



**HAL**  
open science

# Weighting the transitivity of undirected weighted social networks with triadic edge dissimilarity scores

Guillaume Péron

► **To cite this version:**

Guillaume Péron. Weighting the transitivity of undirected weighted social networks with triadic edge dissimilarity scores. *Social Networks*, 2023, 73, pp.1-6. 10.1016/j.socnet.2022.12.001 . hal-03905679v2

**HAL Id: hal-03905679**

**<https://hal.science/hal-03905679v2>**

Submitted on 5 Dec 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# **Weighting the transitivity of undirected weighted social networks with triadic edge dissimilarity scores**

**Running title:** Triadic edge dissimilarity scores for undirected graphs

Guillaume Péron

## **Abstract**

When an individual is socially connected to two others, the resulting triplet can be closed (if the two social partners are themselves connected) or open (if they are not connected). The proportion of closed triplets, referred to as the binary network transitivity, is a classic measure of the level of interconnectedness of a social network. However, in any given triplet, if the closing link is weak, or indeed if any of the links in the triplet is weak, then the triplet should not contribute as much to network transitivity as if all three links were equally strong. I propose two ways to weight the contribution of each triplet according to the dissimilarity between the three links in the triplet. Empirically, the resulting new metrics conveyed information not picked up by any other network-level metric. I envision that this approach could prove useful in studies of triadic mechanisms, i.e., situations where pre-existing social ties influence the interactions with third parties. These metrics could also serve as repeatable synthetic variables that summarize information about the variability of the strength of social connections.

Key-words: animal association network; indirect interactions; clustering coefficient; valence; graph theory; social style; intransitivity; socio-spatial structure

## 1 Introduction

2 In undirected social networks, whenever an individual A is separately connected to two other  
 3 individuals B and C, the resulting triplet can be closed (if B and C are also connected) or open (if B  
 4 and C are not connected) (Fig. 1). The proportion of closed triplets measures the level of  
 5 interconnectedness and is referred to as the binary network transitivity (Newman, 2008). Introducing  
 6 the variable  $\delta_t = 1$  if the triplet  $t$  is closed and  $\delta_t = 0$  if the triplet  $t$  is open, the binary network  
 7 transitivity  $\mathcal{C}_0$  corresponds to the sum of  $\delta_t$  across all the triplets (Newman, 2008).

Eq. 1

$$\mathcal{C}_0 = \frac{1}{\mathcal{T}} \sum_{t=1}^{\mathcal{T}} \delta_t$$

8  $\mathcal{T}$  is the number of triplets in the network.

9 The binary network transitivity has become one of the most used metrics to characterize the  
 10 structure of social networks. Many properties of social networks depend on transitivity, e.g., the  
 11 occurrence of clusters of interconnected individuals, the expected length of the minimum path  
 12 between a given pair of individuals, the expected number of different paths that connect a given pair  
 13 of individuals. However, in the binary definition of transitivity, all the triplets that are closed  
 14 contribute equally to the transitivity of the network. Yet, in many instances, the connections  
 15 (synonyms: links, edges, bonds) vary in strength (synonyms: weight, valence). For example, not all  
 16 connections are necessarily active at any given time. Intuitively, if the closing link is weak or rarely  
 17 active, the triplet does not contribute as much to network transitivity compared to if the closing link  
 18 is strong or frequently active. To address this issue, one might use the strength of the closing links to  
 19 weight the contribution of the triplets to network transitivity (Eq. 2).

Eq. 2

$$\mathcal{C}_{0,closing} = \frac{1}{\mathcal{T}} \sum_{t=1}^{\mathcal{T}} \gamma_t$$

20  $\gamma_t$  is the strength of the closing link in triplet  $t$ , rescaled to vary between 0 and 1. If the network was  
 21 dichotomized by setting all the link strengths to 0 for a missing link or to 1 for a link with non null  
 22 strength, then  $\gamma_t = \delta_t$ , and Eq. 2 would reduce to Eq. 1.

23 Alternatively, following Opsahl and Panzarasa (2009), one can weigh the contributions of the triplets  
 24 according to the strength of their first two links, yielding the weighted network transitivity  $\mathcal{C}_1$ . There  
 25 are several variations (Eqs. 3a-3c).

Eq. 3a	$\mathcal{C}_{1,geom} = \frac{\sum_{t=1}^{\mathcal{T}} \sqrt{\alpha_t \beta_t} \cdot \delta_t}{\sum_{t=1}^{\mathcal{T}} \sqrt{\alpha_t \beta_t}}$
Eq. 3b	$\mathcal{C}_{1,min} = \frac{\sum_{t=1}^{\mathcal{T}} \min(\alpha_t, \beta_t) \cdot \delta_t}{\sum_{t=1}^{\mathcal{T}} \min(\alpha_t, \beta_t)}$
Eq. 3c	$\mathcal{C}_{1,max} = \frac{\sum_{t=1}^{\mathcal{T}} \max(\alpha_t, \beta_t) \cdot \delta_t}{\sum_{t=1}^{\mathcal{T}} \max(\alpha_t, \beta_t)}$

26  $\alpha_t$  and  $\beta_t$  are the strengths of the first two links in triplet  $t$  (Fig. 1). If the network was dichotomized  
 27 by setting all the link strengths to 0 or to 1, then Eqs. 3a-3c would, like Eq. 2, reduce to Eq. 1.

28 Eqs. 3a-3c only use two of the three links in each triplet; Eq. 2 only used one. There does not exist yet  
 29 a method that takes all three links into account. To address that issue, I quantified the dissimilarity  
 30 between the three links in each triplet. First, I reordered and renamed the three link strengths  
 31  $\alpha_t, \beta_t, \gamma_t$  into  $a_t, b_t, c_t$  with  $a_t \geq b_t \geq c_t$  (Fig. 1). The notation change means that  $c_t$  does not  
 32 compulsorily correspond to the closing link. I then computed the pairwise proportional difference in  
 33 each triplet  $t$  (Eq. 4a) (Péron, 2022a).

Eq. 4a	$\omega_t = \frac{a_t - c_t}{a_t} + \frac{b_t - c_t}{a_t} - 1 = \frac{b_t - 2c_t}{a_t}$
--------	---

34  $\omega$  compares the weakest link  $c_t$  to the other two links.  $\omega = 0$  for a false triplet ( $b = c = 0$ ; Fig. 1),  
 35  $\omega = 1$  for an open triplet made of two equally strong links ( $a = b$  and  $c = 0$ ), and  $\omega = -1$  for a  
 36 closed triangle made of three equally strong links ( $a = b = c$ ) (Fig. 1). If the network was  
 37 dichotomized, the largest link  $a_t$  would be set to 1 and  $b_t$  and  $c_t$  would take either value 0 or 1

38 (instead of varying between 0 and 1). Then the only possible values for  $\omega_t$  would be -1, 0, or 1

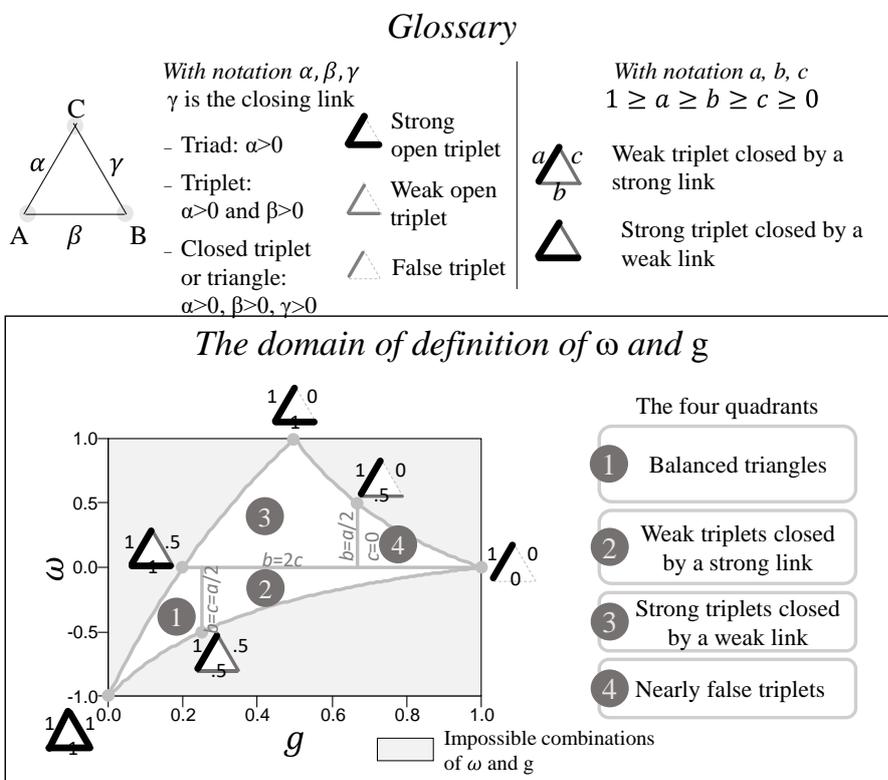
39 (instead of varying between -1 and +1).

40 Next, I computed the Gini coefficient of the three link strengths (Péron, 2022a).

Eq. 4b	$g_t = \frac{1(a_t - b_t) + (b_t - c_t) + (a_t - c_t)}{2(a_t + b_t + c_t)} = \frac{a_t - c_t}{a_t + b_t + c_t}$
--------	---

41 The Gini coefficient is a classic measure of inequality (Gini, 1936), often used to quantify how much  
 42 larger is the largest quantile of a distribution. Applied to three values,  $g = 1$  for a false triplet ( $b =$   
 43  $c = 0$ ),  $g = 0.5$  for an open triplet made of two equally strong links ( $a = b$  and  $c = 0$ ), and  $g = 0$   
 44 for a closed triangle made of three equally strong links ( $a = b = c$ ) (Fig. 1). If the network was  
 45 dichotomized, the only possible values for  $g_t$  would be 0, 0.5, or 1 (instead of varying between 0 and  
 46 1). Throughout, the strengths of the links  $a$ ,  $b$ , and  $c$  were rescaled to vary between 0 (no link) and 1  
 47 (strongest recorded link).

48 Fig. 1: Illustrated glossary and representation of the domain of definition of  $\omega$  and  $g$  as well as the  
 49 four quadrants to distinguish between four different types of weighted triadic interactions.



50

51

52 Compared to previous computations of network transitivity (Eqs. 1-3) (Newman, 2008; Opsahl and  
 53 Panzarasa, 2009), my framework uses the terminology in a slightly different way (Fig. 1). (i) In my  
 54 framework, missing links with a strength of 0 are special cases of weak links. By contrast, in Eqs.1 and  
 55 3, as soon as even the most tenuous closing link was recorded, the triplet would change from open  
 56 to closed. (ii) In my framework, Eqs. 4a-4b actually apply to “triads”, meaning that they only require  
 57 one non-missing link (Fig. 1). By contrast, Eqs. 1-3 apply to “triplets” with at least two non-missing  
 58 links. The difference is important. On the one hand “false triplets” (Fig. 1) count as triads but not as  
 59 triplets (Fig. 1). On the other hand, a triangle made of three interconnected individuals and three  
 60 links  $(a,b,c)$  contributes a single triad to the computation of Eq. 4, but three triplets to the  
 61 computation of Eqs 1-3: triplets  $(a,b)$  closed by  $c$ ,  $(a,c)$  closed by  $b$ , and  $(b,c)$  closed by  $a$  (Newman,  
 62 2008; Opsahl and Panzarasa, 2009). (iii) As mentioned earlier, in my framework, I ordered the link  
 63 strengths in each triad  $(1 \geq a \geq b \geq c \geq 0)$  before computing  $\omega$  and  $g$ . Then I introduced phrases  
 64 like “weak triplet closed by a strong link” to refer to specific triangular configurations (Fig. 1). A  
 65 “weak triplet closed by a strong link” is a triangle  $(a,b,c)$  with  $c$  smaller than  $a$  and  $b$ . As mentioned  
 66 above, in Eqs. 2 and 3, that triangle would have contributed three triplets, each with a different  
 67 weight. By contrast, in my framework, these three triplets all have the same weights  $\omega$  and  $g$ . In  
 68 summary, I associated specific meanings to the terms “triplet” and “triad”.

69 By computing  $\omega$  and  $g$  for all the triads, I reduced the 3D space  $\mathbb{A} = \{a, b, c \mid 1 \geq a \geq b \geq c \geq 0\}$   
 70 into a 2D space  $\mathbb{W} = \{\omega, g \mid 1 \geq a \geq b \geq c \geq 0\}$ . Because of the condition that  $a \geq b \geq c$  and  
 71 because of the non independence between the formulae for  $\omega$  and  $g$ , the domain  $\mathbb{W}$  exhibits a  
 72 specific shape (Fig. 1). For interpretation purposes, I then divided the  $\mathbb{W}$  space into four quadrants  
 73 (Fig. 1).

74 1. Quadrant 1 contained the “balanced triangles” where the three links are of similar strengths:  
 75  $a \approx b \approx c$ . In this situation,  $\omega \rightarrow -1$  and  $g \rightarrow 0$ . I drew the limits of quadrant 1 by

- 76 considering the two extreme cases  $b = 2c$  yielding  $\omega = 0$ ; and  $b = c = a/2$  yielding  $g =$   
 77  $1/4$  (Fig. 1).
- 78 2. Quadrant 2 contained the “weak triplets closed by a strong link” where  $a$  is much larger than  
 79 both  $b$  and  $c$ . I drew the limits of quadrant 2 by considering the two extreme cases  $b = 2c$   
 80 yielding  $\omega = 0$ ; and  $b = c = a/2$  yielding  $g = 1/4$  (Fig. 1).
- 81 3. Quadrant 3 contained the “strong triplets closed by a weak link” where  $c$  is much smaller  
 82 than both  $a$  and  $b$ . I drew the limits of quadrant 3 by considering the two extreme cases  $b =$   
 83  $2c$  yielding  $\omega = 0$ ; and  $c = 0$  plus  $b = a/2$  yielding  $g = 2/3$  (Fig. 1).
- 84 4. Quadrant 4 contained the “nearly false triplets” where  $a$  is extremely much larger than  $b$  and  
 85  $c$ . In this situation,  $\omega \rightarrow 0$  and  $g \rightarrow 1$ . I drew the limits of quadrant 4 by considering the two  
 86 extreme cases  $b = 2c$  yielding  $\omega = 0$ ; and  $c = 0$  plus  $b = a/2$  yielding  $g = 2/3$  (Fig. 1).

87 Next, I averaged  $\omega$  and  $g$  over all the triads in the network (Eq. 5a and 5b).

Eq. 5a	$C_{\omega} = \frac{1}{\tilde{\mathcal{T}}} \sum_{t=1}^{\tilde{\mathcal{T}}} \omega_t$
Eq. 5b	$C_g = \frac{1}{\tilde{\mathcal{T}}} \sum_{t=1}^{\tilde{\mathcal{T}}} g_t$

88  $\tilde{\mathcal{T}}$  is the number of triads. The difference between  $\tilde{\mathcal{T}}$  and the number of triplets  $\mathcal{T}$  depends on the  
 89 number of missing links in the network. The R script to compute  $C_{\omega}$  and  $C_g$  is provided in Appendix  
 90 S1.

91  $C_{\omega}$  and  $C_g$  are new network-level metrics. In this paper, I performed the initial evaluation of their  
 92 relevance. In particular, I assessed whether  $C_{\omega}$  and  $C_g$  captured information not captured by other  
 93 metrics. To do so, I simulated theoretical networks and I reanalyzed published animal association  
 94 networks.

## 95 **Material and methods**

## 96 Comparing with other network metrics

97 First I considered three metrics that quantify the shape of the distribution of link strengths:

- 98 1. The variance in link strength denoted  $Var(w)$
- 99 2. The skewness of the link strength distribution denoted  $Skew(w)$
- 100 3. The Gini coefficient of the link strength distribution denoted  $Gini(w)$

101 The Gini coefficient was here computed for all the links in the networks, whereas  $g_t$  from Eq. 4b is  
 102 for three links only. Next I computed five classic network-level summary metrics:

- 103 4. The binary edge density  $\mathcal{D}_0$ , corresponding to the proportion of non-missing links among all  
 104 the possible links in the network.
- 105 5. The binary network transitivity,  $\mathcal{C}_0$  (Eq. 1)
- 106 6. The weighted network transitivity with the geometric mean of link strengths as weight,  
 107  $\mathcal{C}_{1,geom}$  (Eq. 3a)
- 108 7. The network modularity  $\mathcal{M}$ . I delineated modules of individuals that interact more among  
 109 themselves than across modules using the short random walk community-finding algorithm:  
 110 routine `cluster_walktrap` (Pons and Latapy, 2005) from `igraph` (Csardi and Nepusz,  
 111 2006). I took into account the weight of the links when delineating the modules (argument  
 112 `weights = E(graph)$weight`). I then computed the modularity score following the  
 113 usual formula (Newman et al., 2002).
- 114 8. The network fragmentation  $\mathcal{F}$ . I computed the network fragmentation as the overall network  
 115 size divided by the average size of the modules that contained more than one individual. A  
 116 large fragmentation score meant that there were many small modules.

## 117 Simulating theoretical networks

118 To demonstrate the type of information that the new metrics convey in a unique way, I designed  
 119 scenarios corresponding to each of the four quadrants. I predicted that, using the new metrics, the  
 120 resulting networks would be categorized as belonging to the correct quadrant.

- 121 1. Scenario 1 with many balanced triangles. I generated full networks with no missing links.  
122 Then, at random, half of the links were strong (assigned a weight of 1) and half of the links  
123 were weak (assigned a weight drawn at random between 0 and 0.5). This procedure  
124 generated balanced triangles where either all three links were weak or all three links were  
125 strong.
- 126 2. Scenario 2 with many weak triplets closed by a strong link. I generated full networks with no  
127 missing links. Then, at random, 85% of the links were weak (assigned a weight of 0.1) and  
128 15% of the links were strong (assigned a weight drawn at random between 0.75 and 1). This  
129 procedure generated more triangles with two weak links and one strong link than if the link  
130 strength distribution was unimodal for example.
- 131 3. Scenario 3 with many strong triplets closed by a weak link. First I generated networks with a  
132  $\mathcal{C}_0$ -score of 0.8, using the `rguman` routine in R-package `sna` (Butts, 2020). Next, processing  
133 each triangle one by one in a random order, I made one of the three links weaker than the  
134 other two (by assigning a weight drawn at random between 0 and 0.1). This routine  
135 generated more triangles with two strong links and one weak link than if I omitted the  
136 reassignment step. Note that the routine processed each link multiple times because some  
137 links belonged to several triangles.
- 138 4. Scenario 4 with many false triplets. I generated scale-free networks (Barabási and Albert,  
139 1999) and then attributed random weights to the existing links. The weights were drawn  
140 from either a Gaussian or uniform distribution. Scale-free networks feature more false  
141 triplets (as defined in Fig. 1) than full networks with no missing link.

142 In addition to these 4 scenarios, I considered a few classic theoretical networks.

- 143 5. Full networks with no missing links (i.e.,  $\mathcal{D}_0 = 1$ ) in which I attributed random weights to the  
144 existing links. The link strengths were drawn from either a Gaussian or uniform distribution.

- 145 6. Half-full networks in which I removed half of the links at random (i.e.,  $\mathcal{D}_0 = 0.5$ ), and then  
146 assigned random weights to the remaining links. The link strengths were drawn from either a  
147 Gaussian or uniform distribution.
- 148 7. Lattice graphs. In lattice graphs, each individual is connected to a fixed number of immediate  
149 neighbors. I considered 2D, 3D, and 4D lattices. The link strengths were drawn from either a  
150 Gaussian or uniform distribution.
- 151 8. Modular graphs made of a set number of full networks (from 1 to 8 full networks) of which  
152 the link strengths varied between 0.75 and 1. These modules were then connected by weak  
153 links varying between 0 and 0.25, thereby ultimately obtaining a full network with no missing  
154 link, but a high modularity score as defined above.
- 155 9. Star graphs made of a central node connected to all other individuals by strong links varying  
156 between 0.9 and 1, while the remaining links were all weak and varied between 0 and 0.1.

157 I generated 100 examples of these 9 types of theoretical networks using `igraph` for R (Csardi and  
158 Nepusz, 2006). For each simulated network, I computed the two triadic edge dissimilarity scores and  
159 the 8 comparison metrics. I then performed a principal component analysis (PCA) of the 8  
160 comparison metrics to find axes of covariation between them. I assessed whether the triadic edge  
161 dissimilarity scores or the PCA scores could discriminate between the 9 network types, i.e., if one  
162 could determine the network type using only the triadic edge dissimilarity scores or only the PCA  
163 scores. Lastly, using linear models, I computed the proportion of the variance in the new metrics that  
164 was explained by the variation in the 8 comparison metrics.

### 165 **Using the new metrics to assess triadic mechanisms**

166 A “triadic mechanism” hereafter refers to any situation where an existing link influences the  
167 occurrence or the strength of links with third parties (e.g., Perry et al., 2004; Wittig et al., 2014). If  
168 that influence was negative, strong links would be surrounded by weak links, which in the parlance of

169 this paper corresponds to weak triplets closed by a strong link. The corresponding triads would score  
170 into quadrant 2.

171 As an illustration, I reanalyzed the association networks of experimental flocks of captive barnacle  
172 geese *Branta leucopsis* (Kurvers et al., 2013). I expected triadic mechanisms in geese flocks because  
173 bonded individuals are dominant over non-bonded individuals (Black and Owen, 1989; Kurvers et al.,  
174 2013). In free-ranging flocks, bonds correspond to family units or pairs. In the experiments of which I  
175 reanalyzed the data, bonds were artificially created by raising goslings in separate subgroups, before  
176 pooling the groups together, separately for males and females (Kurvers et al., 2013). In both sexes,  
177 after the subgroups were pooled together, the goslings still bonded preferentially with members of  
178 their initial subgroups (Kurvers et al., 2013). The recorded link strengths corresponded to the  
179 frequency at which individuals fed on the same patch during standardized observation sessions in an  
180 artificially patchy environment. There was no missing link in these networks, i.e., every dyad was  
181 recorded to associate at least once. Before analysis, I applied a logit transformation on the link  
182 strengths so they varied between 0 and 1 and had a median value of 0.5.

183 Because males are more aggressive and dominant than females in geese (Black and Owen, 1989), I  
184 predicted that the effect of the triadic mechanism should be stronger in males than females. To  
185 support that hypothesis, I quantified whether the male triads were more often in quadrant 2 or  
186 deeper into quadrant 2 (i.e., further away from the borders with the other quadrants) than the  
187 female triads using Hotelling's T-squared test (Hotelling, 1931).

### 188 **Using the new metrics as synthetic variables**

189 Even without triadic mechanism, the new metrics can prove useful because they summarize  
190 information that would otherwise require many metrics.

191 As an illustration, I reanalyzed and compared 6 datasets of grooming interactions within groups of  
192 rhesus macaques (*Macaca mulatta*), that I obtained through the ASNR repository (Sah et al., 2019).  
193 Rhesus macaques exhibit an "intolerant social style" characterized in particular by strong preference

194 for kin in social interactions such as grooming (Thierry, 2007). Therefore, the kinship structure of the  
195 group is expected to influence who grooms whom in a macaque group. Among the 6 datasets, there  
196 were three groups of nonkin that met each other after weaning age (all from the same study:  
197 Balasubramaniam et al., 2018). Then there were two free-ranging groups with naturally occurring  
198 kinship structure (Griffin and Nunn, 2012; Puga-Gonzalez et al., 2018) (these are from the same  
199 locality but different time periods). Another group was captive but featured locally born adult and  
200 immature offspring, therefore similar to the natural kinship structure (Massen and Sterck, 2013). In  
201 summary, I obtained data from free-ranging groups of kin, captive groups of kin, and captive groups  
202 of nonkin, making it possible to decipher the influence of kinship and captivity. I used Hotelling's T2  
203 statistic for this purpose.

204 Link strength corresponded to the frequency of recorded allogrooming interactions for each dyad  
205 during standardized observation sessions. I did not distinguish who was the groomer and who was  
206 the recipient. These networks featured many missing links, i.e., most of the dyads never groomed  
207 each other. Based on the results of the simulation exercise (see below), this is a situation where the  
208 overwhelming influence of missing links can cause binary metrics such as  $\mathcal{D}_0$  and  $\mathcal{C}_0$  to outperform  
209 the new metrics in terms of discriminatory power. In other words, this exercise represented a  
210 challenge for the new approach. Before analysis, I applied a logit transformation on the link strengths  
211 so they varied between 0 and 1 and had a median value of 0.5.

## 212 **Results**

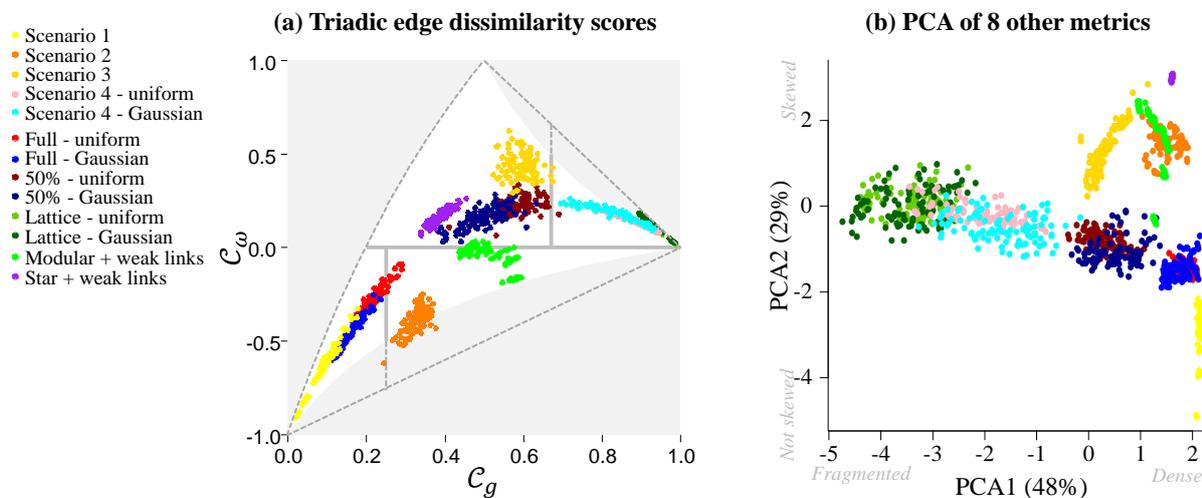
### 213 **Simulation study**

214 Visually, the two new metrics clearly discriminated among the 9 theoretical networks (Fig. 2a). In  
215 particular, Scenarios 1-4 neatly fell in the expected quadrants (Fig. 2a).

216 The PCA of the 8 comparison metrics also discriminated among the 9 theoretical networks (Fig. 2b).  
217 However, there was some overlap that did not occur with the new metrics, in particular between  
218 Scenarios 2 and 3 and between modular networks and Scenario 2. All in all, the main advantage of

219 the new metrics in terms of discriminatory power was the dimension reduction (2 metrics instead of  
 220 8), the removal of the PCA step, and the possibility to interpret the scores using the quadrants.  
 221 29% of the variance in  $\mathcal{C}_\omega$  was left unexplained by any of the 8 comparison metrics. This means that  
 222  $\mathcal{C}_\omega$  unambiguously picked up information not captured by any of the 8 comparison metrics. 4% of the  
 223 variance in  $\mathcal{C}_g$  was left unexplained by any of the 8 comparison metrics. This latter result mostly  
 224 stemmed from the correlation between  $\mathcal{C}_g$  and the edge density  $\mathcal{D}_0$  (Appendix S2). However, for full  
 225 networks with no missing link,  $\mathcal{D}_0$  was always equal to 1 but  $\mathcal{C}_g$  was still very variable. In other  
 226 words, the new metrics were most informative when there were no or few missing links in the study  
 227 networks.

228 Fig. 2: Simulation study results. (a) Network-level triadic edge dissimilarity scores (new method). The  
 229 network-level scores  $\mathcal{C}_\omega$  and  $\mathcal{C}_g$  can fall outside of the domain of definition of the triad-level scores  
 230  $\omega$  and  $g$  ( $\mathbb{W}$ , white shape). However, the network-level scores are always within the dotted lines. (b)  
 231 Principal component analysis (PCA) of 8 other network-level metrics (see list in the main text).



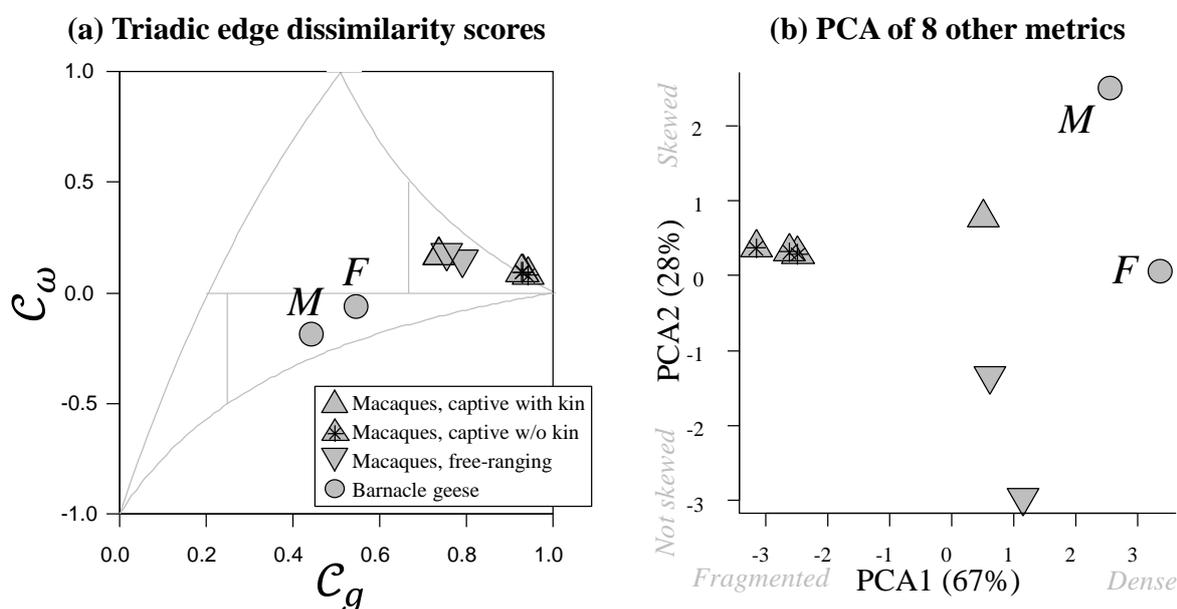
232

### 233 Using the new metrics to assess triadic mechanisms

234 Barnacle goose flocks of both sexes scored in quadrant 2, but as predicted, there were more weak  
 235 triplets closed by a strong link among males than females (Hotelling's T-squared on 2 and 2908  
 236 degrees of freedom: 65.3;  $P < 0.001$ ; Fig. 3a: "M" vs. "F"). This suggests that bonded males were  
 237 more dominant over non-bonded males than bonded females were over non-bonded females.

238 The 8 comparison metrics were also different between the male and female networks (Fig. 3b). In  
 239 particular, the distribution of link strength was more skewed in males than females (5.4 vs. 2.8). This  
 240 suggests that the excess of weak triplets closed by a strong link in males relative to females might  
 241 simply stem from the rarity of strong links among males. Repeat samples would be necessary to  
 242 decipher the effect of the rarity of strong links from the triadic mechanism hypothesis.

243 Fig. 3: Application cases: barnacle geese and rhesus macaques. (a) Network-level triadic edge  
 244 dissimilarity scores (new method). (b) Principal component analysis (PCA) of 8 other metrics (see list  
 245 in the main text). The PCA coefficients and axes are different from Fig. 2b.



246

### 247 Using the new metrics as synthetic variables

248 Among the macaque grooming networks, kinship explained more variation in the triadic edge  
 249 dissimilarity scores than captivity (Fig. 3a; Hotelling's T-squared for kin vs. nonkin: 3793.8 on 2 and  
 250 122780 degrees of freedom; for captive vs. free-ranging: 656.1; both  $P < 0.001$ ).

251 The edge density alone would have been sufficient to discriminate between groups of kin and groups  
 252 of nonkin:  $\mathcal{D}_0 = 0.62 \pm \text{SD } 0.02$  vs  $0.19 \pm 0.02$ . This suggests that in this case, triadic dissimilarities  
 253 depended mostly on the frequency of missing links, and thereby reaffirm the above conclusion that  
 254 the new metrics are most informative when there are no or few of these missing links. This  
 255 conclusion was further verified by permutation exercises (Péron, 2022a): when I permuted the link

256 strengths while keeping the edge density constant, the triadic edge dissimilarity scores and the above  
 257 conclusion remained largely unchanged (not shown). Lastly, the captive groups exhibited more  
 258 skewed distributions of link strength than the free-ranging groups ( $4.2 \pm \text{SD } 0.2$  vs.  $2.8 \pm 0.6$ ; Fig. 3b).  
 259 The triadic edge dissimilarity scores did not pick up this information (Fig. 3a: the symbols for captive  
 260 groups are closer to each other than in Fig. 3b). This means that in the macaque case contrary to the  
 261 goose case, the rarity of strong links did not contribute much to the triadic edge dissimilarities. The  
 262 dissimilarities mostly came from missing links rather than from variation in the strength of non-  
 263 missing links.

## 264 **Discussion**

265 In this paper, I weighted the computation of network transitivity in a way that takes into account the  
 266 strength of all three links in each triplet. Previously, network transitivity was either not weighted for  
 267 link strength, yielding what I denoted the binary transitivity  $\mathcal{C}_0$  (Newman et al., 2002). Alternatively,  
 268 the transitivity was weighted but using only the strength of the first two links, yielding what I  
 269 denoted the weighted transitivity  $\mathcal{C}_1$  (Opsahl and Panzarasa, 2009). Both  $\mathcal{C}_0$  and  $\mathcal{C}_1$  ignore the  
 270 strength of the closing link and the variation between links in the triplets. The new metrics in this  
 271 paper are based on the dissimilarity between the three links. I propose the phrase “triadic edge  
 272 dissimilarity scores”, the notation  $\omega$  and  $g$  for triad-level scores, and  $\mathcal{C}_\omega$  and  $\mathcal{C}_g$  for network-level  
 273 scores. The two set of metrics are complementary and should be interpreted together.

## 274 **What’s the use?**

275 Based on the simulations and case studies, the new metrics are most relevant for full networks with  
 276 no missing links. The main advantages of the new metrics appeared to be:

- 277 1) Ease of interpretation. The new metrics are geared to respond to the occurrence of triadic  
 278 mechanisms. Of course, as illustrated by the geese case, strong inference about the actual  
 279 occurrence of triadic mechanisms requires more than a point comparison. In particular,  
 280 triadic patterns may emerge as a by-product of the rarity of strong links rather than as the

281 consequence of the position of strong links relative to each other (Péron, 2022a).  
282 Nevertheless, in large comparative analyses, the new metrics should help deciphering the  
283 triadic mechanism hypothesis from alternative explanations. Studies into triadic mechanisms  
284 should in any case consider the new triadic edge dissimilarity scores to summarize the  
285 emerging triadic-level properties of the networks.

286 2) Repeatable dimension reduction. In both the geese and the macaque cases, the new metrics  
287 removed to need to explore a wide range of metrics or to train a multivariate model. Most  
288 importantly, in the simulation study, up to 30% of the information that the new metrics  
289 captured was not picked up by any of the 8 alternatives. This means that the new metrics  
290 indicated variation in the structure of weighed networks that the other metrics did not  
291 indicate. This was particularly the case in full networks with no missing link.

292 A major caveat is that, like all weighed network analysis techniques, the new triadic edge dissimilarity  
293 scores are sensitive to the way the link strengths are computed and rescaled. In this paper, I used a  
294 centered logit transformation, so that the link strengths varied between 0 and 1. Selecting another  
295 rescaling method would have changed the point value of the triadic dissimilarity scores, and for  
296 example changed the categorization in terms of quadrants. It is therefore important to standardize  
297 the computation of link strength as much as possible.

298 Another caveat is the influence of missing links on the triadic edge dissimilarity scores. I recommend  
299 sensitivity analyses where missing links are artificially replaced with increasingly strong links, or  
300 where an increasingly high threshold is used to consider a link as missing. In the simulation study,  
301 lattice graphs, which feature a lot of missing links, all grouped together tight into the rightmost  
302 corner of the domain of definition (Fig. 2), providing another illustration of the overwhelming  
303 influence of missing links on the dissimilarity scores. Lattice graphs have applications in behavioral  
304 ecology, e.g., they can describe the interactions between neighbors in territorial species (Péron,  
305 2022b). In this case, the appropriate level of investigation might be the individual rather than the

306 network, e.g., use the weighted degree centrality score, the individual transitivity score, etc. (Barrat  
307 et al., 2004).

## 308 **Conclusion**

309 In this paper, I followed up on an intuition that the contributions of triplets to network transitivity  
310 ought to be weighted according to the strength of all three links in the triplets. In doing so, I created  
311 new network-level metrics. Whether these new metrics will prove useful remains to be established.  
312 Indeed, pre-existing network metrics, if used together, appeared to already capture a lot of the  
313 information conveyed by the new metrics. But not all of it. I envision that the new approach could  
314 prove useful for studies into triadic mechanisms. Another, potentially widespread, use for these  
315 metrics would be to convey the variability of the strength of social connections in a way that  
316 complements the moment theory approach (Fisher, 1930) and the inequality approach (Gini, 1936)  
317 (Appendix S2, Fig. 3).

## 318 **Supplementary material**

319 Appendix S1: R script to compute the new metrics from the adjacency matrix of a social network

320 Appendix S2: Colinarity between the new metrics and the 8 comparison metrics in the simulation

321 study

## 322 **Literature cited**

- 323 Balasubramaniam, K., Beisner, B., Guan, J., Vandeleest, J., Fushing, H., Atwill, E., McCowan, B., 2018.  
324 Social network community structure and the contact-mediated sharing of commensal *E. coli*  
325 among captive rhesus macaques (*Macaca mulatta*). *PeerJ* 6, e4271.  
326 <https://doi.org/10.7717/peerj.4271>
- 327 Barabási, A.-L., Albert, R., 1999. Emergence of Scaling in Random Networks. *Science* 286, 509–512.  
328 <https://doi.org/10.1126/science.286.5439.509>
- 329 Barrat, A., Barthélemy, M., Pastor-Satorras, R., Vespignani, A., 2004. The architecture of complex  
330 weighted networks. *Proceedings of the National Academy of Sciences of the United States of*  
331 *America* 101, 3747–3752. <https://doi.org/10.1073/pnas.0400087101>
- 332 Black, J.M., Owen, M., 1989. Agonistic behaviour in barnacle goose flocks: assessment, investment  
333 and reproductive success. *Animal Behaviour* 37, 199–209. [https://doi.org/10.1016/0003-](https://doi.org/10.1016/0003-3472(89)90110-3)  
334 [3472\(89\)90110-3](https://doi.org/10.1016/0003-3472(89)90110-3)
- 335 Butts, C.T., 2020. sna: Tools for Social Network Analysis. R package version 2.6.
- 336 Csardi, G., Nepusz, T., 2006. The igraph software package for complex network research.  
337 *InterJournal, Complex Systems* 1695, 1–9. <https://doi.org/10.3724/sp.j.1087.2009.02191>

- 338 Fisher, R.A., 1930. The moments of the distribution for normal samples of measures of departure  
 339 from normality. *Proceedings of the Royal Society of London. Series A - Mathematical and*  
 340 *Physical Sciences* 130, 16–28. <https://doi.org/10.1098/rspa.1930.0185>
- 341 Gini, C., 1936. On the Measure of Concentration with Special Reference to Income and Statistics.  
 342 Colorado College Publication, Colorado Springs, General Series 208, 73–79.
- 343 Griffin, R.H., Nunn, C.L., 2012. Community structure and the spread of infectious disease in primate  
 344 social networks. *Evolutionary Ecology* 26, 779–800. <https://doi.org/10.1007/s10682-011-9526-2>
- 346 Hotelling, H., 1931. The Generalization of Student's Ratio. *Ann. Math. Statist.* 2, 360–378.  
 347 <https://doi.org/10.1214/aoms/1177732979>
- 348 Kurvers, R.H.J.M., Adamczyk, V.M.A.P., Kraus, R.H.S., Hoffman, J.I., van Wieren, S.E., van der Jeugd,  
 349 H.P., Amos, W., Prins, H.H.T., Jonker, R.M., 2013. Contrasting context dependence of  
 350 familiarity and kinship in animal social networks. *Animal Behaviour* 86, 993–1001.  
 351 <https://doi.org/10.1016/j.anbehav.2013.09.001>
- 352 Massen, J.J.M., Sterck, E.H.M., 2013. Stability and Durability of Intra- and Intersex Social Bonds of  
 353 Captive Rhesus Macaques (*Macaca mulatta*). *International Journal of Primatology* 34, 770–  
 354 791. <https://doi.org/10.1007/s10764-013-9695-7>
- 355 Newman, M.E.J., 2008. The mathematics of networks, in: Blume, L.E., Durlauf, S.N. (Eds.), *The New*  
 356 *Palgrave Encyclopedia of Economics*. Palgrave Macmillan, Basingstoke, UK, pp. 312–334.
- 357 Newman, M.E.J., Watts, D.J., Strogatz, S.H., 2002. Random graph models of social networks.  
 358 *Proceedings of the National Academy of Sciences of the United States of America* 99, 2566–  
 359 2572.
- 360 Opsahl, T., Panzarasa, P., 2009. Clustering in weighted networks. *Social Networks* 31, 155–163.  
 361 <https://doi.org/10.1016/j.socnet.2009.02.002>
- 362 Péron, G., 2022a. The frequency and position of stable associations offset their transitivity in a  
 363 diversity of vertebrate social networks. *Ethology* in press. <https://doi.org/10.1111/eth.13335>
- 364 Péron, G., 2022b. Reproductive skews of territorial species in heterogeneous landscapes. *Oikos*  
 365 in press. <https://doi.org/10.1111/oik.09627>
- 366 Perry, S., Barrett, H.C., Manson, J.H., 2004. White-faced capuchin monkeys show triadic awareness in  
 367 their choice of allies. *Animal Behaviour* 67, 165–170.  
 368 <https://doi.org/10.1016/j.anbehav.2003.04.005>
- 369 Pons, P., Latapy, M., 2005. Computing communities in large networks using random walks, in:  
 370 *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial*  
 371 *Intelligence and Lecture Notes in Bioinformatics)*. pp. 284–293.  
 372 [https://doi.org/10.1007/11569596\\_31](https://doi.org/10.1007/11569596_31)
- 373 Puga-Gonzalez, I., Ostner, J., Schülke, O., Sosa, S., Thierry, B., Sueur, C., 2018. Mechanisms of  
 374 reciprocity and diversity in social networks: A modeling and comparative approach.  
 375 *Behavioral Ecology* 29, 745–760. <https://doi.org/10.1093/beheco/ary034>
- 376 Sah, P., Méndez, J.D., Bansal, S., 2019. A multi-species repository of social networks. *Scientific Data* 6,  
 377 1–6. <https://doi.org/10.1038/s41597-019-0056-z>
- 378 Thierry, B., 2007. Unity in diversity: Lessons from macaque societies. *Evolutionary Anthropology* 16,  
 379 224–238. <https://doi.org/10.1002/evan.20147>
- 380 Wittig, R.M., Crockford, C., Langergraber, K.E., Zuberbühler, K., 2014. Triadic social interactions  
 381 operate across time: A field experiment with wild chimpanzees. *Proceedings of the Royal*  
 382 *Society B: Biological Sciences* 281. <https://doi.org/10.1098/rspb.2013.3155>
- 383

