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

Guillaume Péron

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

Introduction

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

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

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

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

Eq. 2	$$\mathcal{C}_{0,closing} = \frac{1}{\mathcal{T}} \sum_{t=1}^{\mathcal{T}} \gamma_t$$

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

Alternatively, following Opsahl and Panzarasa (2009), one can weigh the contributions of the triplets according to the strength of their first two links, yielding the weighted network transitivity $\mathcal{C}_1$. There 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	$\mathcal{C}_{1,max} = \frac{\sum_{t=1}^{\mathcal{T}} max(\alpha_t, \beta_t) \cdot \delta_t}{\sum_{t=1}^{\mathcal{T}} max(\alpha_t, \beta_t)}$

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

Eqs. 3a-3c only use two of the three links in each triplet; Eq. 2 only used one. There does not exist yet a method that takes all three links into account. To address that issue, I quantified the dissimilarity between the three links in each triplet. First, I reordered and renamed the three link strengths $\alpha_t, \beta_t, \gamma_t$ 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 compulsorily correspond to the closing link. I then computed the pairwise proportional difference in each triplet *t* (Eq. 4a) (Péron, 2022a).

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

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

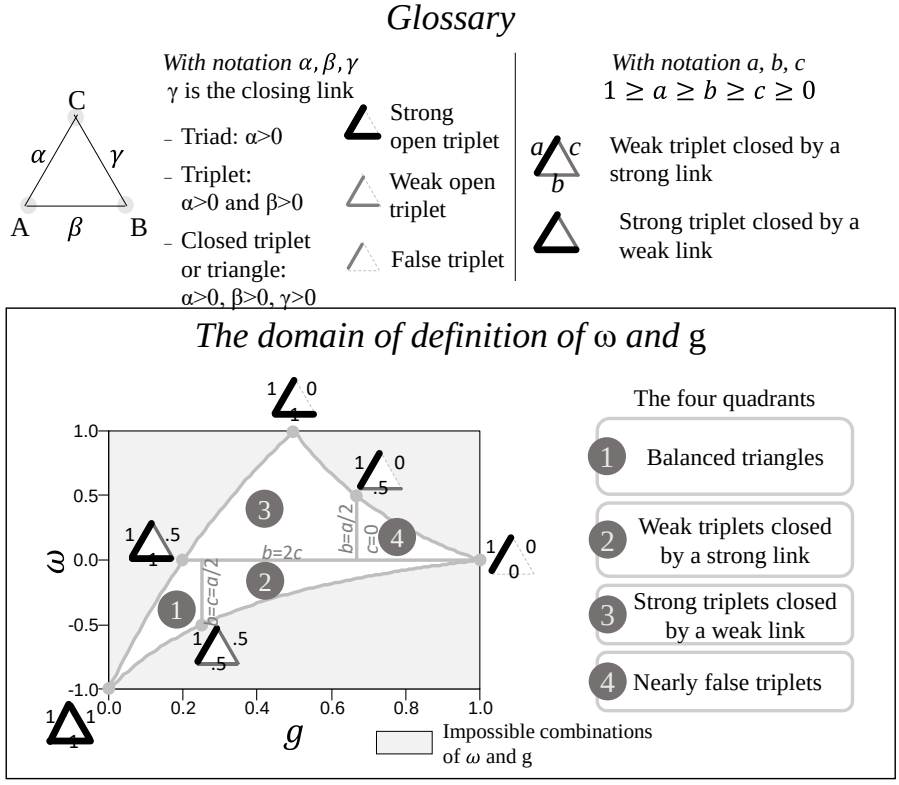
(instead of varying between 0 and 1). Then the only possible values for ω_t would be -1, 0, or 1 (instead of varying between -1 and +1).

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

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

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

Fig. 1: Illustrated glossary and representation of the domain of definition of ω and *g* as well as the four quadrants to distinguish between four different types of weighted triadic interactions.



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

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

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

considering the two extreme cases $b = 2c$ yielding $\omega = 0$; and $b = c = a/2$ yielding $g = 1/4$ (Fig. 1).

2. Quadrant 2 contained the "weak triplets closed by a strong link" where *a* is much larger than both *b* and *c*. I drew the limits of quadrant 2 by considering the two extreme cases $b = 2c$ yielding $\omega = 0$; and $b = c = a/2$ yielding $g = 1/4$ (Fig. 1).
3. Quadrant 3 contained the "strong triplets closed by a weak link" where *c* is much smaller than both *a* and *b*. I drew the limits of quadrant 3 by considering the two extreme cases $b = 2c$ yielding $\omega = 0$; and $c = 0$ plus $b = a/2$ yielding $g = 2/3$ (Fig. 1).
4. Quadrant 4 contained the "nearly false triplets" where *a* is extremely much larger than *b* and *c*. In this situation, $\omega \to 0$ and $g \to 1$. I drew the limits of quadrant 4 by considering the two extreme cases $b = 2c$ yielding $\omega = 0$; and $c = 0$ plus $b = a/2$ yielding $g = 2/3$ (Fig. 1).

Next, I averaged ω and *g* over all the triads in the network (Eq. 5a and 5b).

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

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

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

Material and methods

Comparing with other network metrics

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

1. The variance in link strength denoted *Var*(*w*)
2. The skewness of the link strength distribution denoted *Skew*(*w*)
3. The Gini coefficient of the link strength distribution denoted *Gini*(*w*)

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

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

Simulating theoretical networks

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

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

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

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

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

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

Using the new metrics to assess triadic mechanisms

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

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

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

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

Using the new metrics as synthetic variables

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

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

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

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

Results

Simulation study

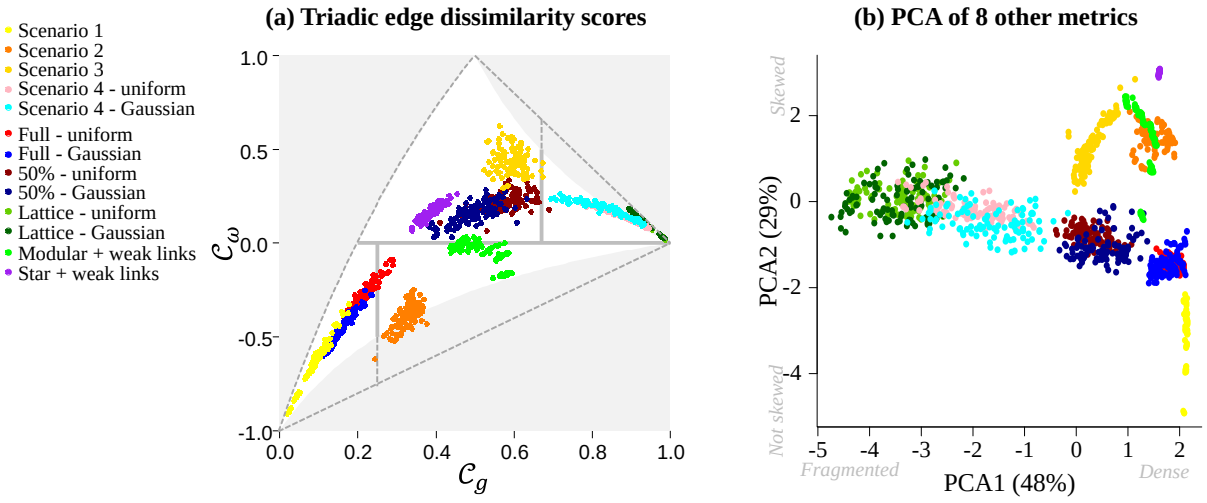
Visually, the two new metrics clearly discriminated among the 9 theoretical networks (Fig. 2a). In particular, Scenarios 1-4 neatly fell in the expected quadrants (Fig. 2a).

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

the new metrics in terms of discriminatory power was the dimension reduction (2 metrics instead of 8), the removal of the PCA step, and the possibility to interpret the scores using the quadrants.

29% of the variance in $\mathcal{C}_\omega$ was left unexplained by any of the 8 comparison metrics. This means that $\mathcal{C}_\omega$ unambiguously picked up information not captured by any of the 8 comparison metrics. 4% of the variance in $\mathcal{C}_g$ was left unexplained by any of the 8 comparison metrics. This latter result mostly stemmed from the correlation between $\mathcal{C}_g$ and the edge density $\mathcal{D}_0$ (Appendix S2). However, for full networks with no missing link, $\mathcal{D}_0$ was always equal to 1 but $\mathcal{C}_g$ was still very variable. In other words, the new metrics were most informative when there were no or few missing links in the study networks.

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

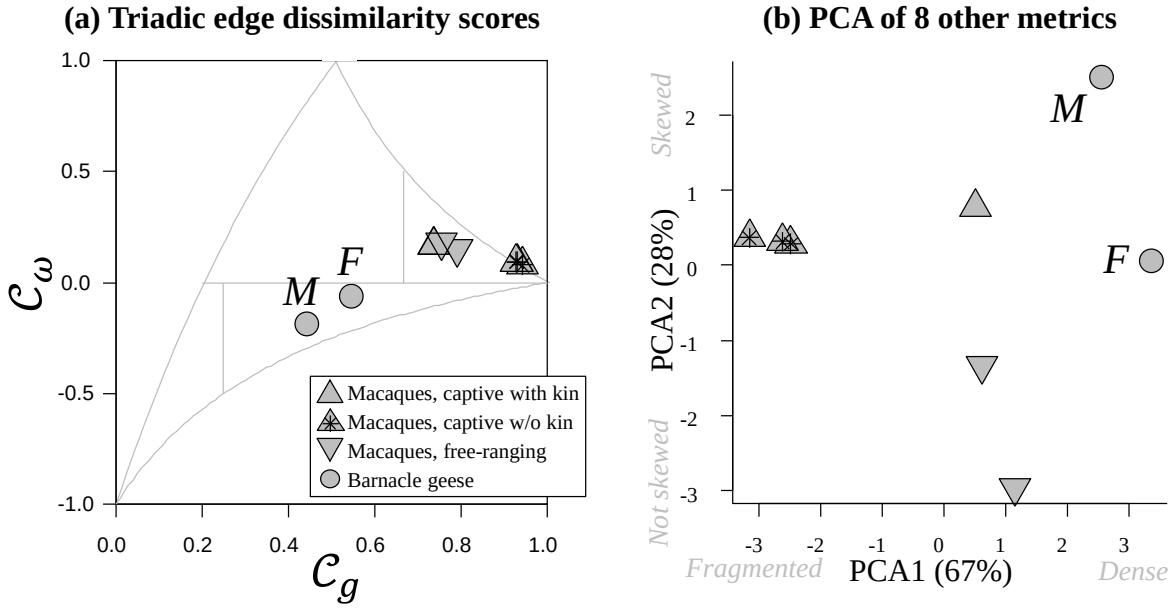


Using the new metrics to assess triadic mechanisms

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

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

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



Using the new metrics as synthetic variables

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

The edge density alone would have been sufficient to discriminate between groups of kin and groups of nonkin: $\mathcal{D}_0$= 0.62 ± SD 0.02 vs 0.19 ± 0.02. This suggests that in this case, triadic dissimilarities depended mostly on the frequency of missing links, and thereby reaffirm the above conclusion that the new metrics are most informative when there are no or few of these missing links. This 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 ± SD 0.2 vs. 2.8 ± 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 ω 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

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

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

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

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

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

Conclusion

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

Supplementary material

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

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

Literature cited

Balasubramaniam, K., Beisner, B., Guan, J., Vandeleest, J., Fushing, H., Atwill, E., McCowan, B., 2018. Social network community structure and the contact-mediated sharing of commensal E. coli among captive rhesus macaques (Macaca mulatta). PeerJ 6, e4271. https://doi.org/10.7717/peerj.4271

Barabási, A.-L., Albert, R., 1999. Emergence of Scaling in Random Networks. Science 286, 509–512. https://doi.org/10.1126/science.286.5439.509

Barrat, A., Barthélemy, 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. https://doi.org/10.1073/pnas.0400087101

Black, J.M., Owen, M., 1989. Agonistic behaviour in barnacle goose flocks: assessment, investment and reproductive success. Animal Behaviour 37, 199–209. https://doi.org/10.1016/0003-3472(89)90110-3

Butts, C.T., 2020. sna: Tools for Social Network Analysis. R package version 2.6.

Csardi, G., Nepusz, T., 2006. The igraph software package for complex network research. InterJournal, Complex Systems 1695, 1–9. https://doi.org/10.3724/sp.j.1087.2009.02191

Fisher, R.A., 1930. The moments of the distribution for normal samples of measures of departure from normality. Proceedings of the Royal Society of London. Series A - Mathematical and Physical Sciences 130, 16–28. https://doi.org/10.1098/rspa.1930.0185

Gini, C., 1936. On the Measure of Concentration with Special Reference to Income and Statistics. Colorado College Publication, Colorado Springs, General Series 208, 73–79.

Griffin, R.H., Nunn, C.L., 2012. Community structure and the spread of infectious disease in primate social networks. Evolutionary Ecology 26, 779–800. https://doi.org/10.1007/s10682-011-9526-2

Hotelling, H., 1931. The Generalization of Student's Ratio. Ann. Math. Statist. 2, 360–378. https://doi.org/10.1214/aoms/1177732979

Kurvers, R.H.J.M., Adamczyk, V.M.A.P., Kraus, R.H.S., Hoffman, J.I., van Wieren, S.E., van der Jeugd, H.P., Amos, W., Prins, H.H.T., Jonker, R.M., 2013. Contrasting context dependence of familiarity and kinship in animal social networks. Animal Behaviour 86, 993–1001. https://doi.org/10.1016/j.anbehav.2013.09.001

Massen, J.J.M., Sterck, E.H.M., 2013. Stability and Durability of Intra- and Intersex Social Bonds of Captive Rhesus Macaques (Macaca mulatta). International Journal of Primatology 34, 770–791. https://doi.org/10.1007/s10764-013-9695-7

Newman, M.E.J., 2008. The mathematics of networks, in: Blume, L.E., Durlauf, S.N. (Eds.), The New Palgrave Encyclopedia of Economics. Palgrave Macmillan, Basingstoke, UK, pp. 312–334.

Newman, M.E.J., Watts, D.J., Strogatz, S.H., 2002. Random graph models of social networks. Proceedings of the National Academy of Sciences of the United States of America 99, 2566–2572.

Opsahl, T., Panzarasa, P., 2009. Clustering in weighted networks. Social Networks 31, 155–163. https://doi.org/10.1016/j.socnet.2009.02.002

Péron, G., 2022a. The frequency and position of stable associations offset their transitivity in a diversity of vertebrate social networks. Ethology in press. https://doi.org/10.1111/eth.13335

Péron, G., 2022b. Reproductive skews of territorial species in heterogeneous landscapes. Oïkos in press. https://doi.org/10.1111/oik.09627

Perry, S., Barrett, H.C., Manson, J.H., 2004. White-faced capuchin monkeys show triadic awareness in their choice of allies. Animal Behaviour 67, 165–170. https://doi.org/10.1016/j.anbehav.2003.04.005

Pons, P., Latapy, M., 2005. Computing communities in large networks using random walks, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). pp. 284–293. https://doi.org/10.1007/11569596_31

Puga-Gonzalez, I., Ostner, J., Schülke, O., Sosa, S., Thierry, B., Sueur, C., 2018. Mechanisms of reciprocity and diversity in social networks: A modeling and comparative approach. Behavioral Ecology 29, 745–760. https://doi.org/10.1093/beheco/ary034

Sah, P., Méndez, J.D., Bansal, S., 2019. A multi-species repository of social networks. Scientific Data 6, 1–6. https://doi.org/10.1038/s41597-019-0056-z

Thierry, B., 2007. Unity in diversity: Lessons from macaque societies. Evolutionary Anthropology 16, 224–238. https://doi.org/10.1002/evan.20147

Wittig, R.M., Crockford, C., Langergraber, K.E., Zuberbühler, K., 2014. Triadic social interactions operate across time: A field experiment with wild chimpanzees. Proceedings of the Royal Society B: Biological Sciences 281. https://doi.org/10.1098/rspb.2013.3155