

Social network dynamics: the importance of distinguishing between heterogeneous and homogeneous changes

Mathias Franz¹ · Susan C. Alberts¹

Received: 13 January 2015 / Revised: 25 October 2015 / Accepted: 26 October 2015 / Published online: 4 November 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract Social network analysis is increasingly applied to understand the evolution of animal sociality. Identifying ecological and evolutionary drivers of complex social structures requires inferring how social networks change over time. In most observational studies, sampling errors may affect the apparent network structures. Here, we argue that existing approaches tend not to control sufficiently for some types of sampling errors when social networks change over time. Specifically, we argue that two different types of changes may occur in social networks, heterogeneous and homogeneous changes, and that understanding network dynamics requires distinguishing between these two different types of changes, which are not mutually exclusive. Heterogeneous changes occur if relationships change differentially, e.g., if some relationships are terminated but others remain intact. Homogeneous changes occur if all relationships are proportionally affected in the same way, e.g., if grooming rates decline similarly across all dyads. Homogeneous declines in the strength of relationships can strongly reduce the probability of observing weak relationships, producing the appearance of heterogeneous network changes. Using simulations, we confirm that failing to differentiate homogeneous and heterogeneous changes can potentially lead to false conclusions about network dynamics. We also show that bootstrap tests fail to distinguish between homogeneous and heterogeneous

changes. As a solution to this problem, we show that an appropriate randomization test can infer whether heterogeneous changes occurred. Finally, we illustrate the utility of using the randomization test by performing an example analysis using an empirical data set on wild baboons.

Keywords Social networks · Social network analysis · Social network dynamics · Sampling errors

Introduction

Social network analysis is increasingly used to study the structures of animal societies (Croft et al. 2008; Wey et al. 2008; Whitehead 2008; Krause et al. 2009; Sueur et al. 2011). While most studies of social networks in animal behavior have focused on describing static network structures, there is an increasing interest in studying how social networks change over time and what determines the stability of social networks (Wittemyer et al. 2005; Flack et al. 2006; Hansen et al. 2009; Ansmann et al. 2012; Blonder et al. 2012; Cantor et al. 2012; Foster et al. 2012; Brent et al. 2013; Gero et al. 2013; Hobson et al. 2013; Boogert et al. 2014; Pinter-Wollman et al. 2014; Wilson et al. 2014). Indeed, inferring changes in social network structures has the potential to provide crucial insights into how social dynamics change over time, for instance, in response to seasonal changes (Wittemyer et al. 2005; Henzi et al. 2009; Brent et al. 2013) or in response to disturbances such as the removal of important individuals (Flack et al. 2006; Barrett et al. 2012).

In most observational studies, only a subset of the social interactions that characterize social networks will actually be observed and recorded. This may be because the duration of a given study is too short to see interactions that are rare, or it may be because interactions occur when observers are not

Communicated by D. P. Croft

Electronic supplementary material The online version of this article (doi:10.1007/s00265-015-2030-x) contains supplementary material, which is available to authorized users.

✉ Mathias Franz
mf144@duke.edu

¹ Department of Biology, Duke University, Durham, NC, USA

observing the group. In either case, particularly weak social relationships are disproportionately likely to remain unrecorded. While this is an obvious problem in field studies, the same problem also occurs in captive studies unless all animals can be observed for the whole time during which social interactions occur. As a consequence, the observed interactions provide only an approximation of the interactions that actually occurred, and the resulting social networks that are inferred are likely to provide an incomplete representation of the actual networks (Farine and Whitehead 2015).

It is widely acknowledged that, to ensure that inferences about the structure of social relationships are robust, it is vital to include an assessment of sampling errors in the analysis of observed social networks (Bejder et al. 1998; Borgatti et al. 2002; Whitehead et al. 2005; Lusseau et al. 2008; Whitehead 2008; James et al. 2009; Croft et al. 2011; Voelkl et al. 2011). However, our examination of the literature indicates that most studies do not generally consider one potential source of sampling error which, if not recognized, can lead to incorrect inferences about changes in social networks. Here, we present an approach that takes into account this type of sampling error.

Specifically, we argue that in analyzing changes in social networks that occur, for instance, after a social disturbance, it is important to differentiate between two different possible types of changes, which we refer to as homogeneous and heterogeneous changes. Homogeneous changes are those in which all relationships change in the same way while the overall pattern of strong and weak relationships remains unchanged (Fig. 1). As an example, a group-wide decrease in interaction rates that affects all dyads in a similar manner would qualify as a homogeneous change. In contrast, heterogeneous changes in relationships are those that affect different relationships in different ways (Fig. 1). For instance, some relationships become weaker or are terminated while other relationships remain unchanged. In other words, heterogeneous changes are changes that result in a “rewiring” of the social relationship network. Both types of change may occur simultaneously. For example, all relationships in a network might become proportionally weaker (a homogeneous change) and, in this process in which all ties weaken, some relationships are additionally terminated (a heterogeneous change).

To further illustrate the differences between both types of changes, we suggest a conceptual distinction between (1) the pattern of variation among relationships in a network and (2) the rate with which these relationships are expressed behaviorally. The pattern of variation among relationships describes how each dyad behaves relative to other dyads; some dyads have strong and some have weak relationships so that, for instance, the grooming rate between individuals *A* and *B* is twice the grooming rate between individuals *A* and *C*. The behavioral expression of this pattern in a given timeframe then

results in instances of observed social behaviors, which would be for instance the number of grooming events between individuals *A* and *B* and between individuals *A* and *C*. Following this distinction, changes in relationships can occur (1) in the pattern of variation among relationships in a network or (2) in the rate with which these relationships are expressed behaviorally. Heterogeneous changes refer to the first case (e.g., if the grooming rate between individuals *A* and *B* changes only moderately while the grooming rate between individuals *A* and *C* is terminated) and homogeneous changes refer to the second case (e.g., if the grooming rate between all individuals in a group are changed by the same factor in a multiplicative way).

Importantly, sampling errors (which might be produced, for instance, by any limit on sampling effort) may cause homogeneous changes to resemble heterogeneous changes. To illustrate this problem, we use a simple example: if all relationships in a network become proportionally weaker in a homogeneous manner (for instance because of a seasonal change in food supply), and if weak relationships are unlikely to be observed, then the total number of unobserved relationships would increase as a result of the homogeneous change. The increased number of unobserved relationships would then lead to a decrease in the mean degree in the observed network (e.g., a decrease in the mean number of interaction partners). A similar decrease in mean degree could be also caused by a heterogeneous change, that is, by a situation in which some relationships were truly terminated whereas others remain stable.

Which type of change occurs or dominates could have profoundly different implications for the interpretation of the observed social dynamics. Specifically, our interpretation of the effect of a social disturbance on a network will vary depending on whether we detect a global (homogeneous) change in interaction levels across the network after that disturbance, or a “rewiring” of the network after that disturbance. While it is possible that both effects co-occur, it might be often the case that one effect is much stronger and dominates the observed social dynamics.

The main aim of this study is thus to raise the awareness that different kinds of changes in social relationships can appear to have similar effects on network structures and that it is important to distinguish between these types of changes. To achieve this aim, we present simulation experiments based on data from wild baboons to illustrate that homogeneous and heterogeneous changes can indeed lead to similar apparent changes in observed social networks. In addition, we show that one commonly used test for the analysis of changes in social network structures, the bootstrap test (e.g., Lusseau et al. 2008; Henzi et al. 2009; Brent et al. 2013), fails to distinguish between homogeneous and heterogeneous changes. We then show that an appropriate

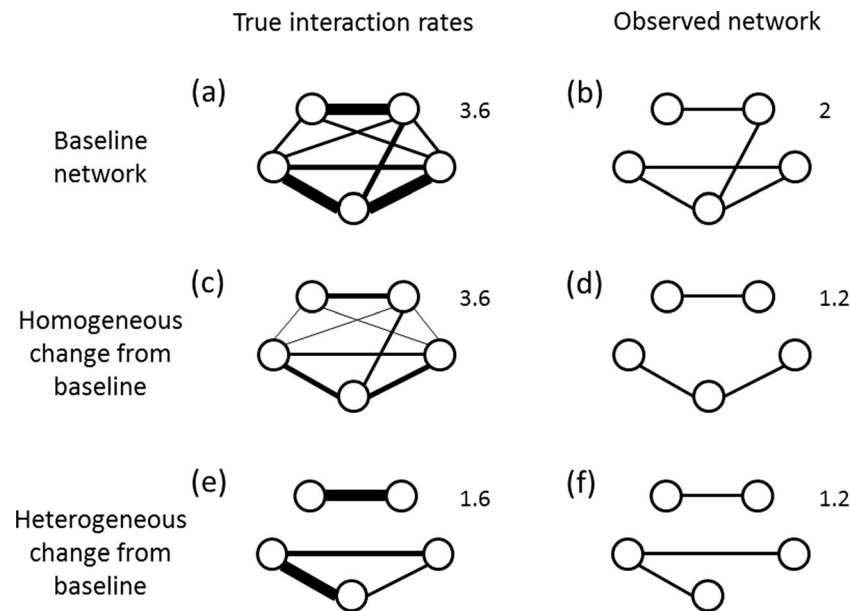


Fig. 1 Illustration of how heterogeneous and homogeneous changes can affect observed social network structures. *Circles* represent individuals. In (a), (c), and (e), edge thickness indicates magnitude of true interaction rates. In (b), (d), and (f), edges indicate whether interactions among two individuals have been observed. *Numbers* indicate mean degree, i.e., the mean number of grooming partners per individual. **a, b** Baseline scenario: low interaction rates are likely to result in undetected relationships. **c, d** Example of a homogeneous change in the baseline scenario: all interaction rates decrease equally, which increases the number of relationships with low interaction rates. Consequently, in this case, the number of dyads for which interactions are observed decreases. Note that although no “true” zero interaction rates occur as a consequence of

homogeneous changes (c), observer error (e.g., simply failing to observe every interaction between all subjects) could produce apparent zero interaction rates (d). **e, f** Example of a heterogeneous change in the baseline scenario: several interaction rates are strongly decreased or set to zero, which also decreases the number of dyads for which interactions are observed. Taken together, this example illustrates how homogeneous and heterogeneous changes can have similar effects on observed networks and associated network measures such as mean degree. Importantly, inferred changes in the observed network correctly approximate the true changes in interaction rates in the case of a heterogeneous change, but not in the case of a homogeneous change

randomization test can be used instead of, or in addition to, the bootstrap, to infer whether heterogeneous changes occurred. In contrast to the bootstrap test, which generates new datasets by randomly selecting observations with replacement, a randomization test in our case involves shuffling of observations between time periods, with constraints (see “Methods” section for details). After describing our proposed application of a randomization test, we perform an example analysis to investigate the effects of the dispersal of the alpha male on the grooming network among adult female baboons. The application of these two different tests illustrates the importance of distinguishing between heterogeneous and homogeneous network changes.

The simulation experiments that we conducted were designed to provide proofs of principles for our main arguments. For that purpose, we focused initially on a few network measures and the simple cases in which either a homogeneous or a heterogeneous change occurs. We present additional analyses in the [Supplementary materials](#) in which we extended this basic approach by (1) considering additional network measures and (2) simulating the simultaneous occurrence of homogenous and heterogeneous changes.

Methods

Simulation of homogeneous and heterogeneous changes

All simulations followed the same conceptual framework. We used empirically observed grooming interaction rates a_x as a baseline that characterized social relationships in a group of individuals at time x (see details of empirical data collection below). Solely for the purposes of our simulations, we assumed these observed grooming rates a_x to be the true, error-free grooming rates (Fig. 1a). This baseline measure of true grooming rates was then used in simulations in which we imposed either homogeneous or heterogeneous changes, which resulted in modified interaction rates a_y that characterized social relationships at time y (represented in Fig. 1 as transitions from panel a to panel c, and from panel a to panel e). Finally, we simulated how interactions rates specified by a_x and a_y resulted in observations of interactions o_x and o_y (represented in Fig. 1 as transitions from panel c to panel d, and from panel e to panel f). Simulated observations were then used to construct social networks and assess how the homogeneous versus heterogeneous changes we imposed affected different network measures (see details below). In addition, pairs of observed interactions o_x and o_y were used as input

to the bootstrap and randomization tests described in the “Differentiating heterogeneous from homogeneous changes” section.

Baseline interaction rates a_x were derived from grooming data collected between January 2008 and June 2008 by the Amboseli Baboon Research Project on one group of yellow baboons, which at that time consisted of 56 individuals. Data on grooming between all possible pairs of individuals were collected ad libitum and during 10-min focal samples (Altmann 1974). Focal samples were conducted in random order on all adult females and juveniles in a given social group. This approach insured that observers continually moved to new locations within the group in a random order, observing all animals on a regular rotating basis. Thus, our procedure for data collection eliminated the possibility that observers spent more time watching particular subsets of the social group, or moved in a biased manner through the group detecting only the most dramatic events.

Based on a total of 1933 grooming events for all dyads in the group during this 6-month window, we calculated grooming rates $a_{x,ij}$ among all individuals i and j for this time period (note—these 1933 events involve only a subset of possible dyads; not all dyads engage in grooming behavior, and as with any observational study some grooming events are inevitably not recorded). Using a relatively large time window of 6 months facilitates detecting weak relationships, which is particularly important for accurately modeling the effect of homogeneous changes. A potential drawback of the large time window is that we ignore potential changes in social relationships that might have occurred during this time window. However, this problem seems to be of minor importance for demonstrating the potential effects of homogeneous and heterogeneous changes.

To simulate heterogeneous network changes, we simulated the complete removal of some grooming relationships (representing an extreme case of heterogeneous changes out of many possible scenarios, represented in Fig. 1 as the transition from panel a to panel e). Relationship removals were implemented by randomly selecting a proportion p of all non-zero rates in a_x and setting them to zero. To simulate homogeneous changes, all baseline grooming rates a_x were multiplied by a factor q (represented in Fig. 1 as the transition from panel a to panel c). For both time periods, observations o_{ij} for all dyads of individuals i and j were simulated by drawing a random number from a Poisson distribution with $\lambda = a_{ij}$ for each a_{ij} (represented in Fig. 1 as the transitions from panel c to panel d, and from panel e to panel f). This assumption is, for instance, well justified in cases where observational data are collected ad libitum (Altmann 1974) and all group members are equally visible to the observer.

Drawing observations o_y from a distribution captures two distinct stochastic processes: (1) the behavioral expression of grooming events, i.e., whether and how many grooming

events take place, and (2) the observation of these grooming events, i.e., whether and how many of the occurred grooming events are observed. As a consequence as values of a_y decrease as a result of homogeneous change from a_x towards 0, the value of o_y becomes increasingly likely to be 0 (i.e., no grooming interactions are observed) even though the value a_y never itself reaches 0 when a_x is greater than 0. This simulates a real-life situation where rare interactions, although present, may never be observed or a situation in which interaction rates become so small that interactions are rarely expressed in the considered time interval.

In simulating heterogeneous changes, we varied the proportion of removed grooming relationships p from 0 to 0.5 in increments of 0.05. In simulations of homogeneous changes, we varied the factor q from 0.5 to 1 in increments of 0.05 to simulate relationships that were homogeneously weakened by varying degrees. In all cases, we conducted 500 independent simulations for each condition.

After each simulation was complete, we constructed an undirected binary network from the simulated grooming observations; in this network, edge weights for all dyads without any grooming interaction were set to 0, and weights for all dyads with at least one interaction were set to 1. The package `igraph` (Csardi and Nepusz 2006) in the statistical software R (R Core Team 2014) was used to calculate for each network two binary network measures: the mean degree (which measures the average number of interaction partners) and the global clustering coefficient (which is a measure of “cliquishness”). In addition, we calculated network entropy, a weighted network measure. Network entropy measures network-wide heterogeneity in interactions by taking into account interaction frequency and directionality. Given a set of observed interactions o , and the corresponding proportions of grooming given $pg_{i,j}$ from i and j for all pairs of individuals where i groomed j at least once, we calculated entropy $H(o)$ as follows:

$$H(o) = -\sum_{\text{all } i,j} pg_{i,j} \log(pg_{i,j}) \quad (1)$$

In additional analyses, presented in the [Electronic supplementary material](#), we also investigated how homogeneous and heterogeneous changes in the network affect weighted clustering coefficients.

Differentiating heterogeneous from homogeneous changes

We compared the performance of a bootstrap test with the performance of a randomization test that allowed us to infer heterogeneous changes in network structures. As input to these tests, we used paired simulated observations of social interactions o_x and o_y , which were derived from unmanipulated baseline interaction rates a_x and manipulated interaction rates a_y . For each pair of networks, we used the bootstrap test

and the randomization test to determine whether mean degree, global clustering coefficient, or entropy significantly changed from simulated observation period x to period y .

The bootstrap test tests the data against the null hypothesis that an observed change in a network measure occurred entirely because of random sampling errors (i.e., the null hypothesis assumes that no real change occurred). To this end, null distributions are generated for each observed set of grooming events o_x and o_y . These distributions describe the expected variation in considered network measures because of random sampling errors.

We implemented the generation of these distributions by (1) resampling single grooming events with replacement from the raw observation data, i.e., observed sets of grooming events o_x and o_y (while keeping the total number of observed grooming events constant between networks), (2) constructing unweighted and weighted networks from the sampled data, and (3) calculating all three network measures (mean degree, global clustering coefficient and entropy). We generated 1000 samples for each time period (based on o_x and o_y), which allowed estimating 95 % confidence intervals for each network measure for each time period. A change in a network measure between the two time periods was assumed to be significant if the confidence intervals did not overlap.

The randomization test that we used here tests the data against the null hypothesis that an observed change in a network measure occurred either because of random sampling errors or because of systematic sampling errors caused by homogeneous changes (i.e., the null hypothesis assumes that no heterogeneous change occurred, but that homogeneous changes could have occurred). To this end, a null distribution is generated that describes the change in a given network measure that is expected either because of homogeneous changes or because of random sampling errors. The randomization test becomes significant if it is sufficiently unlikely that the observed change in a network measure was generated by this expected distribution. Note, because this test controls for potential homogeneous changes, heterogeneous changes can be detected irrespectively of whether homogeneous changes occurred.

We implemented this test by first performing randomizations on the raw observation data. For a single randomization, observations of single grooming events were randomized between the two sets of observed grooming events o_x and o_y at time periods x and time y while retaining the original number of observations for each time period. More specifically, each data point in the input data set corresponds to a single observation of a pairwise grooming interaction. In addition to the information about who groomed whom, each data point contains information on the respective time period (x or y) in which the observation was made. During the randomization procedure, the assigned time period was randomized among all data points (i.e., each data point was reassigned to a time

period without changing the total number of data points that are assigned to each period). This procedure is based on the assumption that the variation in interaction frequencies among individuals are identical in both observation periods (i.e., no heterogeneous change occurred from time x to time y) but that absolute number of interactions might differ as a result of homogeneous changes (which would result in different total numbers of observations at time y relative to time x). After each randomization, unweighted and weighted networks are constructed from the randomized data and network measures are calculated from these networks.

We performed 1000 randomizations and used 0.025 and 0.975 quantiles to estimate 95 % confidence intervals of expected changes. A change in a network measure between the two time periods was assumed to be significant if the observed change was outside the 95 % confidence interval of expected changes. We further investigated the tests described above using a large number of artificially created social networks in which we (1) varied network size and other network properties and (2) investigated cases in which homogeneous and heterogeneous changes occurred simultaneously (see [Electronic supplementary material](#)). In addition, we provide the R code that implements the randomization test (see [Electronic supplementary material](#)).

Note that the randomization approach we used, which involved retaining the original number of observations for each time period, is of key importance for the functioning of this method, but it also imposes some constraints. Retaining the original numbers of observations for each time period implies that mean interaction frequencies in each time period remain unchanged. This property ensures that the randomization test is able to estimate expected changes in network measures if homogeneous changes occur. However, another consequence of retaining the original numbers of observations for each time period is that it will not be possible to detect changes in group-level mean strength (also sometimes referred to as weighted degree), particularly if dyad-specific interaction frequencies are used as edge weights (as we have done here). That is, if one retains the original numbers of observations for each time period, randomizations do not affect interaction frequencies (which are equivalent to group-level mean strength) and therefore cannot generate meaningful null distributions for expected changes of this measure. For this reason, we also chose not to include investigations of mean strength in our analysis.

Example analysis: effects of the dispersal of an alpha male

To further illustrate the importance of considering homogeneous and heterogeneous changes, we applied the bootstrap test and the randomization test to investigate how grooming networks of adult female baboons changed following the dispersal of an alpha male. The specific dispersal event that we investigated occurred in December 2005 in one of Amboseli

study groups, which contained 13 adult females at this time. In our analysis, we compared grooming data collected within 30 days before and 30 days after the dispersal event. We applied the bootstrap and the randomization test using mean degree, global clustering coefficient, and network entropy as test statistics. For the bootstrap test, we generated 1000 samples to estimate 95 % confidence intervals for each network measure for each time period, and for the randomization test we performed 1000 randomizations to estimate 95 % confidence intervals.

Results

As we predicted, both a simulated homogeneous change and a simulated heterogeneous change tended to decrease the mean degree in a social network (Fig. 2a, d). Importantly, we also found the same effect for the global clustering coefficient (Fig. 2b, e), network entropy (Fig. 2c, f), and weighted

clustering coefficients (see [Electronic supplementary material](#)). This result shows that homogeneous changes can systematically affect not only mean degree but also other network measures, including weighted network measures such as entropy and weighted clustering coefficients.

We next examined the proportion of cases in which the difference between baseline and modified grooming patterns produced statistically significant changes in network structure. Specifically, we asked whether increasingly large declines in overall grooming rates (in simulations of homogeneous changes) or increasingly large numbers of terminated relationships (in simulations of heterogeneous changes) produced increasingly larger fractions of cases in which the network structure changed significantly (Fig. 3). We asked this question for three different network measures (mean degree, global clustering coefficient, and entropy), changing in response to two different types of change (simulated heterogeneous and simulated homogeneous changes).

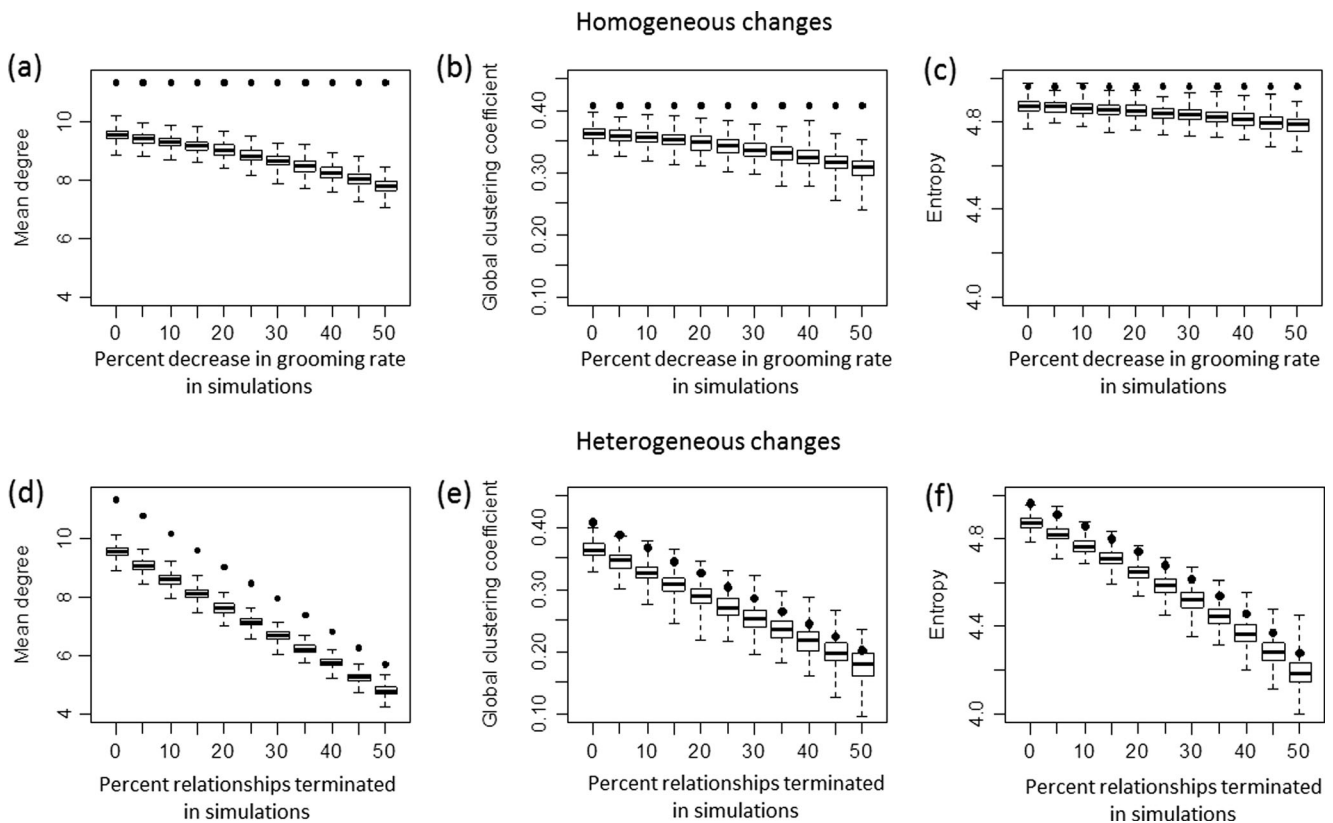


Fig. 2 Homogeneous changes (a, b, c) and heterogeneous changes (d, e, f) can result in similar apparent changes in network measures, leading to potentially incorrect inferences about how relationships have changed over time. Dots indicate the mean network measures of networks based on true interaction rates (a_x). Boxplots indicate network measures (max, 75th percentile, median, 25th percentile, minimum) of networks based on simulated observations of interaction rates (o_x). In (a), (b), and (c), our simulations reduced all interaction rates to the same extent, but sampling error produced changes that appear similar to heterogeneous changes, the simulations for which are shown in (d), (e), and (f). Importantly, only in

the cases of heterogeneous changes, but not in cases of homogeneous change, inferred changes in the observed network structures correctly approximate the true changes in interaction rates. This effect is not only true for the straightforward case of mean degree (a, d), but it also applies to global clustering coefficient (b, e) and entropy (c, f). Note that the percent changes in heterogeneous and homogeneous changes are not directly comparable, but (a) shows how decreasing grooming rates relate to a decrease in mean degree. This decrease in mean degree is caused by a decrease in the number of observed relationships, which directly reflects the increase in unobserved relationships

In three of these six contexts, specifically those in which we simulated homogeneous changes, the results from the bootstrap tests differed strongly from those of the randomization tests. Specifically, as overall grooming rates decreased, the bootstrap test reported strong increases in proportion of significant results (Fig. 3a–c). In contrast, for the randomization test the proportion of significant results did not increase with increasing homogeneous changes and never exceeded the expected type I error rate (Fig. 3a–c).

In the case of simulated heterogeneous changes, both the randomization and the bootstrap tests performed in a similar way: the proportion of significant results increased with increasing heterogeneous changes (Fig. 3d–f). This result shows that although these tests strongly differ in their reaction to homogeneous changes (Fig. 3a–c), both tests are able to detect heterogeneous changes with comparable success (Fig. 3d–f).

In other words, the bootstrap test reported changes in mean degree, global clustering coefficient, and entropy even when such changes did not occur, implying that it could not distinguish between changes in network measures caused by homogeneous changes from those caused by heterogeneous changes. The randomization test provides a solution to this problem as it controls for changes in network measures caused by

homogeneous changes. These findings were confirmed by additional analyses with artificially created networks in which we vary network size and other network properties (see [Electronic supplementary material](#)). Our additional analyses furthermore confirmed that the randomization test can detect heterogeneous changes in the presence of homogeneous changes, even in cases in which homogeneous and heterogeneous changes have similar effects on observed network structures (see [Electronic supplementary material](#)).

Results from our example analysis further illustrate the differences between the bootstrap and the randomization test and the utility of using the randomization test instead or in addition to the bootstrap test. With the application of the bootstrap test, the mean degree and network entropy in the female network significantly decreased following the dispersal of the alpha male (Fig. 4a, c), but no significant change was detected for the global clustering coefficient (Fig. 4b). An additional analysis revealed that individual grooming rates also decreased after the dispersal event (Fig. 4d). This information about individual grooming rates is ignored by the bootstrap test. In contrast, the randomization test automatically accounts for changes in grooming rates when calculating the distribution of expected changes in network measures. The

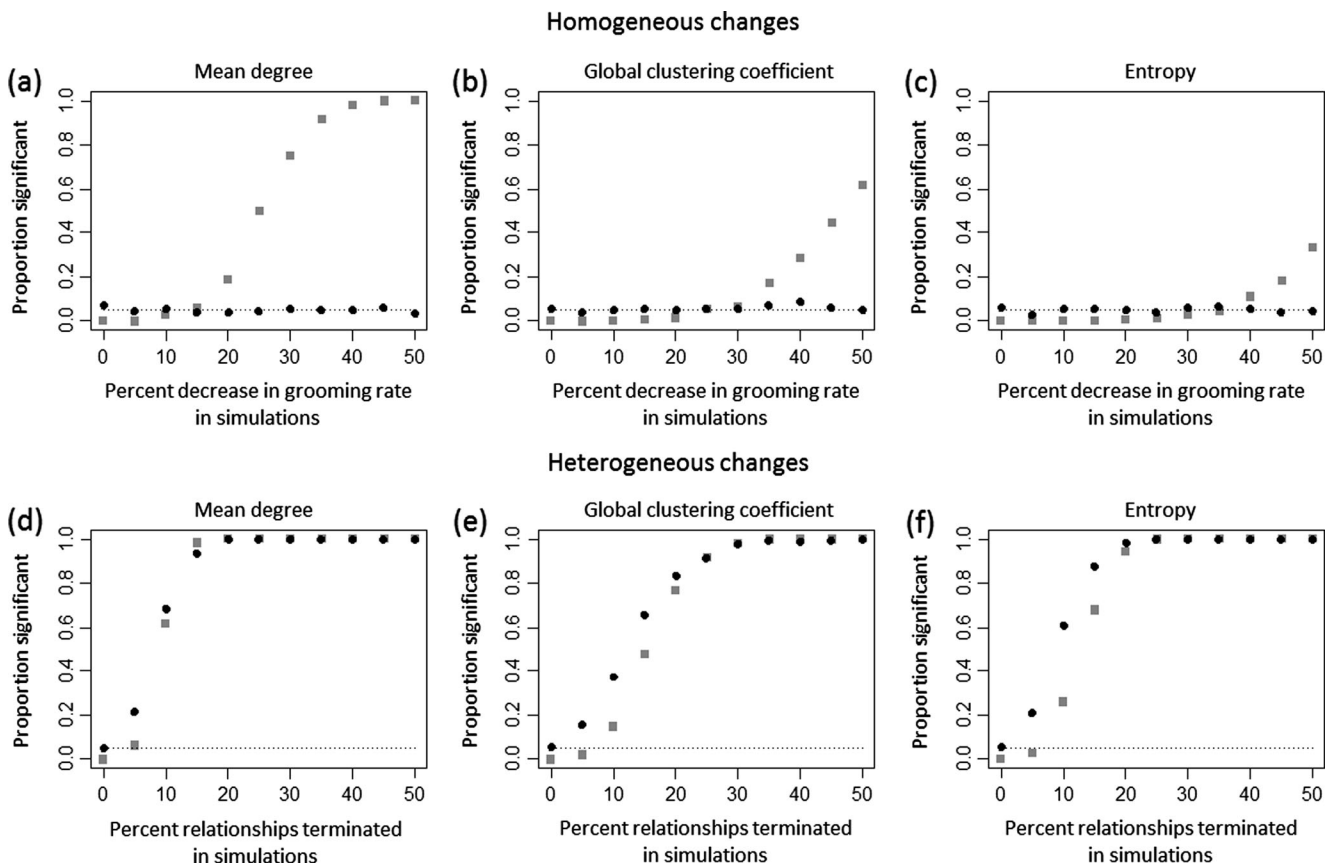


Fig. 3 Effects of simulated homogeneous changes (a, b, c) and simulated heterogeneous changes (d, e, f) on the proportion of significant changes reported in three network measures: mean degree (a, d), global clustering

coefficient (b, e), and entropy (c, f). *Gray squares* indicate results using a bootstrap test; *black circles* indicate results using the randomization test. The *dotted line* indicates the expected type I error rate

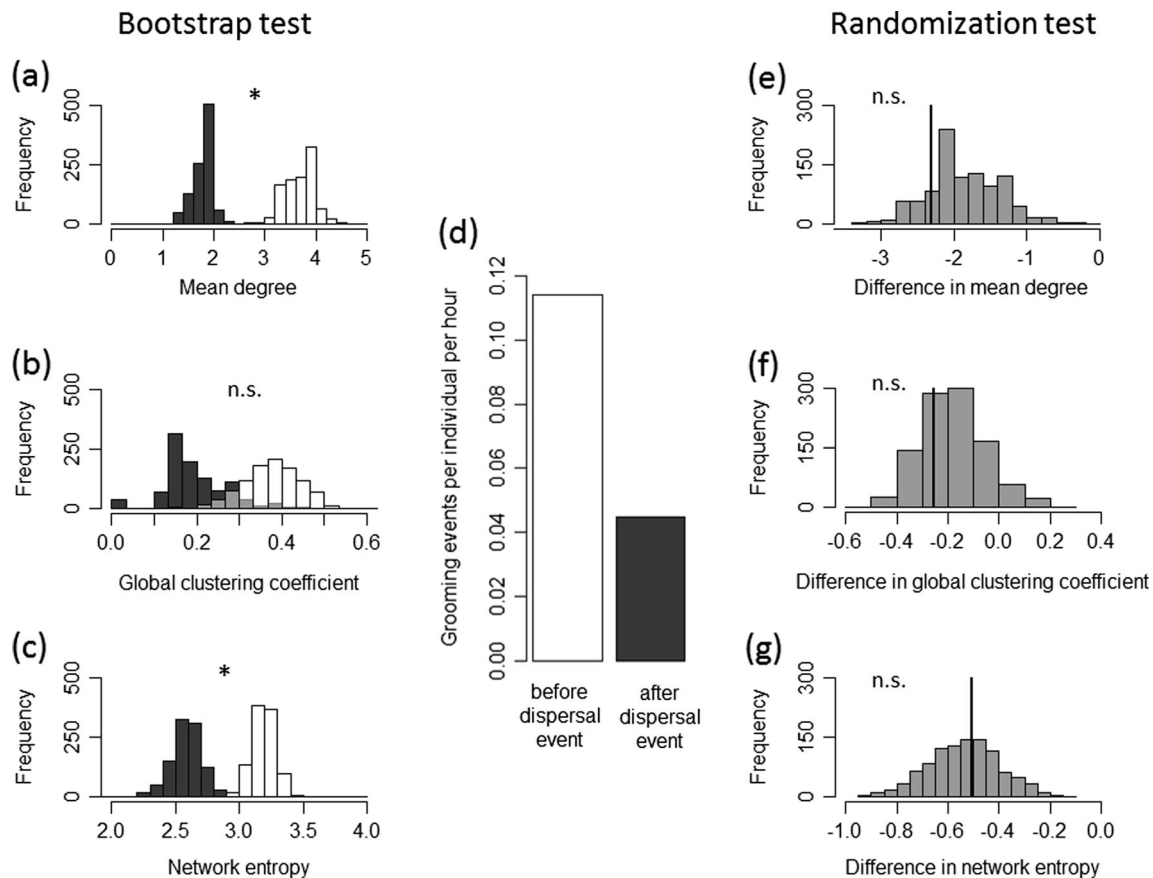


Fig. 4 The bootstrap test (**a**, **b**, **c**) reports significant changes in the grooming networks of adult female baboons after the dispersal of an alpha male in Amboseli, Kenya, while the randomization test (**d**, **e**, **f**) does not. *Black bars* and *white bars* (**a**, **b**, **c**) show the distributions of bootstrapped values for network measures before and after the dispersal event, respectively; *gray bars* indicate where the two distributions overlap. Changes in mean degree and network entropy were reported to be significant using the bootstrap test; the change in global clustering coefficient was not significant. Panel (**d**) illustrates that individual grooming rates decreased after the dispersal event. In contrast to the bootstrap test, the randomization test takes this change in grooming rates into account when testing for a change in a network measure.

application of the randomization test revealed no significant changes in the female network after male dispersal (Fig. 4e–g). The differences between the results of the two tests indicate that the significant changes detected by the bootstrap test were mainly driven by a homogeneous decrease in grooming rates. Based on this finding, we can furthermore conclude that heterogeneous changes in the structure of grooming preferences among adult female baboons did not occur, or had no or only a minor influence on the observed changes in network measures. In other words, it seems that following the dispersal of the alpha male females groomed each other less, without changing their preferences whom to groom. Note, however, that this finding was specific to the investigated dispersal event and does not represent a pattern that generally occurs across events in which high ranking males disperse or die (Franz et al. 2015).

Results from the randomization tests (**e**, **f**, **g**) illustrate, for each network measure, the distribution of expected changes under the null hypothesis of no heterogeneous changes. The *black vertical lines* show the observed changes in the corresponding network measure. In all three cases, the observed change falls well within the distribution of the expected change, which illustrates why the randomization test found no support for the hypothesis that these network measures were affected by heterogeneous changes in social relationships. This finding indicates that the significant changes detected by the bootstrap test (**a**, **c**) were actually the result of a homogeneous decrease in grooming rates (**d**), instead of being the result of changes in the structure of grooming preferences among adult female baboons

Discussion

Here, we emphasized the importance of distinguishing between heterogeneous and homogeneous changes in social relationships as distinct causes of structural changes in social networks (Fig. 1). We have confirmed that heterogeneous and homogeneous changes in relationships can affect observed social network structures in similar ways (Fig. 2). Further, we have shown this is not only true for the straightforward case of mean network degree, but it also applies to other binary and weighted network measures such as global clustering coefficient, network entropy, and weighted clustering coefficients (Fig. 2, [Electronic supplementary material](#)).

We might have expected that the use of weighted network measures would reduce or even completely prevent the systematic influence of homogeneous changes because weighted

network measures use more fine-grained information on interaction or association frequencies than unweighted network measures. However, information on edge weights cannot be recovered if no interactions or associations have been observed. For that reason, binary and weighted network measures are affected by the same fundamental problem: homogeneous changes can alter the probability that any observations (or observations above a certain threshold) are obtained for weak relationships.

This systematic effect of homogeneous changes has far reaching consequences for interpreting results of social network analyses. For instance, commonly applied bootstrap tests are able to infer whether changes in social relationships occurred, but are not able to infer the nature of this change, i.e., whether homogeneous or heterogeneous changes occurred (Fig. 3). This limits the inferences that can be drawn from studies that apply bootstrap tests or other tests that do not allow distinguishing between homogeneous and heterogeneous changes.

Our example analysis illustrates this problem in the context of the potentially disruptive effects of dispersal by an alpha male in a primate group (Fig. 4). The differences between the results of the two applied tests are mainly explained by a general decrease in grooming rates (Fig. 4d). Taken together, these results show that the changes in network measures detected by the bootstrap test are mainly a *side effect* of a homogeneous change in grooming rates; any additional heterogeneous changes in overall preferences of whom to groom must have been absent or relatively weak.

Similar issues to those illustrated by this example analysis might exist in other studies that analyzed temporal dynamics in social networks. For instance, Flack and colleagues (Flack et al. 2005, 2006) conducted a well-known and particularly innovative study of social relationships in which they tested how a specific conflict intervention behavior referred to as “policing” affected the stability of social behavior and social networks in a captive group of pigtailed macaques (*Macaca nemestrina*). They temporarily removed high ranking males, which were identified as the most important “policers,” and investigated how social network structures changed after the “knockouts” of these individuals. In Flack et al. (2005), the researchers reported that the “knockouts” of key “policers” led to an average increase in association rates and a decrease in average grooming rates (Flack et al. 2005). In a subsequent network analysis, Flack et al. (2006) also reported a decrease in mean degree in the grooming network after the knockouts and an increase in the global clustering coefficient in the association network (these effects refer to changes in social networks that only contain the same subset of non-knockout individuals before and after the knockout). These changes were interpreted as indicators of a less open, integrated society after the knockouts and Flack et al. (2006) concluded that “policing” behavior is important for maintaining stable primate

societies. However, Flack et al. (2006) did not investigate whether the observed changes in network structures were caused by homogeneous or heterogeneous changes in social relationships. Heterogeneous changes seem to be an obvious explanation. However, homogeneous changes are a plausible alternative in this case. Specifically, increases in average association rates and decreases in average grooming rates (reported in Flack et al. 2005) are consistent with the possibility that homogeneous changes caused the observed increase in association clustering coefficient and decrease in mean grooming degree (Fig. 2).

Heterogeneous and homogeneous changes would have fundamentally different implications in this case. For instance, if a heterogeneous change occurred (with or without a simultaneous homogeneous change), then a change in the overall interaction patterns would be an effect that occurred *in addition* to changes in average association and average grooming rates. In contrast, if only a homogeneous change occurred then the pattern of variation in relationships among dyads would have remained identical and the changes in observed social network structures would be a side effect of changes in association and grooming rates. As a consequence, if networks experience homogeneous as opposed to heterogeneous changes, the network seems likely to rebound more quickly from a perturbation. Similar issues can also arise in studies on the influence of seasonal changes on social networks (e.g., Henzi et al. 2009; Brent et al. 2013) or in studies of natural knockouts where knockouts effects could be conflated with seasonal effects (Barrett et al. 2012).

A potential solution to this problem could be to intensify observations to a point where even the weakest relationships have a very high detection probability. In this case, homogeneous changes would not easily be confused with heterogeneous ones. However, in many cases, this will not be feasible, particularly in captive studies that occur within short timeframes, or in wild studies that allow only limited observation effort. As a solution to this problem, we have shown that a randomization test can control for potential homogeneous changes (Figs. 3 and 4, [Electronic supplementary material](#)). However, while the randomization test allows the detection of heterogeneous changes, this test does not allow any inference of whether homogeneous changes also took place.

A potential approach that might allow the combined quantification of homogeneous and heterogeneous changes would be the use of random graph models (Robins et al. 2007) and actor-based models (Snijders et al. 2010), which explicitly model relationships dynamics. However, these approaches do not yet consider observation errors. Neither do they consider that relationships dynamics can be affected by different kinds of changes, which means that random graph models do not yet allow researchers to differentiate between homogeneous and heterogeneous changes.

In our analyses, we focused on three network measures that measure properties of the global network structure. Whether and to what extent a network measure is affected by homogeneous changes in any specific case is difficult to predict. Nonetheless, we recommend the use of the applied randomization test, instead or in addition to a bootstrap test, to control for potential homogeneous effects. As shown in our application using network entropy, our test can be applied to any network measure including weighted and directed measures. For reasons of practicality, we focused in our simulations on one extreme case of heterogeneous changes: the complete termination of randomly selected relationships. It is important to note that the randomization test we used is not restricted to detecting this special kind of heterogeneous changes. The test itself is not based on any assumption about specific heterogeneous changes. Instead, the test will indicate a significant change if any kind of heterogeneous change (including less extreme, more gradual changes) led to a pronounced enough change in the investigated network measure.

To apply the randomization test we used here, two important assumptions need to be fulfilled: (1) networks of the same set of individuals need to be compared, and if sets of individuals change over time then only subsets of consistently present individuals can be compared (e.g., see Flack et al. 2006) and (2) all individuals must be equally well sampled. If the latter condition is not fulfilled, e.g., in because association data is analyzed, the randomization procedure might be adapted according to the sampling protocol (e.g., Whitehead 2008). As noted above, a more general approach would be desirable that allows relaxing these assumptions and that can quantify the separate contributions of homogeneous and heterogeneous changes.

Acknowledgments We thank two anonymous reviewers, Damien Farine, Daniel van der Post, and Emily McLean, for helpful suggestions and discussion. We thank the Kenya Wildlife Services, Institute of Primate Research, National Museums of Kenya, National Council for Science and Technology, members of the Amboseli-Longido pastoralist communities, Tortillis Camp, Ker & Downey Safaris, Air Kenya, and Safarilink for their cooperation and assistance in Kenya. Thanks also to R.S. Mututua, S. Sayialel, J.K. Warutere, V. Somen, and T. Wango in Kenya, and to J. Altmann, K. Pinc, N. Learn, L. Maryott, and J. Gordon in the US. This research was approved by the IACUC at Princeton University and at Duke University and adhered to all the laws and guidelines of Kenya.

Compliance with Ethical Standards

Funding The National Science Foundation (most recently BCS 0323553, DEB 0846286, and IOS 0919200) and the National Institute on Aging (R01AG034513 and P01AG031719) for the majority of the data presented here. M. Franz was supported by the German Research Foundation (DFG) and by Duke University.

Conflict of interest The authors declare that they have no competing interests.

Ethical approval All applicable international, national, and/or institutional guidelines for the care and use of animals were followed. All procedures performed in studies involving animals were in accordance with the ethical standards of the institution or practice at which the studies were conducted.

References

- Altmann J (1974) Observational study of behavior: sampling methods. *Behaviour* 49:227–267
- Ansmann IC, Parra GJ, Chilvers BL, Lanyon JM (2012) Dolphins restructure social system after reduction of commercial fisheries. *Anim Behav* 84:575–581
- Barrett L, Henzi SP, Lusseau D (2012) Taking sociality seriously: the structure of multi-dimensional social networks as a source of information for individuals. *Philos T Roy Soc B* 367:2108–2118
- Bejder L, Fletcher D, Bräger S (1998) A method for testing association patterns of social animals. *Anim Behav* 56:719–725
- Blonder B, Wey TW, Dornhaus A, James R, Sih A (2012) Temporal dynamics and network analysis. *Methods Ecol Evol* 3:958–972
- Boogert NJ, Farine DR, Spencer KA (2014) Developmental stress predicts social network position. *Biol Lett* 10:20140561
- Borgatti SP, Everett MG, Freeman LC (2002) Ucinet for windows: software for social network analysis. <https://sites.google.com/site/ucinetsoftware/home>
- Brent L, MacLarnon A, Platt ML, Semple S (2013) Seasonal changes in the structure of rhesus macaque social networks. *Behav Ecol Sociobiol* 67:349–359
- Cantor M, Wedekin LL, Guimaraes PR, Daura-Jorge FG, Rossi-Santos MR, Simoes-Lopes PC (2012) Disentangling social networks from spatiotemporal dynamics: the temporal structure of a dolphin society. *Anim Behav* 84:641–651
- Croft DP, James R, Krause J (2008) Exploring animal social networks. Princeton University Press, Princeton
- Croft DP, Madden JR, Franks DW, James R (2011) Hypothesis testing in animal social networks. *Trends Ecol Evol* 26:502–507
- Csardi G, Nepusz T (2006) The igraph software package for complex network research. *Int J Complex Syst* 1695:1–9, <http://igraph.org>
- Farine DR, Whitehead H (2015) Constructing, conducting and interpreting animal social network analysis. *J Anim Ecol* 84:1144–1163
- Flack JC, Krakauer DC, de Waal FBM (2005) Robustness mechanisms in primate societies: a perturbation study. *Proc R Soc Lond B* 272:1091–1099
- Flack JC, Girvan M, de Waal FBM, Krakauer DC (2006) Policing stabilizes construction of social niches in primates. *Nature* 439:426–429
- Foster EA, Franks DW, Morrell LJ, Balcomb KC, Parsons KM, van Ginneken A, Croft DP (2012) Social network correlates of food availability in an endangered population of killer whales, *Orcinus orca*. *Anim Behav* 83:731–736
- Franz M, Altmann J, Alberts SC (2015) Knockouts of high-ranking males have limited impact on baboon social networks. *Curr Zool* 61:107–113
- Gero S, Gordon J, Whitehead H (2013) Calves as social hubs: dynamics of the social network within sperm whale units. *Proc R Soc B* 280:20131113
- Hansen H, McDonald DB, Groves P, Maier JAK, Ben-David M (2009) Social networks and the formation and maintenance of river otter groups. *Ethology* 115:384–396

- Henzi SP, Lusseau D, Weingrill T, van Schaik CP, Barrett L (2009) Cyclicity in the structure of female baboon social networks. *Behav Ecol Sociobiol* 63:1015–1021
- Hobson EA, Avery ML, Wright TF (2013) An analytical framework for quantifying and testing patterns of temporal dynamics in social networks. *Anim Behav* 85:83–96
- James R, Croft DP, Krause J (2009) Potential banana skins in animal social network analysis. *Behav Ecol Sociobiol* 63:989–997
- Krause J, Lusseau D, James R (2009) Animal social networks: an introduction. *Behav Ecol Sociobiol* 63:967–973
- Lusseau D, Whitehead H, Gero S (2008) Incorporating uncertainty into the study of animal social networks. *Anim Behav* 75:1809–1815
- Pinter-Wollman N, Hobson EA, Smith E et al (2014) The dynamics of animal social networks: analytical, conceptual, and theoretical advances. *Behav Ecol* 25:242–255
- R Core Team (2014) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, <https://www.r-project.org>
- Robins G, Pattison P, Kalish Y, Lusher D (2007) An introduction to exponential random graph (p*) models for social networks. *Soc Netw* 29:173–191
- Snijders TA, Van de Bunt GG, Steglich CE (2010) Introduction to stochastic actor-based models for network dynamics. *Soc Networks* 32:44–60
- Sueur C, Jacobs A, Amblard F, Petit O, King AJ (2011) How can social network analysis improve the study of primate behavior? *Am J Primatol* 73:703–719
- Voelkl B, Kasper C, Schwab C (2011) Network measures for dyadic interactions: stability and reliability. *Am J Primatol* 73:731–740
- Wey T, Blumstein DT, Shen W, Jordan F (2008) Social network analysis of animal behaviour: a promising tool for the study of sociality. *Anim Behav* 75:333–344
- Whitehead H (2008) Analyzing animal societies: quantitative methods for vertebrate social analysis. University of Chicago Press, Chicago
- Whitehead H, Bejder L, Andrea Ottensmeyer C (2005) Testing association patterns: issues arising and extensions. *Anim Behav* 69:e1
- Wilson AD, Krause S, James R, Croft DP, Ramnarine IW, Borner KK, Clement RJ, Krause J (2014) Dynamic social networks in guppies (*Poecilia reticulata*). *Behav Ecol Sociobiol* 68:915–925
- Wittmeyer G, Douglas-Hamilton I, Getz WM (2005) The socioecology of elephants: analysis of the processes creating multitiered social structures. *Anim Behav* 69:1357–1371