Research Article

Statistical Analysis of Weighted Networks

I. E. Antoniou and E. T. Tsompa

Mathematics Department, Aristotle University, 54124 Thessaloniki, Greece

Correspondence should be addressed to E. T. Tsompa, etsompa@math.auth.gr

The purpose of this paper is to assess the statistical characterization of weighted networks in terms of the generalization of the relevant parameters, namely, average path length, degree distribution, and clustering coefficient. Although the degree distribution and the average path length admit straightforward generalizations, for the clustering coefficient several different definitions have been proposed in the literature. We examined the different definitions and identified the similarities and differences between them. In order to elucidate the significance of different definitions of the weighted clustering coefficient, we studied their dependence on the weights of the connections. For this purpose, we introduce the relative perturbation norm of the weights as an index to assess the weight distribution. This study revealed new interesting statistical regularities in terms of the relative perturbation norm useful for the statistical characterization of weighted graphs.

1. Introduction

Complex systems may also [1] emerge from a large number of interdependent and interacting elements. Networks have proven to be effective models of natural or man-made complex systems, where the elements are represented by the nodes and their interactions by the links. Typical well-known examples include communication and transportation networks, social networks, and biological networks [2–5].

Although the statistical analysis of the underlying topological structure has been very fruitful [2–5], it was limited due to the fact that in real networks the links may have different capacities or intensities or flows of information or strengths. For example, weighted links can be used for the Internet to represent the amount of data exchanged between two hosts in the network. For scientific collaboration networks, the weight depends on the number of coauthored papers between two authors. For airport networks, it is either the number of available seats on direct flight connections between airports i and j or the actual number of passengers that travel from airport i to j. For neural networks, the weight is the number

of junctions between neurons and for transportation networks it is the Euclidean distance between two destinations.

The diversification of the links is described in terms of weights on the links. Therefore, the statistical analysis has to be extended from graphs to weighted complex networks. If all links are of equal weight, the statistical parameters used for unweighted graphs are sufficient for the statistical characterization of the network. Therefore, the statistical parameters of the weighted graphs should reduce to the corresponding parameters of the conventional graphs if all weights are put equal to unity.

Complex graphs are characterized by three main statistical parameters, namely, the degree distribution, the average path length, and the clustering coefficient. We will briefly mention the definitions for clarity and for a better understanding of the proposed extensions of these parameters for weighted graphs.

The structure of a network with N nodes is represented by an $N \times N$ binary matrix $A = \{a_{ij}\}$, known as adjacency matrix, whose element a_{ij} equals 1, when there is a link joining node i to node j and 0 otherwise ($i, j = 1, 2, \ldots, N$).

In the case of undirected networks with no self-links (links connecting a node to itself), the adjacency matrix is symmetric ($a_{ij} = a_{ji}$) and all elements of the main diagonal equal 0 ($a_{ii} = 0$).

The degree k_i of a node i is defined as the number of its neighbors, that is, the number of links incident to node i:

$$k_i = \sum_{j \in \Pi(i)} a_{ij}, \tag{1.1}$$

where a_{ij} the elements of the adjacency matrix A and $\Pi(i)$ the neighborhood of node i.

The degree distribution is the probability that some node has k connections to other nodes. Many real complex networks have been found to be described by a power law degree, $P(k) \sim k^{-\gamma}$, with $2 \leq \gamma \leq 3$.

The characteristic path length of a network is defined as the average of the shortest path lengths between any two nodes:

$$L = \frac{2}{N(N-1)} \sum_{i,j} d_{ij}, \tag{1.2}$$

where d_{ij} is the shortest path length between i and j, defined as the minimum number of links traversed to get from node i to node j.

In many real networks, it is found that the existence of a link between nodes i and j and between nodes i and k enhances the probability that node j will also be connected to node k. This tendency of the neighbors of any node i to connect to each other is called clustering and is quantified by the clustering coefficient C_i, which is the fraction of triangles in which node i participates, to the maximum possible number of such triangles:

$$C_i = \frac{n_i}{k_i(k_i - 1)} = \frac{\sum_{j,k} a_{ij} a_{jk} a_{ki}}{k_i(k_i - 1)}, \quad k_i \neq 0, 1, \tag{1.3}$$

where $(1/2)n_i = (1/2)\sum_{j,k} a_{ij} a_{jk} a_{ki}$ is the actual number of triangles in which node i participates, that is, the actual number of links between the neighbors of node i, and $k_i(k_i-1)/2$

is the maximum possible number of links, when the subgraph of neighbors of node i is completely connected.

The clustering coefficient C_i equals 1 if node i is the center of a fully interconnected cluster, and equals 0 if the neighbors of node i are not connected to each other.

In order to characterize the network as a whole, we usually consider the average clustering coefficient C over all the nodes. We may also consider the average clustering coefficient $C(k)$ over the node degree k.

Studies of real complex networks have shown that their connection topology is neither completely random nor completely regular, but lies between these extreme cases. Many real networks share features of both extreme cases. For example, the short average path length, typical of random networks, comes along with large clustering coefficient, typical of regular lattices. The coexistence of these attributes defines a distinct class of networks, interpolating between regular lattices and random networks, known today as small world networks [3–6]. Another class of networks emerges when the degree distribution is a power law (scale free) distribution, which signifies the presence of a nonnegligible number of highly connected nodes, known as hubs. These nodes, with very large degree k compared to the average degree $\langle k \rangle$, are critical for the network's robustness and vulnerability. These networks are known today as scale-free networks [2–4, 7].

The purpose of this paper is to assess the statistical characterization of weighted networks in terms of proper generalizations of the relevant parameters, namely, average path length, degree distribution, and clustering coefficient. After reviewing the definitions of the weighted average path length, weighted degree distribution and weighted clustering coefficient in Section 2, we compare them in Section 3. Although the degree distribution and the average path length admit straightforward generalizations, for the clustering coefficient several different definitions have been proposed. In order to elucidate the significance of different definitions of the weighted clustering coefficient, we studied their dependence on the weights of the connections in Section 4, where we introduce the relative perturbation norm as an index to assess the weight distribution. This study revealed new interesting statistical regularities in terms of the relative perturbation norm useful for the statistical characterization of weighted graphs.

2. Statistical parameters of weighted networks

The weights of the links between nodes are described by an $N \times N$ matrix $W = \{w_{ij}\}$. The weight w_{ij} is 0 if the nodes i and j are not linked. We will consider the case of symmetric positive weights ($w_{ij} = w_{ji} \geq 0$), with no self-links ($w_{ii} = 0$).

In order to compare different networks or different kinds of weights, we usually normalize the weights in the interval $[0, 1]$, by dividing all weights by the maximum weight. The normalized weights are $w_{ij} / \max(w_{ij})$.

The statistical parameters for weighted networks are defined as follows.

The node degree $k_i = \sum_{j \in \Pi(i)} a_{ij}$, which is the number of links attached to node i, is extended directly to the strength or weighted degree, which is the sum of the weights of all links attached to node i:

$$s_i = \sum_{j \in \Pi(i)} w_{ij}. \tag{2.1}$$

The strength of a node takes into account both the connectivity as well as the weights of the links.

The degree distribution is also extended for the weighted networks to the strength distribution $P(s)$, which is the probability that some node's strength equals s. Recent studies indicate power law $P(s) \sim s^{-a}$ [8–10].

There are two different generalizations of the characteristic path length in the literature, applicable to transportation and communication networks. In the case of transportation networks, the weighted shortest path length d_{ij} between the nodes i and j is defined as the smallest sum of the weights of the links throughout all possible paths from node i to node j [11, 12]:

$$d_{ij}^{\text{tr}} = \min_{\gamma(i,j) \in \Gamma(i,j)} \left[\sum_{m,n \in \gamma(i,j)} w_{mn} \right], \tag{2.2}$$

where $\gamma(i,j)$ is a path from node i to node j and $\Gamma(i,j)$ is the class of paths from i to j.

The weight describes physical distances and/or cost usually involved in transportation networks. The capacity, intensity, strength, or efficiency of the connection is inversely proportional to the weight.

However, this definition is not suitable for communication networks, where the efficiency of the communication channel between two nodes is proportional to the weight. The shortest path length in case of communication networks is defined as the smallest sum of the inverse weights of the links throughout all possible paths from node i to node j:

$$d_{ij}^{\text{com}} = \min_{\gamma(i,j) \in \Gamma(i,j)} \left[\sum_{m,n \in \gamma(i,j)} \frac{1}{w_{mn}} \right]. \tag{2.3}$$

The weighted characteristic path length for both cases is the average of all shortest path lengths and it is calculated by formula (1.2).

We found in the literature seven proposals for the definitions of the weighted clustering coefficient, which we will review.

(i) Zhang and Horvath's [13] definition is as follows:

$$C_{w,i}^{Z} = \frac{\sum_j \sum_k w_{ij} w_{jk} w_{ki}}{\left(\sum_j w_{ij} \right)^2 - \sum_j w_{ij}^2}. \tag{2.4}$$

The weights in this definition are normalized. The idea of the generalization is the substitution of the elements of the adjacency matrix by the weights in the nominator of formula (1.3); as for the denominator, the upper limit of the nominator is obtained in order to normalize the coefficient between 0 and 1. The definition originated from gene coexpression networks.

As shown by Kalna and Higham [14], an alternative formula that may apply for this definition is

$$C_{w,i}^{K} = \frac{\sum_j \sum_k w_{ij} w_{jk} w_{ki}}{\sum_j \sum_{k \neq j} w_{ij} w_{ik}}. \tag{2.5}$$

(ii) Lopez-Fernandez et al.'s [15] definition is

$$C_{w,i}^L = \sum_{j,k \in \Pi(i)} \frac{w_{jk}}{k_i(k_i - 1)}. \tag{2.6}$$

The weights in this definition are not normalized. The idea of the generalization is the substitution of the number of links that exist between the neighbors of node i in formula (1.3) by the weight of the link between the neighbors j and k. The definition originated from an affiliation network for committers (or modules) of free, open source software projects.

(iii) Onnela et al.'s [16] definition is

$$C_{w,i}^O = \frac{\sum_{j,k} (w_{ij} w_{jk} w_{ki})^{1/3}}{k_i(k_i - 1)}. \tag{2.7}$$

The weights in this definition are normalized. The quantity $I(g) = (w_{ij} w_{jk} w_{ki})^{1/3}$ is called "intensity" of the triangle ijk. The concept for this generalization is to substitute the total number of the triangles in which node i participates by the intensity of the triangle, which is geometric mean of the links' weights.

(iv) Barrat et al.'s [8] definition is

$$C_{w,i}^B = \frac{1}{s_i(k_i - 1)} \sum_{j,k} \frac{w_{ij} + w_{ik}}{2} a_{ij} a_{jk} a_{ki}. \tag{2.8}$$

The weights in this definition are not normalized. The idea of the generalization is the substitution of the elements of the adjacency matrix in formula (1.3), by the average of the weights of the links between node i and its neighbors j and k with respect to normalization factor $s_i(k_i - 1)$ which ensures that $0 \le C_{w,i}^B \le 1$. This definition was used for airport and scientific collaboration networks.

(v) Serrano et al.'s [17] definition is

$$C_{w,i}^S = \frac{\sum_j \sum_k w_{ij} w_{ik} a_{kj}}{s_i^2(1 - Y_i)}, \tag{2.9}$$

where $Y_i = \sum_j (w_{ij}/s_i)^2$ has been named "disparity."

The weights in this definition are not normalized. This formula is used for the generalization of the average clustering coefficient with degree k, which has a probabilistic interpretation just as the unweighted clustering coefficient.

(vi) Holme et al.'s [18] definition is

$$C_{w,i}^H = \frac{\sum_j \sum_k w_{ij} w_{jk} w_{ki}}{\max(w_{ij}) \sum_j \sum_{k \ne j} w_{ij} w_{ik}}. \tag{2.10}$$

The only difference between formulas (2.4) and (2.10) is that (2.10) is divided by $\max(w_{ij})$.

We will not discuss this definition in the comparison because the essence of the comparison is already addressed by definition (2.4).

(vii) Li et al.'s [19] definition of the weighted clustering coefficient is another version of the Lopez-Fernandez proposal (2.6).

3. The relation between the different weighted clustering coefficients

(1) All definitions reduce to the clustering coefficient (1.3), when the weights w_{ij} are replaced by the adjacency matrix elements.

(2) All weighted clustering coefficients reduce to 0 when there are no links between the neighbors of node i, that is, when $a_{jk} = w_{jk} = 0$.

(3) In the other extreme, all weighted clustering coefficients take the value 1 when all neighbors of node i are connected to each other. Formulas (2.4) and (2.6) take the value 1 if the weights between the neighbors of the node i are 1, independently of the weights of the other links. Formula (2.7) takes the value 1, if and only if all the weights are equal to 1. Formulas (2.8) and (2.9) take the value 1 for all fully connected graphs, independently of all the weights.

These calculations are presented in Appendix A.

(4) We calculated the values of the weighted clustering coefficients of node i participating in a fully connected triangle. Formulas (2.4) and (2.6) take the value w_{jk} of the weight of the link between neighbors j and k of node i. Formula (2.7) becomes equal to the intensity of the triangle $C_{w,i}^{O} = (w_{ij}w_{jk}w_{ki})^{1/3}$ for all nodes of the triangle. Formulas (2.8) and (2.9) take the value 1 for all fully connected graphs, independently of all weights.

These calculations are presented in Appendix B.

4. The dependence of the weighted clustering coefficients on the weights

In order to understand the meaning of the different proposals or definitions (2.4), (2.6), (2.7), (2.8), and (2.9) of the weighted clustering coefficient, we will examine their dependence on the weights, without alteration of the topology of the graph. We simply examine the values of these definitions for different distributions of weights, substituting the nonzero elements of the adjacency matrix A by weights normalized between 0 and 1.

A way to distinguish and compare different weight distributions over the same graph is in terms of the relative perturbation norm $\|A - W\|/\|A\|$, which gives the percentage of the perturbation of the adjacency matrix introduced by the weights. For simplicity, we considered the L_2 norm.

Although the weight distributions of certain real weighted networks have been found to be inhomogeneous like the exponential [8–10, 12, 19], we will examine the dependence of the weighted clustering coefficient with respect to the relative perturbation norms for the exponential as well as for the uniform and normal distributions, for different graphs.

We have examined many networks from 20 up to 300 nodes with different topologies that were generated by the networks software PAJEK [20]. The weights examined are randomly generated numbers following uniform, normal, or exponential distributions with several parameter values, so that the percentages of the perturbations scale from 0–90% increasing by 10% at each perturbation. All simulations gave rise to the same results, Figures 3 and 4, representing the typical trends of random and scale free networks, Figures 1 and 2. The related statistical trend analysis is reported in Appendix C.

Observations on the dependence of the weighted clustering coefficients on the relative perturbation norm are defined as follows.
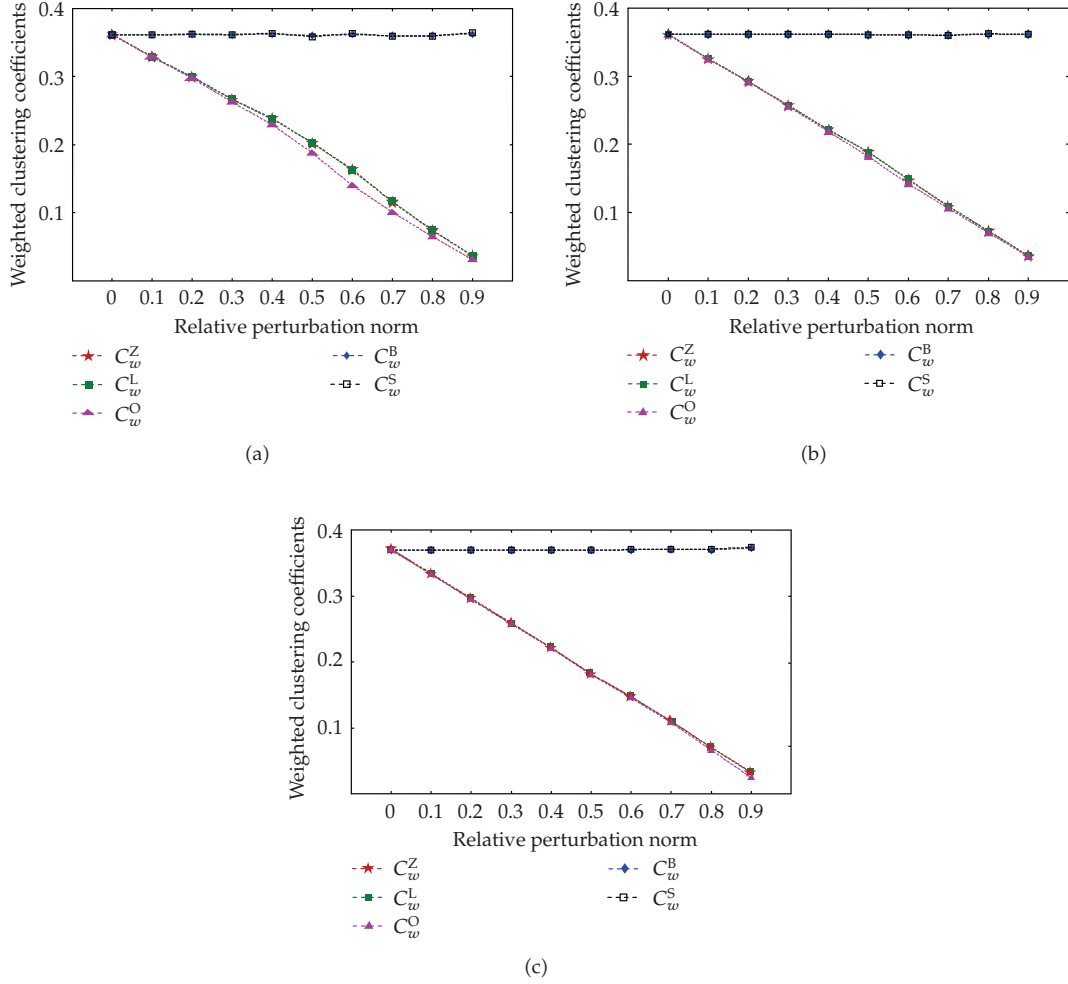
(1) For the weighted clustering coefficients (2.4), (2.6), and (2.7), there is a clear linear dependence of their values, in terms of the relative perturbation norm of the weighted network. It is remarkable that no dependence is observed on the values of weights on specific links.
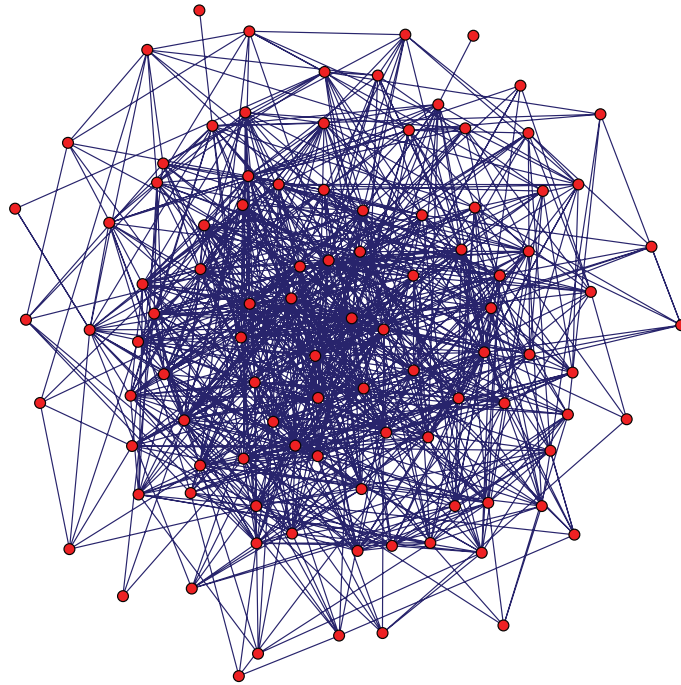
Figure 1: The random network (Erdos-Renyi model) examined consists of 100 nodes and was generated by the networks software PAJEK [20]. The clustering coefficient for the unweighted network is 0.3615.
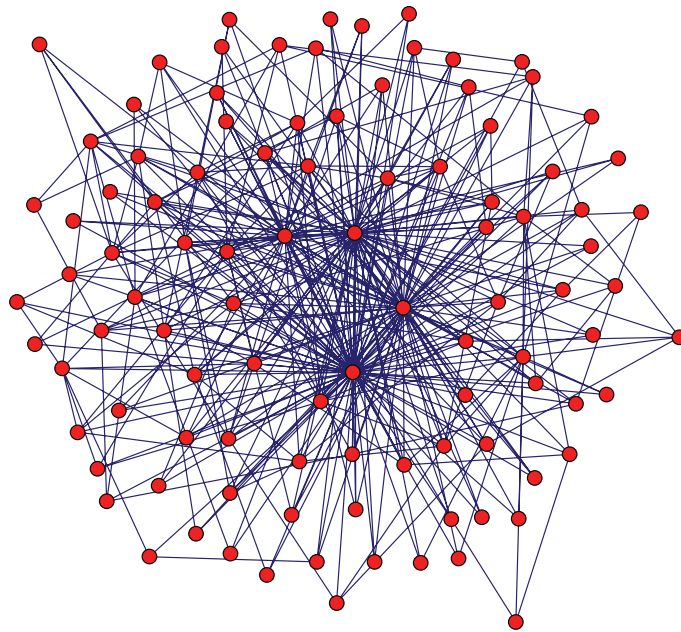
Figure 2: The scale-free network (Barabasi-Albert extended model) examined consists of 100 nodes and was generated by the networks software PAJEK [20]. The clustering coefficient for the unweighted network is 0.6561.
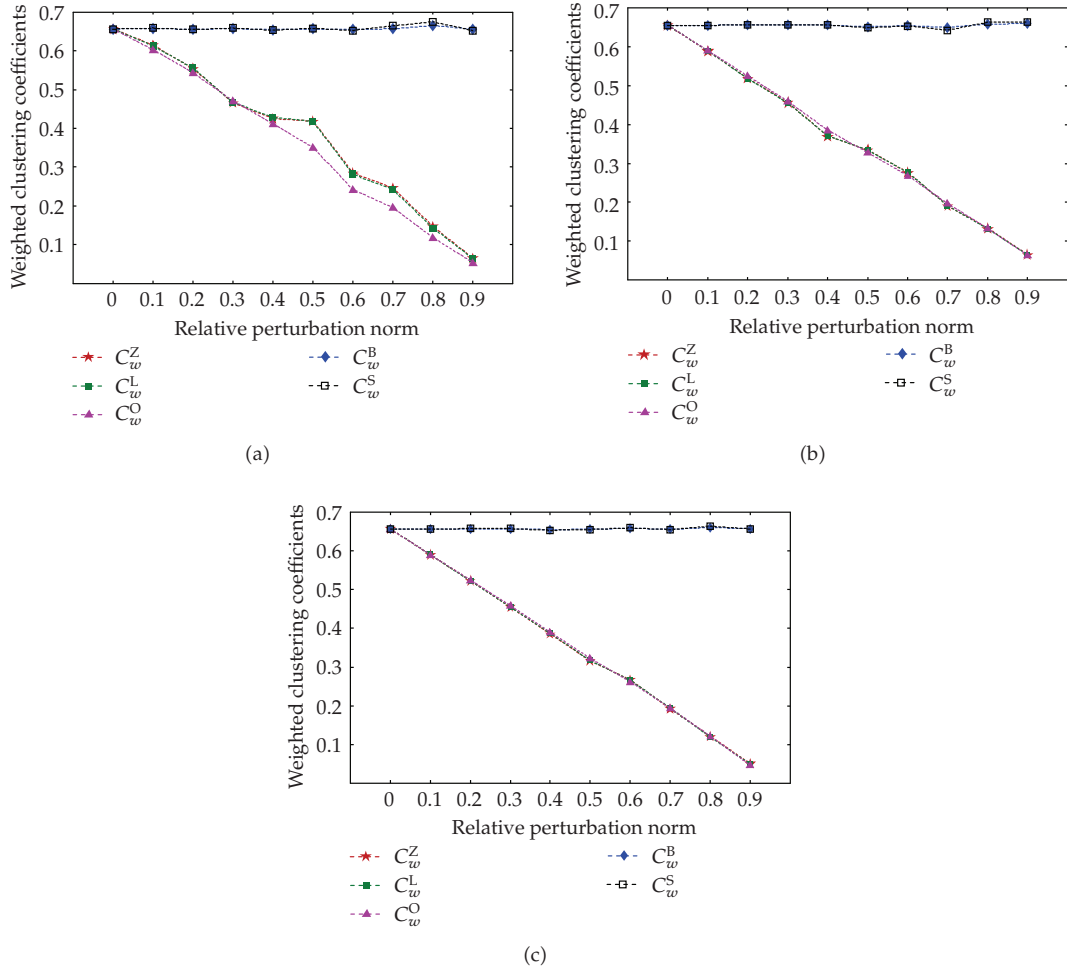
(a)

(b)

(c)

Figure 3: The values of all five weighted clustering coefficients Zhang and Horvath$C_{w,i}^Z$ ($\bigstar$), Lopez-Fernandez et al. $C_{w,i}^L$ ($\blacksquare$), Onnela et al. $C_{w,i}^O$ ($\blacktriangle$), Barrat et al. $C_{w,i}^B$ ($\blacklozenge$), and Serrano et al. $C_{w,i}^S$ ($\square$), in terms of the relative perturbation norm for the random network (Erdos-Renyi model) with 100 nodes. (a) The weights are randomly generated numbers following the uniform distribution; (b) the weights are randomly generated numbers following the normal distribution; (c) the weights are randomly generated numbers following the exponential distribution.

The Zhang and Horvath (2.4), Lopez-Fernandez et al. (2.6), and Onnela et al. (2.7) weighted clustering coefficients follow the same trend, decaying smoothly as the relative perturbation norm increases. More specifically the trends of Zhang and Horvath (2.4) and Lopez-Fernandez et al. (2.6) almost coincide, as expected (Section 3), while the trend of Onnela et al. (2.8) varies slightly from the other two.

(2) The resulting linear models indicate that the values of the weighted clustering coefficients decrease by 10% of the value of the unweighted clustering coefficient, as the relative perturbation norm increases by 10%.

Figure 4: The values of all five weighted clustering coefficients Zhang and Horvath $C_{w,i}^Z$ (★), Lopez-Fernandez et al. $C_{w,i}^L$ (■), Onnela et al. $C_{w,i}^O$ (▲), Barrat et al. $C_{w,i}^B$ (♦), and Serrano et al. $C_{w,i}^S$ (□), in terms of the relative perturbation norm for the scale-free network (Barabasi-Albert extended model) with 100 nodes. (a) The weights are randomly generated numbers following the uniform distribution; (b) the weights are randomly generated numbers following the normal distribution; (c) the weights are randomly generated numbers following the exponential distribution.

(3) The dependence of the weighted clustering coefficients (2.8) and (2.9) on the weights is not significant. The weighted clustering coefficients of Barrat et al. (2.8) and Serrano et al. (2.9) do not change (variations appear after the first two decimal digits), regardless of the size of the network or the distribution of the weights. As mentioned in Section 3, these coefficients are independent of the weights when the graph is completely connected. We observe here that the weighted clustering coefficients (2.8) and (2.9) are independent of the weights for graphs that are not completely connected. We conjecture therefore that this is a general fact.

5. Concluding remarks

(1) From the observed dependence of the values of all five weighted clustering coefficients on the relative perturbation norm, we conjecture that the proposed relative perturbation norm is a reliable index of the weight distribution. The meaning of the decaying trend of weighted clustering coefficients defined by Zhang and Horvath (2.4), Lopez-Fernandez et al. (2.6), and Onnela et al. (2.7), with respect to the increase of the relative perturbation norm, is quite natural. The clustering decreases almost linearly as the weights "decrease."

(2) We presented in Appendices A and B the calculations demonstrating that all definitions reduce to the clustering coefficient (1.3), when the weights w_{ij} are replaced by the adjacency matrix elements. The values of the weighted clustering coefficients of node i participating in a fully connected triangle are presented for completeness because we did not found them in the literature.

Appendices

A. Calculations on the weighted clustering coefficient

The definitions (2.4)–(2.9) reduce to the clustering coefficient (1.3), when the weights w_{ij} are replaced by the adjacency matrix elements.

(1) Zhang and Horvath's is [13]

$$C_{w,i}^{Z} = \frac{\sum_{j}\sum_{k} w_{ij} w_{jk} w_{ki}}{\left(\sum_{j} w_{ij}\right)^{2} - \sum_{j} w_{ij}^{2}}. \tag{A.1}$$

The proof is presented by the authors.

For example, for a fully connected network with four nodes,

$$
\begin{aligned}
C_{w,1}^{Z} &= \frac{\sum_{j\neq1}^{4}\sum_{k\neq1}^{4} w_{1j} w_{jk} w_{k1}}{\left(\sum_{j\neq1}^{4} w_{1j}\right)^{2} - \sum_{j\neq1}^{4} w_{1j}^{2}} = \frac{\sum_{j\neq1}^{4} w_{1j}\left(w_{j2} w_{21} + w_{j3} w_{31} + w_{j4} w_{41}\right)}{\left(w_{12} + w_{13} + w_{14}\right)^{2} - w_{12}^{2} - w_{13}^{2} - w_{14}^{2}} \\
&= \frac{w_{12} w_{23} w_{31} + w_{12} w_{24} w_{41} + w_{13} w_{32} w_{21} + w_{13} w_{34} w_{41} + w_{14} w_{42} w_{21} + w_{14} w_{43} w_{31}}{2\left(w_{12} w_{13} + w_{13} w_{14} + w_{12} w_{14}\right)} \\
&= \frac{w_{12} w_{23} w_{31} + w_{13} w_{34} w_{41} + w_{12} w_{24} w_{41}}{w_{12} w_{13} + w_{13} w_{14} + w_{12} w_{14}},
\end{aligned} \tag{A.2}
$$

$C_{w,1}^{Z} = 1$ when $w_{23} = w_{34} = w_{24} = 1$.

(2) Lopez-Fernandez et al.'s is [15]

$$C_{w,i}^{L} = \sum_{j,k\in\Pi(i)} \frac{w_{jk}}{k_{i}(k_{i}-1)}; \tag{A.3}$$

Table 1: A comparison between random network and scale-free network (NM: network model, WD: weight distribution, WCC: weighted clustering coefficient, R^2: squared correlation coefficient, F | sig: F-ratio and corresponding significance, b_0, b_1: estimated parameters of the linear model (constant and coefficient of the predictor variable relative perturbation norm) with corresponding t | sig (p-value)).

NM	WD	WCC	R^2	F | sig	b_0	t | sig	b_1	t | sig
Random Network Erdos-Renyi model	Uniform (Figure 3(a))	Zhang	0.994	1290.026 | 0.000	0.372	69.092 | 0.000	-0.362	-35.917 | 0.000
		Lopez-Fernandez	0.994	1364.860 | 0.000	0.371	71.072 | 0.000	-0.362	-36.944 | 0.000
		Onnela	0.997	2975.038 | 0.000	0.370	100.354 | 0.000	-0.376	-54.544 | 0.000
		Barrat	0.013	0.101 | 0.758	0.362	657.630 | 0.000	0.000	-0.318 | 0.758
		Serrano	0.000	0.001 | 0.972	0.362	301.844 | 0.000	0.000	-0.036 | 0.972
	Normal (Figure 3(b))	Zhang	0.999	13629.467 | 0.000	0.364	219.958 | 0.000	-0.362	-116.745 | 0.000
		Lopez-Fernandez	0.999	13711.182 | 0.000	0.364	220.768 | 0.000	-0.362	-117.094 | 0.000
		Onnela	1	35769.842 | 0.000	0.363	352.133 | 0.000	-0.366	-189.129 | 0.000
		Barrat	0.001	0.011 | 0.920	0.361	1763.897 | 0.000	0.000	-0.104 | 0.920
		Serrano	0.027	0.224 | 0.649	0.362	888.635 | 0.000	0.000	-0.473 | 0.649
	Exponential (Figure 3(c))	Zhang	1	102303.279 | 0.000	0.362	599.496 | 0.000	-0.362	-319.849 | 0.000
		Lopez-Fernandez	1	109645.313 | 0.000	0.362	619.969 | 0.000	-0.362	-331.127 | 0.000
		Onnela	1	24935.117 | 0.000	0.363	291.250 | 0.000	-0.369	-157.909 | 0.000
		Barrat	0.441	6.305 | 0.036	0.361	903.747 | 0.000	0.002	2.511 | 0.036
		Serrano	0.592	11.622 | 0.009	0.361	700.433 | 0.000	0.003	3.409 | 0.009
Scale free Network Barabasi Albert Extended model	Uniform (Figure 4(a))	Zhang	0.982	425.202 | 0.000	0.679	40.511 | 0.000	-0.647	-20.620 | 0.000
		Lopez-Fernandez	0.982	427.643 | 0.000	0.680	40.421 | 0.000	-0.651	-20.680 | 0.000
		Onnela	0.996	2269.81 | 0.000	0.672	87.611 | 0.000	-0.685	-47.643 | 0.000
		Barrat	0.106	0.945 | 0.360	0.655	356.437 | 0.000	0.003	0.972 | 0.360
		Serrano	0.096	0.854 | 0.382	0.655	162.161 | 0.000	0.007	0.924 | 0.382
	Normal (Figure 4(b))	Zhang	0.998	3375.047 | 0.000	0.652	109.145 | 0.000	-0.650	-58.095 | 0.000
		Lopez-Fernandez	0.998	3510.742 | 0.000	0.652	111.183 | 0.000	-0.651	-59.252 | 0.000
		Onnela	1	25056.870 | 0.000	0.656	296.796 | 0.000	-0.655	-158.294 | 0.000
		Barrat	0.016	0.129 | 0.729	0.655	368.559 | 0.000	0.001	0.359 | 0.729
		Serrano	0.018	0.150 | 0.709	0.655	170.560 | 0.000	0.003	0.387 | 0.709
	Exponential (Figure 4(c))	Zhang	0.999	13635.348 | 0.000	0.656	215.181 | 0.000	-0.667	-116.770 | 0.000
		Lopez-Fernandez	0.999	13883.081 | 0.000	0.656	216.885 | 0.000	-0.668	-117.826 | 0.000
		Onnela	1	21450.540 | 0.000	0.658	269.459 | 0.000	-0.670	-146.460 | 0.000
		Barrat	0.119	1.077 | 0.330	0.656	745.786 | 0.000	0.002	1.038 | 0.330
		Serrano	0.088	0.771 | 0.405	0.655	388.730 | 0.000	0.003	0.878 | 0.405

this formula can be expressed as

$$C_{w,i}^L = \frac{\sum_{j,k} w_{jk} a_{ij} a_{ik}}{k_i(k_i - 1)}. \tag{A.4}$$

It is obvious that the formula reduces to the unweighted (1.3), when w_{jk} are substituted by a_{jk}.

(3) Onnela et al.'s [16]

$$C_{w,i}^O = \frac{\sum_{j,k} (w_{ij} w_{jk} w_{ki})^{1/3}}{k_i(k_i - 1)} \tag{A.5}$$

reduces to the unweighted definition (1.3), when w_{jk} are substituted by a_{jk}:

$$(a_{ij})^{1/3} = a_{ij}, \quad \text{hence } (w_{ij} w_{jk} w_{ki})^{1/3} = (a_{ij} a_{jk} a_{ki})^{1/3} = a_{ij} a_{jk} a_{ki},$$

$$C_{w,i}^O = \frac{\sum_j \sum_k (a_{ij} a_{jk} a_{ki})^{1/3}}{k_i(k_i - 1)} = \frac{\sum_j \sum_k a_{ij} a_{jk} a_{ki}}{k_i(k_i - 1)}. \tag{A.6}$$

(4) Barrat et al.'s [8]

$$C_{w,i}^B = \frac{1}{s_i(k_i - 1)} \sum_{j,k} \frac{w_{ij} + w_{ik}}{2} a_{ij} a_{jk} a_{ki} \tag{A.7}$$

reduces to the unweighted definition (1.3), when w_{ij} and w_{ik} are substituted by the adjacency matrix elements:

$$s_i = \sum_{j \in \Pi(i)} w_{ij} = \sum_{j \in \Pi(i)} a_{ij} = k_i, \qquad a_{ij}^2 = a_{ij},$$

$$C_{w,i}^B = \frac{1}{k_i(k_i - 1)} \sum_{j,k} \frac{a_{ij} + a_{ik}}{2} a_{ij} a_{jk} a_{ki} = \frac{1}{k_i(k_i - 1)} \sum_{j,k} \frac{a_{ij} a_{ij} a_{jk} a_{ki} + a_{ik} a_{ij} a_{jk} a_{ki}}{2}$$

$$= \frac{1}{k_i(k_i - 1)} \sum_{j,k} \frac{a_{ij}^2 a_{jk} a_{ki} + a_{ij} a_{jk} a_{ki}^2}{2} = \frac{1}{k_i(k_i - 1)} \sum_{j,k} \frac{a_{ij} a_{jk} a_{ki} + a_{ij} a_{jk} a_{ki}}{2} \tag{A.8}$$

$$= \frac{1}{k_i(k_i - 1)} \sum_{j,k} a_{ij} a_{jk} a_{ki}.$$

(5) Serrano et al.'s [17] formula can be expressed as

$$C_{w,i}^S = \frac{\sum_j \sum_k w_{ij} w_{ik} a_{kj}}{s_i^2(1 - Y_i)} = \frac{\sum_j \sum_k w_{ij} w_{ik} a_{kj}}{s_i^2(1 - \sum_j (w_{ij}/s_i)^2)} = \frac{\sum_j \sum_k w_{ij} w_{ik} a_{kj}}{s_i^2(1 - (1/s_i^2) \sum_j w_{ij}^2)}$$

$$= \frac{\sum_j \sum_k w_{ij} w_{ik} a_{kj}}{s_i^2 - \sum_j w_{ij}^2} = \frac{\sum_j \sum_k w_{ij} w_{ik} a_{kj}}{(\sum_j w_{ij})^2 - \sum_j w_{ij}^2}. \tag{A.9}$$

It is obvious that the formula reduces to the unweighted (1.3), when w_{jk} are substituted by a_{jk}.

B. The values of the weighted clustering coefficients of some node i participating in a fully connected triangle

We calculate the weighted clustering coefficient of node 1.

(1) Zhang and Horvath's is [13]

$$
C_{w,1}^Z = \frac{\sum_{j\neq 1}^3 \sum_{k\neq 1}^3 w_{1j}w_{jk}w_{k1}}{\left(\sum_{j\neq 1}^3 w_{1j}\right)^2 - \sum_{j\neq 1}^3 w_{1j}^2} = \frac{\sum_{j\neq 1}^3 w_{1j}\left(w_{j2}w_{21} + w_{j3}w_{31}\right)}{\left(w_{12} + w_{13}\right)^2 - w_{12}^2 - w_{13}^2}
$$

$$
= \frac{w_{12}w_{23}w_{31} + w_{13}w_{32}w_{21}}{2w_{12}w_{13}} = \frac{2w_{12}w_{23}w_{31}}{2w_{12}w_{13}} = w_{23}.
$$

(B.1)

(2) Lopez-Fernandez et al.'s is [15]

$$
C_{w,1}^L = \frac{\sum_{k\neq 1}^3 \sum_{j\neq 1}^3 w_{jk}}{k_1\left(k_1 - 1\right)} = \frac{\sum_{k\neq 1}^3 \left(w_{2k} + w_{3k}\right)}{2(2-1)} = \frac{w_{23} + w_{32}}{2} = w_{23}.
$$

(B.2)

(3) Onnela et al.'s is [16]

$$
C_{w,1}^O = \frac{\sum_j \sum_k \left(w_{1j}w_{jk}w_{k1}\right)^{1/3}}{k_1\left(k_1 - 1\right)} = \frac{\sum_j \left(w_{1j}\right)^{1/3}\sum_k \left(w_{jk}w_{k1}\right)^{1/3}}{2(2-1)}
$$

$$
= \frac{1}{2}\sum_j \left(w_{1j}\right)^{1/3}\left(\left(w_{j2}w_{21}\right)^{1/3} + \left(w_{j3}w_{31}\right)^{1/3}\right)
$$

$$
= \frac{1}{2}\left[\left(w_{12}\right)^{1/3}\left(w_{23}w_{31}\right)^{1/3} + \left(w_{13}\right)^{1/3}\left(w_{32}w_{21}\right)^{1/3}\right]
$$

$$
= \frac{1}{2}\left[\left(w_{12}w_{23}w_{31}\right)^{1/3} + \left(w_{13}w_{32}w_{21}\right)^{1/3}\right] = \left(w_{12}w_{23}w_{31}\right)^{1/3} \leq 1,
$$

$$
C_{w,1}^O = C_{w,2}^O = C_{w,3}^O = \left(w_{12}w_{23}w_{31}\right)^{1/3} \leq 1.
$$

(B.3)

(4) Barrat et al.'s is [8]

$$
C_{w,i}^B = \frac{1}{s_i(k_i - 1)}\sum_{j,k}\frac{w_{ij} + w_{ik}}{2}a_{ij}a_{jk}a_{ki}.
$$

(B.4)

Degree of node 1: $k_1 = \sum_{j\in\Pi(1)} a_{1j} = 2$.

Strength of node 1: $s_1 = \sum_{j\in\Pi(1)} w_{1j} = w_{12} + w_{13}$;

$$
C_{w,1}^B = \frac{1}{s_1(k_1 - 1)}\sum_{j,k}\frac{w_{1j} + w_{1k}}{2}a_{1j}a_{jk}a_{k1}
$$

$$
= \frac{1}{s_1(2-1)}\sum_j\left(\frac{w_{1j} + w_{12}}{2}a_{1j}a_{j2}a_{21} + \frac{w_{1j} + w_{13}}{2}a_{1j}a_{j3}a_{31}\right)
$$

$$
= \frac{1}{s_1}\left(\frac{w_{13} + w_{12}}{2}a_{13}a_{32}a_{21} + \frac{w_{12} + w_{13}}{2}a_{12}a_{23}a_{31}\right)
$$

$$
= \frac{1}{w_{12} + w_{13}}\left(w_{12} + w_{13}\right)a_{12}a_{23}a_{31} = a_{12}a_{23}a_{31} = 1,
$$

(B.5)

since $a_{12} = a_{23} = a_{31} = 1$.

We also prove that Barrat et al.'s definition for the weighted clustering coefficient is independent of all weights for all fully connected networks;

$$
\begin{aligned}
C_{w,i}^{B} &= \frac{1}{s_i(k_i - 1)} \sum_{j,h} \frac{w_{ij} + w_{ih}}{2} a_{ij} a_{jh} a_{hi} = \frac{1}{s_i(k_i - 1)} \sum_{j}^{k_i} \sum_{h}^{k_i} \frac{w_{ij} + w_{ih}}{2} a_{ij} a_{jh} a_{hi} \\
&= \frac{1}{s_i(k_i - 1)} \sum_{h}^{k_i} \left(\frac{w_{i1} + w_{ih}}{2} a_{i1} a_{1h} a_{hi} + \frac{w_{i2} + w_{ih}}{2} a_{i2} a_{2h} a_{hi} + \cdots + \frac{w_{ik_i} + w_{ih}}{2} a_{ik_i} a_{k_i h} a_{hi} \right) \\
&= \frac{1}{s_i(k_i - 1)} \left[\left(\frac{w_{i1} + w_{i1}}{2} a_{i1} a_{11} a_{1i} + \frac{w_{i2} + w_{i1}}{2} a_{i2} a_{21} a_{1i} + \cdots + \frac{w_{ik_i} + w_{i1}}{2} a_{ik_i} a_{k_i 1} a_{1i} \right) \right. \\
&\quad + \left(\frac{w_{i1} + w_{i2}}{2} a_{i1} a_{12} a_{2i} + \frac{w_{i2} + w_{i2}}{2} a_{i2} a_{22} a_{2i} + \cdots + \frac{w_{ik_i} + w_{i2}}{2} a_{ik_i} a_{k_i 2} a_{2i} \right) + \cdots \\
&\quad \left. + \left(\frac{w_{i1} + w_{ik_i}}{2} a_{i1} a_{1k_i} a_{k_i i} + \frac{w_{i2} + w_{ik_i}}{2} a_{i2} a_{2k_i} a_{k_i i} + \cdots + \frac{w_{ik_i} + w_{ik_i}}{2} a_{ik_i} a_{k_i k_i} a_{k_i i} \right) \right].
\end{aligned}
$$

(B.6)

For a fully connected network: $a_{ij} = 1, \forall\, i, j = 1, 2, \ldots, k_i$ and $a_{ii} = 0$, so

$$
\begin{aligned}
C_{w,i}^{B} &= \frac{1}{s_i(k_i - 1)} \left[\left(0 + \frac{w_{i2} + w_{i1}}{2} + \cdots + \frac{w_{ik_i} + w_{i1}}{2} \right) \right. \\
&\quad + \left(\frac{w_{i1} + w_{i2}}{2} + 0 + \cdots + \frac{w_{ik_i} + w_{i2}}{2} \right) + \cdots \\
&\quad \left. + \left(\frac{w_{i1} + w_{ik_i}}{2} + \frac{w_{i2} + w_{ik_i}}{2} + \cdots + 0 \right) \right] \\
&= \frac{1}{s_i(k_i - 1)} \left[\left(\frac{w_{i1} + w_{i2} + \cdots + w_{ik_i}}{2} + (k_i - 2) \frac{w_{i1}}{2} \right) \right. \\
&\quad + \left(\frac{w_{i1} + w_{i2} + \cdots + w_{ik_i}}{2} + (k_i - 2) \frac{w_{i2}}{2} \right) + \cdots \\
&\quad \left. + \left(\frac{w_{i1} + w_{i2} + \cdots + w_{ik_i}}{2} + (k_i - 2) \frac{w_{ik_i}}{2} \right) \right] \\
&= \frac{1}{s_i(k_i - 1)} \left(k_i \frac{w_{i1} + w_{i2} + \cdots + w_{ik_i}}{2} + (k_i - 2) \frac{w_{i1} + w_{i2} + \cdots + w_{ik_i}}{2} \right) \\
&= \frac{1}{s_i(k_i - 1)} (2k_i - 2) \frac{w_{i1} + w_{i2} + \cdots + w_{ik_i}}{2} = \frac{w_{i1} + w_{i2} + \cdots + w_{ik_i}}{s_i} = 1,
\end{aligned}
$$

(B.7)

since $s_i = w_{i1} + w_{i2} + \cdots + w_{ik_i}$.

(5) Serrano et al.'s is [17]

$$
\begin{aligned}
C_{w,1}^{S} &= \frac{\sum_{j \neq 1}^{3} \sum_{k \neq 1}^{3} w_{1j} a_{jk} w_{k1}}{\left(\sum_{j \neq 1}^{3} w_{1j} \right)^2 - \sum_{j \neq 1}^{3} w_{1j}^2} = \frac{\sum_{j \neq 1}^{3} w_{1j} (a_{j2} w_{21} + a_{j3} w_{31})}{(w_{12} + w_{13})^2 - w_{12}^2 - w_{13}^2} \\
&= \frac{w_{12} a_{23} w_{31} + w_{13} a_{32} w_{21}}{2 w_{12} w_{13}} = \frac{2 w_{12} a_{23} w_{31}}{2 w_{12} w_{13}} = 1.
\end{aligned}
$$

(B.8)

C. Trend analysis. Linear regression

The trend analysis was performed with SPSS [21], using the least squares method. We also tried quadratic and cubic fitting but the nonlinear regression gave zero for the coefficients corresponding to the quadratic and cubic terms for all cases of networks and weighted clustering coefficients.

We observe the presence of decreasing linear correlation between the weighted clustering coefficients and the relative perturbation norm for all distributions (homogeneous and symmetric (uniform, normal) and inhomogeneous and nonsymmetric (exponential)).

Acknowledgments

References

[1] I. Prigogine, *From Being to Becoming*, Freeman, New York, NY, USA, 1980.

[2] A.-L. Barabasi, *Linked: The New Science of Networks*, Perseus, Cambridge, Mass, USA, 2002.

[3] S. N. Dorogovtsev and J. F. F. Mendes, "Evolution of networks," *Advances in Physics*, vol. 51, no. 4, pp. 1079–1187, 2002.

[4] M. E. J. Newman, "The structure and function of complex networks," *SIAM Review*, vol. 45, no. 2, pp. 167–256, 2003.

[5] D. J. Watts, *Six Degrees: The Science of a Connected Age*, W. W. Norton, New York, NY, USA, 2003.

[6] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[7] A.-L. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[8] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani, "The architecture of complex weighted networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 11, pp. 3747–3752, 2004.

[9] M. Barthélemy, A. Barrat, R. Pastor-Satorras, and A. Vespignani, "Characterization and modeling of weighted networks," *Physica A*, vol. 346, no. 1-2, pp. 34–43, 2005.

[10] M. Meiss, F. Menczer, and A. Vespignani, "On the lack of typical behavior in the global Web traffic network," in *Proceedings of the 14th International Conference on World Wide Web*, pp. 510–518, ACM, Chiba, Japan, May 2005.

[11] H. P. Thadakamalla, R. Albert, and S. R. T. Kumara, "Search in weighted complex networks," *Physical Review E*, vol. 72, no. 6, Article ID 066128, 8 pages, 2005.

[12] X.-J. Xu, Z.-X. Wu, and Y.-H. Wang, "Properties of weighted complex networks," *International Journal of Modern Physics C*, vol. 17, no. 4, pp. 521–529, 2006.

[13] B. Zhang and S. Horvath, "A general framework for weighted gene co-expression network analysis," *Statistical Applications in Genetics and Molecular Biology*, vol. 4, no. 1, article 17, 2005.

[14] G. Kalna and D. J. Higham, "Clustering coefficients for weighted networks," Tech. Rep. 3, University of Strathclyde Mathematics, Glasgow, UK, 2006.

[15] L. Lopez-Fernandez, G. Robles, and J. M. Gonzalez-Barahona, "Applying social network analysis to the information in CVS repositories," in *Proceedings of the 1st International Workshop on Mining Software Repositories (MSR '04)*, pp. 101–105, Edinburgh, Scotland, UK, May 2004.

[16] J.-P. Onnela, J. Saramäki, J. Kertész, and K. Kaski, "Intensity and coherence of motifs in weighted complex networks," *Physical Review E*, vol. 71, no. 6, Article ID 065103, 4 pages, 2005.

[17] M. A. Serrano, M. Boguñá, and R. Pastor-Satorras, "Correlations in weighted networks," *Physical Review E*, vol. 74, no. 5, Article ID 055101, 4 pages, 2006.

[18] P. Holme, S. M. Park, B. J. Kim, and C. R. Edling, "Korean university life in a network perspective: dynamics of a large affiliation network," *Physica A*, vol. 373, pp. 821–830, 2007.

[19] M. Li, Y. Fan, J. Chen, L. Gao, Z. Di, and J. Wu, "Weighted networks of scientific communication: the measurement and topological role of weight," *Physica A*, vol. 350, no. 2–4, pp. 643–656, 2005.

[20] V. Batagelj and A. Mrvar, Pajek. http://vlado.fmf.uni-lj.si/pub/networks/pajek/.

[21] http://www.spss.com/spss/.