

Comparison of Different Generalizations of Clustering Coefficient and Local Efficiency for Weighted Undirected Graphs

Yu Wang

Eshwar Ghumare

eshwargoraknath.ghumare@kuleuven.be

Laboratory for Cognitive Neurology, Department of Neurosciences, KU Leuven, Leuven 3000, Belgium

Rik Vandenberghe

rik.vandenberghe@uzleuven.be

Laboratory for Cognitive Neurology, Department of Neurosciences, KU Leuven, Leuven 3000, Belgium, and Alzheimer Research Centre, KU Leuven, Leuven Institute for Neuroscience and Disease, KU Leuven, Leuven 3000, Belgium

Patrick Dupont

patrick.dupont@kuleuven.be

Laboratory for Cognitive Neurology, Department of Neurosciences, KU Leuven, Leuven 3000, Belgium; Alzheimer Research Centre, KU Leuven, Leuven Institute for Neuroscience and Disease, KU Leuven, Leuven 3000, Belgium; and Medical Imaging Research Center, KU Leuven and University Hospitals Leuven, Leuven 3000, Belgium

Binary undirected graphs are well established, but when these graphs are constructed, often a threshold is applied to a parameter describing the connection between two nodes. Therefore, the use of weighted graphs is more appropriate. In this work, we focus on weighted undirected graphs. This implies that we have to incorporate edge weights in the graph measures, which require generalizations of common graph metrics. After reviewing existing generalizations of the clustering coefficient and the local efficiency, we proposed new generalizations for these graph measures. To be able to compare different generalizations, a number of essential and useful properties were defined that ideally should be satisfied. We applied the generalizations to two real-world networks of different sizes. As a result, we found that not all existing generalizations satisfy all essential properties. Furthermore, we determined the best generalization for the clustering coefficient and local efficiency based on their properties and the performance when applied to two networks. We found that the best generalization of the clustering coefficient is $C^{M,lm}$, defined in Miyajima and Sakuragawa (2014), while the best generalization of the local

efficiency is E_{loc}^P proposed in this letter. Depending on the application and the relative importance of sensitivity and robustness to noise, other generalizations may be selected on the basis of the properties investigated in this letter.

1 Introduction

A complex system can be modeled as a graph or network, which is composed of nodes and edges connecting them. Analysis over a wide range of complex systems has led to a fundamental insight: many complex systems often share certain topological characteristics, and these can be captured by graph-theoretical metrics (Barabasi & Oltvai, 2004; Amaral & Ottino, 2004; Zhang & Horvath, 2005; Bullmore & Sporns, 2009; Fornito, Zalesky, & Breakspear, 2013). The small-world topology for example, has been found in many real-world networks (He, Chen, & Evans, 2007; Opsahl & Panzarasa, 2009; Batalle et al., 2012; Vandenberghe et al., 2013), and is an indication of the cost-efficiency of these networks.

While traditional graph analysis uses binary edges to enhance contrast between strong and weak connections, there is an increasing demand for using edge weight, which entails potentially important information. Incorporating edge weights in the graph analysis calls for generalizations of the graph metrics. While some of these measures can be naturally generalized to a weighted version (e.g., node degree to node strength), others cannot be generalized in a straightforward way. The generalization of clustering coefficient and local efficiency, used to quantify the small-world topology (Watts & Strogatz, 1998; Achard and Bullmore, 2007; Batalle et al., 2012), is far from trivial.

The clustering coefficient reflects the tendency that neighbors of a node are also neighbors to each other (Rubinov & Sporns, 2010). The clustering coefficient is high in small-world networks compared to random networks (Watts & Strogatz, 1998). Local efficiency is a measure for the fault tolerance of the system: it measures how efficient the communication is between neighbors of a node when that node is removed (Latora & Marchiori, 2003). A small-world network features a local efficiency intermediate to that of regular (lattice) and random network (Achard & Bullmore, 2007; Batalle et al., 2012). The two measures are related in a way that the clustering coefficient in an undirected network is found to be a reasonable approximation of local efficiency (Latora & Marchiori, 2003).

Although it is straightforward to find the neighbors of a node, the question of how to define their weighted surrogates is far from obvious. Several generalizations have been proposed (Barrat, Barthelemy, Pastor-Satorras, & Vespignani, 2004; Onnela, Saramäki, Kertész, & Kaski, 2005; Zhang & Horvath, 2005; Saramäki, Kivelä, Onnela, Kaski, & Kertesz, 2007; Opsahl & Panzarasa, 2009; Miyajima & Sakuragawa, 2014; Rubinov & Sporns, 2010).

Different definitions capture slightly different aspects of the network, yet some of the generalizations are not designed for fully weighted networks (Rubinov & Sporns, 2010; Barrat et al., 2004; Onnela et al., 2005). These generalizations require the removal of the weak or noisy connections beforehand. A preferable solution is to adapt the equations in such a way that they can be used for a fully weighted network (Zhang & Horvath, 2005; Saramäki et al., 2007; Opsahl & Panzarasa, 2009; Miyajima & Sakuragawa, 2014).

In this letter, we first define a number of essential and useful properties that ideally should be satisfied when using a generalized graph measure and explain how we will evaluate them. Then we review the existing generalizations, and we propose new generalizations for the local efficiency for fully weighted undirected networks with no self-connections. Finally, we make a thorough comparison of the different generalizations and apply them to two real-world networks.

2 Methods

Assume an undirected weighted network with N nodes and an $N \times N$ adjacency matrix \mathbf{A} for which the (i, j) th element a_{ij} is 1 if an edge between i and j exists and 0 otherwise. In this work, we assume that no self-connections are present: $a_{ii} = 0$. For a binary undirected network, the clustering coefficient for node i is given by

$$C(i) = \frac{1}{k_i(k_i - 1)} \sum_{j,h} a_{ij}a_{ih}a_{jh}, \quad (2.1)$$

with k_i the degree of node i . The node degree k_i is defined as the number of nodes connected to node i . The local efficiency is defined as

$$E_{loc}(i) = \frac{1}{k_i(k_i - 1)} \sum_{j,h; j \neq h} a_{ij}a_{ih}[d_{jh}(N_i)]^{-1}, \quad (2.2)$$

in which N_i is the subgraph consisting of the neighbors of i excluding node i itself, and $d_{jh}(N_i)$ is the length of the shortest path between nodes j and h containing only neighbors of i . If no path containing these neighbors is found, $d_{jh}(N_i) = \infty$.

In a weighted network, we define the weight matrix \mathbf{W} in which each element represents the weight w_{ij} between node i and node j . All weights are assumed to be positive. If no connection is present, the weight is 0. In this work, we focus on networks without self-connections: $\forall i, w_{ii} = 0$. The node degree k_i is calculated based on the presence of connections with nonzero weights irrespective of the amplitude. In case of a fully connected

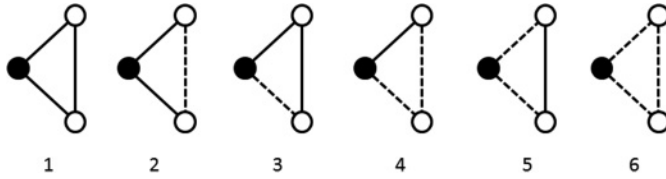


Figure 1: Six possible triangle configurations for node i filled in black. Normal and dashed lines represent strong and weak edges, respectively.

weighted network with N nodes and no self-connections, $k_i = N - 1$. The node strength, which takes into account the weight of the connection, is defined by

$$s_i = \sum_j w_{ij}, \quad (2.3)$$

Note that for a binary network in which $w_{ij} = 1$ if a connection is present between nodes i and j and 0 otherwise, $s_i = k_i$.

2.1 Properties of Generalized Graph Measures. A generalization $G_w(\mathbf{W})$ of a binary graph measure $G_b(\mathbf{A})$ should ideally satisfy some properties. Some of these properties are essential, while others may depend on the application. The essential properties are:

- *General versatility.* If the input is given as a binary network, the output of the generalizations should give the same results as the binary version: $G_w = G_b$ when $\mathbf{W} = \mathbf{A}$ (Miyajima & Sakuragawa, 2014).
- *Continuity.* The graph measure should be continuous, that is, $\lim_{\epsilon \rightarrow 0} G_w(\mathbf{W}_\epsilon) = G_w(\mathbf{W})$ in which $\mathbf{W}_\epsilon = \mathbf{W}$ except for one connection weight, which differs by ϵ (Miyajima & Sakuragawa, 2014).
- *Sensitivity.* The graph measure should be able to make a distinction between different cases for which the graph measure is designed. Because the clustering coefficient and the local efficiency are defined using triangles, we will evaluate the six possible cases in which the weight is either low or high in one of the connections of the triangle (see Figure 1).
- *Robustness to noise.* The graph measure $G_w(\mathbf{W})(i)$ for node i should be robust when adding noise to the weights of the connections. If we assume a noise matrix v , which represents additive noise on each connection weight $w_{ij}^v = w_{ij} + v_{ij}$, and define the mean relative error (in percent) with respect to the noise-free measures as

$$\Delta(\nu) = \frac{100}{N} \sum_{i=1}^N \left| \frac{G_w(\mathbf{W} + \nu)(i) - G_w(\mathbf{W})(i)}{G_w(\mathbf{W})(i)} \right|, \quad (2.4)$$

then a small value of $\|\nu\|$ should lead to a small error value $\Delta(\nu)$.

Additional useful properties can be defined, but these properties depend on the application under investigation and should not be considered essential:

- *Weight-scale invariance.* The graph measure is invariant to a global scale factor $\lambda > 0$ for all edges: $G_w(\lambda\mathbf{W}) = G_w(\mathbf{W})$ (Miyajima & Sakuragawa, 2014).
- *Applicable to fully weighted networks.* In some applications, it is beneficial to avoid any thresholding of the weights of the connections in a network. As a result, every node is connected with every other node, although the weights may be very small. A generalization of the clustering coefficient and the local efficiency should be applicable to such cases.

2.2 Generalizations. A good review of existing generalizations of the clustering coefficient and local efficiency can be found in Miyajima and Sakuragawa (2014). The authors investigated general versatility, weight-scale invariance, and continuity. Here, we investigate all criteria listed in the previous section. We first give a short description of the different generalizations currently available.

Barrat et al. (2004) proposed a generalization by including the node strength s_i and the weights of the connections from node i :

$$C^B(i) = \frac{1}{s_i(k_i - 1)} \sum_{j,h} \frac{w_{ij} + w_{ih}}{2} a_{ij} a_{ih} a_{jh}. \quad (2.5)$$

If the network is fully weighted, $C^B(i) = 1$ for all nodes i .

Onnela et al. (2005) proposed a weighted clustering coefficient for node i based on the weighted triangles, and this algorithm is used in the brain connectivity toolbox (Rubinov & Sporns, 2010):

$$C^O(i) = \frac{1}{k_i(k_i - 1)} \sum_{j,h} \frac{(w_{ij} w_{ih} w_{jh})^{1/3}}{\max(W)}. \quad (2.6)$$

Zhang and Horvath (2005) have defined, in their gene co-expression network analysis, a clustering coefficient for node i for a fully weighted network. The original equation can be rewritten as (Saramäki et al., 2007):

$$C^Z(i) = \frac{1}{\max(W)} \frac{\sum_{j,h} w_{ij} w_{ih} w_{jh}}{\sum_{j,h; j \neq h} w_{ij} w_{ih}}. \quad (2.7)$$

Holme, Park, Kim, and Edling (2007) have defined a similar definition as the previous one except that in the denominator, $j \neq h$ is not required:

$$C^H(i) = \frac{1}{\max(W)} \frac{\sum_{j,h} w_{ij} w_{ih} w_{jh}}{\sum_{j,h} w_{ij} w_{ih}}. \quad (2.8)$$

Opsahl and Panzarasa (2009) proposed a generalization as

$$C^{Op}(i) = \frac{\sum_{j,h} f(w_{ij}, w_{ih}) a_{ij} a_{ih} a_{jh}}{\sum_{j,h; j \neq h} f(w_{ij}, w_{ih}) a_{ij} a_{ih}}. \quad (2.9)$$

They proposed four methods to calculate $f(w_{ij}, w_{ih})$: arithmetic mean ($C^{Op, \text{mean}}$), geometric mean ($C^{Op, \text{gm}}$), and maximum ($C^{Op, \text{max}}$) and minimum ($C^{Op, \text{min}}$) of the weights of the edges. We refer to the set of these four methods by C^{Op} .

Miyajima and Sakuragawa (2014) extended the generalization of the cluster coefficient of node i to the case of weighted directed networks using different functions (multiplication, geometric mean, minimum, and harmonic mean). The case of multiplication in the context of an undirected network leads to the same generalization as the one from Holme et al. (2007). For undirected networks, we give their other extensions below:

Geometric mean:

$$C^{M, \text{gm}}(i) = \frac{\sum_{j,h} \sqrt{\sqrt{w_{ij} w_{ih} w_{jh}}}}{\sum_{j,h; j \neq h} \sqrt{\sqrt{w_{ij} w_{ih} \max(W)}}} \quad (2.10)$$

Minimum:

$$C^{M, \text{min}}(i) = \frac{\sum_{j,h} \min(w_{ij}, w_{ih}, w_{jh})}{\sum_{j,h; j \neq h} \min(w_{ij}, w_{ih})} \quad (2.11)$$

Harmonic mean:

$$C^{M, \text{hm}}(i) = \frac{\sum_{j,h; j \neq h} \frac{2}{\frac{1}{2} + \frac{1}{w_{jh}}} a_{ij} a_{ih}}{\sum_{j,h; j \neq h} \frac{2}{\frac{1}{2} + \frac{1}{\max(W)}} a_{ij} a_{ih}} \quad (2.12)$$

We refer to the set of these three methods by C^M .

A generalization of the local efficiency is given by Rubinov and Sporns (2010):

$$E_{loc}^R(i) = \frac{1}{k_i(k_i - 1)} \sum_{j,h; j \neq h} (w_{ij}w_{ih}[d_{jh}^w(N_i)]^{-1})^{1/3}, \tag{2.13}$$

in which N_i is the subgraph consisting of the neighbors of i excluding node i itself, and $d_{jh}^w(N_i)$ is the shortest distance between nodes j and h containing only neighbors of i . The distance between nodes j and h in a weighted network is defined as

$$d_{jh}^w = \sum_{\substack{k,l \\ a_{kl} \in \mathcal{G}_{j \leftrightarrow h}^W}} \frac{1}{w_{kl}}, \tag{2.14}$$

with $\mathcal{G}_{j \leftrightarrow h}^W$ the shortest weighted path between j and h . If no path containing these neighbors is found, $d_{jh}^w = \infty$.

As is the case for the clustering coefficient, multiple generalizations can be defined for the weighted local efficiency. We define three possible extensions that we compare with the current generalization E_{loc}^R defined by Rubinov and Sporns (2010).

The first alternative for the weighted local efficiency is given by

$$E_{loc}^W(i) = \frac{1}{\max(W)} \frac{\sum_{j,h; j \neq h} ((\min(w_{ij}, w_{ih}))^3 [d_{jh}^w(\tilde{N}_i)]^{-1})}{\sum_{j,h; j \neq h} (\min(w_{ij}, w_{ih}))^2}, \tag{2.15}$$

where $d_{jh}^w(\tilde{N}_i)$ is the adapted shortest distance between j and h . This adapted shortest distance was calculated as the shortest distance in the network \tilde{N}_i containing all neighbors of i excluding node i after replacing the weight of edge (j, h) to \tilde{w}_{jh} :

$$\tilde{w}_{jh} = \frac{w_{jh}w_{ij}w_{ih}}{(\max(W))^3}. \tag{2.16}$$

If no path containing nodes j and h is found in \tilde{N}_i , $d_{jh}^w(\tilde{N}_i) = \infty$.

Weighting by $\max(W)$ makes the shortest distance invariant to the weight scale.

The second alternative for the weighted local efficiency is given by

$$E_{loc}^P(i) = \frac{\sum_{j,h; j \neq h} w_{ij} \cdot w_{ih} [d_{jh}^w(\tilde{N}_i)]^{-1}}{\sum_{j,h; j \neq h} w_{ij} \cdot w_{ih}}, \tag{2.17}$$

and the third alternative, more closely related to the extension defined in equation 2.13 is

$$E_{loc}^Z(i) = \frac{1}{\max(W)^{1/3}} \frac{\sum_{j,h} w_{ij}^{1/3} w_{ih}^{1/3} [d_{jh}^w(N'_i)]^{-1}}{\sum_{j,h} w_{ij}^{1/3} w_{ih}^{1/3}}, \quad (2.18)$$

where $d_{jh}^w(N'_i)$ is the adapted shortest distance between j and h . This adapted shortest distance was calculated as the shortest distance in the network N'_i containing all neighbors of i excluding node i after replacing the weight of edge (j, h) to

$$w'_{jh} = w_{jh}^{1/3}. \quad (2.19)$$

If no path containing nodes j and h is found in N'_i , $d_{jh}^w(N'_i) = \infty$.

2.3 Application to Two Real-World Networks. A number of essential (general versatility, continuity) and useful (weight-scale invariance and applicability to fully weighted networks) properties are evaluated on a theoretical basis. The essential property sensitivity will be assessed by studying the cases shown in Figure 1, while the fourth essential property robustness to noise will be assessed using two real-world networks.

2.3.1 The Associative-Semantic Network. An empirical network was taken to test the properties of the generalization. The associative-semantic network is constructed from functional interactions between brain regions activated during the associative-semantic task (Vandenberghe et al., 2013; Wang et al., 2014) measured using fMRI. The network has 57 nodes, which are mostly located in the left hemisphere. The data are taken from one elderly subject who participated in an fMRI measurement on a 3T Philips system (Wang et al., 2014). The nodes were defined based on a previous experiment (Vandenberghe et al., 2013). Based on the average time series in these nodes, partial correlations were calculated. The weight of the edge between node i and j is calculated from the Fisher r-to-z transform z_{ij} of the partial correlation as

$$w_{ij} = (2\Phi(|z_{ij}|) - 1)^4, \quad (2.20)$$

where Φ is the cumulative distribution function of the standard normal distribution. The weights are a measure of the functional interaction between brain regions.

We will evaluate the robustness to noise of the different generalizations by studying the situation in which we add noise to the Z -value (before the transformation to weights) of each (nonzero) connection using a standard

normal distribution as noise model multiplied by some constant $f > 0$ to model the amount of noise. After adding the noise, weights are calculated as before using equation 2.20. We will study the case of $f = 0.001, 0.01,$ and 0.1 corresponding to different levels of noise ranging from weak to strong noise. The average error will be calculated over 10 noise realizations.

Since this is a fully weighted network, we will also calculate the correlation of the different versions of the clustering coefficient and local efficiency between this fully weighted network and the soft-thresholded network (thresholded such that the density is 80%, 60%, or 40%). Often a threshold is used to remove connections with low weight. This is referred to as the creation of soft-thresholded weighted networks. We hypothesize that weighted graph measures calculated for the original weighted network and for the soft-thresholded weighted network are highly correlated, especially when the density of the latter network is high.

2.3.2 The Resting State fMRI Network. The second network is a functional connectivity network constructed from correlations between regional fMRI times series, measured in 638 nodal locations from 27 healthy volunteers scanned in resting state on a Siemens 3T scanner. The details of this experiment can be found in Crossley et al. (2013). The data of this network are publicly available.¹ Unfortunately, the connection strengths (expressed as Z-scores) were already thresholded, and as a result, only the connections with $Z > 0.6$ were available. Therefore, we can only study the robustness to noise in this case. We will do this in the same way as the first real-world network, including the transformation of Z-values to weights using equation 2.20.

3 Results

3.1 General Versatility. All the generalizations of the clustering coefficient, except C^H , lead to the same equation in the case of binary undirected networks, as can be easily seen from equations 2.5 to 2.7 and 2.9 to 2.12, where $w_{ij} = a_{ij}, a_{ij} = 0$ or 1 and $s_i = k_i$. The generalization C^H does not show general versatility because in the denominator, there is no requirement that the two sides of a triangle should be different (i.e., when the triplet is not a triangle but a line).

The generalization of the local efficiency, introduced in equation 2.13, does not show general versatility because of the power $-\frac{1}{3}$ of the distance compared to -1 in the binary case. In contrast, the generalizations of the local efficiency proposed in this letter—equations 2.15, 2.17, and 2.18—do show general versatility.

¹<https://sites.google.com/site/bctnet/datasets> as GroupAverage_rsfMRI_matrix.mat.

Table 1: General Versatility and Continuity for Fully Weighted Networks of the Different Generalizations.

Method	General Versatility	Continuity
C^B	Yes	No
C^O	Yes	No
C^Z	Yes	Yes
C^H	No	Yes
$C^{Op,mean}$	Yes	No
$C^{Op,gm}$	Yes	No
$C^{Op,max}$	Yes	No
$C^{Op,min}$	Yes	No
$C^{M,gm}$	Yes	Yes
$C^{M,min}$	Yes	Yes
$C^{M,hm}$	Yes	Yes
E_{loc}^R	No	No
E_{loc}^W	Yes	Yes
E_{loc}^P	Yes	Yes
E_{loc}^Z	Yes	Yes

The results of each generalization for this criterion are summarized in Table 1.

3.2 Continuity. The node degree k_i is a discontinuous function for a weighted network since any node with a nonzero weight is considered a neighbor irrespective of the amplitude. This implies that the value for the node degree k_i will differ with 1 between the case in which an arbitrary small weight is present for an edge between nodes i and j and the case in which this edge is not present (i.e., has zero weight). As a result, C^B , C^O , and E_{loc}^R are not continuous. C^{Op} is also not continuous since in the numerator, only closed triangles will contribute, no matter how small the weight of the third connection in the triangle is and the contribution depends only on the weights of the two other connections in the triangle. The other extensions— C^Z , C^H , C^M , E_{loc}^W , E_{loc}^P , and E_{loc}^Z —are all continuous.

The results of each generalization for this criterion are summarized in Table 1.

3.3 Sensitivity. In Table 2, we show how far each generalization is able to detect differences between each of the six possible cases shown in Figure 1 for the clustering coefficient and the local efficiency. We have given a weak connection the value 0.1 and a strong connection the value 1 while assuming that $\max(W) = 1$.

Table 2: Sensitivity of the Clustering Coefficient and Local Efficiency for the Cases Presented in Figure 1.

Method	Number of Different		
	Values (max = 6)	Minimum	Maximum
C^B	1	1	1
C^O	4	0.1	1
C^Z	4	0.1	1
C^H	4	0.0165	0.5
$C^{Op,mean}$	1	1	1
$C^{Op,gm}$	1	1	1
$C^{Op,max}$	1	1	1
$C^{Op,min}$	1	1	1
$C^{M,gm}$	2	0.3162	1
$C^{M,min}$	2	0.1	1
$C^{M,hm}$	4	0.1818	1
E_{loc}^R	4	0.1	1
E_{loc}^W	5	0.0001	1
E_{loc}^P	5	0.001	1
E_{loc}^Z	2	0.4642	1

3.4 Robustness to Noise. When adding different amounts of gaussian noise to both real-world networks, we observe that the generalizations $C^{M,gm}$ and $C^{M,hm}$ of the clustering coefficient perform best for both networks across all levels of noise used in this study with a mean error within 5%. Most of the other generalizations also perform reasonably well, with mean errors within 10% (see Figure 2).

All generalizations of the local efficiency have an acceptable mean error (i.e., less than 5%) in both networks when noise is not too large (see Figure 3). However, when noise increases, only E_{loc}^Z and E_{loc}^R have mean errors below 10% in both networks.

3.5 Weight-scale Invariance and Applicability to Fully Weighted Networks. Weight-scale invariance is satisfied for all generalizations of the clustering coefficient as can be derived mathematically by multiplying every weight by a factor λ and observing that the result is independent of this factor. The generalizations of the local efficiency E_{loc}^W , E_{loc}^P and E_{loc}^Z also show weight-scale invariance, but this is not the case for E_{loc}^R .

From equation 2.5 for C^B and equation 2.9 for C^{Op} , we see that for fully weighted networks, the clustering coefficient equals 1 for all nodes, and therefore we consider these generalizations not suitable for fully weighted networks (see Table 3). All other generalizations for the clustering coefficient and the local efficiency can be used for fully weighted networks.

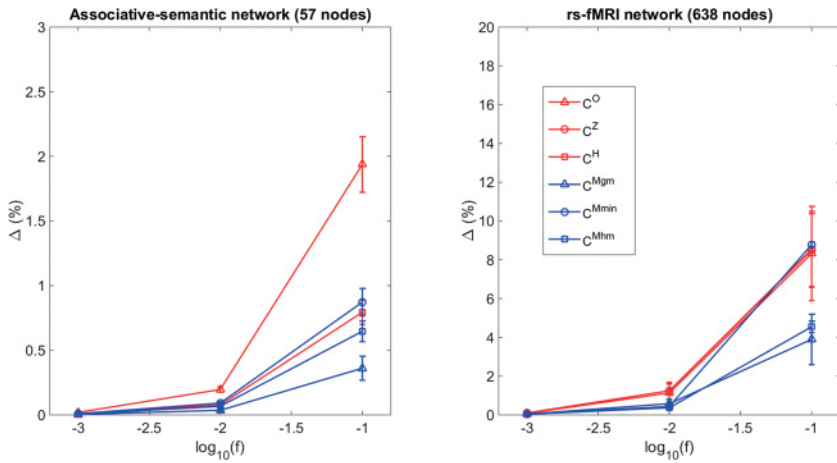


Figure 2: Mean relative error Δ of the robustness to noise of the clustering coefficient for different generalizations as a function of the amount of gaussian noise for the associative-semantic network (left) and the resting-state fMRI network (right) on the nonzero connections. The noise fraction f is defined as the factor that is multiplied with samples from the standard normal noise distribution. In the left panel, the lines of C^Z and C^H almost coincide, and this is also the case for C^O , C^Z , and C^H in the right panel.

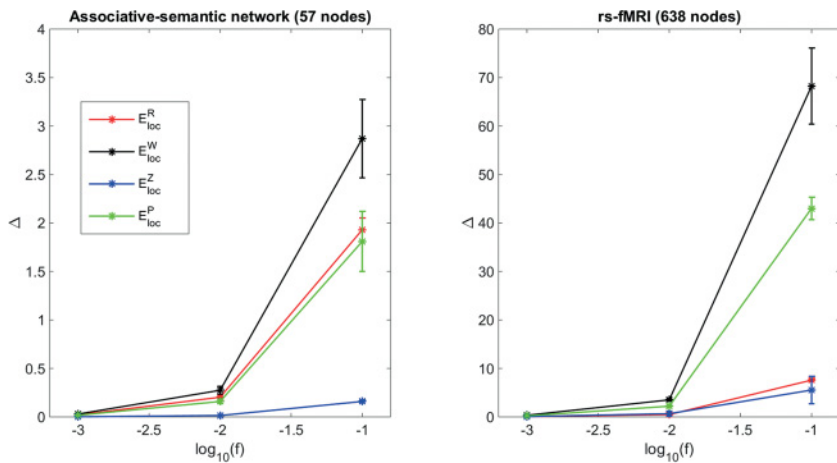


Figure 3: Mean relative error Δ of the robustness to noise of the local efficiency for different generalizations as a function of the amount of gaussian noise for the associative-semantic network (left) and the resting-state fMRI network (right). The noise fraction f is defined as the factor that is multiplied with samples from the standard normal noise distribution.

Table 3: Weight Scale Invariance and Applicability to Fully Weighted Networks of the Different Generalizations.

Method	Weight Scale Invariance	Suitable for Fully Weighted Networks
C^B	Yes	No
C^O	Yes	Yes
C^Z	Yes	Yes
C^H	Yes	Yes
$C^{Op,mean}$	Yes	No
$C^{Op,gm}$	Yes	No
$C^{Op,max}$	Yes	No
$C^{Op,min}$	Yes	No
$C^{M,gm}$	Yes	Yes
$C^{M,min}$	Yes	Yes
$C^{M,hm}$	Yes	Yes
E_{loc}^R	No	Yes
E_{loc}^W	Yes	Yes
E_{loc}^P	Yes	Yes
E_{loc}^Z	Yes	Yes

Table 4: Summary of the Main Findings.

Property	Clustering Coefficient	Local Efficiency
General versatility	$C^{M,hm}, C^{M,gm}, C^{M,min}, C^Z, C^B, C^O, C^{Op,gm}, C^{Op,mean}, C^{Op,max}, C^{Op,min}$	$E_{loc}^P, E_{loc}^W, E_{loc}^Z$
Continuity	$C^{M,hm}, C^{M,gm}, C^{M,min}, C^Z, C^H$	$E_{loc}^P, E_{loc}^W, E_{loc}^Z$
Sensitivity ^a	$C^{M,hm}, C^Z, C^H, C^O, C^{M,gm}, C^{M,min}, C^{Op,gm}, C^{Op,mean}, C^{Op,max}, C^{Op,min}, C^B$	$E_{loc}^P, E_{loc}^W, E_{loc}^R, E_{loc}^Z$
Robustness to noise ^b	$C^{M,gm}, C^{M,hm}, C^Z, C^H, C^{M,min}, C^O$	$E_{loc}^Z, E_{loc}^R, E_{loc}^P, E_{loc}^W$
Weight scale invariance	All generalizations	$E_{loc}^P, E_{loc}^W, E_{loc}^Z$
Applicable to fully weighted networks	$C^{M,hm}, C^{M,gm}, C^{M,min}, C^Z, C^H, C^O$	All generalizations

Note: The generalizations listed satisfy the property or are ranked according to performance.

^aListed based on their overall ranking (see Table 2) from best to worst.

^bListed based on their overall ranking (see Figures 2 and 3) from best to worst.

A summary table with the main findings for the essential and useful properties is given in Table 4.

3.6 Soft-Thresholded Weighted Networks. For generalizations of the clustering coefficient that can be applied to fully weighted networks, we

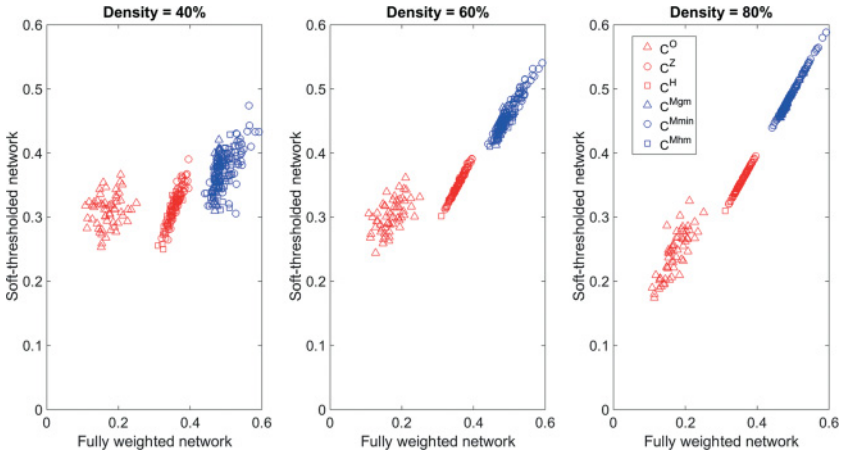


Figure 4: Scatter plot of the nodal clustering coefficient for different generalizations for nodes in the fully weighted network and the soft-thresholded weighted associative-semantic network.

Table 5: Correlation between the Clustering Coefficient or Local Efficiency Calculated in Each Node for the Fully Weighted Network and the Soft-Thresholded Weighted Network in the Associative-Semantic Network.

Density (%)	40	60	80
C^O	0.21	0.57	0.81
C^Z	0.91	1.00	1.00
C^H	0.91	1.00	1.00
$C^{M, gm}$	0.52	0.85	0.96
$C^{M, min}$	0.71	0.97	1.00
$C^{M, hm}$	0.55	0.92	1.00
E_{loc}^R	0.51	0.68	0.82
E_{loc}^W	0.97	1.00	1.00
E_{loc}^P	0.92	0.99	1.00
E_{loc}^Z	0.65	0.89	0.98

compared the values obtained in a fully weighted network with those obtained in a soft-threshold network. Since we have only unthresholded data for the associative-semantic network, the analysis is limited to this network. The hypothesis is that there will be a high correlation between both cases, especially when the density of the soft-thresholded network is high. In Figure 4 and Table 5, the results are shown for different soft-threshold values.

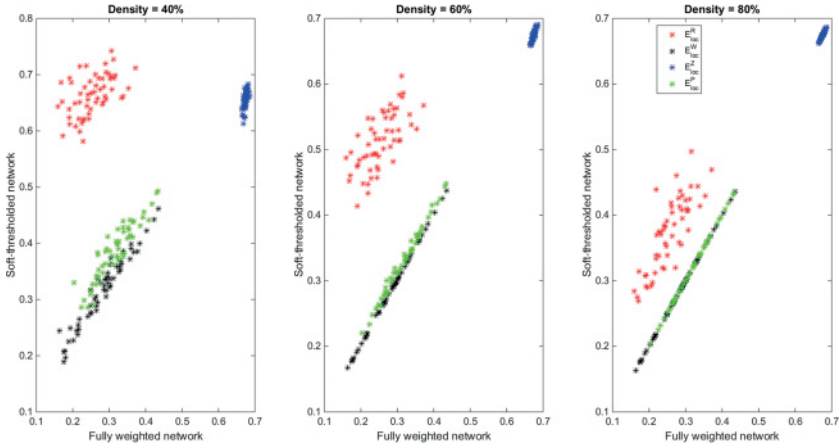


Figure 5: Scatter plot of the nodal local efficiency for different generalizations for nodes in the fully weighted network and the soft-thresholded weighted associative-semantic network.

Second, we performed the same analysis for the generalizations of the local efficiency. The results are shown in Figure 5 and Table 5.

4 Discussion

In this letter, we have defined a set of essential and useful properties that should ideally be satisfied for a generalization of a graph measure when extending from a binary network to a (fully) weighted network. We have compared all of these properties for the generalizations for the clustering coefficient and the local efficiency found in the literature, as well as for new generalizations. Some of the generalizations are especially suited in the case of (fully) weighted undirected graphs.

4.1 Essential Properties of Generalizations of Binary Graph Measures.

Generalizations from binary graph measures that are applicable to (fully) weighted graphs should ideally satisfy a number of properties. These properties can be subdivided into essential properties and useful properties. The latter class of properties depends on the application and should be considered relevant only in those applications.

The first essential property is general versatility, which refers to the fact that when applying the generalization on a binary graph, the result should be the same as the corresponding binary graph measure. This property is not satisfied for the generalization C^H of the clustering coefficient and the generalization E_{loc}^R of the local efficiency. The second essential property is continuity, which means that an infinitesimal small change in one of the

weights should lead to an infinitesimal small change in the graph measure. This is not the case for C^B , C^O , and C^{Op} , and it is also not satisfied for E_{loc}^R . Important to note is that expressions for the local clustering coefficient C^O (see equation 2.6) and the local efficiency E_{loc}^R (see equation 2.13), currently used in the brain connectivity toolbox in case of weighted undirected graphs, are not satisfying continuity (Onnela et al., 2005; Rubinov & Sporns, 2010). The third essential property is sensitivity to capture different cases for which the graph measures are designed. We have evaluated this property by looking at six possible cases for a simple triangle since both clustering coefficient and local efficiency are based on triangles. We found that the most sensitive generalizations for the clustering coefficient are C^O , C^Z , C^H , and $C^{M,hm}$, which could distinguish four of six cases. The best generalizations for the local efficiency are E_{loc}^W and E_{loc}^P , which were able to distinguish five of six cases. The fourth essential property is robustness to noise. We have investigated the robustness against different amounts of gaussian noise for two different real-world networks of different sizes. We found that $C^{M,gm}$ and $C^{M,hm}$ were the most robust generalizations for the clustering coefficient and E_{loc}^Z and E_{loc}^R are the most robust generalizations for the local efficiency.

4.2 Useful Properties of Generalizations of Binary Graph Measures.

Weight-scale invariance means that the graph measure is invariant to a global scale factor for all edges. In some cases, only relative connection strengths can be determined, and this property is especially useful in such cases. All generalizations of the clustering coefficient and the local efficiency are weight-scale-invariant except the generalization E_{loc}^R for the local efficiency. Furthermore, all generalizations of the local efficiency can be applied to fully weighted networks. This is also the case for most generalizations of the clustering coefficient except for C^B and C^{Op} .

4.3 Fully Weighted Undirected Graphs.

Most studies on graphs in neuroscience are related to binary undirected graphs (Sporns, Honey, & Kötter, 2007; He et al., 2007; He, Chen, & Evans, 2008; Van Wijk, Stam, & Daffertshofer, 2010; Vandenberghe et al., 2013). These graphs have either a connection or not between a pair of nodes, and they are easy to analyze. However, in order to obtain a binary graph, some measure of connectivity (often continuous) between nodes needs to be calculated and then thresholded on either amplitude or significance. The results of this procedure critically depend on the threshold used and do not take into account the strength of the connection. As a result, true connections that do not survive the threshold are removed, while false connections may sometimes be included. An alternative is the use of weighted graphs. In order to avoid taking into account spurious noisy connections, a soft threshold is sometimes applied, which removes these connections. The other connections are weighted (Wang, Li, Metzack, He, & Woodward, 2010; van den Heuvel,

Mandl, Stam, Kahn, & Pol, 2010). This reduces the problem described above but does not solve it completely. The advantage of fully weighted graphs is that no thresholding is required and all connections are taken into account (Mumford et al., 2010). The weight should then reflect not only the strength of the actual underlying biological connection but also the probability of being a true connection. Therefore, it might be necessary to apply a transformation from connection strength (e.g., defined by the (partial) correlation between two nodes in fMRI based functional networks) to weights. In case of the associative-semantic network, we have shown that the generalizations that are applicable to fully weighted networks show a high correlation with the network, which is soft-thresholded, especially when the density of the latter is high.

4.4 Choice of the Best Generalization. The choice of the best generalization is not always easy. The essential requirements of general versatility and continuity can be proven mathematically, and they are completely independent of the application. The relative importance of the two other essential properties, sensitivity and robustness to noise, depend on the application and most likely requires a trade-off between these two properties. In this study, we have tried to quantify sensitivity based on the values obtained for six possible triangles and robustness to noise by evaluating the behavior when adding different amounts of gaussian noise. Based on our results, we propose that the best generalization of the clustering coefficient is $C^{M,lm}$, which is more robust to noise compared to C^Z . The choice of the best generalization of the local efficiency is E_{loc}^P , but if robustness to noise is very important, E_{loc}^Z can be selected; sensitivity, however, will clearly be lower compared to E_{loc}^P . If the noise in an application is nongaussian, a similar approach, as we have shown, can be taken for the noise model, which is more appropriate for the application. Satisfying the useful properties can be important but depends on the application. Despite the difficulty in selecting the optimal generalization, we believe that readers can use the assessment of all properties for all generalizations to select the best generalization for the application under study.

5 Conclusion

In this letter, we have focused on the comparison of different generalizations for the clustering coefficient and local efficiency to the case of (fully) weighted networks by looking at different properties of these graph measures and studying the performance in two real-world networks of different sizes. The best generalization of the clustering coefficient is $C^{M,lm}$, defined in Miyajima and Sakuragawa (2014), while the best generalization of the local efficiency is E_{loc}^P proposed in our work. Depending on the application and the relative importance of sensitivity and robustness to noise, other

generalizations may be selected on the basis of the properties investigated in this letter.

Acknowledgments

This work was supported by Research Foundation Flanders (FWO; G0660.09 and G0A0913N to R.V. and P.D.), KU Leuven (OT/12/097 to R.V. and P.D.), Federaal Wetenschapsbeleid belpo (IAP-VII P7/11), and Stichting voor Alzheimer Onderzoek (SAO11020 and 13007). Y.W. has a grant from the Chinese Scholarship Council, and R.V. is a senior clinical investigator. We thank both reviewers for their constructive comments, which greatly improved the letter.

References

- Achard, S., & Bullmore, E. (2007). Efficiency and cost of economical brain functional networks. *PLoS Computational Biology*, 3(2), e17.
- Amaral, L. A., & Ottino, J. M. (2004). Complex networks. *European Physical Journal B—Condensed Matter and Complex Systems*, 38, 147–162.
- Barabasi, A.-L., & Oltvai, Z. N. (2004). Network biology: Understanding the cell's functional organization. *Nature Reviews Genetics*, 5, 101–113.
- Barrat, A., Barthelemy, M., Pastor-Satorras, R., & Vespignani, A. (2004). The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 3747–3752.
- Batalle, D., Eixarch, E., Figueras, F., Muñoz-Moreno, E., Bargallo, N., Illa, M., ... Gratacos, E. (2012). Altered small-world topology of structural brain networks in infants with intrauterine growth restriction and its association with later neurodevelopmental outcome. *NeuroImage*, 60, 1352–1366.
- Bullmore, E., & Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.*, 10, 186–198.
- Crossley N. A., Mechelli, A., Vrtes, P. E., Winton-Brown, T. T., Patel, A. X., Ginestet, C. E., ... Bullmore, E. T. (2013). Cognitive relevance of the community structure of the human brain functional coactivation network. *Proc. Natl. Acad. Sci. USA*, 110, 11583–11588.
- Fornito, A., Zalesky, A., & Breakspear, M. (2013). Graph analysis of the human connectome: Promise, progress, and pitfalls. *NeuroImage*, 80, 426–444.
- He, Y., Chen, Z. J., & Evans, A. C. (2007). Small-world anatomical networks in the human brain revealed by cortical thickness from MRI. *Cereb. Cortex*, 17(10), 2407–2419.
- He, Y., Chen, Z., & Evans, A. (2008). Structural insights into aberrant topological patterns of large-scale cortical networks in Alzheimer's disease. *Journal of Neuroscience*, 28, 4756–4766.
- Holme, P., Park, S. M., Kim, B. J., & Edling, C. R. (2007). Korean university life in a network perspective: Dynamics of a large affiliation network. *Physica A: Statistical Mechanics and Its Applications*, 373, 821–830.

- Latora, V., & Marchiori, M. (2003). Economic small-world behavior in weighted networks. *European Physical Journal B—Condensed Matter and Complex Systems*, *32*, 249–263.
- Miyajima, K., & Sakuragawa, T. (2014). *Continuous and robust clustering coefficients for weighted and directed networks*. arXiv:1412.0059
- Mumford, J. A., Horvath, S., Oldham, M. C., Langfelder, P., Geschwind, D. H., & Poldrack, R. A. (2010). Detecting network modules in fMRI time series: A weighted network analysis approach. *NeuroImage*, *52*, 1465–1476.
- Onnela, J.-P., Saramäki, J., Kertész, J., & Kaski, K. (2005). Intensity and coherence of motifs in weighted complex networks. *Phys. Rev. E*, *71*, 065103.
- Opsahl, T., & Panzarasa, P. (2009). Clustering in weighted networks. *Social Networks*, *31*, 155–163.
- Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*, *52*, 1059–1069.
- Saramäki, J., Kivela, M., Onnela, J.-P., Kaski, K., & Kertesz, J. (2007). Generalizations of the clustering coefficient to weighted complex networks. *Physical Review E*, *75*, 027105.
- Sporns, O., Honey, C. J., & Kötter, R. (2007). Identification and classification of hubs in brain networks. *PLoS One*, *2*(10), e1049.
- van den Heuvel, M. P., Mandl, R. C., Stam, C. J., Kahn, R. S., & Pol, H. E. H. (2010). Aberrant frontal and temporal complex network structure in schizophrenia: A graph theoretical analysis. *Journal of Neuroscience*, *30*(47), 15915–15926.
- Van Wijk, B. C., Stam, C. J., & Daffertshofer, A. (2010). Comparing brain networks of different size and connectivity density using graph theory. *PLoS One*, *5*(10), e13701.
- Vandenbergh, R., Wang, Y., Nelissen, N., Vandenbulcke, M., Dhollander, T., Sunaert, S., & Dupont, P. (2013). The associative-semantic network for words and pictures: Effective connectivity and graph analysis. *Brain Lang.*, *127*(2), 264–272.
- Wang, L., Li, Y., Metz, P., He, Y., & Woodward, T. S. (2010). Age-related changes in topological patterns of large-scale brain functional networks during memory encoding and recognition. *NeuroImage*, *50*, 862–872.
- Wang, Y., Nelissen, N., Adamczuk, K., De Weer, A.-S., Vandenbulcke, M., Sunaert, S., . . . Dupont, P. (2014). Reproducibility and robustness of graph measures of the associative-semantic network. *PLoS One*, *9*(12), e115215.
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of small-world networks. *Nature*, *393*(6684), 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*(1), 1–43.

Received August 27, 2015; accepted September 5, 2016.