Clustering in weighted networks[☆]

Tore Opsahl[*]

Pietro Panzarasa

Queen Mary University of London, School of Business and Management,
Mile End Road, E1 4NS London, UK

Abstract

In recent years, researchers have investigated a growing number of weighted networks where ties are differentiated according to their strength or capacity. Yet, most network measures do not take weights into consideration, and thus do not fully capture the richness of the information contained in the data. In this paper, we focus on a measure originally defined for unweighted networks: the global clustering coefficient. We propose a generalization of this coefficient that retains the information encoded in the weights of ties. We then undertake a comparative assessment by applying the standard and generalized coefficients to a number of network datasets.

Key words: clustering, transitivity, weighted networks

[☆]We wish to give very special thanks to Filip Agneessens, Stephen Borgatti, Carter Butts, and Tom Snijders for their valuable feedback on earlier versions of this paper. We are also grateful to participants of the 3rd Conference on Applications of Social Network Analysis, the 27th Sunbelt Social Networks Conference, and NetSci 2008 for their helpful comments and suggestions.

[*]Corresponding author. Tel.: +44-20-7882-6984; fax: +44-20-7882-3615.

Email addresses: t.opsahl@qmul.ac.uk (Tore Opsahl), p.panzarasa@qmul.ac.uk (Pietro Panzarasa)

1. Introduction

While a substantial body of recent research has investigated the topological features of a variety of networks (Barabási et al., 2002; Ingram and Roberts, 2004; Kossinets and Watts, 2006; Uzzi and Spiro, 2005; Watts and Strogatz, 1998), relatively little work has been conducted that moves beyond merely topological measures to take explicitly into account the heterogeneity of ties (or edges) connecting nodes (or vertices) (Barrat et al., 2004). In a number of real-world networks, ties are often associated with weights that differentiate them in terms of their strength, intensity or capacity (Barrat et al., 2004; Wasserman and Faust, 1994). On the one hand, Granovetter (1973) argued that the strength of social relationships in social networks is a function of their duration, emotional intensity, intimacy, and exchange of services. On the other, for non-social networks, weights often refer to the function performed by ties, e.g., the carbon flow (mg/m^2/day) between species in food webs (Luczkowich et al., 2003; Nordlund, 2007), the number of synapses and gap junctions in neural networks (Watts and Strogatz, 1998), or the amount of traffic flowing along connections in transportation networks (Barrat et al., 2004). In order to fully capture the richness of the data, it is therefore crucial that the measures used to study a network incorporate the weights of the ties.

A measure that has long received much attention in both theoretical and empirical research is concerned with the degree to which nodes tend to cluster together. Evidence suggests that in most real-world networks, and especially social networks, nodes tend to cluster into densely connected groups (Feld, 1981; Friedkin, 1984; Holland and Leinhardt, 1970; Louch, 2000;

Simmel, 1923; Snijders, 2001; Snijders et al., 2006; Watts and Strogatz, 1998). In particular, the problem of network clustering can be investigated from a two-fold perspective. On the one hand, it involves determining whether and to what extent clustering is a property of a network or, alternatively, whether nodes tend to be members of tightly knit groups (Luce and Perry, 1949). On the other, it is concerned with the identification of the groups of nodes into which a network can be partitioned. This can be obtained, for example, by applying algorithms for community detection that assess and compare densities within and between groups (Newman, 2006; Newman and Girvan, 2004; Roswall and Bergstrom, 2008), or by using the image matrix in blockmodeling for grouping nodes with the same or similar patterns of ties and uncovering connections between groups of nodes (Doreian et al., 2005).

In this paper, we focus our attention only on the problem of determining whether clustering is a property of a network. More specifically, to address this problem one may ask: If there are three nodes in a network, i, j, and k, and i is connected to j and k, how likely is it that j and k are also connected with each other? In real-world networks, empirical studies have shown that this likelihood tends to be greater than the probability of a tie randomly established between two nodes (Barabási et al., 2002; Davis, Yoo and Baker, 2003; Ebel et al., 2002; Holme et al., 2004; Ingram and Roberts, 2004; Newman, 2001; Uzzi and Spiro, 2005; Watts and Strogatz, 1998). For social networks, scholars have investigated the mechanisms that are responsible for the increase in the probability that two people will be connected if they share a common acquaintance (Holland and Leinhardt, 1971; Simmel, 1923; Snijders, 2001; Snijders et al., 2006). The nature of these mechanisms can be cognitive, as in the case of individuals' desire to maintain balance among ties with others (Hallinan, 1974; Heider, 1946), social, as in the case of third-part referral (Davis, 1970), or can be explained in other ways, such as in terms of focus constraints (Feld, 1981; Kossinets and Watts, 2006; Louch, 2000; Monge et al., 1985) or the differing popularity among individuals (Feld and Elmore, 1982a,b). While clustering is likely to result from a combinations of all these mechanisms, network studies have offered no conclusive theoretical explanation of its causes, nor have they concentrated as much on its underpinning processes as on the measures to formally detect its presence in real-world networks (Levine and Kurzban, 2006).

Traditionally, the two main measures developed for testing the tendency of nodes to cluster together into tightly knit groups are the local clustering coefficient (Watts and Strogatz, 1998) and the global clustering coefficient (Feld, 1981; Karlberg, 1997, 1999; Louch, 2000; Newman, 2003). The local clustering coefficient is based on ego's network density or local density (Scott, 2000; Wasserman and Faust, 1994). For node i, this is measured as the fraction of the number of ties connecting i's neighbors over the total number of possible ties between i's neighbors. To create an overall local coefficient for the whole network, the individual fractions are averaged across all nodes.

Despite its ability to capture the degree of social embeddedness that characterizes the nodes of a network, nonetheless the local clustering coefficient suffers from a number of limitations. First, in its original formulation, it does not take into consideration the weights of the ties in the network. As a result, the same value of the coefficient might be attributed to networks that share the same topology but differ in terms of how weights are distributed across ties and, as a result, may be characterized by different likelihoods to befriend the friends of one's friends. Second, the local clustering coefficient does not take into consideration the directionality of the ties connecting a node to its neighbors (Wasserman and Faust, 1994).[1] Recently, there have been a number of attempts to extend the local clustering coefficient to the case of weighted networks (Barrat et al., 2004; Lopez-Fernandez, Robles and Gonzalez-Barahona, 2004; Onnela et al., 2005; Zhang and Horvath, 2005). However, the issue of directionality still remains mainly unresolved (Caldarelli, 2007), thus making the coefficient suitable primarily for undirected networks.

Moreover, the local clustering coefficient, even in its weighted version, is biased by correlations with nodes' degrees: a node with more neighbors is likely to be embedded in relatively fewer closed triplets, and therefore have a smaller local clustering than a node connected to fewer neighbors (Ravasz et al., 2002; Ravasz and Barabási, 2003). An additional bias might stem from degree-degree correlations. When nodes preferentially connect to others with similar degree, local clustering is positively correlated with nodes' degree (Ravasz and Barabási, 2003; Ravasz et al., 2002; Soffer and Vàzquez, 2005). Lack of comparability between values of clustering of nodes with different degrees thus makes the average value of local clustering sensitive with respect to how degrees are distributed across the whole network.

Unlike the local clustering coefficient, the global clustering coefficient is

[1]Node i's neighbor might be: 1) a node that has directed a tie toward i; 2) a node to which i has directed a tie; or 3) a node that has directed a tie toward i and, at the same time, has received a tie from i.

based on transitivity, which is a measure used to detect the fraction of triplets that are closed in directed networks (Wasserman and Faust, 1994, pg. 243). It is not an average of individual fractions calculated for each node, and, as a result, it does not suffer from the same type of correlations with nodes' degrees as the local coefficient. Despite its merits, however, in its original formulation, the global coefficient applies only to networks where ties are unweighted. To address this limitation, and make the coefficient suitable also to networks where ties are weighted, researchers have typically introduced an arbitrary cut-off level of the weight, and then dichotomized the network by removing ties with weights that are below the cut-off, and then setting the weights of the remaining ties equal to one (Doreian, 1969; Wasserman and Faust, 1994). The outcome of this procedure is a binary network consisting of ties that are either present (i.e., equal to 1) or absent (i.e., equal to 0) (Scott, 2000; Wasserman and Faust, 1994). For example, Doreian (1969) studied clustering in a weighted network by creating a series of binary networks from the original weighted network using different cut-offs. To address potential problems arising from the subjectivity inherent in the choice of the cut-off, a sensitivity analysis was conducted to assess the degree to which the value of clustering varies depending on the cut-off (Doreian, 1969). However, this analysis tells us little about the original weighted network, apart from the fact that the value of clustering changes at different levels of the cut-off.

In this paper, we focus on the global clustering coefficient, and propose a generalization that explicitly takes weights of ties into consideration and, for this reason, does not depend on a cut-off to dichotomize weighted networks. In what follows, we start by discussing the existing literature on the global clustering coefficient in undirected and unweighted networks. In Section 3, we propose our generalized measure of clustering. We then turn our attention to directed networks, and discuss the current literature on clustering in those networks. We extend our generalized measure of clustering to cover weighted and directed networks. In Section 5, we empirically test our proposed measure, and compare it with the standard one, by using a number of weighted network datasets. Finally, in Section 6 we summarize and discuss the main results.

2. Clustering coefficient

The global clustering coefficient is concerned with the density of triplets of nodes in a network. A triplet can be defined as three nodes that are connected by either two (open triplet) or three (closed triplet) ties. A triangle consists of three closed triplets, each centered on one node. The global clustering coefficient is defined as the number of closed triplets (or 3 $\times$ triangles) over the total number of triplets (both open and closed). The first attempt to measure the coefficient was made by Luce and Perry (1949). For an undirected network, they showed that the total number of triplets could be found by summing the non-diagonal cells of a squared binary matrix. The number of closed triplets could be found by summing the diagonal of a cubed matrix. For clarity, we will refer to the global clustering coefficient as the *standard* clustering coefficient C:

$$C = \frac{3 \ \times \text{number of triangles}}{\text{number of triples}} = \frac{\sum \tau_\Delta}{\sum \tau} \tag{1}$$

where $\sum \tau$ is the total number of triplets and $\sum \tau_{\Delta}$ is the subset of these triplets that are closed as a result of the addition of a third tie. The coefficient takes values between 0 and 1. In a completely connected network, $C = 1$ as all triplets are closed, whereas in a classical random network $C \to 0$ as the network size grows. More specifically, in a classical random network, the probabilities that pairs of nodes have of being connected are, by definition, independent (Erdős and Rényi, 1959; Solomonoff and Rapoport, 1951). Therefore, C is equal to the probability of a tie in these networks (Newman, 2003).

A major limitation of the clustering coefficient is that it cannot be applied to weighted networks. As a result, the same outcome might be attributed to networks that differ in terms of distribution of weights and that, for this reason, might be characterized by different likelihoods of one's neighbors being connected with each other. This limitation could therefore bias the analysis of the network structure. In order to overcome this shortcoming, in the following section we will propose a generalization of the clustering coefficient that explicitly captures the richness of the weights attached to ties, while at the same time it produces the same results as the standard clustering coefficient when ties are unweighted.

3. Generalized Clustering Coefficient

We can generalize the clustering coefficient, C, to take weights of ties into consideration, by rewriting Equation 1 in terms of a triplet value, ω. To this end, it is vital to choose an appropriate method for defining the triplet value as this impacts on the value of the coefficient. The method should be chosen based on the research question, and should reflect the way in which the weights of ties are defined. First, the triplet value, ω, can be defined as the arithmetic mean of the weights of the ties that make up the triplet. This is the simplest method of calculating the triplet value. However, this method does not take into account differences between weights, and is not robust against extreme values of weights. Second, ω can be defined as the geometric mean of the weights of ties. This method overcomes some of the sensitivity issues of the arithmetic mean. A triplet made up by a tie with a low weight and a tie with a high weight value will have a lower value by using the geometric mean than would be the case if the arithmetic mean were used. Third, ω can be defined as the maximum or minimum value of the weights of the ties. These two methods, however, represent extreme ones. On the one hand, using the maximum weight makes ω insensitive with respect to small weights. As a result, two triplets, one with a strong tie and a weak tie and the other with two strong ties, may well be assigned the same value. Conversely, using the minimum weight makes ω insensitive with respect to large values of weights. In this case, two triplets, one consisting of a strong tie and weak tie and the other of two weak ties may well be assigned the same value. The advantages and shortcomings of each of these four methods should be evaluated based on the research question and the type of network dataset at hand. For example, in a network where the weights corresponds to the level of flow, and a weak tie would act as a bottleneck, the minimum method might be most appropriate to use. By contrast, when ties are weighted in terms of costs or time, it may be more suitable to apply the maximum method so as not to underestimate the values of triplets. Table 1 highlights the differences between the methods for defining triplet values. We will explore them further at the end of Section 5.

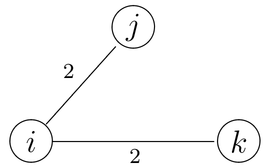
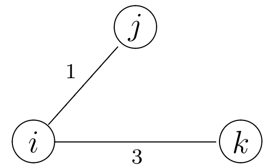
	Triplet value ω of	
Method		
Arithmetic mean	$(2+2)/2 \quad = 2$	$(1+3)/2 \quad = 2$
Geometric mean	$\sqrt{2 \times 2} \quad = 2$	$\sqrt{1 \times 3} \quad = 1.73$
Maximum	$\max(2,2) \quad = 2$	$\max(1,3) \quad = 3$
Minimum	$\min(2,2) \quad = 2$	$\min(1,3) \quad = 1$

Table 1: Methods for calculating the triplet value, ω.

Formally, we propose to generalize the clustering coefficient as follows:

$$C_\omega = \frac{\textit{total value of } \text{closed triplets}}{\textit{total value of } \text{triplets}} = \frac{\sum_{\tau_\Delta} \omega}{\sum_{\tau} \omega} \tag{2}$$

The generalized clustering coefficient produces the same result as the standard version C when it is applied to a binary network. This is because all triplets have the same value ($\omega = 1$), irrespective of the method used to calculate triplet values. In addition, the generalized coefficient has the same properties as C. It still ranges between 0 and 1 because neither numerator nor denominator of the fraction can be negative. Moreover, all weights that are part of the numerator are also part of the denominator. In a completely connected network, all triplets are closed as the third tie will always be present (e.g., between node j and node k in Table 1). Therefore, the same triplets are part of both the numerator and denominator, and thus $C_\omega = \frac{1}{1} = 1$. To test whether $C_\omega \to 0$ as the size of a classical random network increases or, more specifically, whether C_ω equals the probability that two randomly chosen nodes are connected with each other, we created a set of networks of different size, but with a fixed average degree. Since classical random networks are binary, we assigned to each tie a random weight between 1 and 10 to create weighted random networks. We applied the generalized clustering coefficient to these networks, and found that $C_\omega \to 0$ as the network size increases. In particular, we found that C_ω is very close to the probability of a tie in classical random networks.[2] Furthermore, to assess sensitivity with respect to weights, we tested the generalized clustering coefficient on networks where the structure was kept invariant, but where the weights were randomly assigned to the ties. We found $C_\omega \approx C_{GT0}$, where C_{GT0} is C calculated on networks dichotomized by setting equal to 1 all weights that are greater than 0.[3]

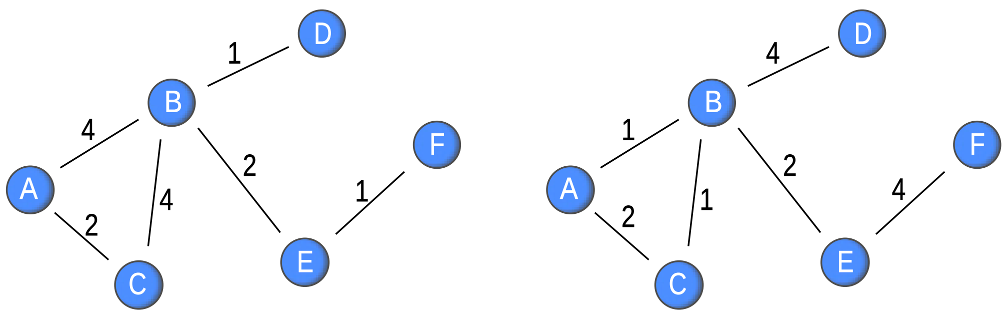
In this paper, we assume that weights can take on only positive values. Moreover, we will use the absolute values of weights without normalizing them (e.g., by dividing them by their maximum or average) as this would have no effect on the results of our analysis. This is due to the fact that in Equation 2 the total value of closed triplets is divided by the total value of all triplets. In addition, our generalizied clustering coefficient is applicable primarily to networks in which weights are measured on a ratio scale. When

[2]These findings are based on ensembles of classical random networks with 50, 100, 200, 400, 800, and 1,600 nodes and an average degree of 10. Each ensemble contains 1,000 realizations. The four methods for defining triplet values did not lead to significantly different results.

[3]This finding is based on the empirical networks presented in Section 5. For each network, we reshuffled the weights among the ties (1,000 realizations), and found that C_ω was not statistically significantly different from C_{GT0}. C_{GT0} was calculated on networks where all tie weights with positive values were set to present, i.e., equal to 1.

weights are measured on an ordinal scale (e.g., in social networks where weights represent the ranks of different levels of friendship), special care should be taken when assessing the value of the coefficient, because the same differences between weights may not have the same meaning. In this case, it would be advisable to transform the ordinal scale into a ratio scale.

To illustrate the applicability of the generalized clustering coefficient, Figure 1 shows two sample networks, each with six nodes and six weighted ties. In network a, the ties between the nodes that form the triangle have higher weights than the average tie weight in the network, whereas the reverse is true in network b.



(a) Network where stronger triplets are closed

(b) Network where weaker triplets are closed

Figure 1: Two weighted networks.

Both networks have the same clustering coefficient C_{GT0} when they are transformed by setting ties with weights greater than 0 to present:

$$C_{GT0} = \frac{3 \times 1}{9} = 0.33 \tag{3}$$

However, if, for example, the two sample networks represented social networks in which ties refer to friendship between individuals, we believe that it would not be accurate to claim that both these networks show the same tendency of one's friends to be friends themselves. Being friends refers to a social relationship that can be assessed by using the same criteria (duration, emotional intensity, intimacy, and exchange of services) that Granovetter (1973) proposed for classifying tie weights. The generalized clustering coefficient helps highlight the difference between the two sample networks. More specifically, for networks a and b in Figure 1, the generalized clustering coefficients obtained by using the geometric mean method (gm) for defining triplet values are, respectively:

$$C_{\omega,gm} = 0.44 \tag{4a}$$

$$C_{\omega,gm} = 0.23 \tag{4b}$$

The difference in values stems from the fact that the generalized clustering coefficient captures more information than C_{GT0}. In particular, the difference between 4a and 4b reflects the differences in tie weights in the two sample networks. If, for example, the tie weights in the sample networks were to represent duration, we might reasonably argue that the nodes in network a are investing more time on interactions with other nodes that are themselves connected than is the case with network b.

Following Barrat et al. (2004), in our proposed generalization of the clustering coefficient we do not take into account the weight of the closing tie of a triplet. This is because the aim of the clustering coefficient is to assess the likelihood of the occurrence of a tie that closes a triplet, and not the strength of this tie. A triplet must be created prior to the closing tie. In other words, as networks evolve over time by the creation and removal of ties, clustering occurs when a triplet exists, and a newly created third tie closes the triplet. Nevertheless, when we observe a triangle in a cross-sectional network dataset, we do not know which of the three triplets that make up the triangle occurred in the first place. In effect, this means that the weight of the closing tie of a triplet is taken into account since it is part of the values of the other two triplets in the triangle.

4. Directed networks

In directed networks, connections between nodes are described as ties that originate from one node and point toward another (Wasserman and Faust, 1994). The weight of a tie directed from node i to node j is expressed as x_{ij}. In a binary network, the weight of a present tie is set equal to 1, whereas the weight of an absent tie is 0. We define the triplet consisting of the two directed ties, x_{ji} and x_{ik}, as $\tau_{ji,ik}$, and the value of this triplet as $\omega_{ji,ik}$.

The standard clustering coefficient as stated in Equation 1 cannot be applied to directed data. A more refined measure to calculate closure in directed networks is called transitivity, T (for a review, see Wasserman and Faust, 1994, pg. 243). Transitivity produces the same results as the standard clustering coefficient if applied to an undirected network (Feld, 1981; Newman, 2003). It also shares the same properties. In fact, $0 \leq T \leq 1$: in a completely connected network, we have: $T = 1$; in a classical random network, $T \to 0$ as the network size grows. T takes the direction of the ties between nodes into consideration by using a more sophisticated definition of a triplet. A triplet τ centered on node i must have one incoming and one outgoing tie, i.e., $x_{ki} = x_{ij} = 1$ or $x_{ji} = x_{ik} = 1$, as shown by the solid lines in Figure 2. Wasserman and Faust (1994) termed triplets that do not fulfill the above condition as vacuous. These triplets are not part of the numerator nor of the denominator of the fraction in Equation 1. More specifically, when we are dealing with directed data, there can be four basic configurations of a triplet around an individual node i: $\tau_{ij,ik}$, $\tau_{ij,ki}$, $\tau_{ji,ik}$, and $\tau_{ji,ki}$. The configurations $\tau_{ji,ki}$ and $\tau_{ij,ik}$ form, respectively, an in- and out-star, and therefore are vacuous and not part of the fraction in Equation 1. Conversely, the configurations $\tau_{ij,ki}$ and $\tau_{ji,ik}$ are non-vacuous. These triplets can be either transitive or intransitive.

Figure 2: Non-vacuous triplets centered around node i

Triplets defined according to Wasserman and Faust (1994) form chains of nodes. These triplets have been termed 2-path as they form chains of two directed ties between three nodes (Luce and Perry, 1949). A triplet is transitive if a tie is present from the first node to the last node of the chain. For the two triplets shown in Figure 2, transitivity would imply $x_{kj} = 1$ and $x_{jk} = 1$, respectively.

Transitivity suffers from the same limitation as the standard clustering coefficient in that it cannot be applied to networks where the ties are weighted. To overcome this shortcoming, here we extend our proposed generalization also to directed and weighted networks by using the same definition of a triplet, τ, as transitivity. The triplet value, ω, is calculated by using the same methods as stated in Section 3. This generalization produces the same results as transitivity if applied to binary directed networks, and the same results as the standard clustering coefficient if applied to binary and undirected networks. Moreover, it still ranges between 0 and 1. In a completely connected network, we would still obtain $C_\omega = 1$, whereas in a classical random network, $C_\omega \to 0$ as the network size grows. In particular, once again we found that C_ω approximated the probability of a tie in a classical random network.[4]

To clarify which triplets are transitive and non-vacuous, Table 2 illustrates configurations of triplets centered on node i. The first four rows show the basic configurations mentioned above. The remaining rows show configurations of triplets where ties are reciprocated. In these cases, each additional tie doubles the number of triplets. Moreover, the table shows which triplets are transitive under different conditions, and which triplet values should be included in the fraction of Equation 2.

5. Empirical test of the generalized clustering coefficient

We now test the proposed generalization of clustering on a number of network datasets. We also compare the generalized coefficient with the stan-

[4]These findings are based on ensembles of classical random networks with 50, 100, 200, 400, 800, and $1,600$ nodes and an average degree of 10. Each ensemble contains $1,000$ realizations. The four methods for defining triplet value were not significantly different.

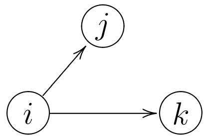
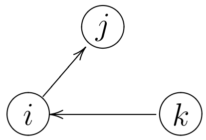
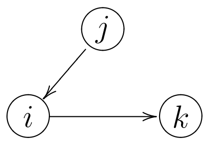
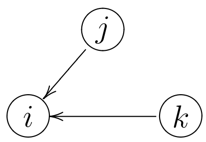
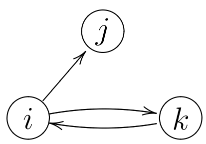
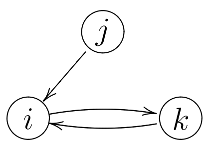
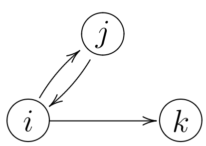
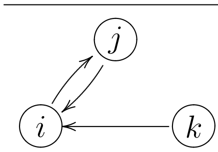
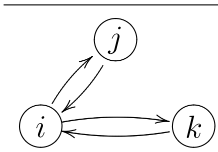
	Triplets	Denominator of Eq. 2	Numerator of Eq. 2			
			if $w_{jk} = 0$ $w_{kj} = 0$	if $w_{jk} > 0$ $w_{kj} = 0$	if $w_{jk} = 0$ $w_{kj} > 0$	if $w_{jk} > 0$ $w_{kj} > 0$
	$\tau_{ij,ik}$	...	...	...	...	...
	$\tau_{ij,ki}$	$\omega_{ij,ki}$	0	0	$\omega_{ij,ki}$	$\omega_{ij,ki}$
	$\tau_{ji,ik}$	$\omega_{ji,ik}$	0	$\omega_{ji,ik}$	0	$\omega_{ji,ik}$
	$\tau_{ji,ki}$	...	...	...	...	...
	$\tau_{ij,ik}$ $\tau_{ij,ki}$	... $\omega_{ij,ki}$	... 0	... 0	... $\omega_{ij,ki}$	... $\omega_{ij,ki}$
	$\tau_{ji,ik}$ $\tau_{ji,ki}$	$\omega_{ji,ik}$ ...	0 ...	$\omega_{ji,ik}$ ...	0 ...	$\omega_{ji,ik}$ ...
	$\tau_{ij,ik}$ $\tau_{ji,ik}$	... $\omega_{ji,ik}$	... 0	... $\omega_{ji,ik}$	... 0	... $\omega_{ji,ik}$
	$\tau_{ij,ki}$ $\tau_{ji,ki}$	$\omega_{ij,ki}$ ...	0 ...	0 ...	$\omega_{ij,ki}$ ...	$\omega_{ij,ki}$ ...
	$\tau_{ij,ik}$ $\tau_{ij,ki}$ $\tau_{ji,ik}$ $\tau_{ji,ki}$	... $\omega_{ij,ki}$ $\omega_{ji,ik}$ ...	... 0 0 ...	... 0 $\omega_{ji,ik}$ ...	... $\omega_{ij,ki}$ 0 ...	... $\omega_{ij,ki}$ $\omega_{ji,ik}$ ...

Table 2: Triplets (τ) and triplet values (ω) in a directed graph. $i \neq j \neq k$.

dard one measured with different cut-offs.[5] Table 3 summarizes the empirical results.

The first dataset we consider is Freeman's EIES networks (Freeman and Freeman, 1979), also used in Wasserman and Faust (1994). This dataset was collected in 1978 and contains three networks of researchers working on social network analysis. The first is an acquaintance network including 48 researchers, and in which relationships were recorded at the beginning of the study (time 1). The second network is similar, but the data were recorded at the end of the study (time 2). The third is a frequency matrix of the number of messages sent among 32 of the researchers that used an electronic communication tool. In the two acquaintance networks, all relationships have a weight between 0 and 4. 4 represents a close personal friend of the researcher's; 3 represents a friend; 2 represents a person the researcher has met; 1 represents a person the researcher has heard of, but not met; and 0 represents a person unknown to the researcher. In the frequency matrix, the average tie weight is 33.7 and the maximum weight is 559. The three networks are highly connected, with densities of 0.34, 0.40, and 0.46, respectively. They also exhibit a fairly large tendency toward clustering: C_{GT0} for the three networks is 0.7627, 0.8131, and 0.6386, respectively. When the proposed generalization of clustering is applied to the three networks, clus-

[5]For the standard clustering coefficient, the networks are dichotomized with different values X of the cut-off. More specifically, C_{GTX} refers to Equation 1 where ties with weights that are greater than X are set to present and ties with weights that are lower than, or equal to, X are removed. Unless otherwise specified, ties are set to present if their weights are greater than 0. Moreover, in our empirical analysis, we adopt the generalized coefficient $C_{\omega,gm}$ that uses the geometric mean method gm for defining triplet value ω. A program to calculate the standard and generalized clustering coefficient using R or Matlab is available upon request from the authors.

network	C_{ω}				C					
	$C_{\omega,am}$	$C_{\omega,gm}$	$C_{\omega,\min}$	$C_{\omega,\max}$	C_{GT0}	C_{GT2}	C_{GT4}	C_{GT6}	C_{GT8}	C_{GT10}
Freeman EIES (time 1)	0.7702	0.7708	0.7722	0.7688	0.7627	0.4027				
Freeman EIES (time 2)	0.8214	0.8218	0.8228	0.8204	0.8131	0.4275				
Freeman EIES (messages)	0.7378	0.7332	0.7250	0.7411	0.6386	0.6251	0.5677	0.5631	0.5377	0.5031
Online community	0.0646	0.0638	0.0626	0.0653	0.0547	0.0360	0.0306	0.0269	0.0287	0.0298
Consulting (advice)	0.7130	0.7168	0.7271	0.7054	0.6932	0.5184	0.4398			
Consulting (value)	0.6852	0.6857	0.6885	0.6827	0.6764	0.6466	0.5478			
Research team (advice)	0.7127	0.7209	0.7383	0.7000	0.6848	0.5534	0.3730			
Research team (awareness)	0.6957	0.6977	0.7064	0.6880	0.6785	0.6532	0.5970			
101^{st} US Congress	0.7630	0.7639	0.7639	0.7627	0.7219	0.4987	0.3649	0.2902	0.2760	0.3216
C.elegans' neural network	0.2403	0.2210	0.2028	0.2518	0.1843	0.1357	0.1182	0.1785	0.1954	0.2936
US airport network*	0.4765	0.5066	0.5366	0.4586	0.3514	0.4125	0.3969	0.3032	0.2940	0.1793

Table 3: Comparison between the generalized and the standard clustering coefficients.
*Due to a large range of tie weights in the US airport network, we have divided the tie weights by $100,000$. This operation has no impact on the generalised coefficient, but it enables us to conduct the sensitivity analysis (i.e., in the original dataset, the minimum tie weight is equal to 17, and therefore a sensitivity analysis with values of the cut-off lower than 17 would be meaningless).

tering increases. More specifically, $C_{\omega,gm}$ takes the value of 0.7708, 0.8218, and 0.7332, respectively. Thus, for the acquaintance networks, clustering increases of 1.1%, whereas for the frequency matrix it shows a relatively higher increase of 14.8%.

Figure 3 shows Freeman's third EIES network, in which the size of a node is proportional to the number of messages sent by the researcher, and the width of a tie between two nodes corresponds to the number of messages exchanged between the two researchers. As shown in the figure, all researchers at the center are connected with one another, whereas this is not the case for researchers located in the outer ring. Moreover, the strongest ties in the network tend to connect the researchers in the center with one another and with nodes at the periphery. This implies that stronger ties are more likely to be part of triangles than weaker ties. For example, Nick Mullins is strongly connected to Sue Freeman and Barry Wellman, who are in turn connected with each other. By contrast, Phipps Arabie is weakly connected to Ev Rogers and Carol Barner-Barry, who are not connected with each other. This tendency of strongly connected researchers to establish a tie with the same third party is responsible for the increased value of clustering when measured with our generalized coefficient.

The second dataset is a network created from an online community (Panzarasa et al., 2009). This network dataset covers the period from April to October 2004. It includes $1,899$ nodes that represent students at the University of California, Irvine. During the observation period, students sent a total number of $59,835$ online messages. A directed tie is established from one student to another if one or more messages have been sent from the for-

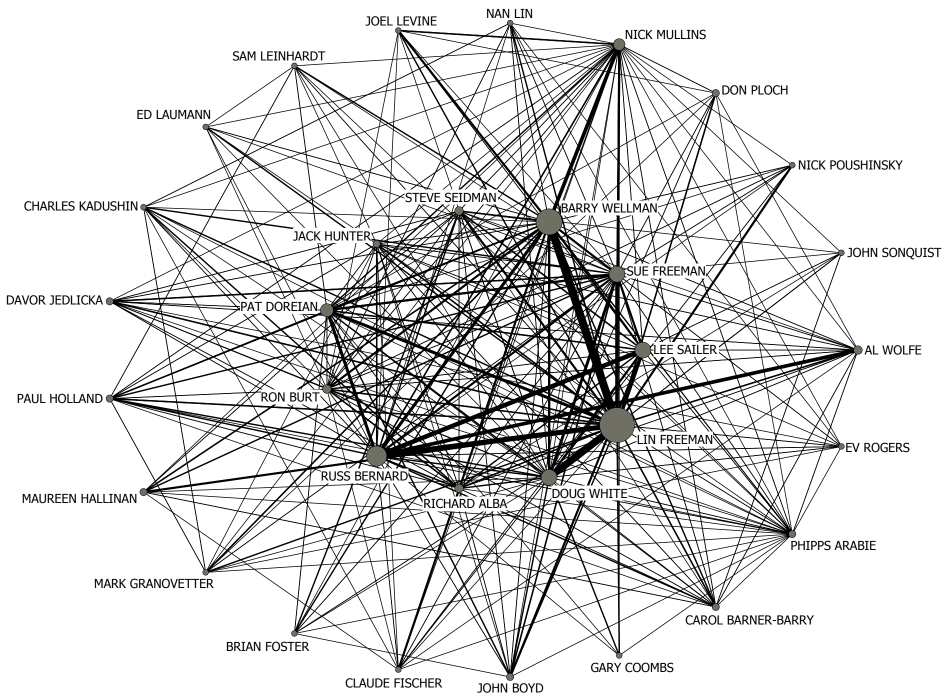


Figure 3: Freeman's third EIES network. The size of a node is proportional to the total number of messages sent by the corresponding scientist, and the width of a tie to the number of messages exchanged among the two connected scientists. All possible ties among the scientists in the inner circle are present.

mer to the latter. The weight of a tie is defined as the number of messages sent. The maximum and average tie weight are 98 and 2.95, respectively. This network exhibits a density of 0.0056, and an average degree of 10.69. In this network, we found $C_{GT0} = 0.0547$ and $C_{\omega,gm} = 0.0638$. Thus, when the generalized coefficient is applied, there is an increase in clustering of 16.8%.

The third dataset contains four organizational networks, two from a consulting company and two from a research team in a manufacturing company (Cross and Parker, 2004).[6] The consulting company had 46 employees that

[6]We thank Andrew Parker at Stanford University for supplying this dataset.

are the nodes in the first two networks. The ties in the first network are differentiated in terms of frequency of information or advice requests, whereas the ties in the second network are differentiated in terms of the value placed on the information or advice received. In both these networks, ties are weighted on a scale from 0 to 5. The company had offices both in Europe and in the US. The US employees were divided into two tightly knit groups, whereas this did not occur with the European employees. The other two networks are concerned with a research team in a manufacturing company. The nodes in these networks are the 77 employees. The ties in the first network are differentiated in terms of advice, whereas in the second network in terms of the employees' awareness of knowledge and skills. In both these networks, ties are weighted on a scale from 0 to 6. Moreover, for both networks, data collection took place after an organizational restructuring operation that combined four separate units in different European countries. The research team was partitioned into strong communities based on the employees' previous geographical location (Cross and Parker, 2004, pg. 15-17). Thus, focus constraint might have been partly responsible for a high value of clustering (Feld, 1981). All four networks do in fact exhibit a high clustering coefficient: C_{GT0} ranges between 0.6764 and 0.6932, and $C_{\omega,gm}$ between 0.6857 and 0.7209. The data thus exhibit an average increase in clustering of 3.2% when the generalised coefficient is applied.

The fourth dataset is a network of political support in the US Senate ((101^{st} Congress, 1989/1990; see Skvoretz, 2002).[7] The network includes 102 nodes that represent senators. Ties among senators reflect co-sponsorship of

[7]We thank John Skvoretz for making this dataset available to us.

bills. This network has a density of 0.58 and an average degree of 59. As the network is well connected, it is difficult to draw conclusions from C_{GT0}. We found: $C_{GT0} = 0.7219$. Weights of ties represent the number of bills that the connected senators have co-sponsored. The average tie weight is 2.68 and the maximum weight is 29. The large difference between the mean and the maximum weights signals that many of the ties are relatively weak. This is an indication that a cut-off higher than zero might be more appropriate for dichotomizing the network. In Table 3, we list C_{GTX} calculated using different values X of the cut-off. When we applied the generalized coefficient, we found: $C_{\omega,gm} = 0.7639$. This represents an increase in clustering of 5.8%. This increase in clustering is likely to be influenced by the fact that party membership and ideologies place a constraint on the strength of ties among senators. In particular, senators belonging to different parties are likely to co-sponsor a limited number of bills, which inevitably affects the total value of closed triplets connecting senators from different parties.

The fifth dataset is the neural network of the Caenorhabditis elegans worm. This network was studied in Watts and Strogatz (1998)[8]. The network contains 306 nodes that represent neurons. A tie joins two neurons if they are connected by either a synapse or a gap junction. The weight of a tie represents the number of these synapses and gap junctions. The average tie weight is 3.74, and the maximum tie weight is 70. The density is 0.0253 and the average degree is 7.7. We found: $C_{GT0} = 0.1843$, and $C_{\omega,gm} = 0.2210$. Thus, the generalized coefficient is 19.9% higher than the binary one.

[8]This dataset was obtained from the Collective Dynamics Group's (Duncan Watts) website: `http://smallworld.sociology.columbia.edu/cdg/datasets/`

The sixth dataset is the network of the 500 busiest commercial airports in the United States (Colizza et al., 2007; Opsahl et al., 2008)[9]. In this network, two airports are connected if a flight was scheduled between them in 2002. The weight of a tie between two airports corresponds to the number of seats available on the scheduled flights. Although air transportation networks are directed by nature, they are also highly symmetric (Barrat et al., 2004). Therefore, we analyse this network as an undirected one. On average, each airport is connected to 11.92 other airports (i.e., density is 0.0239). For the average route, 152,320 seats were scheduled. In this network, the standard and generalised clustering coefficients were well-above the randomly expected value: $C_{GT0} = 0.3514$ and $C_{\omega,gm} = 0.5066$. The generalised coefficient is 44.16% larger than the standard one. This suggests that airports with busy routes are part of transitive triplets.

A number of observations are now in order. First, for all the networks, the standard clustering coefficient, C_{GTX}, generally decreases as the value X of the cut-off increases. However, the rate of decrease differs considerably among the networks. Moreover, for each network, there is variation in the rate of decrease between different values of the cut-off. Despite an average dicreasing trend, we also found that, in certain networks, the clustering coefficient increases in correspondence of increasing levels of the cut-off. In addition, the reliability of the results when large cut-offs are used should be questioned, for in the networks there remain only few triplets and triangles when those cut-offs are used. Thus, these findings from a sensitivity analysis

[9]We thank Vittoria Colizza for making this dataset available: `http://cxnets.googlepages.com/usairtransportationnetwork`

of the standard clustering coefficient do not lend themselves to unequivocal interpretation.

Second, there are variations in the values of the generalized clustering coefficient, C_ω, when different methods for defining the triplet value ω are used. For most of the networks, the highest C_ω is obtained when the minimum method is used, whereas the lowest outcome is obtained when the maximum method is used. Given two triplets with the same average weight, the minimum method assigns a lower value to the triplet with a higher dispersion of weights than to the triplet with a lower dispersion. The reverse is true for the maximum method. Since, for most networks, the minimum (maximum) method produces the highest (lowest) value of clustering, the triplets consisting of ties with a lower (higher) variation in weight are more likely to be closed (open) than the triplets with a larger (lower) variation. This means that triplets consisting of two ties with approximately the same weight are likely to be closed.[10]

Third, for all networks, the generalized clustering coefficient is higher than the standard coefficient. When networks are dichotomized by setting ties with weights greater than 0 to present, the standard clustering coefficient can be used as a benchmark for the generalized one. As shown by simulations

[10] This observation does not apply to three of our networks: Freeman's frequency matrix, the Online Community, and C.elegans' neural network. For these networks, the maximum method is associated with the highest level of C_ω, and vice versa. A possible reason for this is that these networks have a relatively high variation of tie weights. The fact that in these networks $C_{\omega,\max}$ is higher than $C_{\omega,\min}$ signals the tendency of triplets with large variation in weights to be closed. Moreover, variation in tie weights might translate into variation between triplet values, which makes clustering sensitive to individual triplets. For example, in a network with a single extremely strong triplet, the value of the generalized coefficient will depend heavily on whether or not this triplet is closed.

in Section 3, when the weights are reshuffled among the ties, $C_\omega \approx C_{GT0}$. Thus, by comparing C_ω with C_{GT0}, we can assess whether strong triplets are more likely to be closed than weak triplets. More specifically, if the generalized clustering coefficient is significantly higher than the standard clustering coefficient, strong triplets are more likely to be closed than weak ones, whereas if the reverse were the case, weak triplets would be more likely to be closed than strong ones.

6. Conclusions and Discussion

Relationships among unique people are unique. We live in an increasingly connected world with an increasing number of contacts to whom we relate in different ways, with different frequencies, and for different reasons. Each social relationship bears a special meaning to us, and it would be overly simplistic and grossly unfair to treat every contact in the same manner. Therefore, it is important to capture differences among relationships when mapping and studying social networks. In particular, social network measures should reflect the richness of the information that the weights of relationships convey. However, despite the fact that there are a large number of network datasets where the weights of the relationships are recorded (see Section 5, but also Barrat et al., 2004; Ebel et al., 2002; Holme et al., 2004; Kossinets and Watts, 2006; Panzarasa et al., 2009), only a limited number of measures take weights into account (among others, Barrat et al., 2004; Burt, 1992; Freeman et al., 1991; Nordlund, 2007; Opsahl et al., 2008; Yang and Knoke, 2001). Therefore, most measures can only be calculated on network data that are binary.

Among the measures that suffer from this shortcoming is the clustering coefficient. In this paper, we focused on this measure, and offered a generalization that takes the weight of ties explicitly into account by attaching a value to each triplet. The standard coefficient divides the number of closed triplets by the total number of triplets, whereas the generalized coefficient divides the total value of the closed triplets by the total value of all triplets. In particular, the generalized clustering coefficient produces the same result as the standard coefficient when applied to a binary network.

We measured and compared the standard and generalized clustering coefficients on a number of network datasets where the weights of ties are recorded. First, we found that the value of the standard coefficient generally decreases as the value of the cut-off increases. However, as the rate of decrease varies across datasets, it is difficult to interpret this result. Second, we found that there were differences among the outcomes when different methods for defining the triplet value were used. The generalized coefficient based on the minimum method yielded mostly the highest value, whereas when the maximum method was used, the lowest outcome was generally attained. This suggests that similarity in tie weights in a triplet increases the chance of closure of that triplet. Third, we found that, in all social networks studied, the value of the generalized coefficient was greater than the value of the standard one. These findings thus provide support in favour of Granovetter's (1973) claim that in social networks strong ties are more likely to be part of transitive triplets than weak ones.

Being able to produce values of clustering that are positively affected by the tendency of strong ties to be part of transitive triplets is a distinct property of our method as well as an advantage over alternative methods for applying binary measures to weighted networks. For example, we adopted Anhert et al.'s (2007) method for converting a weighted network into an ensamble of binary networks, and calculated the average standard clustering coefficient on these networks. Drawing on Freeman's third EIES network (see Figure 3), we produced $1,000$ binary networks in which the probability of a tie was obtained by dividing its weight by the maximum weight in the network. The average standard clustering coefficient found on this ensamble is 0.1288. This value is not only much lower than the one found with our method (i.e., 0.7332), but also lower than the value obtained with C_{GT0} (i.e., 0.6386).[11] Thus, despite the fact that, as suggested by Figure 3, in Freeman's third EIES network, strong ties tend to be part of transitive triplets, the results obtained by using Anhert et al.'s method would in fact suggest the opposite.

Our generalized clustering coefficient is consistent with the local weighted clustering coefficient proposed by Barrat et al. (2004). For example, in the US airport network, both measures produce values that are higher than the values of the corresponding binary measures. However, the weighted local clustering coefficient is inevitably biased by the fact that it builds explicitly on the local binary coefficient. This is likely to constrain the measure in two ways. First, as the binary measure, the weighted one is not applicable to directed networks. Second, it still suffers from negative correlation between the degree of nodes and their likelihood of being embedded in closed triplets. For example, in the US airport network, we found a negative correlation of

[11] This might be due to the fact that the density of the binary networks tend to be much smaller than the density of the weighted network.

-0.24 between node degree and weighted local clustering. Unlike our global meansure, the weighted local clustering coefficient is therefore affected by the way degrees are distributed across the nodes in a network.

One of the advantages of the generalized clustering coefficient is also a limitation. Unlike what is normally done with the standard clustering coefficient, our measure does not require ties in weighted networks to be transformed. This becomes an issue when all possible ties within a network are assigned a weight, even a very small one. In these circumstances, the network is fully connected, and the generalized clustering coefficient is 1. The standard clustering coefficient does not have this shortcoming as ties with a small weight are set to absent and, therefore, the network does not become fully connected. An example of a weighted, fully connected network is a network consisting of cities, where the ties between cities are assigned a weight that reflects the distance between the two connected cities. Here, all possible ties are present and assigned a weight. The standard clustering coefficient overcomes this issue by setting weak relations, i.e., those characterized by long distances, to absent. A possible solution when applying the generalized coefficient, which does not normally transform the data, is to carry out precisely this transformation and filter the data by setting weak relations, with distances smaller than a fixed cut-off, to absent. However, the suitability and appropriateness of this solution depends on the data, the context in which the data were collected, and the research question.

More generally, researchers should operationalize variables with care when dealing with research questions concerned with tie weights. Marsden and Campbell (1984) conducted a comparative analysis of Granovetter's (1973, pg. 1361) four criteria for defining tie weights. They found that emotional intensity was a better indicator of strength of friendship than the other three criteria. Researchers should choose the appropriate measures of tie strength depending on the nature of the nodes and ties and, more generally, on the context of the research setting. In addition, the scale of the weights should be carefully defined. The scale should be consistent with the chosen criteria. For example, a typical network question often used in studies of advice networks is:

> *Please indicate how often you have turned to this person for information or advice on work-related topics in the past three months.*

with the ordinal scale: 0 (Do not know this person); 1 (Never); 2 (Seldom); 3 (Sometimes); 4 (Often); 5 (Very Often).[12] In this case, answers are inevitably subject to the bias that comes from the different ways in which different people assess duration and define the meaning of the time-related scale. One way to overcome this problem is to transform the ordinal scale into a ratio scale that describes reality more consistently across people. For example, a more appropriate scale could be: 0 (Never); 1 (Once); 3 (Monthly); 6 (Bi-weekly); 12 (Weekly). In turn, this scale, when compared to the former, is likely to yield a network dataset that is richer in information, more robust against potential inaccuracies emanating from subjective judgments, and more suitable to investigations that rely on generalized measures, such as our proposed clustering coefficient.

[12] Cross and Parker (2004) used this question to create the advice network in the consulting company used in Section 5.

References

Barabási, A.-L., Jeonga, H., Néda, Z., Ravasz, E., Schubert, A., Vicsek, T., 2002. Evolution of the social network of scientific collaborations. Physica A 311, 590-614.

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

Burt, R.S., 1992. Structural Holes. Harvard University Press, Cambridge, MA.

Caldarelli, G., 2007. Scale-Free Networks: Complex Webs in Nature and Technology. Oxford Univiversity Press, Oxford.

Colizza, V., Pastor-Satorras, R., Vespignani, A., 2007. Reaction-diffusion processes and metapopulation models in heterogeneous networks. Nature Physics 3, 276-282.

Cross, R., Parker, A., 2004. The Hidden Power of Social Networks. Harvard Business School Press, Boston, MA.

Davis, J.A., 1970. Clustering and hierarchy in interpersonal relations: Testing two graph theoretical models on 742 sociomatrices. American Sociological Review 35(5), 843-851.

Davis, G.F., Yoo, M., Baker, W.E., 2003. The small world of the American corporate elite, 1982-2001. Strategic Organization 1(3), 301-326.

Doreian, P., 1969. A note on the detection of cliques in valued graphs. Sociometry 32(2), 237-242.

Doreian, P., Batagelj, V., Ferligoj, A., 2005. Generalized Blockmodeling. Cambridge University Press, New York, NY.

Ebel, H., Mielsch, L.-I., Bornholdt, S., 2002. Scale-free topology of e-mail networks. Physical Review E 66(3), 035103.

Erdős, P., Rényi, A., 1959. On random graphs. Publicationes Mathematicae 6, 290-297.

Feld, S.L., 1981. The focused organization of social ties. American Journal of Sociology 86(5), 1015-1035.

Feld, S.L., Elmore, R.G., 1982a. Patterns of sociometric choices: transitivity reconsidered. Social Psychology Quarterly 45, 77-85.

Feld, S.L., Elmore, R.G., 1982b. Processes underlying patterns of sociometric choice: Response to Hallinan. Social Psychology Quarterly 45, 9092.

Freeman, L.C., Borgatti, S.P., White, D.R., 1991. Centrality in valued graphs: A measure of betweenness based on network flow. Social Networks 13(2), 141-154.

Freeman, S.C., Freeman, L.C., 1979. The networkers network: A study of the impact of a new communications medium on sociometric structure. Social Science Research Reports 46. University of California, Irvine, CA.

Friedkin, N.E., 1984. Structural cohesion and equivalence explanations of social homogeneity. Sociological Methods and Research 12, 235-261.

Granovetter, M., 1973. The strength of weak edges. American Journal of Sociology 78(6), 1360-1380.

Hallinan, M.T., 1974. A structural model of sentiment relations. American Journal of Sociology 80, 364378.

Heider, F., 1946. Attitudes and cognitive organization. Journal of Psychology 21, 107-112.

Holland, P.W., Leinhardt, S., 1970. A method for detecting structure in sociometric data. American Journal of Sociology 76, 492513.

Holland, P.W., Leinhardt, S., 1971. Transitivity in structural models of small groups. Comparative Group Studies 2, 107124.

Holme, P., Edling, C.R., Liljeros, F., 2004. Structure and time evolution of an Internet dating community. Social Networks 26(2), 155-174.

Ingram, P., Roberts, P.W., 2000. Friendships among competitors in the Sydney hotel industry. American Journal of Sociology 106(2), 387-423.

Karlberg, M., 1997. Testing transitivity in graphs. Social Networks 19(4), 325-343.

Karlberg, M., 1999. Testing transitivity in digraphs. Sociological Methodology 29, 225-251.

Kossinets, G., Watts, D.J., 2006. Empirical analysis of an evolving social network. Science 311, 88-90.

Levine, S.S., Kurzban, R., 2006. Explaining clustering in social networks: Toward an evolutionary theory of cascading benefits. Managerial and Decision Economics 27, 173-187.

Lopez-Fernandez, L., Robles, G., Gonzalez-Barahona, J.M., 2004. Applying social network analysis to the information in CVS repositories. Proceedings of the 1st International Workshop on Mining Software Repositories (MSR2004), 101-105.

Louch, H., 2000. Personal network integration: Transitivity and homophily in strong-tie relations. Social Networks 22(1), 45-64.

Luce, R.D., Perry, A.D., 1949. A method of matrix analysis of group structure. Psychometrika 14(1), 95-116.

Luczkowich, J.J., Borgatti, S.P., Johnson, J.C., Everett, M.G., 2003. Defining and measuring trophic role similarity in food webs using regular equivalence. Journal of Theoretical Biology 220, 303321.

Marsden, P.V., Campbell, K.E., 1984. Measuring tie strength. Social Forces 63(2), 482-501.

Monge, P., Rothman, L., Eisenberg, E., Miller, K., Kirste, K., 1985. The dynamics of organizational proximity. Management Science 31, 1129-1141.

Newman, M.E.J., 2001. Scientific collaboration networks: II. Shortest paths, weighted networks, and centrality. Physical Review E 64(1), 016132.

Newman, M.E.J., 2003. The structure and function of complex networks. SIAM Review 45, 167-256.

Newman, M.E.J., 2006. Modularity and community structure in network. Proceedings of the National Academy of Sciences USA 103, 8577-8582.

Newman, M.E.J., Girvan, M., 2004. Finding and evaluating community structure in networks. Physical Review E 69, 026113.

Nordlund, C., 2007. Identifying regular blocks in valued networks: A heuristic applied to the St. Marks carbon flow data, and international trade in cereal products. Social Networks 29(1), 59-69.

Onnela, J.-P., Saramki, J., Kertsz, J., Kaski, K., 2005. Intensity and coherence of motifs in weighted complex networks. Physical Review E 71(6), 065103.

Opsahl, T., Colizza, V., Panzarasa, P., Ramasco, J.J., 2008. Prominence and control: The weighted rich-club effect. Physical Review Letters 101, 168702.

Panzarasa, P., Opsahl, T., Carley, K.M., forthcoming. Patterns and dynamics of users' behavior and interaction: Network analysis of an online community. Journal of the American Society for Information Science and Technology 60(4).

Ravasz, E., Barabási, A.-L., 2003. Hierarchical organization in complex networks. Physical Review E 67(2), 026112.

Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N., Barabási, A.-L., 2002. Hierarchical organization of modularity in metabolic networks. Science 297(5586), 1551-1555.

Rosvall, M., Bergstrom, C.T., 2008. Maps of random walks on complex networks reveal community structure. Proceeding of the National Academy of Sciences USA 105, 1118-1123.

Scott, J., 2000. Social Network Analysis: A Handbook. Sage Publications, London, UK.

Simmel, G., 1923. The Sociology of Georg Simmel (KH Wolff, trans.). Free Press, New York, NY.

Skvoretz, J., 2002. Complexity theory and models for social networks. Complexity 8, 47-55.

Snijders, T.A.B., 2001. The statistical evaluation of social network dynamics. Sociological Methodology 31, 361-395.

Snijders, T.A.B., Pattison, P., Robins, G., Handcock, M.S., 2006. New specifications for exponential random graph models. Sociological Methodology 36(1), 99-153.

Soffer, S.N., Vàzquez, A., 2005. Network clustering coefficient without degree-correlation biases. Physical Review E 71, 057101.

Solomonoff, R., Rapoport, A., 1951. Connectivity of random nets, Bulletin of Mathematical Biophysics 13, 107-117.

Uzzi, B., Spiro, J., 2005. Collaboration and creativity: The small world problem. American Journal of Sociology 111(2), 447504.

Wasserman, S., Faust, K., 1994. Social Network Analysis: Methods and Applications. Cambridge University Press, New York, NY.

Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of small-world networks. Nature 393, 440-442.

Yang, S., Knoke, D. 2001. Optimal connections: Strength and distance in valued graphs. Social Networks 23(4), 285-295.

Zhang, B., Horvath, S. 2005. A general framework for weighted gene co-expression network analysis. Statistical Applications in Genetics and Molecular Biology 4(1), 17.