

This file was downloaded from the institutional repository BI Brage - <http://brage.bibsys.no/bi> (Open Access)

***Considering clustering measures: third ties, means, and triplets***

**Binh Phan**  
**BI Norwegian Business School**

**Kenth Engø-Monsen**  
**Telenor Group**

**Øystein D. Fjeldstad**  
**BI Norwegian Business School**

This is the authors' final, accepted and refereed manuscript to the article published in

***Social Networks*, 35(2013)3: 300-308**

DOI: <http://dx.doi.org/10.1016/j.socnet.2013.02.007>

The publisher, Elsevier, allows the author to retain rights to “post a revised personal version of the text of the final journal article (to reflect changes made in the peer review process) on your personal or institutional website or server for scholarly purposes, incorporating the complete citation and with a link to the Digital Object Identifier (DOI) of the article”. (Publisher’s policy 2011).

# Considering Clustering Measures: Third Ties, Means, and Triplets<sup>1</sup>

by

Binh Phan<sup>2</sup>, Kenth Engø-Monsen<sup>3</sup>, and Øystein D. Fjeldstad<sup>2</sup>

Abstract

*Measures that estimate the clustering coefficients of ego and overall social networks are important to social network studies. Existing measures differ in how they define and estimate triplet clustering with implications for how network theoretic properties are reflected. In this paper, we propose a novel definition of triplet clustering for weighted and undirected social networks that explicitly considers the relative strength of the tie connecting the two alters of the ego in the triplet. We argue that our proposed definition better reflects theorized effects of the important third tie in the social network literature. We also develop new methods for estimating triplet, local and global clustering. Three different types of mathematical means, i.e. arithmetic, geometric, and quadratic, are used to reflect alternative theoretical assumptions concerning the marginal effect of tie substitution.*

**Key words:** social network, network clustering, triplet clustering, clustering measure, and tie strength

## 1. Introduction

Clustering is a major structural property of social networks (Newman, 2003; Uzzi et al., 2007). Clustering can constrain network actors as well as provide them with network closure benefits. Local measures capture the clustering of ego networks, whereas global measures capture the clustering of overall social networks. Actors are affected by both local and global network properties (Coleman, 1988; Fleming et al., 2007; Snijders et al., 2010; Uzzi and Spiro, 2005; Xiao and Tsui, 2007). The way clustering is defined and measured has implications for how network theoretic properties are reflected in a study. Table 1 presents well-established local and global network clustering measures. A number of these are derived from measurement of network triplets.

---

<sup>1</sup> This paper has been accepted for publication in Social Networks.

<sup>2</sup> Department of Strategy and Logistics, BI Norwegian Business School, 0484 Oslo, Norway. Emails: [t.binh.phan@bi.no](mailto:t.binh.phan@bi.no) and [oystein.fjeldstad@bi.no](mailto:oystein.fjeldstad@bi.no)

<sup>3</sup> Telenor Group, Research and Future Studies, Snarøyveien 30, N-1331 Fornebu, Norway. Email: [Kenth.Engo-Monsen@telenor.com](mailto:Kenth.Engo-Monsen@telenor.com)

-----  
Insert Table 1 approximately here  
-----

A triplet is defined as consisting of a focal actor (or ego) and two alters. In a triplet, there are two ties (T1 and T2) connecting the ego with its two alters, and a possible third tie (T3) connecting the two alters (Luce and Perry, 1949; Opsahl and Panzarasa, 2009; Watts and Strogatz, 1998). The ties may vary in strength. See Figure 1.

-----  
Insert Figure 1 approximately here  
-----

The existence and properties of a third tie connecting an ego's alters are given much attention in the social network literature (c.f. Burt (1992) and Granovetter (1973)). However, its theoretical characteristics are not fully reflected in the existing clustering measures. In this paper, we propose a definition of triplet clustering for weighted social networks that explicitly considers the relative strength of the third tie and develop corresponding local and global clustering measures. We argue that the proposed definition and measures better reflect the theoretical effects that clustering is intended to capture. We only consider undirected networks and leave the extension to directed networks for future research.

The clustering coefficient of a network is estimated by aggregating the clustering values of all triplets in the network. A triplet is the appropriate unit of analysis for studying clustering in a network. The clustering value of a triplet is an indicator of the focal actor's embeddedness, opportunities and constraints. Local clustering measures take into account all triplets in which the ego is the focal actor (Barrat et al., 2004; Opsahl, 2009; Watts and Strogatz, 1998). Global clustering measures estimate the clustering coefficient of overall social networks by aggregation of triplets (Opsahl and Panzarasa, 2009).<sup>4</sup>

Existing clustering measures differ in how they estimate the relative clustering value of triplets. The differences are rooted in the definitions of triplet clustering and the type of mean used to estimate triplet value.

-----  
Insert Table 2 approximately here  
-----

---

<sup>4</sup> The below review of clustering measures draws on Engø-Monsen and Canright (2011)

Existing clustering measures can be classified into three main groups, See Table 2. The first binary group defines triplet clustering as a function of triplet connectedness (Luce and Perry, 1949; Watts and Strogatz, 1998). A triplet does not have clustering if T3 is missing and vice versa clustered if T3 is present. The second group of clustering measures, define triplet clustering as a function of triplet connectedness and the mean of the strength of all three ties in a triplet – hereafter denoted as the strength of T1/2/3 (c.f. Grindrod (2002), Holme, Min Park, Kim and Edling (2007), Kalna and Higham (2007), Latora and Marchiori (2003), Zhang and Horvath (2005)). The third group defines triplet clustering as a function of triplet connectedness and the mean of the strengths of T1 and T2 – hereafter denoted the strength of T1/2 (Barrat et al., 2004; Opsahl, 2009; Opsahl and Panzarasa, 2009). In summary, the first group of clustering measures does not address tie strength. The second group averages the three ties when present, whereas the third group considers only the strength of T1/2.

The strength of T1/2 reflects the time and resources that the focal actor has invested in the triplet, whereas the strength of T3 affects the relative position of triplet members and thus the focal actor’s network benefits and constraints (Burt, 1992; Coleman, 1988; Granovetter, 1973; Uzzi, 1996; Uzzi and Ryon, 2003) . The type of mathematical mean used in computing clustering yields different marginal rates of substitution among tie strengths.<sup>5</sup>

The geometric, quadratic and arithmetic means respectively yield diminishing, increasing and constant marginal rates of substitution among tie strengths. Hence, different types of means can be used to reflect different network theoretic properties e.g. embeddedness, opportunities, and constraints (c.f. Burt (2001), Uzzi (1996)). Figure 2 illustrates the indifference curves for geometric, arithmetic and quadratic means.

-----  
Insert Figure 2 approximately here  
-----

We claim three contributions to the social network literature. First, we offer new clustering definitions and measures that explicitly consider the relative strength of the third tie in a triplet. Second, we show how different types of mathematical means can be used to reflect important network theoretic differences. Third, we show by way of analysis, empirical testing, and simulation how the measures affect clustering coefficients. The second and third contributions imply that

---

<sup>5</sup> Marginal rate of substitution among tie strengths refer to the rate at which the amount of the strength of one or more ties needs to decrease in response to a 1-unit increase of another tie’s strength so that the joint effect of all these ties on the social actor remains the same.

choice of clustering measures in social network studies should be made contingent on the network theoretic properties under consideration.

The remainder of the paper is organized as follows. The second section analyzes the clustering properties of triplets, triads, ego networks and overall social networks, and proposes a new definition of triplet clustering. The third section refines the proposed method for estimating triplet clustering values, by using three different types of mathematical mean and develops the corresponding specific local and global clustering measures. In the fourth section, we provide analysis and empirical tests. The fifth section outlines the parameters of the proposed measure in a Bernoulli random graph. The final section discusses contributions, limitations and implications for research.

## **2. Clustering properties of triplets, triads, ego networks and social networks**

An undirected social network includes a set of social actors and a set of undirected relations (ties) that connect the actors (Boccaletti et al., 2006). Ties vary in strength frequently measured by the intensity of interactions between pairs of actors (Barrat et al., 2004; Granovetter, 1973; Newman, 2001). The strength of a tie reflects the amount of time and resources that the actors have invested in it. Below we review common clustering properties and measurements using a small illustrative weighted network of five actors (A, B, C, D, and E). In this network, depicted by Figure 3b, the strength of ties is denoted on the vertices of the graph.

-----  
Insert Figure 3 approximately here  
-----

A social network contains a number of ego networks, i.e. networks consisting of a focal node (ego) and the nodes directly connected to it. In Figure 3, there are 5 ego networks. Within an ego network, there are a number of triads – subsets of three actors and possible ties among them (Wasserman and Faust, 1994: 19). For example, in the ego network in which B is the ego, there are 6 triads, specifically, A-B-C, B-C-E, A-B-D, A-B-E, B-C-D, and B-D-E.

*Between-triad differences:* Triads vary in their degree of clustering. Triad clustering definitions incorporate connectedness among actors within a triad as well as the strengths of ties connecting them. Among these six triads, the triads A-B-D, A-B-E, B-C-D, and B-C-E are incomplete and have no clustering since a tie is missing, whereas the triads A-B-C and B-D-E are complete and clustered since all their three actors are connected to each other. Between the two complete triads, A-B-C is

more clustered than B-D-E because the average strength of the ties in the former is greater than in the average strength of the latter.

*Within-triad difference:* Inside a triad, the three actors differ in the amount of time and resources that they have invested in the triad. For example, in the A-B-C triad, A and C have invested more time and resources, and are therefore more embedded in the triad than B. Such within triad structural differences are captured by triplets, which consist of a focal actor, two alters, the ties (T1 and T2), connecting the focal actor with the two alters and a possible tie (T3) connecting the two alters (Opsahl and Panzarasa, 2009). A complete triad constitutes three closed triplets with each of the triad's members being the focal actor in one of the triplets. An incomplete triad with only two ties constitutes an open triplet in which the actor connecting the two alters is the focal actor and T3 is missing. For instance, the complete triad A-B-C constitutes the three closed triplets, A-B-C, B-A-C and A-C-B (the middle node is the focal actor of the triplet). The incomplete triad A-B-D constitutes one open triplet A-B-D in which B is the focal actor and the tie between A and D, T3, is missing.

We define triplet clustering as a function of three factors: (i) triplet connectedness, (ii) the strength of T1/2, and (iii) the relative strength of T3 to the strength of T1/2. We discuss the role of each factor below.

- i. Triplet connectedness differentiates closed triplets from open triplets. An open triplet, where T3 is missing, has no clustering and creates no network closure benefits for the focal actor. The focal actor is not embedded in the triplet. Instead, in an open triplet, the focal actor occupies the brokering position, is relatively powerful in the triad, and can obtain brokering benefits by playing the two alters against one another or being the third party (Burt, 1992, 2005; Cook and Emerson, 1978; Emerson, 1962). By contrast, all closed triplets are clustered and can create network closure benefits for their focal actors.
- ii. The strength of T1/2 is an indicator of the focal actors investment of time and resources in the triplet and hence the focal actor's embeddedness. In the complete triad A-B-C, A and C have invested more time and resources, and are therefore more embedded in the triad than B. This is indicated by the strength of the closed B-A-C and A-C-B triplets' T1/2s being greater than the strength of T1/2 of the closed A-B-C triplet. Hence, the B-A-C and A-C-B triplets are more clustered than the A-B-C triplet.
- iii. The focal actor is also affected by the relative position of the focal actor in relation to its two alters, of which the relative strength of T3 to the strength of T1/2 is an indicator. The strength of T3 being greater than T1/2 indicates that the two alters have invested more time and resources in

the triad than the focal actor, and focus on each other more than on the focal actor. Hence, the focal actor faces significant disadvantages. By contrast, if  $T1/2$  is greater than  $T3$ , the focal actor is potentially more powerful than the two alters in the triplet, but the power is constrained by the existence of the connection between the two alters (Burt, 1992; Grindrod, 2002; Holme et al., 2007; Onnela et al., 2005).

Taking into account the relative strength of  $T3$  to  $T1/2$  differentiates the proposed measure from the definitions of the second and third groups (given in Table 2) of weighted clustering measures. The second group of clustering measures define triplet clustering as a function of (i) tie connectedness and (ii) the strengths of  $T1/2/3$ . There are two issues with this triplet clustering definition. First, it does not discriminate between triplets where the focal actors have invested different amounts of time and resources in the triad. For example, in the A-B-C triad, A-B-C, B-A-C, and A-C-B, would be considered equivalent even if A and C have invested more time and resources in the triad than B. Moreover, the strength of  $T3$  and the strength of  $T1/2$  play the same role in computing the clustering of the triplet and hence their respective effects on the focal actor are assumed to be similar. This is unsatisfactory, because although  $T3$  does not have a direct effect on the focal actor, it affects the focal actor indirectly by determining the relative position of the focal actor vis-à-vis the two alters. Our triplet clustering definition reflects our suggestion for altering the role of the third tie in the second group's definition in ways that capture the indirect effect of this tie on the focal actor. The third group of clustering measures define triplet clustering as a function of (i) triplet connectedness, (ii) the strength of  $T1/2$ . This definition does not account for the indirect effect of the third tie. Our definition extends the third group's definition to include the relative strength of  $T3$  to  $T1/2$ .

### **3. A New triplet value estimation method and new clustering measures**

#### *3.1. Triplet value estimation method with three different types of mathematical mean*

In this section we develop a new method for estimating triplet clustering with three different types of mean based on our proposed triplet clustering definition discussed above.

As shown in Figure 2, the geometric, arithmetic and quadratic means have different marginal rates of substitution among tie strengths. The geometric mean yields a diminishing rate of substitution among tie strengths. The geometric mean emphasizes the role of weak ties in the network. In contrast, the quadratic mean yields an increasing rate of substitution. It emphasizes the role of strong ties in the network. The arithmetic mean yields constant marginal rates of substitution among tie strengths, and therefore considers the value of a tie being a linear function of its strength. Hence,

different types of mean should be considered when estimating triplet clustering values and clustering coefficients for networks being examined from different theoretical properties. Two major theoretical approaches to the study of actor benefits from social networks postulate contrary effects of ego-network clustering and ascribe different effects to tie strength ((Burt, 1992, 2001; Coleman, 1988; Granovetter, 1973). From an opportunity perspective, social actors obtain non-redundant information and resources, as well as brokering benefits from sparse networks in which weak ties play an important role (Burt, 1992; Granovetter, 1978; Granovetter, 1973). The geometric mean accentuates the importance of weak ties in a network. In contrast, from an embeddedness perspective, social actors obtain trust and reliable support from strong ties and closure (Coleman, 1988; Xiao and Tsui, 2007). The quadratic mean diminishes the role of weak ties in a network. Research on ‘small world’ properties of large scale networks has shown complementary benefits in network topographies that combine embeddedness and sparseness (Schilling and Phelps, 2007; Uzzi and Spiro, 2005).

In the below development of our proposed measures, we start with the geometric mean and then extend it to the arithmetic and quadratic means.

The proposed method for estimating triplet clustering based on a geometric mean is:

$$V_{i,t} = \sqrt{\sqrt{w_{t,1}w_{t,2}} \sqrt[3]{w_{t,1}w_{t,2}w_{t,3}}} \quad (1)$$

where,  $i$  is the focal actor,  $w_{t,1}$ ,  $w_{t,2}$ , and  $w_{t,3}$  denote the weights (or strengths) of the ties connecting the focal actor with the alters, and the tie connecting the two alters in Triplet  $t$ . Further,  $V_{i,t}$  denotes the relative clustering value of Triplet  $t$  in which  $i$  is the focal actor.

The factor  $\sqrt{w_{t,1}w_{t,2}}$  in Equation (1) reflects the combined role of the ties T1/2 in computing the clustering of the triplet. The factor  $\sqrt[3]{w_{t,1}w_{t,2}w_{t,3}}$  plays two roles in the formula. First, it captures triplet connectedness that differentiates open triplets from closed triplets. This factor is equal to 0 for all open triplets and greater than 0 for all closed triplets. Second, it allows relating the strength of T3 to the strength of T1/2. Specifically, if the strength of T3,  $w_{t,3}$ , is greater than the geometrically averaged strength of T1/2,  $\sqrt{w_{t,1}w_{t,2}}$ , then the product  $\sqrt{\sqrt{w_{t,1}w_{t,2}} \sqrt[3]{w_{t,1}w_{t,2}w_{t,3}}}$  is greater than the single factor  $\sqrt{w_{t,1}w_{t,2}}$ . By contrast, if  $w_{t,3}$  is weaker than  $\sqrt{w_{t,1}w_{t,2}}$ , then the product of  $\sqrt{\sqrt{w_{t,1}w_{t,2}} \sqrt[3]{w_{t,1}w_{t,2}w_{t,3}}}$  is smaller than the product of  $\sqrt{w_{t,1}w_{t,2}}$ . If  $w_{t,3}$  is equal to  $\sqrt{w_{t,1}w_{t,2}}$ , then the product of  $\sqrt{\sqrt{w_{t,1}w_{t,2}} \sqrt[3]{w_{t,1}w_{t,2}w_{t,3}}}$  is identical to  $\sqrt{w_{t,1}w_{t,2}}$ . Summing up: using

equation (1), the triplet clustering value increases disproportionately with the ratio of the strength of T3 to T1/2. Intuitively, it would capture that, all else equal, a strengthening of T3 disproportionately increases the constraint on the focal actor. Descriptively, by taking into account the relative strength of T3 compared to the strength of T1/2, this triplet estimation method is able to differentiate among closed triplets in a complete triad where the focal actors have invested different amounts of time and resources in the triad.

See Appendix A for the corresponding method for estimating triplet clustering values based on arithmetic and quadratic means.

### 3.2. New Local Clustering Measure

Local clustering measures consider triplets in which the ego is the focal actor (Barrat et al., 2004; Luce and Perry, 1949; Newman, 2003; Opsahl, 2009). For instance, in the illustrative network in Figure 3d, the ego network in which B is the ego has 6 triads A-B-C, A-B-D, A-B-E, B-C-D, B-C-E, and B-D-E. Among these triads, the two complete triads A-B-C and B-D-E constitute 6 closed triplets, B-A-C, A-B-C, A-C-B, B-D-E, D-B-E and B-E-D, whereas the four other incomplete triads constitute 4 open triplets, A-B-D, A-B-E, C-B-D and D-B-E. The existing local clustering measures estimate the clustering coefficient of this ego network with an aggregated clustering value of all triplets in which the ego, B, is the focal actor, specifically, A-B-C, D-B-E, A-B-D, A-B-E, C-B-D and D-B-E.

In an ego network, a complete triad consists of three closed triplets. In these triplets, the ego is the focal actor and in the two other triplets, the ego's alters are the focal actors. The exclusion of triplets, from an ego network, in which the ego is not the focal actor, ignores their specific T3 effects in the calculation of ego network clustering coefficients.

Therefore, we propose local clustering measures which take into account all triplets in the ego network. The formula of the proposed local clustering measure using geometric mean is as follows:

$$c_i = \frac{\sum_t \sqrt{\sqrt{w_{t,1}w_{t,2}} \sqrt[3]{w_{t,1}w_{t,2}w_{t,3}}}}{\sum_t^c \sqrt{\sqrt{w_{t,1}w_{t,2}} \sqrt[3]{w_{t,1}w_{t,2}w_{t,3}}} + \sum_t^o \sqrt{w_{t,1}w_{t,2}}} \quad (2)$$

Where,  $c_i$  denotes the clustering coefficients of the ego network of actor  $i$ .  $t$  denotes triplet.  $c$  and  $o$  in the denominator denote closed and open triplets respectively. In the numerator, all open triplets have no clustering value. The numerator is thus the aggregated clustering value of all closed triplets. The denominator is a normalization factor which ensures that the clustering coefficient varies between 0 and 1. It is a combination of the aggregated clustering value of all closed triplets and the

aggregated value of open triplets. The first factor  $\sum_t^c \sqrt{\sqrt{w_{t,1}w_{t,2}} \sqrt[3]{w_{t,1}w_{t,2}w_{t,3}}}$  is used to estimate the value of closed triplets. Due to the absence of T3, the value of open triplets is estimated with the second factor  $\sum_t^o \sqrt{w_{t,1}w_{t,2}}$ , which is the mean of the strengths of T1 and T2 (or the strength of T1/2) (Opsahl and Panzarasa, 2009). See Appendix B for the corresponding local clustering measures using arithmetic and quadratic means.

### 3.3. New Global Clustering Measure

Global clustering measures estimate the clustering coefficients of overall social networks. Existing global clustering measures estimate the clustering coefficient of a social network with either (i) an averaged clustering coefficient over all ego networks (Barrat et al., 2004; Newman, 2003), which tends to inflate the clustering coefficient of a network (Opsahl and Panzarasa, 2009) or (ii) an aggregated clustering value of all its triplets, divided by a denominator normalizing the clustering coefficient to be in the 0-1 range (Eng -Monsen and Canright, 2011; Opsahl and Panzarasa, 2009). In the following we apply the proposed triplet clustering framework to the second method.

$$C = \frac{\sum_t \sqrt{\sqrt{w_{t,1}w_{t,2}} \sqrt[3]{w_{t,1}w_{t,2}w_{t,3}}}}{\sum_t^c \sqrt{\sqrt{w_{t,1}w_{t,2}} \sqrt[3]{w_{t,1}w_{t,2}w_{t,3}}} + \sum_t^o \sqrt{w_{t,1}w_{t,2}}} \quad (3)$$

Where, C denotes the clustering coefficient of a social network.

The difference between this global clustering measure (Equation 3) and its local version (Equation 2) is the triplets that are taken into account. The local version takes into account all triplets in the ego network, while the global version takes into account all triplets in the network.

See Appendix C for the corresponding global clustering measures using arithmetic and quadratic means.

## 4. Differences among clustering measures

The theme of this paper is how to assess clustering properties for weighted networks. In the presence of tie strength variance, the suggested, new clustering measures score the clustering properties differently, depending on the type of mean implemented in the clustering formulae. We compare the three networks in Figure 3 to illustrate this. Overall, the three networks have the same aggregated tie strength, and vary only in tie strength variance. This variance is quantified as 0, 0.4, and 0.6 for the networks in Figures 3a, 3b, and 3c, respectively. The different clustering measures are summarized in Table 3. Consistently, Opsahl and Panzarasa (2009) gives the greater value, Eng -Monsen and Canright (2011) gives the smaller value, and the new measure comes out in the

middle for the geometric mean versions. Using the quadratic mean reverses the order. Table 4 presents the clustering coefficients of various undirected and weighted social networks estimated by the new global measures using the three different types of mean (Equations 3, 8, and 9) and the representative measures of the three existing groups of clustering measure (geometric mean). Except for the clustering coefficients for the first and second networks estimated by Engø-Monsen and Canright's weighted measure, the clustering coefficients estimated with the unweighted clustering measures are lower than the corresponding coefficients estimated with the weighted clustering measures.

-----  
 Insert Table 4 approximately here  
 -----

Of the measures using a geometric mean, Opsahl and Panzarasa (2009) and Engø-Monsen and Canright (2011) give the largest and smallest clustering coefficients, respectively, while the corresponding coefficients estimated with the new measure are in between.

## 5. Clustering coefficient distributions of Bernoulli random networks

In this section, we outline the parameters of the proposed measure in a Bernoulli random graph. Figure 4a contains four clustering coefficient distributions of 1000 Bernoulli random networks with the same size of 200 actors and random tie probability of 0.10. All ties are assigned random weights that vary uniformly from 0 to 1. The clustering coefficients of these distributions are estimated with the traditional unweighted clustering measure and the new weighted clustering measure using geometric, arithmetic and quadratic means. The network size and/or tie probability increase to (200, 0.20), (500,0.10) and (500, 0.20) in Figures 4b, 4c and 4d, respectively.

-----  
 Insert Figure 4 approximately here  
 -----

In Figure 4a, the distribution of the unweighted clustering coefficients has its mean equal to the tie probability (0.1000), as expected. In Figures 4b-d, the distribution of the weighted clustering coefficients estimated with the new measure using arithmetic mean is overlapping with that of the unweighted clustering coefficients. Their mean value is 0.1998, 0.0999, 0.1999 in Figures 1b, c, d respectively. The distributions based on the geometric mean and the quadratic mean shift to the left and the right-hand sides of the two former. The mean value of the geometric mean-based

distribution is 0.1935, 0.0964, and 0.1936 in Figures 1b, c, d respectively, whereas the corresponding one of the quadratic mean-based distribution is 0.2030, 0.1017, and 0.2031. These orderings and effects remain and the distributions become more distinct with increased network size and/or tie probability, which reduce the distributions' standard deviations.

Moreover, the Figures 4a,b,c,d show that the measure using the geometric mean gives the smallest clustering coefficients whereas the measure using the quadratic mean give the largest ones. In other words, they respectively accentuate the roles of weak ties and strong ties in the estimation of clustering coefficients. The simulation results are consistent with the results of the application of the measures to empirical networks as presented in Table 4. We have also done a series expansion of the clustering measures using Wolfram|Alpha<sup>6</sup>, which are reported in Appendix D. The results – the same constant and linear term but difference in the quadratic terms – are consistent with the results of the empirical test and the Bernoulli simulation.

## 6. Discussion

Clustering is a major property of social networks, which is used as indicator of network benefits and constraints (Brass et al., 2004; Burt, 1992, 2005; Coleman, 1988). We argue that the existing clustering measures fail to reflect the important, but theoretically distinct, roles that the 'third tie' plays in ego networks. In order to explicitly incorporate specific 'third-tie' roles in triplets, we define triplet clustering as a function of triplet connectedness, the average strength of the ties to the alters (T1/2) and the relative strength of the tie connecting the alters (T3) to T1/2.

Based on our proposed definition, we developed a new method for estimating the relative clustering value of triplets, as well as new local and global clustering measures for weighted, undirected networks. We showed how geometric, arithmetic, and quadratic means can be used to reflect different theoretical assumptions about marginal rate of substitution among tie strengths. For example, a geometric mean may be appropriate when assessing opportunities and structural constraints (Burt, 1992, Granovetter, 1973) whereas a quadratic mean may be appropriate when assessing embeddedness (Coleman, 1988; Granovetter, 1985; Uzzi, 1997). We provided analysis, empirical tests, and outlined the parameters of the proposed measures in a Bernoulli random graph to show when and how measures in their characterization of networks. Our examination of the proposed measures found significant and desired differences relative to existing measures.

We claim three contributions to social network studies. First, by explicitly considering the relative strength of the third tie in a triplet, the new measures better capture its theoretical effects –that is,

---

<sup>6</sup> Wolfram|Alpha: <http://www.wolframalpha.com/>

the third tie affects the focal actor by determining the relative position of the focal actor in relation to the two partners in the triplet (Burt, 1992; Coleman, 1988; Granovetter, 1978). Second, we have shown how the use of different types of mean in the estimation of triplet clustering coefficients can be used to address different social network perspectives. For instance, an information perspective emphasizes the value of weak ties (Burt, 1992; Granovetter, 1978; Hansen, 1999) while a relational perspective emphasizes the value of strong ties (Coleman, 1988; Uzzi, 1997; Xiao and Tsui, 2007). Third, by analysis, empirical tests, and simulation we found support for the new measures reliably and validly discriminating the desired properties at the network level, thus providing social network researchers with new instruments that can be tailored to reflect the theoretical perspectives of their studies.

### *Implications*

Our measures and findings allow incorporation of important theoretical assumptions about the role of the ‘third tie’ in future network studies. Furthermore they imply that researchers should pay greater attention to theoretical assumptions about the effects of tie strength variance within triads as well as the effect of such variance on all actors in the triad, even when the ego is focused.

A triplet is the appropriate unit of analysis for studying clustering in a network. Studying how triplet clustering affects the focal actors’ strategic networking behaviour (e.g., creating, maintaining, destroying ties or triplets) could improve our understanding of the evolution of triplets, triads and social networks. Future research should also assess the empirical implications of the new clustering measures, and empirically evaluate and establish guidelines for the usage of these measures on networks reflecting different theoretical properties.

The proposed method for estimating triplet clustering values and clustering measures is limited to weighted, undirected networks. Ties in social networks are not only weighted, but also often directed. The purpose of developing clustering measures is to better capture all the dimensions of social networks (see Table 1). We leave the challenging, but important, extension of the proposed measures and methods to weighted, directed networks for further research.

### *Conclusions*

There is increased attention to large networks across a wide variety of areas in both academia and practice. Social network analysis is a vibrant field with on-going important theoretical developments and empirical testing. As with any field, the development of instruments is a condition for the advancement of knowledge. The recent upsurge in the availability of data and computing resources necessitates further development of tools and instruments to take advantage of the exiting opportunities presented. We believe the measures proposed and tested in this paper is a

step along the way toward both better instruments and greater flexibility in tailoring the instruments to theoretical lenses.

## **References:**

- Barrat, A., Barthelemy, M., Pastor-Satorras, R., Vespignani, A., 2004. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America* 101, 3747-3752.
- Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, D.-U., 2006. Complex networks: Structure and dynamics. *Physics Reports* 424, 175-308.
- Brass, D.J., Galaskiewicz, J., Greve, H.R., Tsai, W., 2004. Taking Stock of Networks and Organizations: A Multilevel Perspective. *The Academy of Management Journal* 47, 795-817.
- Burt, R.S., 1992. *Structural Holes*. Cambridge, MA: Harvard University Press.
- Burt, R.S., 2001. Structural Holes versus Network Closure as Social Capital, in: Lin, N., Cook, K., Burt, R.S. (Eds.), *Social capital: theory and research*. Aldine Transaction, New Brunswick, N.J., pp. 31-56.
- Burt, R.S., 2005. *Brokerage and closure: an introduction to social capital*. Oxford University Press, Oxford.
- Coleman, J.S., 1988. Social capital in the creation of human capital. *American Journal of Sociology* 94, S95-S120.
- Cook, K.S., Emerson, R.M., 1978. Power, Equity and Commitment in Exchange Networks. *American Sociological Review* 43, 721-739.
- Emerson, R.M., 1962. Power-dependence relations. *American Sociological Review* 27, 31-41.
- Engø-Monsen, K., Canright, G., 2011. *Weighted Clustering Coefficients*, Telenor Report R5. Telenor, Oslo.
- Fleming, L., Mingo, S., Chen, D., 2007. Collaborative Brokerage, Generative Creativity, and Creative Success. *Administrative Science Quarterly* 52, 443-475.
- Granovetter, M., 1978. Threshold Models of Collective Behavior. *The American Journal of Sociology* 83, 1420-1443.
- Granovetter, M., 1985. Economic Action and Social Structure: The Problem of Embeddedness. *The American Journal of Sociology* 91, 481-510.
- Granovetter, M.S., 1973. The strength of weak ties. *American Journal of Sociology* 78, 1360-1380.
- Grindrod, P., 2002. Range-dependent random graphs and their application to modeling large small-world Proteome datasets. *Physical Review E* 66, 066702.
- Hansen, M.T., 1999. The Search-Transfer Problem: The Role of Weak Ties in Sharing Knowledge across Organization Subunits. *Administrative Science Quarterly* 44, 82-111.
- Holme, P., Min Park, S., Kim, B.J., Edling, C.R., 2007. Korean university life in a network perspective: Dynamics of a large affiliation network. *Physica A: Statistical and Theoretical Physics* 373, 821-830.
- Kalna, G., Higham, D.J., 2007. A clustering coefficient for weighted networks, with application to gene expression data. *Ai Communications* 20, 263-271.
- Latora, V., Marchiori, M., 2003. Economic small-world behavior in weighted networks. *The European Physical Journal B - Condensed Matter and Complex Systems* 32, 249-263.
- Luce, R., Perry, A., 1949. A method of matrix analysis of group structure. *Psychometrika* 14, 95-116.
- Newman, M.E.J., 2001. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E* 64, 016132.
- Newman, M.E.J., 2003. The structure and function of complex networks. *Siam Review* 45, 167-256.
- Onnela, J.-P., Saramaki, J., Kertesz, J., Kaski, K., 2005. Intensity and coherence of motifs in weighted complex networks. *Physical Review E* 71, 065103.
- Opsahl, T., 2009. Weighted local clustering coefficient, *Network Science Blog*.
- Opsahl, T., Panzarasa, P., 2009. Clustering in weighted networks. *Social Networks* 31, 155-163.
- Schilling, M.A., Phelps, C.C., 2007. Interfirm Collaboration Networks: The Impact of Large-Scale Network Structure on Firm Innovation. *Management Science* 53, 1113-1126.

- Snijders, T.A.B., van de Bunt, G.G., Steglich, C.E.G., 2010. Introduction to stochastic actor-based models for network dynamics. *Social Networks* 32, 44-60.
- Uzzi, B., 1996. The Sources and Consequences of Embeddedness for the Economic Performance of Organizations: The Network Effect. *American Sociological Review* 61, 674-698.
- Uzzi, B., 1997. Social structure and competition in interfirm networks: The paradox of embeddedness. *Administrative Science Quarterly* 42, 35-67.
- Uzzi, B., Amaral, L.A.N., Reed-Tsochas, F., 2007. Small-world networks and management science research: a review. *Eur Manage Rev* 4, 77-91.
- Uzzi, B., Ryon, L., 2003. Relational Embeddedness and Learning: The Case of Bank Loan Managers and Their Clients. *Management Science* 49, 383-399.
- Uzzi, B., Spiro, J., 2005. Collaboration and creativity: The small world problem. *American Journal of Sociology* 111, 447-504.
- Wasserman, S., Faust, K., 1994. *Social network analysis: methods and applications*. Cambridge University Press, Cambridge.
- Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of 'small-world' networks. *Nature* 393, 440-442.
- Xiao, Z., Tsui, A.S., 2007. When Brokers May Not Work: The Cultural Contingency of Social Capital in Chinese High-tech Firms. *Administrative Science Quarterly* 52, 1-31.
- Zhang, B., Horvath, S., 2005. A General Framework for Weighted Gene Co-Expression Network Analysis. *Statistical Applications in Genetics and Molecular Biology* 4, Article 17.

## Appendix

Appendix A: The development of the method for estimating triplet clustering values based on arithmetic and quadratic means

Arithmetic mean:

$$V_{i,t} = \frac{1}{2} \left( \frac{w_{t,1} + w_{t,2}}{2} + \frac{w_{t,1} + w_{t,2} + w_{t,3}}{3} \right) a_{t,1} a_{t,2} a_{t,3} \quad (4)$$

Quadratic mean:

$$V_{i,t} = \left( \sqrt{\frac{1}{2} \left( \left( \sqrt{\frac{w_{t,1}^2 + w_{t,2}^2}{2}} \right)^2 + \left( \sqrt{\frac{w_{t,1}^2 + w_{t,2}^2 + w_{t,3}^2}{3}} \right)^2 \right)} \right) a_{t,1} a_{t,2} a_{t,3} \quad (5)$$

Here,  $a_{t,1}$ ,  $a_{t,2}$  or  $a_{t,3}$  is equal to 1 if the ties T1, T2, and T3 are present, and is equal to 0 otherwise. The role of the factor  $a_{t,1} a_{t,2} a_{t,3}$ , is to measure triplet interconnectedness. It is equal to 1 if Triplet t is

closed and 0 if it is open. The factor  $\frac{w_{t,1} + w_{t,2}}{2}$  in Equation (4) or  $\sqrt{\frac{w_{t,1}^2 + w_{t,2}^2}{2}}$  in Equation (5)

captures the role of the strength of T1/2, whereas the factor  $\frac{w_{t,1} + w_{t,2} + w_{t,3}}{3}$  in Equation (4) or

$\sqrt{\frac{w_{t,1}^2 + w_{t,2}^2 + w_{t,3}^2}{3}}$  in Equation (5) allows relating the strength of T3 to the strength of T1/2.

Intuitively, it would capture that, all else equal, a weakening of T3 disproportionately decreases the embeddedness of the focal actor.

Appendix B: The formulas of new local clustering measures using arithmetic and quadratic means.

Arithmetic mean:

$$c_i = \frac{\sum_t \frac{1}{2} \left( \frac{w_{t,1} + w_{t,2}}{2} + \frac{w_{t,1} + w_{t,2} + w_{t,3}}{3} \right) a_{t,1} a_{t,2} a_{t,3}}{\sum_t^c \frac{1}{2} \left( \frac{w_{t,1} + w_{t,2}}{2} + \frac{w_{t,1} + w_{t,2} + w_{t,3}}{3} \right) + \sum_t^o \frac{w_{t,1} + w_{t,2}}{2}} \quad (6)$$

Quadratic mean:

$$c_i = \frac{\sum_t \left( \sqrt{\frac{1}{2} \left( \left( \sqrt{\frac{w_{t,1}^2 + w_{t,2}^2}{2}} \right)^2 + \left( \sqrt{\frac{w_{t,1}^2 + w_{t,2}^2 + w_{t,3}^2}{3}} \right)^2 \right)} \right) a_{t,1} a_{t,2} a_{t,3}}{\sum_t^c \sqrt{\frac{1}{2} \left( \left( \sqrt{\frac{w_{t,1}^2 + w_{t,2}^2}{2}} \right)^2 + \left( \sqrt{\frac{w_{t,1}^2 + w_{t,2}^2 + w_{t,3}^2}{3}} \right)^2 \right)} + \sum_t^o \sqrt{\frac{w_{t,1}^2 + w_{t,2}^2}{2}}} \quad (7)$$

Where,  $c_i$  denotes the clustering coefficients of the ego network of actor i. In the denominator, c and o stands for closed and open triplets respectively

Appendix C: The formulas of new global clustering measures using arithmetic and quadratic means.

Arithmetic mean:

$$C = \frac{\sum_t \frac{1}{2} \left( \frac{w_{t,1} + w_{t,2}}{2} + \frac{w_{t,1} + w_{t,2} + w_{t,3}}{3} \right) a_{t,1} a_{t,2} a_{t,3}}{\sum_t^c \frac{1}{2} \left( \frac{w_{t,1} + w_{t,2}}{2} + \frac{w_{t,1} + w_{t,2} + w_{t,3}}{3} \right) + \sum_t^o \frac{w_{t,1} + w_{t,2}}{2}} \quad (8)$$

Quadratic mean:

$$C = \frac{\sum_t \left( \sqrt{\frac{1}{2} \left( \left( \sqrt{\frac{w_{t,1}^2 + w_{t,2}^2}}{2} \right)^2 + \left( \sqrt{\frac{w_{t,1}^2 + w_{t,2}^2 + w_{t,3}^2}}{3} \right)^2 \right)} \right) a_{t,1} a_{t,2} a_{t,3}}{\sum_t^c \sqrt{\frac{1}{2} \left( \left( \sqrt{\frac{w_{t,1}^2 + w_{t,2}^2}}{2} \right)^2 + \left( \sqrt{\frac{w_{t,1}^2 + w_{t,2}^2 + w_{t,3}^2}}{3} \right)^2 \right)} + \sum_t^o \sqrt{\frac{w_{t,1}^2 + w_{t,2}^2}{2}}} \quad (9)$$

Where, C denotes the clustering coefficient of a social network. The only difference between this global clustering measure (Equations 8-9) and its local version (Equations 6-7) stems from which triplets are taken into account. The local version takes into account all triplets in the ego network, while the global version takes into account all triplets in the network.

#### Appendix D: Series expansions of the new clustering measure

Geometric mean:

Series expansion at  $x = 1$  :

$$\frac{3}{5} + \frac{2(x-1)}{25} - \frac{41(x-1)^2}{1000} + O((x-1)^3)$$

Computed by Wolfram|Alpha

Quadratic mean:

Series expansion at  $x = 1$  :

$$\frac{3}{5} + \frac{2(x-1)}{25} + \frac{9(x-1)^2}{1000} + O((x-1)^3)$$

Computed by Wolfram|Alpha

Arithmetic mean:

Series expansion at  $x = 1$  :

$$\frac{3}{5} + \frac{2(x-1)}{25} - \frac{2}{125} (x-1)^2 + O((x-1)^3)$$

Computed by Wolfram|Alpha

## Tables

**Table 1:** Clustering measures

		<b>Binary</b>	<b>Weighted</b>
<b>Local</b>	<b>Undirected</b>	Watts and Strogatz (1998)	Grindrod (2002) Latora and Marchiori (2003) Barrat et al. (2004) Zhang and Horvath (2005) Onnela, Saramaki, Kertesz, and Kaski (2005) Holme, Min Park, Kim and Edling (2007) Kalna and Higham (2007), Opsahl (2009b)
	<b>Directed</b>		
<b>Global</b>	<b>Undirected</b>	Luce and Perry (1949)	Opsahl and Panzarasa (2009) Engø-Monsen and Canright (2011)
	<b>Directed</b>	Holland and Leinhardt (1970, 1971)	Opsahl and Panzarasa (2009)

**Table 2:** Triplet clustering definitions and estimation methods and strengths of clustering measures

	<b>Group 1</b>	<b>Group 2</b>	<b>Group 3</b>	<b>Group 4</b>
<b>Group-representative measures</b>	Classic unweighted measure	Engø-Monsen and Canright (2011)'s weighted clustering measure	Opsahl and Panzarasa (2009)'s weighted clustering measure	New Measure*
<b>Global Clustering Measures</b> (Geometric Mean)	$C = \frac{\sum_t a_{t,1} a_{t,2} a_{t,3}}{\sum_t a_{t,1} a_{t,2}}$	$C = \frac{\sum_t \sqrt[3]{w_{t,1} w_{t,2} w_{t,3}}}{\sum_t \sqrt{w_{t,1} w_{t,2}}}$	$C = \frac{\sum_t \sqrt{w_{t,1} w_{t,2}} a_{t,1} a_{t,2} a_{t,3}}{\sum_t \sqrt{w_{t,1} w_{t,2}} a_{t,1} a_{t,2}}$	$C = \frac{\sum_t \sqrt{\sqrt{w_{t,1} w_{t,2}} \sqrt[3]{w_{t,1} w_{t,2} w_{t,3}}}}{\sum_t^c \sqrt{\sqrt{w_{t,1} w_{t,2}} \sqrt[3]{w_{t,1} w_{t,2} w_{t,3}}} + \sum_t^o \sqrt{w_{t,1} w_{t,2}}}$
<b>Triplet clustering value estimation method</b>	$V_{i,t} = a_{t,1} a_{t,2} a_{t,3}$	$V_{i,t} = \sqrt{w_{i,1} w_{i,2} w_{i,3}}$	$V_{i,t} = \sqrt{w_{t,1} w_{t,2}} a_{t,1} a_{t,2} a_{t,3}$	$V_{i,t} = \sqrt{\sqrt{w_{t,1} w_{t,2}} \sqrt[3]{w_{t,1} w_{t,2} w_{t,3}}}$
<b>Triplet clustering definition</b>	Triplet clustering is a function of: - Triplet connectedness	Triplet clustering is a function of: - Triplet connectedness - The mean of the strength of T1, T2 and T3	Triplet clustering is a function of: - Triplet connectedness - The mean of the strengths of T1 and T2, denoted the strength of T1/2 in the text.	Triplet clustering is a function of: - Triplet connectedness - The mean of the strengths of T1 and T2, denoted the strength of T1/2 in the text. - The relative strength of T3 compared to the strength of T1/2
<b>Strengths</b>	- Ability to differentiate closed triplets from open triplets	- The strength of Group 1 - Ability to differentiate triplets constituted by different triads with different clustering degrees.	- The strengths of Group 2 - Ability to differentiate triplets in a triad where the focal actors are different in the extent to which they are embedded in the triad.	- The strengths of Group 3 - Capturing the theoretic role of the relative strength of T3 to the strength of T1/2.

Notes: -  $i$  is the focal actor.  $j$  and  $h$  are the first and second partners of the focal actor  $i$ .  $t$  denotes Triplet in the network.

-  $a_{t,1}$ ,  $a_{t,2}$ , or  $a_{t,3}$  denotes the presence of the tie connecting actors  $i$  and  $j$ , connecting actor  $i$  and  $h$ , or connecting  $j$  and  $h$ , respectively. It is equal to 1 if the tie is present and equal to 0 if it is absent.

$w_{t,1}$ ,  $w_{t,2}$ , or  $w_{t,3}$  denote the weights of ties connecting actors  $i$  and  $j$ , connecting actor  $i$  and  $h$ , or connecting  $j$  and  $h$ , respectively. It is greater than 0 if the tie is present and equal to 0 if it is absent.

- Clustering coefficients range from 0 to 1.

- \* In the denominator,  $c$  standards for closed triplets while  $o$  stands for open triplets

**Table 3:** Clustering coefficients estimated by different clustering measures for the three example networks in Figure 3:

Types of mean	Measures	Networks and Variance*		
		3A	3B	3C
		0.00	0.40	0.60
Arithmetic Mean	Engø-Monsen and Canright (2011)	0.6000	0.6000	0.6429
	Opsahl and Panzarasa (2009)	0.6000	0.6000	0.6429
	New Measure	0.6000	0.6000	0.6429
Geometric Mean	Engø-Monsen and Canright (2011)	0.6000	0.5957	0.6320
	Opsahl and Panzarasa (2009)	0.6000	0.5977	0.6385
	New Measure	0.6000	0.5971	0.6367
Quadratic Mean	Engø-Monsen and Canright (2011)	0.6000	0.6040	0.6524
	Opsahl and Panzarasa (2009)	0.6000	0.6023	0.6468
	New Measure	0.6000	0.6028	0.6483

**Notes:** \* Variance denotes the tie strength variance within the network.

- In Table 1, the denominator of the new clustering measure is different from that of the existing measures. When the new measure has the same denominator as the existing ones, the difference between the coefficients estimated by the new measure and those estimated by the existing measures is the same in pattern.

**Table 4:** Clustering coefficients of various undirected social networks

Networks		Existing measures					New Clustering Measure					
		Unweighted measure	Engø-Monsen and Canright (2011)'s weighted measure		Opsahl and Panzarasa (2009)'s weighted measure		Geometric mean		Arithmetic mean		Quadratic mean	
			Coefficient	Coefficient	%	Coefficient	%	Coefficient	%	Coefficient	%	Coefficient
1	Freeman EIES (time 1)	0.8417	0.8393	-0.28	0.8593	2.09	0.8572	1.84	0.8556	1.65	0.8548	1.55
2	Freeman EIES (time 2)	0.9010	0.8981	-0.32	0.9146	1.51	0.9135	1.38	0.9126	1.29	0.9121	1.23
3	Freeman EIES (Messages)	0.6569	0.6775	3.13	0.7654	16.51	0.7486	13.95	0.7820	19.04	0.8007	21.88
4	Online community	0.0568	0.0632	11.16	0.0694	22.12	0.0650	14.37	0.0689	21.31	0.0734	29.13
5	Consulting (advice)	0.7242	0.7463	3.05	0.7750	7.02	0.7700	6.33	0.7677	6.00	0.7690	6.19
6	Consulting (value)	0.7062	0.7231	2.40	0.7321	3.67	0.7303	3.41	0.7283	3.14	0.7269	2.94
7	Research team (advice)	0.6723	0.7163	6.55	0.7439	10.65	0.7384	9.84	0.7236	7.64	0.7184	6.86
8	Research team (awareness)	0.6723	0.7077	5.27	0.7200	7.09	0.7174	6.70	0.7103	5.65	0.7057	4.97
9	Scientific collaboration (paper)	0.3596	0.3807	5.88	0.3924	9.12	0.3871	7.64	0.3949	9.83	0.4055	12.76
	<b>Average</b>	<b>0.6212</b>	<b>0.6391</b>	<b>4.09</b>	<b>0.6636</b>	<b>8.87</b>	<b>0.6586</b>	<b>7.27</b>	<b>0.6604</b>	<b>8.39</b>	<b>0.6630</b>	<b>9.73</b>

**Notes:**

- % is the difference between the corresponding weighted clustering coefficient and the classic unweighted clustering coefficient (Column 2), measured in percentage.
- Networks 1, 2 and 3 are Freeman EIES acquaintance networks that include 32 actors. Network 4 is the facebook-like social network used in Opsahl (2009a). Network 5, 6, 7 and 8 are intra-organizational networks used in Cross and Parker (1981). Network 9 is the co-authorship network (paper) used in Newman (2001). The data of Networks 1, 2, 3 are obtained from Wasserman and Faust (1994) whereas the data of the other networks is obtained from Opsahl (2009b). The detailed descriptions of these networks can be seen in Opsahl (2009b) or Opsahl and Panzarasa (2009). Network 9 is originally a two-mode network and transformed into one-mode weighted network by counting the number of common nodes. The remaining networks are originally directed networks and transformed into the corresponding undirected ones by taking the arithmetic mean of the weights of a relationship's two directed ties as the weight of the relation or the undirected tie. The clustering coefficients estimated with the unweighted measure and Opsahl and Panzarasa (2009)'s weighted measure are calculated with Tnet software developed by Opsahl (2009a) whereas the coefficients estimated with the remaining measures are calculated with a software developed by the second author. The coefficients in the columns 2, 3, 4 are calculated with the geometric mean method.
- The new clustering measure has the denominator different from that of the existing measures. When the new measure has the same denominator as the existing ones, the differences between the coefficients estimated by the new measure and those estimated by the other measures remain the same in pattern.