affiliation network. *Physica A*, **373**, 821–830.

Kalna, G., & Higham, D. J. (2006). Clustering coefficients for weighted networks. *University of Strathclyde Mathematics Research Report No. 3.*

Li, Y., Liu, Y., Li, J., Qin, W., Li, K., et al. (2009). Brain anatomical network and intelligence. *PLoS Computational Biology*, **5**(5), e1000395. doi: 10.1371/journal.pcbi.1000395.

McAssey, M. P., Bijma, F., Tarigan, B., van Pelt, J., van Ooyen, A., & de Gunst, M. (2013). A morpho-density approach to estimating neural connectivity. *PLoS ONE*, **9**(1), e86526. doi: 10.1371/journal.pone.0086526.

Newman, M. (2003). The structure and function of complex networks. *SIAM Review*, **45**, 167–256.

Onnela, J.-P., Saramäki, J., Kertész, J., & Kaski, K. (2005). Intensity and coherence of motifs in weighted complex networks. *Physical Review E*, **71**(065103), 1–4.

Saramäki, J., Kivelä, M., Onnela, J.-P., Kaski, K., & Kertész, J. (2007). Generalizations of the clustering coefficient to weighted complex networks. *Physical Review E*, **75**(027105), 1–5.

Szabó, G., Alava, J., & Kertész, J. (2004). In Ben-Naim, et al. (Eds.), *Complex networks*. Springer Lecture Notes in Physics volume 650, (pp. 139–162). Berlin: Springer.

Tijms, B. M., Wink, A. M., de Haan, W., van der Flier, W. M., Stam, C. J., Scheltens, P., & Barkhof, F. (2013). Alzheimers disease: Connecting findings from graph theoretical studies of brain networks. *Neurobiology of Aging*, **34**, 2023–2036.

van den Heuvel, M. P., & Fornito, A. (2014). Brain networks in schizophrenia. *Neuropsychology Review*, **24**, 32–48.

Watts, D., & Strogatz, S. (1998). Collective dynamics of small-world networks. *Nature*, **393**, 440–442.

Zhang, B., & Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology* **4**, 17.