Commentary

The role of weighted and topological network information to understand animal social networks: a null model approach

Robert W. Rankin [a], *, Janet Mann [b, c], Lisa Singh [d], Eric M. Patterson [b], Ewa Krzyszczyk [b], Lars Bejder [a]

[a] Cetacean Research Unit, School of Veterinary and Life Sciences, Murdoch University, Murdoch, WA, Australia
[b] Department of Biology, Georgetown University, Washington DC, U.S.A.
[c] Department of Psychology, Georgetown University, Washington DC, U.S.A.
[d] Department of Computer Science, Georgetown University, Washington DC, U.S.A.

Network null models are important to drawing conclusions about individual- and population-(or graph) level metrics. While the null models of binary networks are well studied, recent literature on weighted networks suggests that: (1) many so-called 'weighted metrics' do not actually depend on weights, and (2) many metrics that supposedly measure higher-order social structure actually are highly correlated with individual-level attributes. This is important for behavioural ecology studies where weighted network analyses predominate, but there is no consensus on how null models should be specified. Using real social networks, we developed three null models that address two technical challenges in the networks of social animals: (1) how to specify null models that are suitable for 'proportion-weighted networks' based on indices such as the half-weight index; and (2) how to condition on the degree- and strength-sequence and both. We compared 11 metrics with each other and against null-model expectations for 10 social networks of bottlenose dolphin, *Tursiops aduncus*, from Shark Bay, Australia. Observed metric values were similar to null-model expectations for some weighted metrics, such as centrality measures, disparity and connectivity, whereas other metrics such as affinity and clustering were informative about dolphin social structure. Because weighted metrics can differ in their sensitivity to the degree-sequence or strength-sequence, conditioning on both is a more reliable and conservative null model than the more common strength-preserving null-model for weighted networks. Other social structure analyses, such as community partitioning by weighted Modularity optimization, were much less sensitive to the underlying null-model. Lastly, in contrast to results in other scientific disciplines, we found that many weighted metrics do not depend trivially on topology; rather, the weight distribution contains important information about dolphin social structure.

© 2015 The Association for the Study of Animal Behaviour. Published by Elsevier Ltd. All rights reserved.

The social network paradigm is increasingly being used to study the behavioural ecology of social animals. It holds the promise of expanding the field from investigations about the presence and fitness consequences of associations to understanding the pattern of associations, including how network structures persist over time or serve ecological functions. For example, whereas researchers have plenty of ideas why animals may be social (e.g. for antipredator defence, foraging) and can demonstrate that one's position in the network can lead to higher fitness (Stanton & Mann, 2012), it is more controversial to posit functional importance to structural properties of networks themselves. Consider bottlenose dolphins *Tursiops* spp., where patterns such as triangle-closure, assortativity by sociality, and the presence of 'social brokers' between different subgroups are features that are more than just individual-level tendencies to have a certain number of associates. Hypotheses about the ecological function of such 'social structure' are few and tentative (Pinter-Wollman et al., 2014), such as facilitating information transmission (Allen, Weinrich, Hoppitt, & Rendell, 2013; Lusseau, 2003). Behavioural ecology remains significantly challenged by the difficulty of defining and measuring social structure. Here, we try to identify which network metrics may be informative about dolphin social structure, and which are redundant to individual-level differences in sociality, such as individual

* Correspondence: R. W. Rankin, Cetacean Research Unit, Department of Veterinary and Life Sciences, Murdoch University, 90 South Street, Murdoch 6150 WA, Australia.
E-mail address: robertw.rankin@gmail.com (R. W. Rankin).

differences in 'degree' (number of associates) and 'strength' (sum of weights of associates).

We employ a null-model approach: to calculate expectations of weighted network metrics while conditioning on individual-level properties, and compare such expectations to the observed metrics. For example, if we observe a network and accept its empirical degree-sequence (the number of connections each individual has in the network), then how are we to interpret other network metrics and judge whether they provide evidence of an underlying organizing structure? What metric values are likely even when there are no true underlying structures like 'clusters' or true organizing principles like 'assortativity'? It is by comparing metric values to their null-model expectations that allows us to find potentially meaningful metrics that actually measure aspects of higher-order social structure, or whether metrics are merely redundant to individual-level attributes.

Despite the simplicity of the null-model approach, there is little consensus on how to calculate expectations of network metrics. Two challenges emerge: (1) one must decide what properties to condition upon (e.g. strength-sequence, degree-sequence, both or others); and (2) one needs a way to calculate expectations without biasing results. Behavioural ecologists primarily address these challenges by conditioning on aspects of survey design, observation error and sociality (Whitehead, 2008), and primarily use permutation methods to calculate expectations under random associations (Bejder, Fletcher, & Bräger, 1998). Outside of behavioural ecology, there is a growing suite of 'random-graph' algorithms (Ansmann & Lehnertz, 2011; Leskovec, Chakrabarti, Kleinberg, Faloutsos, & Ghahramani, 2010; Prettejohn, Berryman, & McDonnell, 2011; Serrano, Boguñá, & Pastor-Satorras, 2006; Watts & Strogatz, 1998) which emphasize core properties such as the degree-sequence, strength-sequence, network size and density; they have shown that unless such properties are held constant across random-graphs, then any conclusions about network properties will just reflect variation in the degree-sequence, strength-sequence, network-size, etc. There is a near consensus about the need to condition on the degree-sequence for binary networks, but the matter is more controversial for weighted-networks, and one's conclusions are sensitive to such conditioning (Garlaschelli & Loffredo, 2009; Mastrandrea, Squartini, Fagiolo, & Garlaschelli, 2014).

This paper follows in the spirit of Garlaschelli and Loffredo (2008), to calculate metric expectations based on null models that assume only basic individual-level properties, and to do so by generating an 'ensemble' of random networks based on the Exponential Random Graph formulation. In contrast, our equations are valid for proportion weighted-networks, $w_{ij} \in [0,1]$ used in behavioural ecology (Cairns & Schwager, 1987). An advantage of this method is its principle of 'maximum entropy' to produce an ensemble of networks that makes the fewest assumptions, thereby ensuring that we have randomized all other topological and weight patterns that could be misconstrued as social structure. The method is similar to permutation-based and random-graph algorithms in that they offer null models conditioned on simple assumptions. However, permutation-based and random-graph algorithms do not necessarily guarantee that their ensembles do not have structural correlations or biases that are mere artefacts of the randomization algorithm (Garlaschelli & Loffredo, 2008).

We specified three different null models that are constrained to the degree-sequence (Topology Null Model, TNM), strength-sequence (Weighted Null Model, WNM), and both (Mixed Null Model, MNM) for proportion-weighted networks. We derive the probability distributions for the TNM, WNM and MNM and apply them to 10 years of association data from a well-studied population of bottlenose dolphins, *Tursiops* cf. *aduncus*, in Shark Bay, Western Australia (Mann, Stanton, Patterson, Bienenstock, & Singh, 2012).

Three aims of this study are:

(1) to compare how well observed individual-level network metrics correspond to null-model expectations, for three null models;
(2) to compare averages of whole network metrics to their null-model expectations, especially as a function of network size N (an ongoing controversy in network science; e.g. see Anderson, Butts, & Carley, 1999);
(3) to compare how inferences about network community structure differ according to the null-model used via Modularity optimization (Squartini & Garlaschelli, 2011).

For aims 1 and 2, we focus on 11 popular node-level metrics used in analyses of animal societies, such as clustering, affinity, centrality, dispersion and connectivity.

The method of Garlaschelli and Loffredo (2009) caused considerable upset in other disciplines. For example, Garlaschelli and Loffredo (2009) discovered that some weighted measures 'inherit' trivially from topological features and called for 'a systematic redefinition of weighted network properties', while Mastrandrea et al. (2014) noted that 'the strength sequence is in general uninformative about the higher-order properties of the network'. The implications for behavioural ecologists are that: (1) many weighted-network metrics may not depend on weights per se and actually depend on the underlying binary, topological patterns; and (2) that many metrics of higher-order structure are not significantly different from (and often highly corrected with) the values one would expect from networks with only individual-level constraints (degree and/or strength).

The above claims were supported over a broad range of networks, such as food-webs, online social networks and financial/trading networks. If the conclusions of Garlaschelli and Loffredo (2009) and Mastrandrea et al. (2014) generalize to animal social networks, then it would be a setback to behavioural ecology studies based on network metrics. For example, if clustering and affinity metrics were merely redundant to individual-level attributes, and did not measure higher-order properties as intended, they would produce misleading conclusions about 'social structure', as defined as higher-order structure that is more than the sum of individuals (Holland & Leinhardt, 1979; Faust, 2006). However, the methods and insights from integer-weighted networks cannot be accepted naively for proportion-weighted networks. We show that the eastern gulf Shark Bay dolphins stand as a contrary case to the many and varied networks considered by Garlaschelli and Loffredo (2009) and Mastrandrea et al. (2014).

METHODS

Data

Our data source is a 31-year long-term study of over 1500 individually identified bottlenose dolphins resident in the eastern gulf of Shark Bay, Western Australia (Mann et al., 2012). Associations among individual dolphins were estimated from opportunistically encountered groups during boat-based surveys, using a 10 m chain rule to define in-situ group membership (Smolker, Richards, Connor, & Pepper, 1992). We truncated the data to include noncalf individuals encountered at least five times each year within a constant spatial and temporal domain. The constant space-time domain was evaluated in the following way: (1) we included surveys that occurred between May and November; (2) per year, we calculated a minimum convex hull (MCH) which enveloped all georeferenced encounters; (3) we used the spatial intersection of all 10 per-year MCHs to define a small region of

consistent spatial surveillance; (4) we smoothed the perimeter of the constant spatial domain by adding a 2500 m spatial buffer. Annual pairwise associations were estimated using the Half-weight index (Cairns & Schwager, 1987) to generate networks for each year between 2003 to 2012 (for a total of 10 networks). In total, 209 individuals were included. Forty-one per cent of individuals occurred in just one network and 4% occurred in all 10 networks. The number of nodes N per year varied from 18 individuals in 2005, to 122 individuals in 2004 (mean 61.4, SD 32). The range of network sizes facilitated our study of the behaviour of empirical and expected metrics under variable N. The number of survey days varied from 36 in 2005 to 98 in 2003 (mean 66.0, SD 19.5).

Null-models

In the following description, we denote a weighted network as being fully described by its weight matrix $\mathbf{W}$ with individual nodes i and the strength of association between nodes i and j denoted $w_{ij} \in [0,1]$. It is symmetric and undirected, $w_{ij} = w_{ji}$. Some key metrics are strength $s_i = \sum_{i \neq j}^{n} w_{ij}$ (the sum of a node's weights given that a node has n neighbours) and the strength-sequence $\vec{s} = (s_1, s_2, ..., s_N)$, as well as degree $k_i = \sum_{i \neq j}^{n} I[w_{ij} > 0]$ (the sum of a node's binary connections), and the degree-sequence $\vec{k} = (k_1, k_2, ..., k_N)$.

The goal of this paper is to calculate an unbiased expectation of a network metric, c_i (e.g. clustering coefficient) here indexed to individual i, but whose calculation may include all pairs or triplets in the network. To do this, we need a method to generate the ensemble of networks $\mathscr{W}$ whose expected strength- and/or degree-sequence are the same as our observed weighted network, i.e. $\vec{\theta}_{obs} = \mathbb{E}[\vec{\theta}]$. The expectation $\mathbb{E}[c_i]$ is calculated over the entire ensemble, whereby each constituent network ($\mathbf{W}'$) in the ensemble contributes a value in proportion to its probability $P(\mathbf{W}' | \vec{\theta})$ conditional on property $\vec{\theta}$. The ensemble is approximated by drawing a large number of random networks from $P(\mathbf{W} | \vec{\theta})$. The expected value of a metric value is approximated by taking the average over the random draw of networks, denoted $\langle c_i \rangle$; i.e. $\mathbb{E}[c_i] = \sum_{W' \in \mathscr{W}} p(W' | \vec{\theta}) c_i(W') \approx \langle c_i \rangle$.

The challenge is to define $P(\mathbf{W} | \vec{\theta})$ and sample from it. The underlying assumption of the distribution is that the initial cost of forming an association is greater than maintaining an existing tie, as developed by Garlaschelli and Loffredo (2008; 2009), but unlike their work, our equations respect the [0,1] bound for weights in a proportion-weighted network. The mathematical details are in the Appendix. The key point is that by specifying the probability distribution according to the Exponential Random Graph model, we ensure that the resulting ensemble of networks is maximally random in all regards aside from the user-specified constraints (e.g. $\vec{s}$, $\vec{k}$). For example, the topology (who is connected to whom) is maximally randomized, as well as the pattern of weights on this topology.

Our analysis was repeated for 10 networks based on annual field surveys conducted during the austral winter and spring. One example network is shown in Fig. 1. The steps of the analyses were as follows. (1) We derived the probability and expectation functions of weights for the three null models (TNM, WNM and MNM; detailed in the Appendix). (2) We specified the maximum entropy ensemble of networks by setting the ensemble average properties (strength- and/or degree-sequence) to the observed values and solving node-specific parameters (known in the field of constrained-optimisation as 'Lagrangians', also detailed in the Appendix). Solving these parameters fully specifies the ensemble

and was repeated for each observed network and each null-model. (3) We approximated the maximum entropy ensemble for each empirical network by taking a large sample (1000) of random networks drawn from $P(\mathbf{W})$, using the probability distribution function of weights from step 2. The construction of each ensemble was performed for each year's empirical network and each null-model, for a total of 30 network ensembles. (4) Metrics' expectations were calculated using the ensembles and were compared to the observed networks' values; we also looked for patterns among the different metrics, their residuals and by sex. The metric expectations were calculated per metric × per individual × per year × per null-model, but we simplified the comparison by looking for global patterns over all years, by metric and null-model, resulting in the 11×3 scatter plots in Figs. 2–4 and correlation coefficients in Table 1. (5) We calculated population-level averages and intervals of metrics and compared these to the population averages and intervals of the observed networks. (6) We compared the communities implied by each null-model by performing community partitioning by Modularity optimization, using the expectations of weights according each null-model. Inferred communities were compared within each year for all pairs of null models.

Node-level Metrics

For each node-per-year, we calculated 11 weighted metrics. We focused on affinity and clustering/transitivity metrics, which are intensely studied second- and third-order properties (Barrat, Barthélemy, Pastor-Satorras, & Vespignani, 2004; Garlaschelli & Loffredo, 2009; Mastrandrea et al., 2014; Serrano et al., 2006). We also calculated some metrics that are recommended for animal societies (Wey, Blumstein, Shen, & Jordán, 2008; Whitehead, 2008) as well as weighted versions of Freeman's centrality indices (Freeman, 1979).

The expectations of the 11 different metrics were calculated for each node-per-year and per null-model (TNM, WNM, MNM). To simplify the comparisons of observed versus expected, we summarize our results by metric and null-model, according to: (1) the Pearson's correlation coefficient R_0; and (2) the partial correlation coefficient $R_{1|year}$ (Cohen, West, Aiken, & Cohen, 2003), after removing variation explained by year (and therefore different network sizes). In both cases, it was individual dolphins (per year) that constitute individual data points for the correlation statistics. The two R statistics helped contextualize the degree to which the null-model expectations were similar to observed values across all individual dolphins (i.e. high R_0 and high $R_{1|year}$) or just for population averages (high R_0 but lower $R_{1|year}$).

We used the following weighted metrics.

C^c weighted closeness centrality, a measure of short paths to all other nodes (Freeman, 1979; Opsahl, Agneessens, & Skvoretz, 2010). We predicted that all null models should be able to predict node-level values of C^c, inasmuch as individuals' strength and degree are inversely related to the path length to other nodes, e.g. if high-degree and high-strength individuals have a greater tendency to have short paths across the network.

C^b weighted betweenness centrality, how well an individual connects different subsets or clusters (Freeman, 1979; Opsahl et al., 2010). High C^b-individuals may be important for global propagation of information, disease, etc. Because randomized graphs should not have substructures, C^b is not expected to be well predicted by any null-model.

C^e eigenvector centrality, a measure of the centrality of a node's neighbours (Butts, 2008). It is the most succinct representation of the total variation in a network along a single dimension (i.e. the dominant social dimension). Inasmuch as a network's total variation is driven by individuals' tendency to associate with others of a

similar degree/strength (i.e. degree/strength-assortativity) C^e should be fairly redundant to degree- and strength-preserving null models.

C^L Laplacian centrality, a measure of the sensitivity of a network to node deletions (Qi, Fuller, Wu, Wu, & Zhang, 2012). It addresses an enduring interest in wildlife conservation about the resilience of animal societies to targeted removal of important individuals (Williams & Lusseau, 2006). For example, odontocetes societies may be particularly resilient to deletions because of the peculiar degree-distribution (flat at small degrees and scale-free otherwise; Lusseau, 2003). If this is the case, then the degree-sequence may be sufficient to produce expected values of C^L close to their observed values.

k^{nn} degree-affinity, the weighted average nearest neighbour degree. It is a second-order measure used to diagnose degree assortativity: whether individuals preferentially associate with those who have a similar degree.

s^{nn} strength-affinity, a second-order measure similar to k^{nn}, but estimates neighbours' strength instead of degree. It is the recommended affinity metric for animal societies (Kasper & Voelkl, 2009; Whitehead, 2008). Mastrandrea et al. (2014) demonstrated the remarkable predictability of k^{nn} and s^{nn} from degree- and strength-sequence null models for a variety of social and nonsocial networks, and we anticipated similar results.

c^H Holme clustering coefficient, a third-order measure of link transitivity, as recommended by Whitehead (2008), which compares the number of closed-triangles versus triplets around a node. Null-models should have no great ability to predict the presence of cliques, clustering, or the lack thereof (unlike the results of Mastrandrea et al., 2014).

c^O Opsahl geometric-mean clustering coefficient, another third-order measure of transitivity. While similar to c^H, it has a different way to score the number of triangles versus triplets around a node, and was designed to serve generally in weighted network analyses (Opsahl & Panzarasa, 2009).

T^* cost-integrated triangle count, another measure of transitivity and the simplest third-order metric. T^* is simply the count of triangles around a node, but is generalized for weighted networks via the technique of cost integration (Ginestet, Nichols, Bullmore, & Simmons, 2011).

Y disparity, a measure of the dispersion of weights around a node. It helps to contextualize whether an individual has just a few strong connections or many weak connections (Barthélemy, Barrat, Pastor-Satorras, & Vespignani, 2005; Whitehead, 2008). The metric compares an individual's strength metric to its degree, which for proportion-weighted weights are inherently related: $s_i \leq k_i$. Therefore, disparity based on proportion-weights is expected to be predictable when both strength- and degree-sequences are set, but not when either is randomized.

r^S reach, a measure of indirect correctness, calculated as the total strength of a node's neighbours (Whitehead, 2008). Because the measure depends heavily on the strength of individuals, it is expected to be well predicted by the strength-preserving models, WNM and MNM, but not TNM.

Metric Correlations

Social structure is often identified (and sometimes defined) by patterns among network metrics, such as k versus k^{nn} for degree-(dis)assortativity. We compared network metrics pairs by inspecting plots and calculating a nonlinear metric of association, the Maximal Information Coefficient (MIC; Reshef et al., 2011) whereby $MIC \in [0,1]$ shows perfect agreement at 1. We also used metric residuals (i.e. the difference between an observed metric and its null-model expectation) as another candidate for identifying structure. Such residuals may be more informative than observed
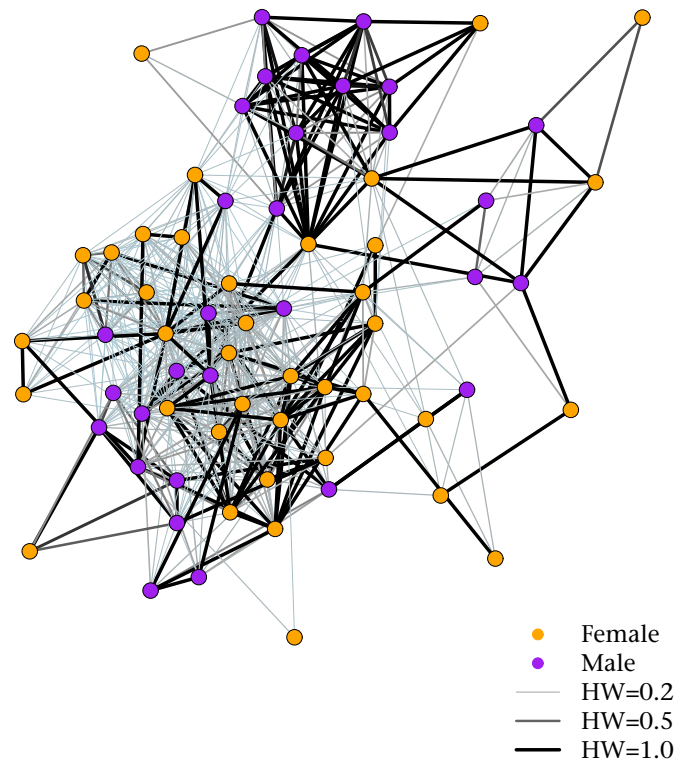
Figure 1. Example visualization of the social network of eastern gulf Shark Bay bottlenose dolphins for the year 2007. Associations are calculated from group encounters using the Half-Weight index.

metrics in that they may partially standardize metrics for different network size and densities. We also explored the distributions of metrics by males and females. We computed the Jensen-Shannon distance ($D_{JS} \in [0,0.69]$) to quantify which metric distributions were strongly different by males and females. We note that the residuals, MIC and D_{JS} are not fundamental to the null-model approach. Rather, the null-model expectations can help uncover patterns that may be difficult to discern among the raw metrics.

Lastly, we also compared the network metrics to other non-network metrics to probe for artefacts of the survey design or sampling procedure, such as comparing metrics to simple indices of over-representation of certain individuals in the data. We used the number of surveys and years in which a dolphin was encountered, i^{survey} and i^{years} respectively.

Model Comparisons

We estimated null-model support by calculating likelihood values and AIC weights (Mastrandrea et al., 2014). The comparisons were made among the three null models, per empirical network. The log-likelihood is conditional on dolphin-specific parameters $\vec{x}, \vec{y}$, also called Lagrangian multipliers, which are used to ensure that each dolphins' degree-/strength-sequence have equal values between their observed and ensemble averages (see Appendix). The log-likelihood is:

$$\log \mathscr{L}(\mathbf{W}|\vec{x}, \vec{y}) = \log P(\mathbf{W}|\vec{x}, \vec{y}) = \sum_{i<j} \log q(w_{ij}|\vec{x}, \vec{y}) \tag{1}$$

Calculation of the AIC is $-2\log \mathscr{L}(\mathbf{W}|\vec{x}, \vec{y}) + 2k$, where k is the number of model parameters. For the TNM and WNM, there is one parameter for each individual in the network ($k = N$), while there
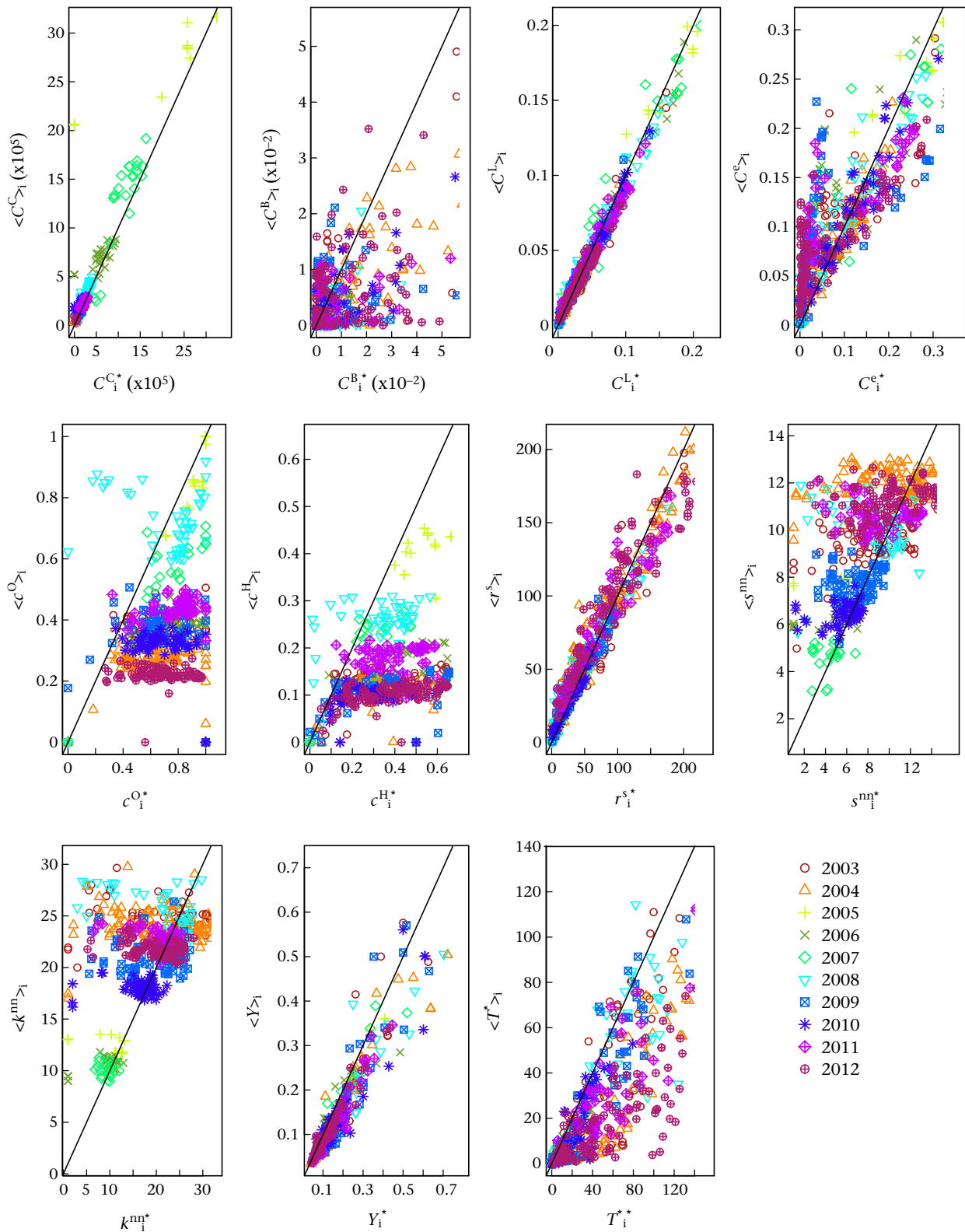
Figure 2. Network metrics of individual dolphins in Shark Bay, observed (x-axis) versus expected values (y-axis) according to the Mixed Null Model, conditioning on strength- and degree-sequence. C^c closeness, C^b betweenness, C^L Laplacian centrality, C^e eigenvector centrality, c^O Opsahl clustering, c^H Holme clustering, r^S reach, s^{nn} strength-affinity, k^{nn} degree-affinity, Y disparity, T^* triangle-count. Different colours and symbols represent different networks (i.e. metrics are calculated within the context of a single network).
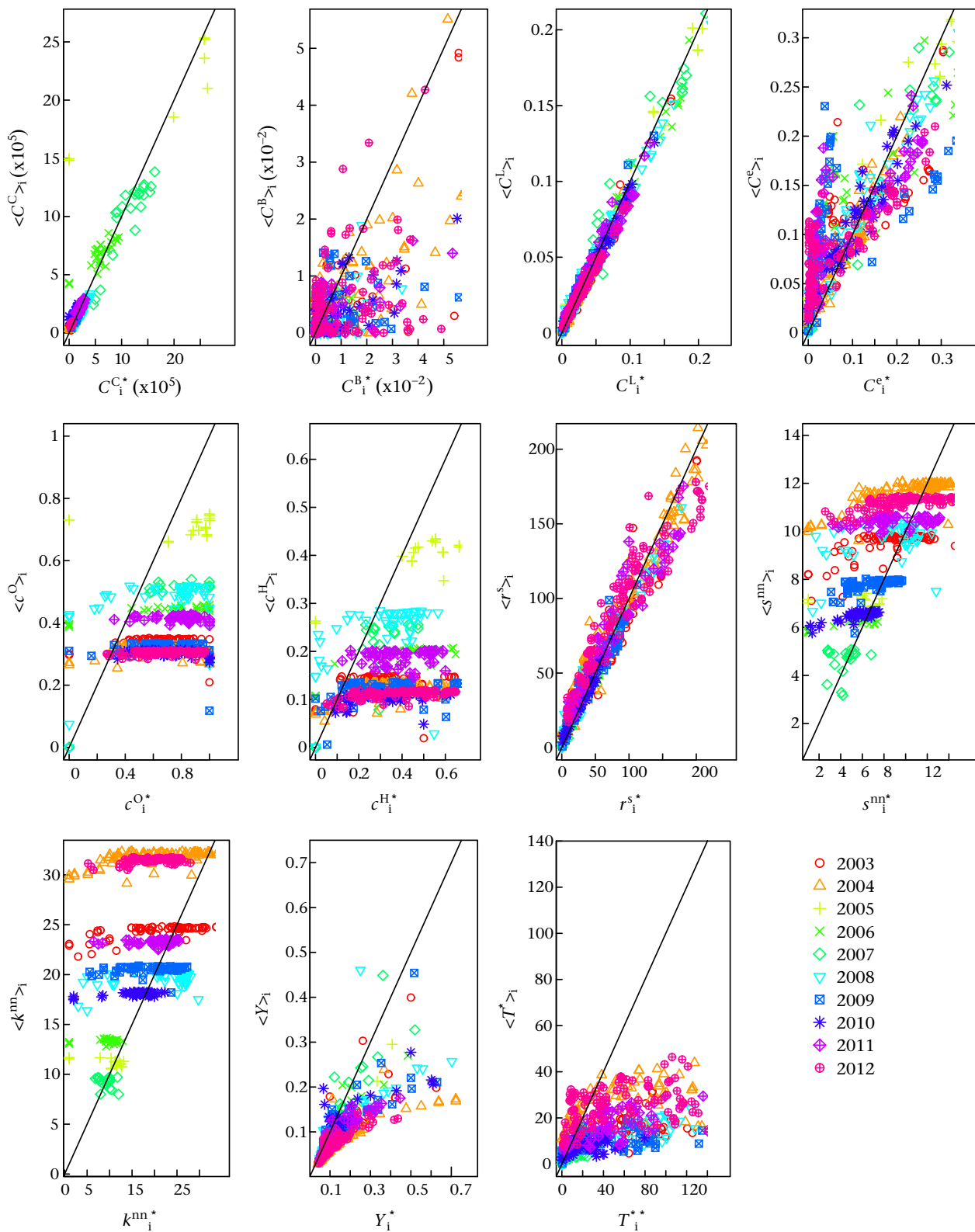
Figure 3. Network metrics of individual dolphins in Shark Bay, observed (x-axis) versus expected values (y-axis) according to the Weighted Null Model, conditioning on strength-sequence. C^c closeness, C^b betweenness, C^L Laplacian centrality, C^e eigenvector centrality, c^O Opsahl clustering, c^H Holme clustering, r^S reach, s^{nn} strength-affinity, k^{nn} degree-affinity, Y disparity, T^* triangle-count. Different colours and symbols represent different networks (i.e. metrics are calculated within the context of a single network).
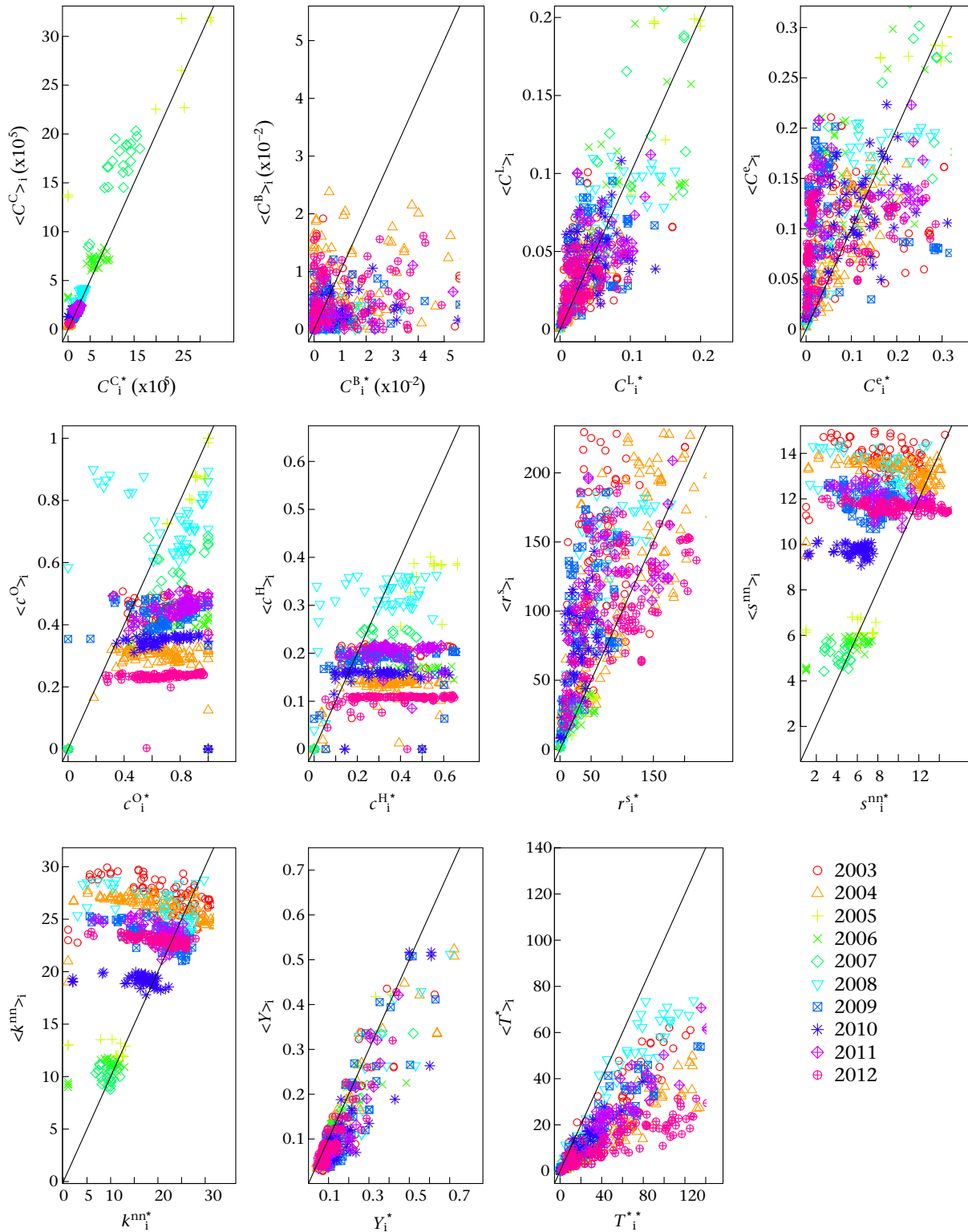
Figure 4. Network metrics of individual dolphins in Shark Bay, observed (x-axis) versus expected values (y-axis) according to the Topology Null Model, conditioning on degree-sequence. C^c closeness, C^b betweenness, C^L Laplacian centrality, C^e eigenvector centrality, c^O Opsahl clustering, c^H Holme clustering, r^S reach, s^{nn} strength-affinity, k^{nn} degree-affinity, Y disparity, T^* triangle-count. Different colours and symbols represent different networks (i.e. metrics are calculated within the context of a single network).

are two parameters for every individual dolphin in the MNM ($k = 2N$). The AIC weights sum to 1 for each empirical network, and high values suggest better agreement between model and observed data, while penalizing model complexity.

Community Structure

To assess how the different null models can influence inferences about fundamental network structure, such as clustering into

Table 1
Correlations between observed and expected node-level metric values, by metric and null-model

Node-level metric	TNM		WNM		MNM	
	R_0^2	$R_{1|year}^2$	R_0^2	$R_{1|year}^2$	R_0^2	$R_{1|year}^2$
Laplacian centrality	**0.633**	0.435	**0.985**	**0.979**	**0.983**	**0.976**
reach	0.38	0.343	**0.941**	**0.93**	**0.936**	**0.923**
closeness centrality	**0.921**	**0.724**	**0.937**	**0.74**	**0.925**	**0.612**
disparity	**0.845**	**0.836**	**0.682**	**0.704**	**0.874**	**0.865**
eigenvector centrality	0.29	0.213	**0.709**	**0.686**	**0.722**	**0.7**
triangle count	**0.664**	**0.753**	0.295	0.288	**0.674**	**0.702**
strength-affinity	0.153	0.017	0.421	0.281	0.346	0.034
betweenness centrality	0.081	0.045	0.317	0.287	0.302	0.27
degree-affinity	0.169	0.156	0.161	0.204	0.204	0.017
clustering (Holme)	0.057	0.16	0.072	0.155	0.142	0.247
clustering (Opsahl)	0.114	0.184	0.078	0.104	0.114	0.181

R_0^2 is the squared zero-order correlation pooled over 10 years of association data, per null-model (TNM, WNM and MNM). $R_{1|year}^2$ is the squared partial-correlation, which removes the trend over years and shows the within-year correlations. The TNM conditions on the degree-sequence, the WNM conditions on the strength-sequence, and MNM conditions on both. Value >0.6 are bold for emphasis.

communities, we performed network partitioning by Modularity optimization, per year, according to four null models: the TNM, WNM, MNM, plus the default 'bilinear' strength-preserving model of Newman (2004). Modularity is a measure of fragmentation of a network into densely connected clusters with sparse connections between clusters, and is a popular optimization criterion for finding discrete communities (Whitehead, 2008). Weighted Modularity can be generalized for any null-model by substituting the expected weights ($\mathbb{E}[w_{ij}] \approx \langle w_{ij} \rangle$) into the Modularity score $Q = \frac{1}{2s_{tot}} \sum_{ij} (w_{ij}^{obs} - \langle w_{ij} \rangle) \delta(c_i, c_j)$ (Squartini & Garlaschelli, 2011), where $\delta(c_i,c_j)=1$ if i and j are in the same community, c is a vector of community membership, and s_{tot} is the total strength of observed network **W**. We used the spectral optimization techniques of Donetti and Muñoz (2004) and Simonsen (2005) for community partitioning. The number of communities (L) was inferred by optimizing Q with respect to L. We also optimised over all algorithm parameters such as: use of the graph Laplacian (Donetti & Muñoz, 2004) vs. the transfer matrix (Simonsen, 2005); the number of eigenvectors; and the clustering criterion (average, complete, McQuitty's, single, and Ward's criterion). The general method is known to have good performance (Fortunato, 2010) and we benchmarked it against exhaustive simulated annealing (Bélisle, 1992).

This analysis resulted in different estimated communities per year and per null-model, for a total of 40 different community partitionings. To calculate the similarity between two partitionings, we used the Corrected Rand Index, which is useful when the number of communities may differ (see R package fpc Hennig, 2014). The statistic takes values 0 to 1, where 1 is perfect agreement between communities. Only communities within each year are comparable to each other.

RESULTS

To examine the performance of the null models, we compared the expected metrics they generated with the observed values (Figs. 2–4). Null-models that generate expected values that strongly correlate with observed metric values suggest that a lot of the variation is due to individual-level attributes. Our general finding is that MNMs, which conditions on both the degree- and strength-sequence, performed best based on the high correlation between observed and expected metric values and high AIC weights (Fig. 2 and Table 1).

Across all null models, some metrics were consistently more predictable than others: Laplacian centrality, reach and disparity showed high predictability, with R_1^2 values >0.8 between observed and expected values across all models. Of the other metrics, correlations were weaker, with nonuniform scattering of residuals. For example, most observed affinity values clustered along the 1:1 line with a long-tailed distribution of lower-than-expected values. Likewise, most clustering values were higher than expected, and rarely fell below the expected values. This last pattern was only true when considering all years combined, and did not hold within years. Within years, the correlation between observed and expected clustering metrics was low, with R_1^2 values <0.3, and the scatter plots showed a near-flat relationship.

Overwhelmingly, the AIC weights suggested greatest support for the MNM. The MNM had $w_{MNM}^{AIC} > 0.99$ for nine out of 10 networks. Only the smallest network of $N = 18$ individuals was supported by the TNM ($w_{TNM}^{AIC} = 0.784$).
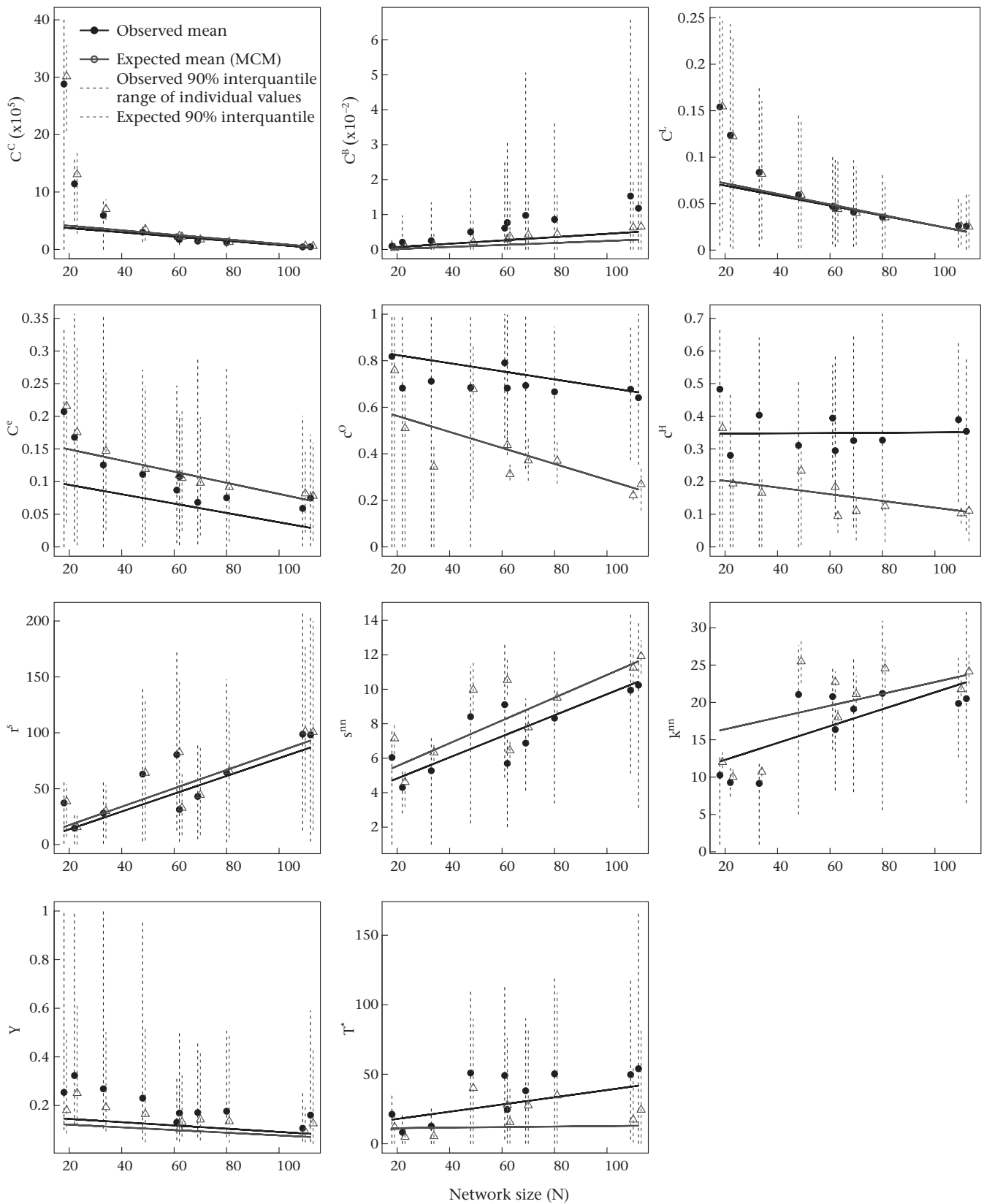
Metrics Versus Network Size

Fig. 5 shows the robust trend lines regressing observed and MNM expected values versus network size, whereby each year had a different network size. Very similar results were obtained when comparing metrics to changes in weighted density (but we omit the results because of the high correlation between N and weighted density, $\hat{\rho}_{Pearson} > 0.95$). Betweenness centrality, degree-affinity, strength-affinity and reach all showed increasing values with increasing network size, consistent between empirical and expected networks. Likewise, closeness centrality, Laplacian centrality, eigenvector centrality and disparity had decreasing values with increasing N, consistent between observed and expected networks. Even when the observed individual metric values were poorly predicted, the MNM none the less made good approximations of the networks' averages as well as the trend over N. In the case of affinity, which was poorly predicted at the individual level, the trend and intercept were nearly identical at the population level. The trend in transitivity measures showed marked deviation from MNM expectations, whereby clustering and triangle counts were higher than expected, and the disagreement between the observed and expected values increased as N increased.

Patterns Among Metrics

Inspecting metrics' residuals (i.e. the observed values minus the expected values) and their relationship among other metrics can lead to insights into social structure. We focused on those metrics that were least predictable by strength and/or degree, such as the cluster coefficients (c^O, c^H), strength- and degree-affinities (s^{nn}, k^{nn}), and eigenvector and betweenness centrality (C^e, C^B). Some residuals seemed to tightly cluster around 0, suggesting that the majority of individuals had values that were predictable solely by degree and strength-sequence alone, while off-zero residuals had erratic patterns on either side of the zero line, such as for s^{nn}, k^{nn}, and especially C^B. Those exceptional individuals that strongly deviated from expectations could be worthy of further investigation. Considering only the MNM, the clustering coefficients had residuals that seemed to be distributed uniformly and randomly with respect to most other metrics and were generally uncorrelated with other metrics, suggesting that clustering uniquely measures some aspect of social biology which all the other metrics do not. For clustering, the largest association by MIC was between the C^B-residual and the c^H-residuals with MIC = 0.363 (where MIC is defined on [0,1]), which was likewise the largest MIC for C^B, suggesting a weakly negative relationship between clustering and betweenness, i.e.

Figure 5. How per-year network averages of metrics vary by network size *N*, both empirically and predicted by the Mixed Null Model. C^c closeness, C^b betweenness, C^L Laplacian centrality, C^e eigenvector centrality, c^O Opsahl clustering, c^H Holme clustering, r^S reach, s^{nn} strength-affinity, k^{nn} degree-affinity, Y disparity, T^* triangle-count.

individuals that are high social brokers tend to cluster less than expected.

In contrast to clustering, the residuals of the affinity metrics showed a strong correlation with other metrics. We observed a concave positive relationship between pairs (s, s^{nn}) and (k, k^{nn}) which indicates strength/degree-assortativity, whereby high-strength/degree individuals tend to associate with other high-strength/degree individuals. However, this regression had an MIC of 0.516 and conditional mutual information (CMI) of 0.701, and ranked lower than the pair $(C^e$ - residuals, s^{nn} - residuals), with MIC 0.644 and CMI 0.758. Together, the residuals revealed a three-tine-star pattern, or three overlapping clusters (Fig. 6): one cluster was a linear relationship between higher-than-expected s^{nn} and higher-than-expected-C^e individuals (i.e. they are the same individuals); a second cluster included individuals with lower-than-expected s^{nn}, low-s, but behaved as expected for C^e; a third cluster had lower-than-expected C^e, biased-low s^{nn} values, and a range of s values. This pattern appeared for each year and for both sexes, and was also evident in C^e - residuals versus k^{nn} - residuals, albeit with a weaker MIC of 0.327. If we recall that the first eigenvector is the best single-dimension representation of the pattern of associations (e.g. the dominant social dimension) then the appearance of strength-assortativity may actually be due to the underlying community structure (as revealed by C^e), where one end of the spectrum constitutes the well-connected, high-s individuals central to the network, and the other end of the spectrum is more diverse, with two different clusters and a mixture of low and high-s individuals

with low s^{nn}. Importantly, this pattern was not related to artefacts of the sampling region as suggested by: (1) the low correlation with measures of preferential sampling (i^{survey} and i^{years} with a maximum MIC of 0.177); and (2) the pattern was strongest for weighted information (s^{nn} versus s), rather than binary information (k^{nn} versus k), i.e. weights should be more robust to sampling artefacts than binary information (Farine, 2014).

Considering sex, few of the metrics showed strong differences between males and females, e.g. the five largest Jensen-Shannon distances [0,0.693] between males and females were 0.171, 0.150, 0.138, 0.128, and 0.121 for c^H, C^e, C^e-residuals, s and c^H-residuals, respectively. Rather, it was only through pairwise bivariate distributions that strong differences manifested between the sexes. The top five bivariate JS-distances were higher at 0.411, 0.411, 0.410, 0.407 and 0.400, for the pairs (c^H, k), (c^H, s), (c^H, T^*), $(c^H, s^{nn}$ - residuals) and $(c^H$ - residuals, $k)$. Males tended to cluster more, had slightly larger affinity values and were slightly more central to the network, but there was considerable overlap and a wide range of values among the sexes.

Modularity

Modularity scores and community partitioning showed high agreement between the three strength-preserving null models (bilinear, WNM and MNM), with nine out of 10 years having identical community partitioning (Table 2). All WNM and MNM partitions were identical. Modularity scores varied by year more than by null-model, with Q values in the range of 0.13−0.28 and averaged 0.21 over all years and methods (i.e. low fragmentation). The TNM resulted in very different communities, with a mean Corrected Rand Index of approximately 0.54 compared to the other three methods. The TNM estimated 145% more communities on average than the other three methods.

DISCUSSION

Network science has a growing influence on many fields of biology, especially behavioural ecology. As the strength of associations between animals is central to theories of social evolution, weighted networks play an important role. Outside of behavioural ecology, some researchers have dismissed weighted information as redundant to binary connections. Using a null-model approach to compare observed weighted metrics versus null-model expectations, similar to Garlaschelli and Loffredo (2008, 2009) and Mastrandrea et al. (2014), we show how weighted information reveals interesting insights into the social lives of dolphins and the behaviour of certain weighted-network metrics.

In behavioural ecology, the null-model approach typically involves application of permutation-based algorithms to generate expectations of metric values under a variety of constraints (Bejder et al., 1998). The contribution of this paper is to focus on how to specify and sample from null models, while meeting three important criteria: (1) the random networks are proportion-weighted w_{ij}
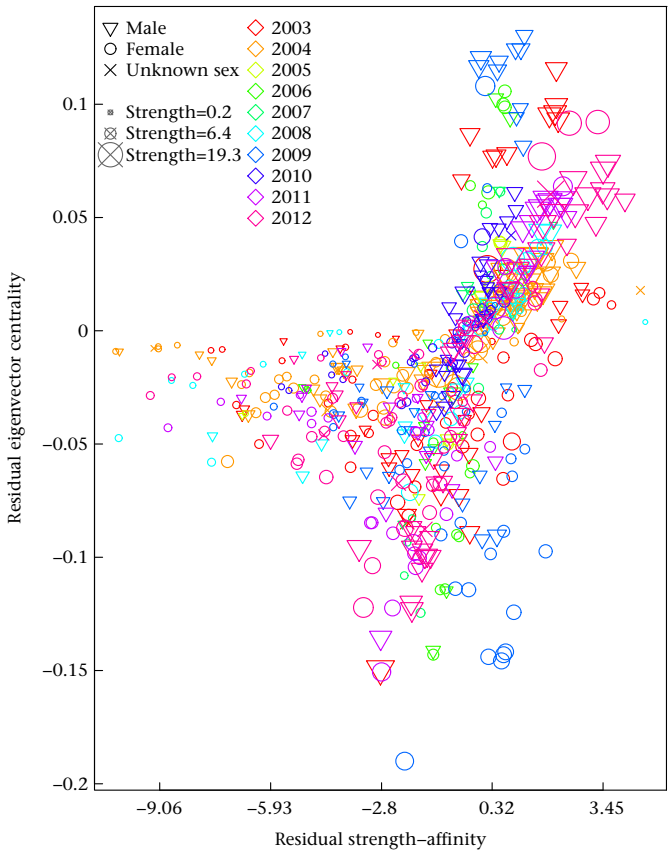
Figure 6. A five-way relationship among metric residuals and individual attributes: residual strength-affinity (x-axis), residual eigenvector centrality (y-axis), strength (size of symbols), sex (shape of symbols) and per year (colours). Residuals are the observed values minus the expected values from a null model (MNM) which conditions on both degree and strength. Residual values around 0 show high concordance between observation and null-model expectation.

Table 2
Agreement between community partitionings implied by four null models over 10 weighted networks

	Bilinear	WNM	MNM	TNM
Bilinear	—			
WNM	0.95	—		
MNM	0.95	1	—	
TNM	0.55	0.54	0.54	—

Communities were inferred by optimizing Modularity, based on the WNM, MNM, TNM, and the original 'bilinear' strength-preserving model by Newman (2004). Values are corrected Rand indices averaged over 10 networks.

∈ [0,1]; (2) we can condition on different individual-level properties, such as degree $\vec{k}$ (sum of binary connections) and strength $\vec{s}$ (sum of weights of connections) as well as size and density of an empirical network; and (3) we ensure that all other properties are maximally randomized, like the topology and weights on the topology, and thereby avoid the risk of confusing artefacts of the randomization algorithm for social structure.

Our analysis compared the null models at three levels of network structure (individual, population averages and subcommunities) with different insights at each level.

(1) Inferences about weighted networks depend on how one conditions a null-model.
(2) Higher-order network properties of dolphin networks differ in their predictability compared to flagship examples used in other scientific disciplines, especially clustering and strength-assortativity.
(3) While clustering and strength-affinity were not trivially redundant to individual-level properties, many other network metrics yielded little extra information over their expectations from strength-preserving null models.
(4) For the eastern gulf Shark Bay bottlenose dolphins, males tend to cluster more than females, and the dominant social dimension seems to be along the lines of strength-assortativity, with high-strength and high-affinity males on one end and a diverse constituency at the other.
(5) Null-models can often make good predictions about the relationship between metrics values and network size N and network density, thus facilitating comparisons between networks across time and space.

These results would theoretically hold for any permutation or randomization-based null-model that could likewise condition on the expected values of individuals' strength and degree. In other words, it is the conditioning that should characterize the null distribution of network metrics, analogous to how a Normal distribution is fully specified by its mean and variance. However, conditioning on both strength and degree is technically difficult, and other permutation or random-graphs methods cannot necessarily guarantee that their algorithms do not artificially induce structural correlations which can be confused for genuine social structure. This is an important point, given the large body of network literature devoted to new algorithms that condition on or reproduce certain properties for inference about network structure and function (Ansmann & Lehnertz, 2011; Leskovec et al., 2010; Prettejohn et al., 2011; Serrano et al., 2006; Watts & Strogatz, 1998). It is the principle of maximum entropy that ensures that all other attributes such as the topology and topological-weight correlations are maximally random.

Inferences Based on Null-models

Of the three null models we studied (constraining the degree-sequence, TNM; the strength-sequence, WNM; and both, MNM), the MNM had the most support by AIC and gave expected values for individual-level metrics that were most inline with observed values. Our results show an important difference to the studies by Garlaschelli and Loffredo (2009) and Mastrandrea et al. (2014), who considered many other nonsocial integer-weighted networks. They emphasised the importance of the degree-sequence and challenged the importance of weight information in weighted metrics. Our results were more equivocal: some weighted metrics were more predictable according to the degree-sequence, while other metrics were more predictable based on the strength-sequence. Therefore, we emphasize that it is more conservative to base inferences on

null models which incorporate both strength- and degree-sequences (i.e. the MNM). Conditioning on both ensures that a larger set of possible metrics are genuinely informative about structure beyond mere individual-level properties, and that our conclusions are not due to an arbitrary decision of whether we conditioned on the strength- or degree-sequence.

Although it may seem obvious that a more complex model would better fit observations than simpler ones, this does not mean that conditioning on more properties is necessarily 'better'. This way of thinking hails from the 'generative model' line of inquiry, such as regression analyses, which try to explain as-best-as-possible how data arose. Instead, the null-model approach is to compare an observed statistic to its expectation under 'no effect', analogous to how one compares t statistics to a null Student-t distribution. Here, we are interested in network metrics that do not behave according to their expectations based on individual-level properties. This line of inference is rooted in the notion that social structure is a phenomenon that is more than the sum of constituent individuals (Holland & Leinhardt, 1979). Therefore, the question of including degree-sequence, strength-sequence or both is a question of conditioning: what features do I want to measure given the values of another? For example, it is well known that network size and density explain the greatest amount of variation in almost all network measurements (Anderson et al., 1999; van Wijk, Stam, & Daffertshofer, 2010), and therefore, one can only look for meaningful deviations in network metrics if they are standardized for size and/or density.

Likewise, the degree-sequence has long been recognized as one such key property for binary networks which demands being conditioned upon. But for weighted networks, serious questions have arisen about the appropriateness of either the degree- or strength-sequence (or both) to serve this crucial conditioning role (Garlaschelli & Loffredo, 2009; Mastrandrea et al., 2014; Newman, 2004). Our analyses suggest that both strength-sequence and degree-sequence together are core properties for some metrics, but not all. For example, if we were to condition on degree (TNM) and we compare dolphins' reach values versus their expectations, we would conclude that a persistent feature of dolphin society is for individuals to have weaker connections among neighbours than expected based on their number of connections. Whereas, if we were to condition on strength (WNM) we would conclude that they were behaving almost exactly according to expectations. If one does not condition on both, then the information in some network measurements will just reflect variations in strength and/or degree-sequence, rather than meaningful higher-order structure.

One level of social structure that was not sensitive to the underlying null-model was community partitioning by Modularity optimization. For example, all three strength-preserving models, the WNM, the MNM and the bilinear model of Newman (2004, which is known to be biased; Squartini & Garlaschelli, 2011) generally resulted in the same conclusions about community structure. Large differences were only observed when comparing the strength-preserving null models versus the degree-preserving model (TNM): the TNM predicted a greater number of communities and much smaller communities than the WNM/MNM/bilinear models. In the case of the Shark Bay dolphins, we are able to a priori motivate the appropriateness of the WNM/MNM/bilinear model partitionings versus the TNM communities, because 'fission-fusion' societies are likely to have many temporary encounters (and therefore a very dense topology) but only a fraction of these putative connection will actually reflect long-term associations.

The community partitioning results underscore the utility of conditioning on both the degree and strength sequence: we do not need to choose between two competing null models and two competing community structures (TNM versus WNM), because the

MNM recovers the 'correct' structure by conditioning on both. The MNM would be especially useful for researchers who cannot a priori motivate the use of either a strength- or degree-preserving null-model for community partitioning.

Higher-order Network Properties

Another insight from our null models is the lack of structural correlations in higher-order network metrics reported in the statistical physics literature. This contrasts sharply with the statements of Garlaschelli and Loffredo (2009) and Mastrandrea et al. (2014) who declared weighted information to be mostly redundant to topology. In our case, the expectations from the WNM and MNM were more similar than those of the TNM. Our findings are particularly striking for proportion-weighted networks, which have a fundamental correlation between strength and degree ($s_i \leq k_i$).

In particular, the second- and third-order properties were not reproduced trivially by any of the null models, unlike what was observed by the comprehensive analyses of Mastrandrea et al. (2014). Instead, clustering and strength/degree-affinity differed strongly from their null-model expectations and provided rich information about the social structure of the eastern gulf Shark Bay dolphins. Most simply, dolphins clustered more than expected and had equal-or-lower affinity values than expected. Males, in particular, cluster more than females (Mann et al., 2012), which is probably due to their long-term pairs and hierarchical alliances, whereas females may associate less strongly based on factors unrelated to 'friendship' (e.g. reproductive status, age of calves, or coercion by other males; Smolker et al., 1992; Scott, Mann, Watson-Capps, Sargeant, & Connor, 2005; Frere et al., 2010; Mann et al., 2012). The information in affinity metrics related strongly to other metrics, especially eigenvector centrality, hinting at a more complex structure than simple strength/degree assortativity: the networks' high-strength individuals with high-strength neighbours constituted one end of the dominant social dimension, and also tended to be male; while at the other end were the low-affinity individuals with diverse strength values (Fig. 6). This was the dominant structure of the network, around which there was a diverse group of individuals with lower-than-expected strength-affinity and highly unpredictable eigenvectors, who may be from other subcommunities which cannot be adequately described by the leading eigenvector. Functionally, dolphins form associations based on shared foraging tactics, sex and kinship (Mann et al., 2012); these and other factors cannot be expressed by network-metrics, may be the cause behind the null models' inability to reproduce the seemingly simple relationship of strength/degree-assortativity.

It is not surprising that our results differ from the comprehensive studies of Mastrandrea et al. (2014), given the wealth of literature on dolphin social alliances, sociality-fitness correlations and emergent network properties (Connor, Heithaus, & Barre, 2001; Frere et al., 2010; Krützen, Barré, Connor, Mann, & Sherwin, 2004; Lusseau, 2003; Stanton & Mann, 2012). It nonetheless underscores the caution that behavioural ecologists must exercise when interpreting the developments in network analysis from other disciplines, whose subjects can span trophic webs to financial networks. First, there is likely a fundamental difference between social animals and the more abstract networks from the physical sciences. Second, behavioural ecology studies are probably more sensitive to observation error and the data-collection process, which may mediate the importance of binary versus weighted-information, e.g. animal networks measured by group affiliation are more likely to have many frivolous binary connections. It remains to be seen whether our conclusions about the redundancy of certain metrics and the informativeness of higher-order metrics will generalize to other taxa and societies.

Comparing Networks

Our null models potentially offer a way to standardize metrics across networks of different size and density. Because open populations of animals may have year-to-year variation in N, the dependence of metric values on network size and density hinders researchers' ability to compare network values over time or across populations. It is especially problematic for researchers who wish to monitor and estimate changes in animals' social lives over time and due to different treatments (e.g. changes in dolphin-watch tourism, resource extraction activities, conservation management, etc.), but who cannot hold population and/or network size constant. This problem does not have a satisfactory answer in general (Anderson et al., 1999), and less for weighted metrics (van Wijk et al., 2010). Even seemingly simple metrics like centrality (where solutions exist for binary measures) have no accepted standardization candidate for weighted metrics (Kasper & Voelkl, 2009). The null models here provided satisfactory per-year population-level summaries, and may offer a possible standardization technique via the difference between observed and expected values. Unfortunately, this does not apply to clustering metrics.

Limitations and Future Considerations

In the broader context of network analysis, the null-model approach has limitations. In our case, its insights are limited to what individual attributes may be important to condition upon (especially strength) and what network metrics may be redundant to individual attributes. We offer no generative model of the true underlying social process, nor do we specify how observation error obfuscates the true social process. This is true of a lot of network analyses using null models and network metrics, whereby metrics are not necessarily representative of the processes that animals in the network are influenced by, nor are they necessarily functionally important. Furthermore, we are generally unable to make conclusions about how network properties arise, such as being the manifestation of many individual-based decisions, or whether there is some higher-order feedback from patterns in the network itself.

Second, we do not handle confounding effects of space, time, artefacts of the survey design or observation error, which are a major preoccupation in others fields of animal ecology. In our case, any spatial processes or observation errors are taken as a given inasmuch as they manifest in individuals' degree or strength. On this issue, permutation and random-graph algorithms may be better able to account for observation error, survey design and inherent problems with inferring associations from group-membership data. Investigators need to judge what may be the greatest source of bias for their particular research questions. Ultimately, the field needs better generative models, such as those common to sociology (e.g. Exponential Random Graphs or Stochastic Actor-Oriented Models, Pinter-Wollman et al. 2014), which can incorporate node-level covariates, exogenous covariates and dynamic processes. While these techniques are widely used for binary and integer-weighted networks, there is currently no obvious formulation for proportion-weighted networks.

The insights from null models can aid in the development of such generative models, especially in terms of finding worthy summary statistics that one may use within approximate-likelihood or simulation-based inference (see for example Ratmann et al., 2007, who specified a generative protein network model using simulation-based, likelihood-free inference). Our results suggest that despite the overwhelming number of possible

network metrics, many of them, including reach, disparity and various centrality measures, are highly redundant to strength- and/ or degree-sequence. Therefore, researchers seeking to develop complex generative models of network formation will probably need to focus on higher-order properties such as betweenness and clustering. In fact, Ratmann et al. (2007) found that using lower-level network metrics could even lead to inconsistent estimates.

Until the time that generative models are developed, inference by null models will remain an important tool. Therefore, it is important for behavioural ecologists to understand how their conclusions depend on the assumptions and constraints of their null models.

Acknowledgments

References

Allen, J., Weinrich, M., Hoppitt, W., & Rendell, L. (2013). Network-based diffusion analysis reveals cultural transmission of lobtail feeding in humpback whales. *Science, 340*, 485−488.

Anderson, B. S., Butts, C., & Carley, K. (1999). The interaction of size and density with graph-level indices. *Social Networks, 21*, 239−267.

Ansmann, G., & Lehnertz, K. (2011). Constrained randomization of weighted networks. *Physical Review E, 84*, 026103. http://dx.doi.org/10.1103/PhysRevE.84.026103.

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.

Barthélemy, M., Barrat, A., Pastor-Satorras, R., & Vespignani, A. (2005). Characterization and modeling of weighted networks. *Physica A: Statistical Mechanics and its Applications, 346*, 34−43.

Bejder, L., Fletcher, D., & Bräger, S. (1998). A method for testing association patterns of social animals. *Animal Behaviour, 56*, 719−725.

Bélisle, C. J. (1992). Convergence theorems for a class of simulated annealing algorithms on Rd. *Journal of Applied Probability*, 885−895.

Butts, C. T. (2008). Social network analysis: a methodological introduction. *Asian Journal of Social Psychology, 11*, 13−41.

Cairns, S. J., & Schwager, S. J. (1987). A comparison of association indices. *Animal Behaviour, 35*, 1454−1469.

Cohen, J., West, S. G., Aiken, L. S., & Cohen, P. (2003). *Applied multiple regression/ correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Connor, R. C., Heithaus, M. R., & Barre, L. M. (2001). Complex social structure, alliance stability and mating access in a bottlenose dolphin 'super-alliance'. *Proceedings of the Royal Society B: Biological Sciences, 268*, 263−267.

Donetti, L., & Muñoz, M. A. (2004). Detecting network communities: a new systematic and efficient algorithm. *Journal of Statistical Mechanics: Theory and Experiment, 2004*, P10012.

Farine, D. R. (2014). Measuring phenotypic assortment in animal social networks: weighted associations are more robust than binary edges. *Animal Behaviour, 89*, 141−153.

Faust, K. (2006). Comparing social networks: size, density, and local structure. *Metodološki zvezki, 3*, 185−216.

Fortunato, S. (2010). Community detection in graphs. *Physics Reports, 486*, 75−174.

Freeman, L. C. (1979). Centrality in social networks: conceptual clarification. *Social networks, 1*, 215−239.

Frere, C. H., Krutzen, M., Mann, J., Connor, R. C., Bejder, L., & Sherwin, W. B. (2010). Social and genetic interactions drive fitness variation in a free-living dolphin population. *Proceedings of the National Academy of Sciences, 107*, 19949−19954.

Garlaschelli, D., & Loffredo, M. (2008). Maximum likelihood: extracting unbiased information from complex networks. *Physical Review E, 78*, 015101.

Garlaschelli, D., & Loffredo, M. I. (2009). Generalized Bose-Fermi statistics and structural correlations in weighted networks. *Physical Review Letters, 102*, 038701.

Ghalanos, A., & Theussl, S. (2012). *Rsolnp: General non-linear optimization using augmented lagrange multiplier method.*

Ginestet, C. E., Nichols, T. E., Bullmore, E. T., & Simmons, A. (2011). Brain network analysis: separating cost from topology using cost-integration. *PLoS One, 6*, e21570.

Hennig, C. (2014). *fpc: Flexible procedures for clustering. R package version 2.1-9.*

Holland, P. W., & Leinhardt, S. (1979). Structural sociometry. In P. W. Holland, & S. Leinhardt (Eds.), *Perspectives on social network research* (pp. 63−83). New York, NY: Academic Press.

Kasper, C., & Voelkl, B. (2009). A social network analysis of primate groups. *Primates, 50*, 343−356.

Krützen, M., Barré, L. M., Connor, R. C., Mann, J., & Sherwin, W. B. (2004). 'o father: where art thou?'—paternity assessment in an open fission−fusion society of wild bottlenose dolphins (*tursiops* sp.) in shark bay, western australia. *Molecular Ecology, 13*, 1975−1990.

Leskovec, J., Chakrabarti, D., Kleinberg, J., Faloutsos, C., & Ghahramani, Z. (2010). Kronecker graphs: an approach to modeling networks. *Journal of Machine Learning Research, 11*, 985−1042.

Lusseau, D. (2003). The emergent properties of a dolphin social network. *Proceedings of the Royal Society B: Biological Sciences, 270*, S186−S188.

Mann, J., Stanton, M. A., Patterson, E. M., Bienenstock, E. J., & Singh, L. O. (2012). Social networks reveal cultural behaviour in tool-using using dolphins. *Nature Communications, 3*, 980.

Mastrandrea, R., Squartini, T., Fagiolo, G., & Garlaschelli, D. (2014). Enhanced reconstruction of weighted networks from strengths and degrees. *New Journal of Physics, 16*, 043022.

Newman, M. E. (2004). Analysis of weighted networks. *Physical Review E, 70*, 056131.

Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: generalizing degree and shortest paths. *Social Networks, 32*, 245−251.

Opsahl, T., & Panzarasa, P. (2009). Clustering in weighted networks. *Social Networks, 31*, 155−163.

Pinter-Wollman, N., Hobson, E. A., Smith, J. E., Edelman, A. J., Shizuka, D., de Silva, S., et al. (2014). The dynamics of animal social networks: analytical, conceptual, and theoretical advances. *Behavioral Ecology, 25*, 242−255.

Prettejohn, B. J., Berryman, M. J., & McDonnell, M. D. (2011). Methods for generating complex networks with selected structural properties for simulations: a review and tutorial for neuroscientists. *Frontiers in Computational Neuroscience, 5*. http://www.frontiersin.org/Computational_Neuroscience/10.3389/fncom.2011.00011/abstract.

Qi, X., Fuller, E., Wu, Q., Wu, Y., & Zhang, C.-Q. (2012). Laplacian centrality: a new centrality measure for weighted networks. *Information Sciences, 194*, 240−253.

R Core Team. (2014). *R: A language and environment for statistical computing.* http://www.r-project.org/.

Ratmann, O., Jørgensen, O., Hinkley, T., Stumpf, M., Richardson, S., & Wiuf, C. (2007). Using likelihood-free inference to compare evolutionary dynamics of the protein networks of *H. pylori* and *P. falciparum*. *PLoS Computational Biology, 3*, e230.

Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., et al. (2011). Detecting novel associations in large data sets. *Science, 334*, 1518−1524.

Scott, E. M., Mann, J., Watson-Capps, J. J., Sargeant, B. L., & Connor, R. C. (2005). Aggression in bottlenose dolphins: evidence for sexual coercion, male-male competition, and female tolerance through analysis of tooth-rake marks and behaviour. *Behaviour, 142*, 21.

Serrano, M., Boguñá, M., & Pastor-Satorras, R. (2006). Correlations in weighted networks. *Physical Review E, 74*.

Simonsen, I. (2005). Diffusion and networks: a powerful combination! *Physica A: Statistical Mechanics and its Applications, 357*, 317−330.

Smolker, R. A., Richards, A. F., Connor, R. C., & Pepper, J. W. (1992). Sex differences in patterns of association among Indian Ocean bottlenose dolphins. *Behaviour, 123*, 38−69.

Squartini, T., & Garlaschelli, D. (2011). Analytical maximum-likelihood method to detect patterns in real networks. *New Journal of Physics, 13*, 083001.

Stanton, M. A., & Mann, J. (2012). Early social networks predict survival in wild bottlenose dolphins. *PLoS One, 7*, e47508.

Watts, D., & Strogatz, S. (1998). Collective dynamics of 'small-world' networks. *Nature, 393*, 440−442.

Wey, T., Blumstein, D. T., Shen, W., & Jordán, F. (2008). Social network analysis of animal behaviour: a promising tool for the study of sociality. *Animal Behaviour, 75*, 333−344.

Whitehead, H. (2008). *Analyzing animal societies: Quantitative methods for vertebrate social analysis.* Chicago, IL: University of Chicago Press.

van Wijk, B. C. M., Stam, C. J., & Daffertshofer, A. (2010). Comparing brain networks of different size and connectivity density using graph theory. *PLoS One, 5*, e13701.

Williams, R., & Lusseau, D. (2006). A killer whale social network is vulnerable to targeted removals. *Biology Letters, 2*, 497−500.

Ye, Y. (1987). *Interior algorithms for linear, quadratic, and linearly constrained non-linear programming* (Unpublished doctoral thesis). Stanford, CA: Stanford University.

APPENDIX

This appendix outlines the method for calculating the expectation of an individual's network metric c_i for an observed proportion-weighted network. We denote an empirical weighted network interchangeably with its symmetric weight matrix **W**, consisting of N nodes/individuals. Associations between individuals i and j are indexed by the elements $w_{ij} \in [0,1]$. The empirical network has properties $\vec{\theta}$ that are used to constrain a null-model,

such as the observed degree-sequence $\vec{k}$, strength-sequence $\vec{s}$, or both.

To calculate a metric's expected value $\mathbb{E}[c_i] = \sum_{W' \in \mathscr{W}} P(\mathbf{W'}|\vec{\theta})c_i(\mathbf{W'})$, the challenge is specify the probability of a weighted-network $P(\mathbf{W}|\vec{\theta})$, conditioned on $\vec{\theta}$, and to sample from it. To do so, we use the maximum entropy ensemble of weighted networks, $\mathscr{W}$, subject to constraints of $\vec{\theta}$, to define this probability distribution. The Exponential Random Graph formulation allows us to specify the 'canonical' ensemble of networks for an empirical network (Garlaschelli & Loffredo, 2008, 2009), whose expected values of θ are the empirical network's values, i.e. $\mathbb{E}[\vec{\theta}] = \vec{\theta}$. In statistical mechanics, this contrasts with the 'microcanonical' ensemble, whereby every constituent random-network $\mathbf{W'} \in \mathscr{W}$ has exactly the same $\vec{\theta}$ as observed empirically, which is most common among permutation-based algorithms. Note that the number of random-networks in the canonical ensemble is at least as great (and likely much greater) than the number of random-networks in the microcanonical ensemble, and therefore has greater entropy, i.e. more randomness among all other topological and weight-distribution characteristics.

We approximate $\mathbb{E}[c_i]$ with $\langle c_i \rangle$, using a large sample of size m of random networks $\mathscr{W}_m$ from the ensemble $\mathscr{W}$, each drawn with probability $P(\mathbf{W'}|\vec{\theta})$.

$$\mathbb{E}[c_i] \approx \left\langle c_i \right\rangle = \frac{\sum_{W' \in \mathscr{W}_m} c_i(W')}{m} \quad (2)$$

The probability of a proportion-weighted network $P(\mathbf{W}|\vec{\theta})$ under constraints is formulated according to the Exponential Random Graph Model:

$$P(\mathbf{W}|\vec{\theta}) = \frac{e^{-H(\mathbf{W}|\vec{\theta})}}{\sum_{\mathscr{W}} e^{-H(\mathscr{W}|\vec{\theta})}} \quad (3)$$

A realization of a weighted network is proportional to e^{-H}, the inverse of the exponential of the Graph Hamiltonian H. The Graph Hamiltonian H specifies the conditioning on observed properties $\vec{\theta}$, i.e. it contains parameters $\vec{a}$ and $\vec{b}$ which can be optimized so as to make the observed quantities $\vec{\theta}$ match the ensemble averages $\langle \vec{\theta} \rangle$. According to Garlaschelli and Loffredo (2009), the Hamiltonian of the TNM, WNM and MNM are:

$$H_{TNM}(\mathbf{W}|\vec{a}) = \sum_i \alpha_i k_i = \sum_{i<j} (\alpha_i + \alpha_j) I(w_{ij}) \quad (4)$$

$$H_{WNM}\left(\mathbf{W}|\vec{b}\right) = \sum_i \beta_i s_i = \sum_{i<j} (\beta_i + \beta_j) w_{ij} \quad (5)$$

$$\begin{aligned} H_{MNM}\left(\mathbf{W}|\vec{a},\vec{b}\right) &= \sum_i (\alpha_i k_i + \beta_i s_i) \\ &= \sum_{i<j} ((\alpha_i + \alpha_j) I(w_{ij}) + (\beta_i + \beta_j) w_{ij}) \end{aligned} \quad (6)$$

where $I(w)$ is the indicator function (if $w > 0$, then 1, otherwise 0), $k_i = \sum_{i \neq j} I(w_{ij})$ is the degree of node i, and $s_i = \sum_{i \neq j} w_{ij}$ is the strength of node i. Often, it is easier to work with reparametrized versions of the Hamiltonian parameters: $x_i = e^{-\alpha_i}$ and $y_i = e^{-\beta_i}$. The parameters $\vec{x}$ and $\vec{y}$ can be related to the probability of a realization of a weighted network $\mathbf{W}$ by considering the probability as the product of probabilities of individual edge weights: $P(\mathbf{W}|x,y) = \prod_{i<j} q(w_{ij}|x,y)$, where $q(w_{ij})$ is the probability that the weight between nodes i and j is w. For the WNM and MNM null models, we can therefore define the probability of a weight w_{ij} as a function of the node parameters x_i, x_j, y_i and y_j:

$$\begin{aligned} q(w_{ij}|x,y) &= \frac{(x_i x_j)^{I(w_{ij})} \left(y_i y_j\right)^{w_{ij}}}{1 + x_i x_j \int_{w^* > 0}^{1} \left(y_i y_j\right)^{w^*} dw^*} \\ &= \frac{(x_i x_j)^{I(w_{ij})} \left(y_i y_j\right)^{w_{ij}} \log\left(y_i y_j\right)}{x_i x_j y_i y_j - x_i x_j + \log\left(y_i y_j\right)} \end{aligned} \quad (7)$$

The TNM is derived similarly from its Hamiltonian. The probability of a nonzero link is therefore $p(w_{ij} > 0) = 1 - q(w_{ij} = 0)$.

The ensembles are defined by specifying the constraints and solving $\vec{x}$ and $\vec{y}$. For the WNM, this means setting the expected value of nodes' strength to their observed values, $\langle s_i \rangle = s_i^{obs}$, where $\langle s_i \rangle$ is the sum of the expected weights $\langle w_{ij} \rangle$. For the TNM, we set the expected values of the nodes' degree to their observed values, $\langle k_i \rangle = k_i^{obs}$, whereby $\langle k_i \rangle$ is the sum of probabilities of a link. For the MNM, we set the expected degree-sequence and the expected strength-sequence to their observed values.

$$\begin{aligned} \left\langle w_{ij} \right\rangle &= \int_{w^*=0}^{1} w^* q(w^*|x,y) dw^* \\ &= \frac{x_i x_j y_i y_j \log\left(y_i y_j\right) - x_i x_j y_i y_j + x_i x_j}{\log^2\left(y_i y_j\right) + \left(x_i x_j y_i y_j - x_j x_i\right)\log\left(y_i y_j\right)} \end{aligned} \quad (8)$$

$$\left\langle k_i \right\rangle = \sum_{j \neq i} p(w_{ij} > 0|x,y) = \sum_{j \neq i} (1 - q(w_{ij} = 0|x,y)) = k_i^{obs} \quad (9)$$

$$\left\langle s_i \right\rangle = \sum_{j \neq i} \left\langle w_{ij} \right\rangle = s_i^{obs} \quad (10)$$

Equations (9) and (10) constitute the system of equations specified by the constraints $\langle s_i \rangle = s_i^{obs}$ and/or $\langle k_i \rangle = k_i^{obs}$ with Lagrangian multipliers $x_i > 0$ and/or $y_i > 0$. We found fast and accurate solutions to our parameters using the package Rsolnp (Ghalanos & Theussl, 2012; Ye, 1987) in the R programming language and environment (R Core Team, 2014). For the MNM, there are $2N$ parameters to solve, constrained by both strength and degree. For the WNM, there are N parameters to solve, whereby strength is preserved by setting $\vec{x} = \mathbf{1}^N$ and re-solving $\vec{y}_{MNM} \neq \vec{y}_{WNM}$. The TNM likewise has N parameters to solve $\vec{x}_{TNM} \neq \vec{x}_{MNM}$. In all cases, both topology and weights are maximally randomized subject to these constraints.

The cumulative distribution function (CDF) can then be used to sample weights w_{ij} from their appropriate distribution. By sampling w_{ij} for all $i < j$ from their appropriate distribution, we also sample random networks from their appropriate distribution $P(\mathbf{W})$. For the MNM and WNM, the CDF has the form:

$$CDF(w^*|x,y) = q(0|x,y) + \frac{x_i x_j \left(y_i^{w^*} y_j^{w^*} - 1\right)}{x_i x_j y_i y_j - x_i x_j + \log\left(y_i y_j\right)} \quad (11)$$

The analysis proceeds by: (1) solving the Lagrangian multipliers $\vec{x}$ and $\vec{y}$; (2) sampling a large number networks from $P(\mathbf{W}|\vec{x}, \vec{y})$; (3) then calculating metric expectations according to equation (2).