

# The Perceived Assortativity of Social Networks: Methodological Problems and Solutions

David N. Fisher, Matthew J. Silk, and Daniel W. Franks

## 1 Introduction

Network theory is a useful tool that can help us explain a range of social, biological and technical phenomena [1–3]. For instance, network approaches have been used to investigate diverse topics such as the global political and social system [4, 5] to the formation of coalitions among individuals [6–9]. Networks can be described using a number of local (related to the individual) and global (related to the whole network) measures. One important global measure is degree correlation or assortativity (we use the latter term for brevity), which was formally defined by Newman [10], although Pastor-Satorras et al. [3] had calculated an analogous measure previously. The assortativity of a network measures how the probability of a connection between two nodes (individuals) in a network depends on the degrees of those two nodes (the degree being the number of connections each node possesses). The measure quantifies whether those with many connections associate with others with many connections (assortative networks or networks with assortativity), or if “hubs” form where well-connected individuals are connected to many individuals with few other connections (disassortative networks or networks with disassortativity). If the

---

D.N. Fisher (✉)

Department of Integrative Biology, University of Guelph, Guelph, ON, N1G 2W1, Canada  
e-mail: [davidnfisher@hotmail.com](mailto:davidnfisher@hotmail.com)

M.J. Silk (✉)

Environment and Sustainability Institute, University of Exeter, Penryn Campus, Penryn,  
TR10 9FE, UK  
e-mail: [matthewsilk@outlook.com](mailto:matthewsilk@outlook.com)

D.W. Franks (✉)

Department of Biology, University of York, York, YO10 5DD, UK

Department of Computer Science, University of York, York, YO10 5DD, UK  
e-mail: [daniel.franks@york.ac.uk](mailto:daniel.franks@york.ac.uk)

tendency for nodes to be connected is independent of each other's degrees a network has neutral assortativity. Assortativity is calculated as the Pearson's correlation coefficient between the degrees of all pairs of connected nodes, and ranges from  $-1$  to  $1$  [10]. The Pastor-Satorras method involves plotting the degree of each node against the mean degree of its neighbours, and judging the network assortative if the slope is positive and disassortative if the slope is negative [3]. In this article we will focus on the Newman measure, as it has been more commonly used by the scientific community, gives a coefficient bounded between  $-1$  and  $1$  rather than a slope and typically is supported by a statistical test, something general absent from the reporting of the Pastor-Satorras method.

Assortativity is a key property to consider when understanding how networks function, especially when considering social networks. A network's robustness to attacks is increased if it is assortative [11, 12]. However, the speed of information transfer and ability to act in synchrony is increased in disassortative networks [13, 14]. Thus the assortativity of a social network in which an individual is embedded can have a substantial impact on that individual.

Networks are typically found to be neutrally or negatively assortative [10, 15, 16]. When considering all networks comprised of eight nodes, Estrada [17] observed that only 8% of over 11,000 possible networks were assortative. Despite this general trend, social networks are often said to differ from other networks by being assortative [15, 18]. This has led to those finding disassortativity in networks of online interactions [19], mythical stories [20] or networks of dolphin interactions [21] to suggest they are different to typical human social networks. While the generality of assortativities in social networks has been questioned [16, 22], a wide variety of recent research still states that this is a property typical of social networks (e.g. [17, 23–28]). With assortativity being a key network property and subject of interest in a range of fields, this topic requires clarification.

In this paper we review assortativity in the networks literature, with emphasis on social networks. We assess the generality of the hypothesis that social networks tend to be distinct from other kinds of networks in their assortativities and explore whether the precise method of social network construction influences this metric. We go on to show how particular methodologies of social network construction could result in falsely assortative networks, and present a number of solutions to this by drawing on advances in sociological and biological fields.

## 2 Assortativity in Social and Other Networks

Random networks should be neutrally assortative [10]. However, simulations by Franks et al. [29] showed that random social networks constructed using group-based methods are assortative unless extensive sampling is carried out. Group-based methods are where links are formed between individuals not when they directly interact, but when they both are found in the same group, or contribute to a joint piece of work, e.g. co-author a paper or appear in a band together. We hypothesised that the suggestion that social networks possess assortativity was due to a prepon-

derance of social networks constructed using group-based methods in the early literature. We thus conducted a literature search, recording the (dis)assortativity of networks, network types (social or non-social) and for the social networks, method of construction (direct interactions or group-based). We expect that social networks built using group-based methods would be more assortative than the other two classes of network, which would be similarly assortative.

## ***2.1 Literature Search: Method***

Assortativity has been calculated for a wide range of networks in the last decade. We compiled a list of the assortativity of published networks based on the table in Whitney and Alderson [16], literature searches with the terms “degree correlation” and “assortativity”, and examining the articles citing Newman [10]. If it could not be determined how the network was constructed, or how large it was, the network was excluded. We did not include the average assortativity when reported from a range of similar networks when the individual scores were not reported. We also did not include assortativities of networks from studies re-analysing existing datasets, to avoid pseudo-replication. Only undirected networks were considered; see Piraveenan et al. [30] for a review of assortativities in directed networks. We then classified these networks as non-social networks, social networks constructed using direct interactions or social networks constructed using group-based methods. We then compared assortativity across network classes using a Kruskal–Wallis test as assortativity scores were not normally distributed. If this revealed significant differences among network classes, we then compared network classes to each other using Wilcoxon rank sum tests and to the a neutral assortativity of zero with Wilcoxon signed rank tests.

## ***2.2 Literature Search: Results***

In published papers we found assortativities for 88 networks that met our criteria for inclusion, see Table 1.52 of these were social networks, of which 25 were constructed using group-based methods. The assortativities for the network classes are shown as boxplots in Fig. 1. The Kruskal–Wallis tests indicated that there were differences among groups (Kruskal–Wallis  $\chi^2 = 26.8$ , d.f. = 2,  $p < 0.001$ ). All networks types were different from each other, with the group-based social networks being more assortative than both other classes of network, and the direct social networks being more assortative than the non-social networks (all Wilcoxon rank sum tests, group-based social networks vs. direct social networks,  $W = 214$ ,  $n(\text{direct}) = 27$ ,  $n(\text{group-based}) = 25$ ,  $p = 0.024$ ; group-based social networks vs. non-social networks,  $W = 783$ ,  $n(\text{group-based social networks}) = 25$ ,  $n(\text{non-social network}) = 36$ ,  $p < 0.001$ ; direct social networks vs. non-social network,  $W = 716$ ,

**Table 1** 88 networks of various types and methods of construction

Network	Size	Type	Assortativity	Method of construction	Source
Beowulf (myth)	74	Social	-0.1	Direct	[20]
Cyworld (online)	12,048,186	Social	-0.13	Direct	[22]
Email address books	16,881	Social	0.092	Direct	[18]
Epinions neg (online)	131,828	Social	-0.022	Direct	[31]
Epinions pos (online)	131,828	Social	0.217	Direct	[31]
Facebook (online)	721,000,000	Social	0.226	Direct	[32]
Flickr (online)	1,846,198	Social	0.202	Direct	[22]
Gnutella P2P (online)	191,679	Social	-0.109	Direct	[22]
Ground squirrels	65	Social	0.82	Direct	[33]
Iliad (myth)	716	Social	-0.08	Direct	[20]
LiveJournal (online)	5,284,457	Social	0.179	Direct	[22]
Mixi (online)	360,802	Social	0.122	Direct	[22]
MySpace (online)	100,000	Social	0.02	Direct	[22]
Nioki (online)	20,259	Social	-0.13	Direct	[19]
Orkut (online)	100,000	Social	0.31	Direct	[22]
Pussokram (online)	29,341	Social	-0.048	Direct	[19]
Slashdot neg (online)	82,144	Social	-0.114	Direct	[31]
Slashdot pos (online)	82,144	Social	0.162	Direct	[31]
Student relationships	573	Social	-0.029	Direct	[18]
Táin (myth)	404	Social	-0.33	Direct	[20]
Twitter (online)	4,317,000	Social	-0.025	Direct	[34]
Whisper (online)	690,000	Social	-0.011	Direct	[34]
Xiaonei (online)	396,836	Social	-0.0036	Direct	[22]
YouTube (online)	1,157,827	Social	-0.033	Direct	[22]
Chinese science citations	81	Social	-0.036	Direct	[35]
Barbary macaque (grooming)	141	Social	0.351	Direct	[36]
GitHub (online)	671,751	Social	-0.0386	Direct	[37]
Australian dolphins	117	Social	0.003	Group	[38]
Biology co-authors	1,520,251	Social	0.13	Group	[10]
Birds (mixed species)	93	Social	0.29	Group	[39]
Company directors	7,673	Social	0.28	Group	[10]
Condensed matter co-authors	36,458	Social	0.18	Group	[40]
Condensed matter co-authors 1995-9	16,729	Social	0.18	Group	[41]
Film actors	449,913	Social	0.21	Group	[18]
Killer whales	7	Social	-0.48	Group	[42]
New Zealand dolphins	64	Social	-0.044	Group	[21]
Maths co-authors	253,339	Social	0.12	Group	[10]
Physics co-authors	52,909	Social	0.36	Group	[10]

(continued)

**Table 1** (continued)

Physics co-author2	16,264	Social	0.18	Group	[10]
Scottish dolphins	124	Social	0.17	Group	[43]
Sticklebacks	94	Social	0.66	Group	[44]
TV series actor collaboration	79,663	Social	0.53	Group	[22]
Grant proposals (accepted)	24,181	Social	-0.1018	Group	[45]
Grant proposals (rejected)	46,567	Social	-0.1145	Group	[45]
Brazilian co-authorship (Humanities)	74,490	Social	0.3737	Group	[46]
Brazilian co-authorship (Linguistics, letters and arts)	15,375	Social	0.3761	Group	[46]
Brazilian co-authorship (Engineering)	15,375	Social	0.0273	Group	[46]
Brazilian co-authorship (Agricultural science)	55,695	Social	0.0769	Group	[46]
Brazilian co-authorship (Biological science)	75,304	Social	0.1404	Group	[46]
Brazilian co-authorship (Exact and earth sciences)	65,221	Social	0.1173	Group	[46]
Brazilian co-authorship (applied social sciences)	48,340	Social	0.1373	Group	[46]
Brazilian co-authorship (health sciences)	114,169	Social	0.1301	Group	[46]
Airports in Pakistan	35	Transport	-0.47	NA	[47]
Berlin U & S Bahn	75	Transport	0.096	NA	[16]
Bike	131	Mechanical	-0.2	NA	[16]
Brain connections	160	Biological	0.058	NA	[48]
Brain connections, Polymicrogyria	160	Biological	0.044	NA	[48]
Car door	649	Mechanical	-0.16	NA	[16]
Domain network (online)	11,174	Technological	-0.17	NA	[49]
Film references	40,008	Commercial	-0.057	NA	[50]
Fresh water food web	92	Biological	-0.33	NA	[18]
Grand piano action key 1	71	Mechanical	-0.32	NA	[16]
Internet	10,697	Technical	-0.19	NA	[18]
Jet engine	60	Mechanical	-0.13	NA	[16]
London underground	92	Transport	0.01	NA	[16]
Marine food web	134	Biological	-0.26	NA	[18]
Metabolic pathways	765	Biological	-0.24	NA	[18]
Moscow subway	51	Transport	0.18	NA	[16]
Moscow subway and regional rail	129	Transport	0.26	NA	[16]
Neural pathways	307	Biological	-0.23	NA	[18]
Power grid	4941	Technical	-0.003	NA	[18]
Proteins	2115	Biological	-0.16	NA	[18]
Pseudomonas strains	37	Biological	-0.553	NA	[51]
Router network	228,298	Technological	-0.01	NA	[49]

(continued)

**Table 1** (continued)

Network	Size	Type	Assortativity	Method of construction	Source
Six speed transmission	143	Mechanical	-0.18	NA	[16]
Software	3162	Technical	-0.016	NA	[18]
Tokyo regional rail	147	Transport	-0.09	NA	[16]
Tokyo regional rail and subway	191	Transport	0.043	NA	[16]
V8 engine	243	Mechanical	-0.27	NA	[16]
World Wide Web	269,504	Technical	-0.067	NA	[18]
Yeast genes	333	Biological	-0.15	NA	[49]
Yeast proteins	1066	Biological	-0.12	NA	[49]
Copenhagen streets	1637	Transport	-0.07	NA	[52]
London streets	3010	Transport	-0.06	NA	[52]
Paris streets	4501	Transport	-0.06	NA	[52]
Manhattan streets	1046	Transport	-0.26	NA	[52]
San Francisco streets	3110	Transport	-0.01	NA	[52]
Toronto streets	2599	Transport	-0.06	NA	[52]

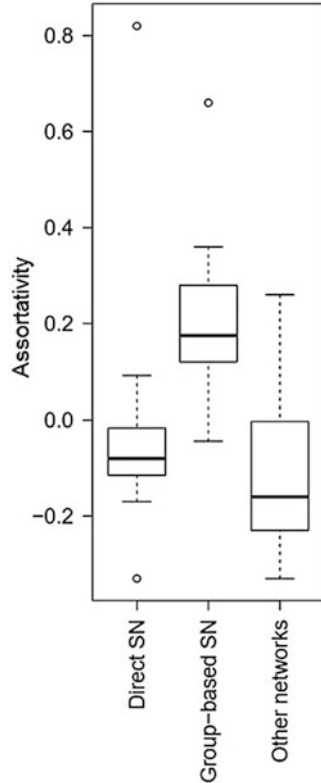
Type indicates the nature of the links between nodes, social are networks based on social interactions, technical are networks of interacting technology systems, biological are networks of some kind of biological process, transport networks are networks of a mode of transport in a particular area, mechanical networks are based on connections between parts in a defined object and links in commercial networks are formed when an employee moved from one company to another. Method of construction for social networks can be either “direct” where one to one interactions are used to build the network, or “group” where interactions are inferred based on shared use of physical space or contribution to a piece of work. All quoted degree correlations are from the Newman [10] method of calculating assortativity

$n(\text{direct social networks} = 27)$ ,  $n(\text{non-social network}) = 36$ ,  $p = 0.001$ ). The direct social networks possessed neutral assortativities (mean = 0.054, Wilcoxon signed rank test,  $V = 213$ ,  $n = 27$ ,  $p = 0.572$ ), while the group-based social networks were assortative (mean = 0.157, Wilcoxon signed rank test,  $V = 288$ ,  $n = 25$ ,  $p < 0.001$ ) and the non-social networks were disassortative (mean = -0.117, Wilcoxon signed rank test,  $V = 94.5$ ,  $n = 36$ ,  $p < 0.001$ ).

### 2.3 Literature Search: Conclusions

To confirm our hypothesis, that a preponderance of group-based methods makes it appear as if social networks typically possess assortativity, we found that group-based social networks were more assortative than direct social networks and non-social networks. However, direct social networks were still more assortative than non-social networks, which showed disassortativity on average. This therefore indicates that social networks only tend to be assortative if they are constructed with group-based methods, and non-social networks are typically disassortative. The

**Fig. 1** Boxplots of assortativities in each class of network. Differences between all classes were statistically significant (all Wilcoxon rank sum tests,  $p < 0.025$  in all cases). Direct social networks did not possess assortativities different from zero (mean = 0.054,  $p = 0.572$ ), while the group-based social networks were assortative (mean = 0.157,  $p < 0.001$ ) and the non-social networks were disassortative (mean = -0.117,  $p < 0.001$ ; all Wilcoxon signed rank tests)



processes that create disassortative networks is a large and active research area (e.g. [53, 54]) and so we will not investigate this result further here. We also note that many of the direct social networks in our table were based on interactions online. This is perhaps problematic if interactions online are fundamentally different to human interactions in the real world. Our choice of networks was simply based on what was available. Therefore, to confirm whether the assortativities of offline and online human social networks made using the same method (i.e. with direct interactions) are different, more offline human networks using direct interactions need to be constructed and published.

The bias towards assortativity when using group-based methods has been used has previously been noted [18, 55], but this appears to have escaped the notice of much of the research community, who continue to state that assortativity is a characteristic quality of a social network (e.g. [17, 23–28] but see [16, 22]). This may be because no clear rationale for this has been presented, nor solutions offered to combat it. We now go on to present why we think this assortativity was being discovered with group-based methods, if they are erroneous, and how to avoid incorrect estimates of assortativity in the future.

### 3 Methodological Pitfalls and False Assortativity

In this section we will discuss the issues that exist when using assortativity to describe the structure of social networks. We explore how methodologies used to sample networks could influence their assortativities, and whether changing how assortativity is calculated may also be important. Understanding the consequences of the method used to sample networks on properties such as assortativity could have important implications for our understanding of networks in general, as well as contributing to the further development of relevant analytical techniques.

#### 3.1 *Group-Based Networks and Assortativity*

In situations where it is not possible or feasible to directly observe social interactions, social networks are constructed using group-based methods. This approach has been applied both in constructing collaboration networks in humans, e.g. jazz musicians, scientific co-authorship networks and film actors [56–58] and also in networks based on co-occurrence in a social group in animals, e.g. song birds, dolphins and sharks [59–61]. Networks built using this method assume that every member of the group is associating with every other member of the group at each sampling census. This seems perfectly reasonable, hence this method for constructing social networks has been used by many studies [55, 62]. The assumption that meaningful social networks can be constructed based on this type of co-occurrence data has been termed “Gambit of the group” in the animal behaviour literature [63], which hints at the risk involved. This broad usage of group-based methods makes it very important that we fully understand the implications of using this method on the network and individual-level metrics calculated.

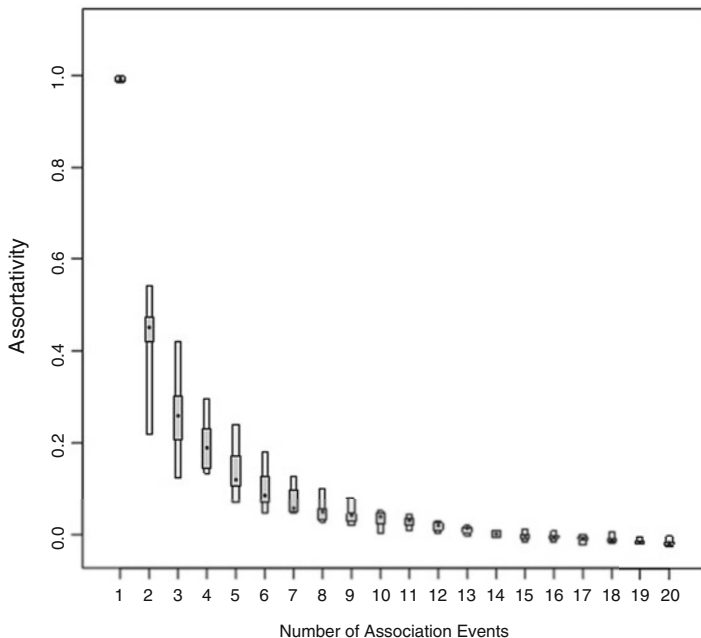
One study has previously investigated the influence of variation in sampling effort of group-based networks on various network-level metrics, including assortativity [29]. Franks et al. [29] investigated the impact of different group-based sampling regimes, changing both the total number of censuses (the number of times a population is sampled using a group-based approach) and the proportions of individuals sampled at each census, in random networks. The assortativity of a random network, which should on average be zero [10], was *always positive* if an insufficient number of censuses were completed before network construction ([29]; our emphasis). Crucially, any network not sampled intensively enough could show assortativity, either due to too few censuses or by not sampling enough of the population. Therefore, the use of group-based methods can produce a sampling bias, and requires further statistical analysis to determine the importance of the results obtained.



### 3.2 *Modeling Group-Based Sampling*

We wanted to extend the findings of Franks et al. [29] to demonstrate how assortativity changes as individuals are recorded in increasing numbers of groups. The aim was to show that in a system where freely moving individuals form social groups randomly (as opposed to a system based on random social networks), assortativity would decline as the effect of two individuals being seen in the same group diminished. We used a simulation-based approach where individuals associated randomly to construct networks using a group-based approach in a simple population of 100. Individuals were allocated randomly between 20 possible groups during each “census” using a symmetric Dirichlet distribution (with a uniform shape parameter) to define the size of each group. This enabled us to generate variation around a fixed expected group size, whilst maintaining a fairly consistent number of groups in each census (occasionally the size of a group could be zero). Any remaining individuals were allocated to a random group, meaning that all individuals were sampled in each census. Association data was collected over 20 censuses, with this process repeated for 10 different repeats of the simulation. After each census cumulative association data was recorded and the networks were dichotomised to create binary networks. In these networks individuals are either connected if they were observed in the same group at least once, or not connected if never observed together. This is necessary as assortativity is an unweighted measure, i.e. only the number of different associations an individual had is counted, not the frequency with which it associated with them. We discuss issues related to this method later. We then calculated assortativity for each set of cumulative association data. Simulations were carried out and degree correlations calculated in R 3.0.1 [64] (<http://www.R-project.org>).

We plot the results of these simulations in Fig. 2. From this plot it is clear that networks possess assortativity at a low number of censuses, and become gradually more neutrally assortative as the number of censuses increases (Fig. 2). This pattern emerges because of how group-based approaches are used in network construction. At low numbers of censuses, individuals that have been found in the same groups will both be connected and have similar degree, thus giving the network assortativity [15, 18]. As the number of censuses increases; however, this connection will gradually break down. Therefore, our simple simulation model shows that the assortativity found using group-based sampling approaches is highly dependent on the number of censuses completed. Indeed a low number of censuses can lead to many network measures being distorted [29, 65]. This is therefore an important consideration when deciding the sampling regimes used when constructing social networks using this method. The significance of the assortativity can only be established by comparison to appropriate null networks that truly highlight what aspects of the real network are interesting. This is achieved by randomly resampling observations using the correct group-size distribution [66], a method that will be outlined in more detail in Sect. 3.2.



**Fig. 2** The relationship between sampling effort (in terms of number of association events observed) and assortativity as a result of random social interactions in a simulated population. For each census the *black point* represents the median value, the *dark grey box* the interquartile range and the *light grey box* the range from ten runs of the simulation

### 3.3 Filtering Networks

Filtering networks involves taking a weighted (or valued) network, which contains information about how strong associations are as well as whether they are present or absent, and removing edges below a certain weight, i.e. infrequent or unimportant associations. This results in a binary network, with only edges above a certain weight present. The effect of dichotomising networks to transform them from weighted to binary has been studied for a number of network-level metrics [29]. The effect of this on assortativity is striking. While there was a limited effect of filtering at a low threshold (and therefore removing few edges), filtering at a high threshold (removing relatively many edges) had a considerable impact on the assortativity even when a high number of censuses on the simulated networks were completed. While for unfiltered, weighted random networks the assortativity reduced to zero as expected when a sufficient number of censuses were completed, this was not the case for networks filtered with a high threshold. These networks typically remained assortative at all levels of sampling (Figure 3 of [29]), despite being based on networks that were originally random and would be expected to be neutrally assortative. For example, removing all edges with a weight of less than 0.5 (out of 1) meant that the assortativity reached 0.4 after 10 censuses, compared

to 0.1 after 10 censuses if the threshold weight was set at 0.2. Many studies that have used group-based approaches to construct social networks have filtered the social networks produced before analysis [55], and it seems likely that the use of this approach may have played an additional role in inflating the degree correlation calculated for social networks.

## 4 Solutions

There are a number of solutions available to reduce the occurrence of erroneous assortativity that address each of the major issues outlined above. These methods have been developed in different parts of the social network literature, and if used collectively can greatly improve our understanding of the true variation in assortativity in all kinds of social networks.

### 4.1 *Increased Sampling*

When using group-based approaches, increasing the number of censuses above a threshold should produce a more accurate measure of the network's true assortativity [29]. Thus a suitable minimum number of censuses completed is required whenever using group-based approaches. If 80% of the population of interest can be sampled, then 10 censuses should be sufficient information on social interactions to accurately calculate network metrics. If only 40% of the population can be sampled, then 15–20 censuses may be required [29]. Collaboration networks, a key example of social networks that possess assortativity, typically do not have sampling periods. Instead all papers published over an agreed time period are looked at. To allow better comparison with networks and techniques that do have sampling periods, collaborations could be examined over particular, regular time periods, then each time period used as an "observation" from which to generate associations and so networks. However, if altering the sampling regime is not feasible, ensuring erroneous conclusions are not reached requires more rigorous statistical analysis.

### 4.2 *Use of Null Models*

Socio-biologists typically use resampling procedures to account for group-based sampling (and other, random) effects in their animal social networks [63, 66]. Resampling the observations used to construct the observed network using the original group-size data, and potentially additional biologically meaningful constraints, provides a large number of null networks against which to compare the observed network and test relevant hypotheses [66]. This method has been continually revised to take into account the non-independence of group sightings [67] and the risk of producing biased group compositions [68]. This approach has been used in a number

of studies (e.g. [69, 70]) and means that assortativity would only be considered interesting if the null model does not also display assortativity. We found one example of this approach being used in human social networks, in a study performed on a network of a board of directors [15]. However, no statistical tests were used for this comparison, and the method does not seem to have been universally adopted by the wider social network community. The importance of adopting this method is further highlighted the results of Franks et al. [29] which show that assortativity can be expected in networks where individuals randomly interact.

Random permutations are used by other network analysis techniques such as quadratic assignment procedure (QAP) to overcome problems with structural autocorrelation [71], but the Bejder et al. [66] method and extensions go further by randomly permuting the raw data, rather than the network matrix, to account for the distribution of group sizes and any other biases in sampling. Appropriate null models like these are the most effective way to make statistical inferences about networks constructed using group-based approaches, where controlling for effects such as group size is of considerable importance. There have been some recent examples of appropriate null models being used to test various hypotheses in the wider social networks literature (e.g. [72, 73]) and we add to their calls for this to be more widely embraced.

### 4.3 *Analysing Weighted Networks*

Weak interactions are potentially important (e.g. [74, 75]), and removing them can further increase error in the degree correlations calculated [29, 41]. If a population is observed for an extended period of time, and a large number of censuses are performed, the probability that two individuals are never observed to associate approaches zero. In a binary network, this would result in all individuals having the same, maximum, degree (which would be the population size  $-1$ ), and therefore the assortativity would increase before becoming undefined. Therefore, it is preferable to use weighted, unfiltered networks. For example, if collaboration networks are based on some measure of the information contained in an email [76], or the relative roles of those involved [77], you could then produce weighted networks that would be more informative to study and be less likely to be falsely assortative. The analysis of weighted networks is becoming increasingly manageable (e.g. [78]), reducing the need to filter networks and analyse them as binary data. As such, this approach should only be used when absolutely required [79], preferably with filtering using low thresholds (few edges removed). The filtering of some networks is unavoidable, for example, in brain imaging data which is likely to be very noisy [80], but we still recommend the lowest amount of filtering possible while accounting for edge uncertainty. Additionally, it would also be highly beneficial to continue to develop social network analysis in weighted networks to further reduce the requirement for binary network data, for example, by continuing the development of exponential random graph models that can be used in weighted networks [81, 82].

#### ***4.4 Using Diadic Over Group-Based Approaches***

The use of group-based methods is instrumental in creating some of the problems we suggest solutions for in previous subsections. If possible, alternative methods to group-based approaches should be used when constructing a network. For example, instead of paper co-authorship being used to define interactions between scientists, direct, diadic interactions based on correspondence such as emails could be used [76, 77]. Human social networks based on communication data such as with mobile phones or online are also examples of diadic interactions, and are becoming increasingly common [22, 76, 83–85]. In animal networks, this can be achieved by focussing on observation of suitable behavioural interactions [33, 86] or using reality mining methodologies [87], which are more fully outlined in Sect. 4.5. Social networks constructed using direct interactions on average possess neutral assortativity (see above), indicating they are not prone to the same problems with false assortativities that networks built using group-based approaches are.

#### ***4.5 Modern Technology***

Missing associations, and therefore edges in the network, is a problem largely unique to social networks. This can be prevented in human social networks, especially collaboration networks, by examining information-rich communications data. Information-rich interpersonal communications such as texts, tweets and emails can be used to construct networks on a very large scale with a high degree of accuracy [83]. In animal social networks a solution is to use reality mining [87], the concept of which has been borrowed from the sociology social network literature [88]. In these studies, modern technologies such as GPS trackers or proximity loggers are used to track animal movements and monitor interactions or accurately infer associations. For example, proximity loggers have been used to automatically record associations based on spatial and temporal overlap between individuals in animals such as cows, Tasmanian devils, crows and badgers [89–91], although see [92, 93] for discussion of the potential problems associated with using these devices.

These methods substantially reduce the number of missing edges in the network. Additionally, even when they don't allow us to move towards constructing networks using direct interactions, they can make it much more feasible to complete a large number of censuses if a group-based approach is still required. Furthermore, with careful thought technologies such as these could be used to complement the study of human social networks, potentially combined with data gathered using other methods. For instance, in the workplace, are those an employee physically associates with, e.g. at the water cooler or on a cigarette break, the same as those they communicate with electronically? What about different types of electronic communication, e.g. Facebook messages compared to work emails? Such application of modern technologies will only enhance the ability of scientists to measure social interactions in a wide range of networks.

## 4.6 *Alternatives to the Newman Degree Correlation Measure*

Recent research has suggested that calculating the correlation coefficient proposed by Newman [10] may not be appropriate for large networks, as it tends to misestimate the assortativity, especially in disassortative networks [25]. Litvak and van der Hofstad [25] suggested that using a method that ranks the degree of nodes, rather than using their absolute values (like a Spearman's rank correlation coefficient), may produce more valid results. Using this measure, the assortativity was consistent across a wide range of network sizes and typically consistently different from those calculated using Newman's assortativity coefficient (Figure 1 of [25]). It may be that changing the way that assortativity is calculated could reduce the differences between group-based networks, other social networks and different network types. This can be combined with the Bejder et al. [66] method to control for both group-size variation and biases associated with whole network size.

Alternatively, several authors have proposed individual-orientated metrics to quantify the tendency for well-connected individuals to connect to other individuals. The "Rich-clubs" of Zhou and Mondragon [94, 95], and the number of "differences" in node degree between neighbours of Thedchanamoorthy et al. [28] may both be robust to these pitfalls. Understanding the consequences of using different statistical approaches for calculating degree correlations should be the subject of further modeling work, in order to determine whether this can have an important influence to the properties with which a network is prescribed.

## 5 Conclusions

Assortativity has often been suggested to be a property of social networks that makes them different from non-social networks. We show that it is likely that this phenomenon may be driven by the data collection methods, resulting in previous studies overestimating the extent to which social networks possess assortativity. With alterations to methods or analyses, however, this problem with erroneous assortativities being described can be mitigated.

Through the use of methods initially developed in biology and sociology, we suggest that careful application of sampling and statistics can afford us increased confidence that finding assortativity in social networks is genuine, and not an artefact of the methodology used to construct them. It may well be that social networks are unique amongst types of network; there are certainly good reasons why they could have developed to be more assortative than non-social networks, e.g. increased robustness [11, 12]. However, given that information transfer and the ability to act in synchrony are greater in networks that show disassortativity [13, 14], it would perhaps be surprising if this property is as universal as has previously been described [10, 15].

Many of the methods discussed have been developed in one part of the social networks literature, and would benefit researchers in other fields greatly. Furthermore, following the integration of methods used in these different fields, it is likely that new insights will become available and further progress made in our ability to make inferences about assortativity in networks more generally.

## References

1. Girvan, M., Newman, M.E.J.: Community structure in social and biological networks. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 7821–7826 (2002)
2. Krause, J., Croft, D.P., James, R.: Social network theory in the behavioural sciences: potential applications. *Behav. Ecol. Sociobiol.* **62**, 15–27 (2007)
3. Pastor-Satorras, R., Vázquez, A., Vespignani, A.: Dynamical and correlation properties of the Internet. *Phys. Rev. Lett.* **87**, 258701 (2001)
4. Nemeth, R.J., Smith, D.A.: International trade and world-system structure: a multiple network analysis. *Rev. (Fernand Braudel Cent)*. **8**, 517–560 (2010)
5. Snyder, D., Kick, E.L.: Structural position in the world system and economic growth, 1955–1970: a multiple-network analysis of transactional interactions. *Am. J. Sociol.* **84**, 1096–1126 (1979)
6. Kapferer, B.: In: Boissevain, J., Mitchell, J.C. (eds.) *Norms and the Manipulation of Relationships in a Work Setting*, pp. 83–110. *Netw. Anal. Stud. Hum. Interact.* Mouton, Paris (1969)
7. Thurman, B.: In the office: networks and coalitions. *Soc. Networks.* **2**, 47–63 (1979)
8. Voelkl, B., Kasper, C.: Social structure of primate interaction networks facilitates the emergence of cooperation. *Biotechnol. Lett.* **5**, 462–464 (2009)
9. Zachary, W.W.: An information flow model for conflict and fission in small groups. *J. Anthropol. Res.* **473**, (1977)
10. Newman, M.: Assortative mixing in networks. *Phys. Rev. Lett.* **2**, 1–5 (2002)
11. Hasegawa, T., Konno, K., Nemoto, K.: Robustness of correlated networks against propagating attacks. *Eur. Phys. J. B.* **85**, 1–9 (2012)
12. Jing, Z., Lin, T., Hong, Y., et al.: The effects of degree correlations on network topologies and robustness. *Chin. Phys.* **16**, 3571–3580 (2007)
13. Di Bernado, M., Garofalo, F., Sorrentino, F.: Effects of degree correlation on the synchronization of networks of oscillators. *Int. J. Bifurcation Chaos.* **17**, 3499–3506 (2007)
14. Gallos, L., Song, C., Makse, H.: Scaling of degree correlations and its influence on diffusion in scale-free networks. *Phys. Rev. Lett.* **100**, 248701 (2008)
15. Newman, M., Park, J.: Why social networks are different from other types of networks. *Phys. Rev. E.* **68**, 036122 (2003)
16. Whitney, D., Alderson, D.: Are technological and social networks really different? *Unifying Themes Complex Syst.* **6**, 74–81 (2008)
17. Estrada, E.: Combinatorial study of degree assortativity in networks. *Phys. Rev. E.* **84**, 047101 (2011)
18. Newman, M.: Mixing patterns in networks. *Phys. Rev. E.* **67**, 026126 (2003)
19. Holme, P., Edling, C.R., Liljeros, F.: Structure and time evolution of an Internet dating community. *Soc. Networks.* **26**, 155–174 (2004)
20. Mac Carron, P., Kenna, R.: Universal properties of mythological networks. *EPL Europhys. Lett.* **99**, 28002 (2012)
21. Lusseau, D., Newman, M.: Identifying the role that animals play in their social networks. *Proc. R. Soc. B Biol. Sci.* **271**, S477–S481 (2004)
22. Hu, H.-B., Wang, X.-F.: Disassortative mixing in online social networks. *EPL Europhys. Lett.* **86**, 18003 (2009)

23. Araújo, E.B., Moreira, A.A., Furtado, V., et al.: Collaboration networks from a large CV database: dynamics, topology and bonus impact. *PLoS One*. **9**, e90537 (2014)
24. Furtenbacher, T., Arendás, P., Mellau, G., Császár, A.G.: Simple molecules as complex systems. *Sci. Rep.* **4**, 4654 (2014)
25. Litvak, N., van der Hofstad, R.: Uncovering disassortativity in large scale-free networks. *Phys. Rev. E*. **87**, 022801 (2013)
26. Mac Carron P, Kenna R. A quantitative approach to comparative mythology. [nestor.coventry.ac.uk](http://nestor.coventry.ac.uk) (2013)
27. Palathingal, B., Chirayath, J.: Clustering similar questions in social question answering services. In: Shan, L.P., Cao, T.H. (eds.) *The 16th Pacific Asia Conference on Information Systems (PACIS)*, USA, 13–15 July 2012 (2012)
28. Thechanamoorthy, G., Piraveenan, M., Kasthuriratna, D., Senanayake, U.: Node assortativity in complex networks: an alternative approach. *Proc. Comput. Sci.* **29**, 2449–2461 (2014)
29. Franks, D.W., Ruxton, G.D., James, R.: Sampling animal association networks with the gambit of the group. *Behav. Ecol. Sociobiol.* **64**, 493–503 (2009)
30. Piraveenan, M., Prokopenko, M., Zomaya, A.: Assortative mixing in directed biological networks. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **9**, 66–78 (2012)
31. Ciotti, V., Bianconi, G., Capocci, A., et al.: Degree correlations in signed social networks. *Phys. A Stat. Mech. Appl.* **422**, 25–39 (2015)
32. Ugander J, Karrer B, Backstrom L, Marlow C. The anatomy of the facebook social graph. *arXiv Prepr arXiv*. 1–17 (2011)
33. Manno, T.G.: Social networking in the Columbian ground squirrel, *Spermophilus columbianus*. *Anim. Behav.* **75**, 1221–1228 (2008)
34. Wang G, Wang B, Wang T, et al.. Whispers in the dark. In: *Proceedings of the 2014 Conference on Internet Measurement Conference—IMC '14*, ACM Press, New York, NY, pp. 137–150 (2014)
35. Shan, W., Liu, C., Yu, J.: Features of the discipline knowledge network: evidence from China. *Technol. Econ. Dev. Econ.* **20**, 45–64 (2014)
36. Sosa, S.: Structural architecture of the social network of a non-human primate (*Macaca sylvanus*): a study of its topology in La Forêt des Singes, Rocamadour. *Folia Primatol. (Basel)*. **85**, 154–163 (2014)
37. Lima A, Rossi L, Musolesi M. Coding Together at Scale: GitHub as a Collaborative Social Network. In: *Proceedings of 8th AAAI International Conference on Weblogs and Social Media (ICWSM)* (2014)
38. Wiszniewski, J., Lusseau, D., Möller, L.M.: Female bisexual kinship ties maintain social cohesion in a dolphin network. *Anim. Behav.* **80**, 895–904 (2010)
39. Farine, D.R., Aplin, L.M., Sheldon, B.C., Hoppitt, W.: Interspecific social networks promote information transmission in wild songbirds. *Proc. R. Soc. B.* **282**, 20142804 (2015)
40. Illenberger, J., Flötteröd, G.: Estimating network properties from snowball sampled data. *Soc. Networks.* **34**(4), 701–711 (2012)
41. Kossinets, G.: Effects of missing data in social networks. *Soc. Networks.* **28**, 247–268 (2006)
42. Whitehead, H.: *Analysing Animal Societies: Quantitative Methods for Vertebrate Social Analysis*. The University Chicago Press, Chicago (2008)
43. Lusseau, D., Wilson, B., Hammond, P.S., et al.: Quantifying the influence of sociality on population structure in bottlenose dolphins. *J. Anim. Ecol.* **75**, 14–24 (2006)
44. Croft, D.P., James, R., Thomas, P.O.R., et al.: Social structure and co-operative interactions in a wild population of guppies (*Poecilia reticulata*). *Behav. Ecol. Sociobiol.* **59**, 644–650 (2006)
45. Tsouchnika, M., Argyrakis, P.: Network of participants in European research: accepted versus rejected proposals. *Eur. Phys. J. B.* **87**, 292 (2014)
46. Mena-Chalco, J.P., Digiampietri, L.A., Lopes, F.M., Cesar, R.M.: Brazilian bibliometric coauthorship networks. *J. Assoc. Inf. Technol.* **65**, 1424–1445 (2014)
47. Mohman, Y.T., Wang, A., Chen, H.: Statistical analysis of the airport network of Pakistan. *Pramana.* **85**, 173–183 (2015)



48. Im, K., Paldino, M.J., Poduri, A., et al.: Altered white matter connectivity and network organization in polymicrogyria revealed by individual gyral topology-based analysis. *Neuroimage*. **86**, 182–193 (2014)
49. Lee, I., Kim, E., Marcotte, E.M.: Modes of interaction between individuals dominate the topologies of real world networks. *PLoS One*. **10**, e0121248 (2015)
50. Spitz, A., Horvát, E.-Á.: Measuring long-term impact based on network centrality: unraveling cinematic citations. *PLoS One*. **9**, e108857 (2014)
51. Aguirre-von-Wobeser, E., Soberón-Chávez, G., Eguiarte, L.E., et al.: Two-role model of an interaction network of free-living  $\gamma$ -proteobacteria from an oligotrophic environment. *Environ. Microbiol.* **16**, 1366–1377 (2014)
52. Jiang, B., Duan, Y., Lu, F., et al.: Topological structure of urban street networks from the perspective of degree correlations. *Environ. Plan. B Plan. Des.* **41**, 813–828 (2014)
53. Mussmann S, Moore J, Pfeiffer JJ, Neville J. Assortativity in Chung Lu random graph models. In: Proceedings of the 8th Workshop on Social Network Mining and Analysis—SNAKDD'14, ACM Press, New York, NY, pp. 1–8 (2014)
54. Yang R. Modifying network assortativity with degree preservation. In: 29th International Conference on Computers and Their Applications. CATA 2014, International Society for Computers and Their Applications, Winona, MN, pp. 35–40 (2014)
55. Croft, D.P., James, R., Krause, J.: Exploring Animal Social Networks. Princeton University Press, Oxford (2008)
56. Gleiser, P.M., Danon, L.: Community structure in jazz. *Adv. Complex Syst.* **06**, 565–573 (2003)
57. Newman, M.E.: The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 404–409 (2001)
58. Watts, D.J., Strogatz, S.H.: Collective dynamics of “small-world” networks. *Nature*. **393**, 440–442 (1998)
59. Aplin, L.M., Farine, D.R., Morand-Ferron, J., Sheldon, B.C.: Social networks predict patch discovery in a wild population of songbirds. *Proc. Biol. Sci.* **279**, 4199–4205 (2012)
60. Lusseau, D.: The emergent properties of a dolphin social network. *Proc. Biol. Sci.* **270**, 186–188 (2003)
61. Mourier, J., Vercelloni, J., Planes, S.: Evidence of social communities in a spatially structured network of a free-ranging shark species. *Anim. Behav.* **83**, 389–401 (2012)
62. Newman, M.: Coauthorship networks and patterns of scientific collaboration. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 5200–5205 (2004)
63. Whitehead, H., Dufault, S.: Techniques for analyzing vertebrate social structure using identified individuals: review and recommendations. *Adv. Study Behav.* **28**, 33–73 (1999)
64. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna (2013)
65. Perreault, C.: A note on reconstructing animal social networks from independent small-group observations. *Anim. Behav.* **80**, 551–562 (2010)
66. Bejder, L., Fletcher, D., Bräger, S.: A method for testing association patterns of social animals. *Anim. Behav.* **56**, 719–725 (1998)
67. Sundaresan, S.R., Fischhoff, I.R., Dushoff, J.: Avoiding spurious findings of nonrandom social structure in association data. *Anim. Behav.* **77**, 1381–1385 (2009)
68. Krause, S., Mattner, L., James, R., et al.: Social network analysis and valid Markov chain Monte Carlo tests of null models. *Behav. Ecol. Sociobiol.* **63**, 1089–1096 (2009)
69. Aplin, L.M., Farine, D.R., Morand-Ferron, J., et al.: Individual personalities predict social behaviour in wild networks of great tits (*Parus major*). *Ecol. Lett.* **16**, 1365–1372 (2013)
70. Wey, T.W., Burger, J.R., Ebensperger, L.A., Hayes, L.D.: Reproductive correlates of social network variation in plurally breeding degus (*Octodon degus*). *Anim. Behav.* **85**, 1407–1414 (2013)
71. Krackhardt, D.: Predicting with networks: nonparametric multiple regression analysis of dyadic data. *Soc. Networks.* **10**, 359–381 (1988)

72. Hanhijarvi S, Garriga GC, Puolmakai K. Randomization techniques for graphs. In: Proceedings of the 9th SIAM International Conference on Data Mining (SDM '09), pp. 780–791 (2009)
73. La Fond T, Neville J. Randomization tests for distinguishing social influence and homophily effects. In: Proceedings of the 19th international conference on World wide web—WWW '10, ACM Press, New York, NY, p. 601 (2010)
74. Farine, D.R.: Measuring phenotypic assortment in animal social networks: weighted associations are more robust than binary edges. *Anim. Behav.* **89**, 141–153 (2014)
75. Granovetter, M.: The strength of weak ties. *Am. J. Sociol.* **78**, 1360–1380 (1973)
76. Garton, L., Haythornthwaite, C., Wellman, B.: Studying online social networks. *J. Comput. Mediated Commun.* (2006). doi:10.1111/j.1083-6101.1997.tb00062.x
77. Rowe R, Creamer G, Hershkop S, Stolfo SJ. Automated social hierarchy detection through email network analysis. In: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis—WebKDD/SNA-KDD '07, ACM Press, New York, NY, pp. 109–117 (2007)
78. Opsahl, T., Agneessens, F., Skvoretz, J.: Node centrality in weighted networks: Generalizing degree and shortest paths. *Soc. Networks.* **32**, 245–251 (2010)
79. Noldus, R., Van Mieghem, P.: Assortativity in complex networks. *J. Complex Networks.* **3**, 507–542 (2015)
80. Iturria-Medina, Y., Canales-Rodríguez, E.J., Melie-García, L., et al.: Characterizing brain anatomical connections using diffusion weighted MRI and graph theory. *Neuroimage.* **36**, 645–660 (2007)
81. Krivitsky, P.: Exponential-family random graph models for valued networks. *Electron. J. Stat.* **6**, 1100–1128 (2012)
82. Krivitsky P, *ergm.count*: Fit, simulate and diagnose exponential-family models for networks with count edges. The Statnet Project (2015). <http://www.statnet.org>. R package version 3.2.2. <http://CRAN.R-project.org/package=ergm.count>
83. De Choudhury M, Mason W. Inferring relevant social networks from interpersonal communication. In: Proceedings of the 19th International Conference on World wide web, pp. 301–310 (2010)
84. Expert, P., Evans, T.S., Blondel, V.D., Lambiotte, R.: Uncovering space-independent communities in spatial networks. *Proc. Natl. Acad. Sci. U. S. A.* **7663–7668** (2011)
85. Peruani, F., Tabourier, L.: Directedness of information flow in mobile phone communication networks. *PLoS One.* **6**, e28860 (2011)
86. Wey, T.W., Blumstein, D.T.: Social cohesion in yellow-bellied marmots is established through age and kin structuring. *Anim. Behav.* **79**, 1343–1352 (2010)
87. Krause, J., Krause, S., Arlinghaus, R., et al.: Reality mining of animal social systems. *Trends Ecol. Evol.* **28**, 1–11 (2013)
88. Eagle, N., Pentland, A.: Reality mining: sensing complex social systems. *Pers. Ubiquitous Comput.* **10**, 255–268 (2005)
89. Böhm, M., Hutchings, M.R., White, P.C.L.: Contact networks in a wildlife-livestock host community: identifying high-risk individuals in the transmission of bovine TB among badgers and cattle. *PLoS One.* **4**, e5016 (2009)
90. Hamede, R.K., Bashford, J., McCallum, H., Jones, M.: Contact networks in a wild Tasmanian devil (*Sarcophilus harrisi*) population: using social network analysis to reveal seasonal variability in social behaviour and its implications for transmission of devil facial tumour disease. *Ecol. Lett.* **12**, 1147–1157 (2009)
91. Rutz, C., Burns, Z.T., James, R., et al.: Automated mapping of social networks in wild birds. *Curr. Biol.* **22**, R669–R671 (2012)
92. Boyland, N.K., James, R., Mlynski, D.T., et al.: Spatial proximity loggers for recording animal social networks: consequences of inter-logger variation in performance. *Behav. Ecol. Sociobiol.* **67**, 1877–1890 (2013)

93. Drewe, J.A., Weber, N., Carter, S.P., et al.: Performance of proximity loggers in recording intra- and inter-species interactions: a laboratory and field-based validation study. *PLoS One*. **7**, e39068 (2012)
94. Zhou, S., Mondragón, R.: The rich-club phenomenon in the Internet topology. *IEEE Commun. Lett.* 1–3 (2004)
95. Colizza, V., Flammini, A., Serrano, M.A., Vespignani, A.: Detecting rich-club ordering in complex networks. *Nat. Phys.* **2**, 1–18 (2006)