Behav Ecol Sociobiol (2012) 66:1437–1447
DOI 10.1007/s00265-012-1376-6

METHODS

Wildlife contact analysis: emerging methods, questions, and challenges

Paul C. Cross · Tyler G. Creech · Michael R. Ebinger · Dennis M. Heisey · Kathryn M. Irvine · Scott Creel

Received: 26 April 2012 / Revised: 12 June 2012 / Accepted: 13 June 2012 / Published online: 12 July 2012
© Springer-Verlag (outside the USA) 2012

Abstract Recent technological advances, such as proximity loggers, allow researchers to collect complete interaction histories, day and night, among sampled individuals over several months to years. Social network analyses are an obvious approach to analyzing interaction data because of their flexibility for fitting many different social structures as well as the ability to assess both direct contacts and indirect associations via intermediaries. For many network properties, however, it is not clear whether estimates based upon a sample of the network are reflective of the entire network. In wildlife applications, networks may be poorly sampled and boundary effects will be common. We present an alternative approach that utilizes a hierarchical modeling framework to assess the individual, dyadic, and environmental factors contributing to variation in the interaction rates and allows us to estimate the underlying process variation in each. In a disease control context, this approach will allow managers to focus efforts on those types of individuals and environments that contribute the most toward super-spreading events. We account for the sampling distribution of proximity loggers and the non-independence of contacts among groups by only using contact data within a group during days when the group membership of proximity loggers was known. This allows us to separate the two mechanisms responsible for a pair not contacting one another: they were not in the same group or they were in the same group but did not come within the specified contact distance. We illustrate our approach with an example dataset of female elk from northwestern Wyoming and conclude with a number of important future research directions.

Keywords Contact · Disease transmission · Hierarchical model · Proximity logger · Social network · Super-spreading

Communicated by M. Festa-Bianchet

Paul C. Cross and Tyler G. Creech contributed equally.

Electronic supplementary material The online version of this article (doi:10.1007/s00265-012-1376-6) contains supplementary material, which is available to authorized users.

P. C. Cross (✉) · K. M. Irvine
U.S. Geological Survey,
Northern Rocky Mountain Science Center,
2327 University Way, Suite 2,
Bozeman, MT 59715, USA
e-mail: pcross@usgs.gov

T. G. Creech
Department of Ecology, Montana State University,
310 Lewis Hall,
Bozeman, MT 59717, USA

M. R. Ebinger · S. Creel
Department of Ecology, Montana State University,
302 Lewis Hall,
Bozeman, MT 59717, USA

D. M. Heisey
U.S. Geological Survey, National Wildlife Health Center,
6006 Schroeder Road,
Madison, WI 53711, USA

Present Address:
T. G. Creech
Department of Fisheries and Wildlife, Oregon State University,
104 Nash Hall,
Corvallis, OR 97331, USA

Introduction

Quantifying interactions among individuals is central to several fields of ecology, such as animal behavior and infectious disease dynamics. Many early studies of animal contact rates directly observed individuals (e.g., Goodall 1963). The development of very high-frequency (VHF) radiotracking, global positioning system (GPS) devices, and mark–recapture techniques extended the ability of ecologists to study animal contacts in species where individuals

are hard to identify or are difficult to directly observe (Electronic supplementary material (ESM) 1). In cases where rare and brief encounters play an important role (e.g., disease transmission), these traditional approaches may be biased because either individuals cannot be observed simultaneously day and night or the spatiotemporal resolution of the data is too coarse. In addition, interaction data are problematic to analyze with traditional statistical approaches because they are non-independent (i.e., pairwise and often clustered within groups) and non-normally distributed (Kenny et al. 2006; Whitehead 2008; Croft et al. 2011). Recent technological advances, such as proximity loggers that record when individuals are within a specified distance, open many new opportunities for ecologists, epidemiologists, social and behavioral scientists, but statistical approaches that realize the full utility of these advances still need development.

Proximity loggers use ultra-high-frequency transceivers to continuously record when sampled individuals are within a user-specific distance (currently adjustable from 0.5 to 100 m). Disease ecologists and epidemiologists often refer to "contacts" as interactions among individuals where pathogen transmission may occur even without physical contact. We refer to interactions and contacts interchangeably, but note that contacts do not necessarily imply physical touch. At present, two vendors make proximity loggers: Sirtrack Ltd. (now owned by Lotek Wireless Inc.) and Vectronic Aerospace GmbH. Proximity loggers have been used to study intraspecific contact rates of brushtail possums (Ji et al. 2005), European wild rabbits (Marsh et al. 2010), Tasmanian devils (Hamede et al. 2009), raccoons (Prange et al. 2006, 2011), elk (Creech et al. 2012; Vander Wal et al. 2012a, b), and white-tailed deer (Walrath et al. 2011) and interspecific contact rates between European badgers and cattle (Böhm et al. 2010). One particularly promising application of proximity loggers is explaining the differences in contact rate among individuals and habitats. Variation in the interaction rates among categories of individuals based on characteristics such as age, sex, or social rank is a well-studied phenomenon (e.g., Pereira 1988; Creel et al. 1992; Bradley et al. 2004; Wolf et al. 2007). The amount of variation that is attributable to particular individuals within these categories, however, remains poorly understood and is rarely estimated despite being a common feature of human and animal populations (Bansal et al. 2007; Clay et al. 2009; Marsh et al. 2010). Here, we discriminate between process and sampling variation, where process variation is the predictable differences among individuals, pairs, or environments rather than unpredictable stochastic events that lead an individual to be highly connected at one point in time.

In the context of disease dynamics, Woolhouse et al. (1997) proposed a "20/80 rule" as a general feature of animal populations, whereby 20 % of individuals are responsible for 80 % of disease transmission in a population. Lloyd-Smith et al. (2005) related this pattern of transmission to super-spreading events, which are uncommon but important situations in which a small number of individuals have large effects on transmission. Super-spreading events are due to individual variation in infectiousness and susceptibility, and variation in contact rates that is driven by individual and environmental factors. Proximity logger data can be used to identify those factors driving variation in the interaction rate, and because the loggers repeatedly sample individuals and pairs, we can estimate the process variation.

We propose a hierarchical modeling framework for analyzing contact data that estimates the individual, dyadic, and environmental factors contributing to variation in contact rates and controls for the sampling distribution and group structure of many social species. Our approach differs from the more common social network analyses (SNA) that are applied to interaction data (e.g., Carrington et al. 2005; Wasserman and Faust 1994; Krause et al. 2007; Wey et al. 2008), so we first review some of the issues and challenges associated with applying SNA to wildlife datasets. In many systems where only a percentage of the population can be sampled, we believe that hierarchical models can answer a number of important ecological questions that would be problematic for SNA. We illustrate this approach using proximity logger data on 150 female elk from northwestern Wyoming as an example. Several important issues remain that require further development, so we conclude with a discussion of future statistical and ecological research directions.

Social network analyses

Recent studies have used SNA to explore a variety of topics in epidemiology and animal behavior (ESM 1; Krause et al. 2007). Social networks represent individuals as nodes and the connections between them as edges, and there are numerous metrics to describe the properties and topology of the network. Wey et al. (2008) describe three levels of organization for network metrics: individual-level metrics describing the properties of a focal node (e.g., node degree), intermediate-level metrics describing the subgroup structure within a network (e.g., clustering coefficient, cliquishness), and group-level metrics describing the properties of the entire network (e.g., density, diameter). It is also useful to distinguish between metrics that are influenced only by direct connections between nodes (e.g., node degree) and metrics that also account for indirect connections between nodes separated by more than one edge (e.g., average path length). SNA has, at least, two main benefits over traditional approaches. First, networks, and the various metrics describing those networks, account for both direct and indirect

connections among individuals. Second, the flexibility of networks can accommodate any social structure, whereas alternative frameworks often require researchers to make, sometimes arbitrary, decisions to fit their species into that framework. For example, what constitutes a group and the membership of that group may be unclear. Social network analyses, however, also involve a number of statistical and sampling challenges, which we outline below.

Sampling networks

Quantifying animal contact rates usually requires sampling a subset of individuals from the population of interest. Methods that require capturing and outfitting individuals with recording devices (proximity loggers, GPS, or VHF telemetry) typically limit researchers to sampling a small fraction of the total population of interest because of the costs associated with purchasing and deploying these devices. Direct observation methods often allow a much greater fraction of the population to be sampled over the course of a study, but may be limited to just those individuals that are uniquely identifiable, which may not always be a representative sample of the population. Low temporal resolution generally results in the omission of edges, while incomplete sampling of individuals results in the omission of nodes and the edges that would have been associated with them (Fig. 1). The implications of a proportion of the population sampled (hereafter referred to as "sampling intensity") are rarely discussed explicitly in animal contact studies, but may be critical when inferences about the full contact network are desired.

Effects of incomplete data on network properties have received considerable attention in the human social sciences literature (Marschall 2007). The most common approach has been to randomly remove nodes or edges from simulated networks (e.g., random, scale-free, small world) and observe the resulting changes in network metrics. Some metrics will be biased in a predictable direction by random sampling of a network. For instance, the mean node degree will be equal or lower in a randomly sampled network than a full network because a portion of each node's neighbors are omitted from the sampled network (Stumpf et al. 2005). Indirect metrics may be especially vulnerable to sampling effects because the omission of any one node or edge potentially affects many distant nodes. Failing to include even a single node, for instance, may dramatically increase the diameter of the observed network if the omitted node provided an important link between otherwise distantly connected nodes (Marschall 2007). Finally, some indirect metrics are not calculable for networks consisting of unconnected components (e.g., average path length).

Borgatti et al. (2006) found predictable declines in the accuracy of centrality measures due to random sampling of networks. Frantz et al. (2009) found large differences between five model networks in the robustness of centrality metrics to sampling and concluded that network topology has a greater effect on metric accuracy than other network properties such as size or density. Both of these studies simulated error rates (i.e., percentages of omitted nodes and edges) of up to 50 %. Field studies of wildlife populations will often fail to obtain sampling intensities as good as these studies' worst-case scenarios. Studies of how

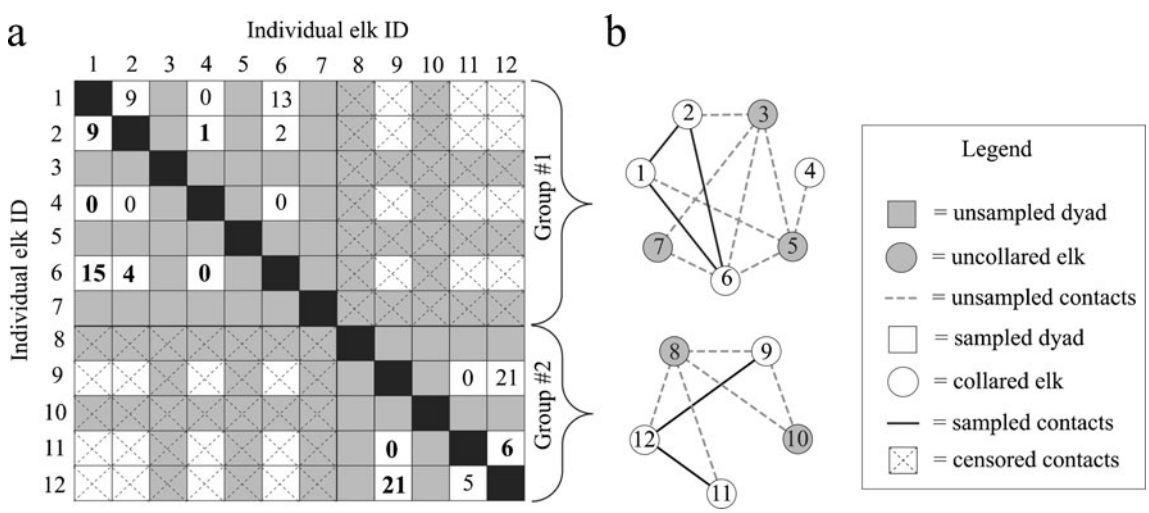

Fig. 1 Schematic of the who-contacts-whom matrix (**a**) of contact data for a given period of time used in the statistical modeling of contact rate within a group. The network representation (**b**) of the contact data illustrates the differences between the sampled and unsampled edges (*lines*) and nodes (*circles*). *Numbers within the circles* correspond to the rows and columns of the matrix. *Bold* numbers are the high counts of the number of contacts for a given pair used in the statistical analysis

incomplete data affect the estimation of network properties are relatively rare for empirically based networks (but see Costenbader and Valente 2008; Wey et al. 2008). The effects of incomplete data are even more poorly understood for wild animal social networks than for human social networks, for two reasons. First, sampling intensity is sometimes not known in wildlife studies because precise estimates of population size are difficult to obtain. Second, studies of incomplete data have typically removed data at random. Removing a percentage of interactions will often have a lesser effect on binary network metrics than the removal of nodes or individuals, which is more equivalent to sampling only a proportion of the population. Furthermore, highly influential nodes may be rare and unlikely to be sampled in the first place. Thus, even if the sampled network appears robust to subsampling nodes, we still have only limited confidence that the same is true for the entire network. Finally, Lee et al. (2006) found that subsampling a dataset has different effects on network metrics such as average path length and clustering coefficient when sampling occurs via random selection of nodes versus random selection of edges.

Static versus dynamic networks

Proximity loggers provide detailed temporal data over several months to years depending on battery life and available memory. However, a common method for analyzing social networks is to collapse data over relatively long time periods in order to capture connections that would go undetected over shorter time periods due to sampling limitations (e.g., Lusseau and Newman 2004). Static networks generated in this manner rely on the assumption that network structure is constant through time or that temporal variation would not affect inferences about the question of interest, but as Wey et al. (2008) note, "not all of the relationships represented may have existed at the same time, nor indeed may have all the individuals been together simultaneously."

While static networks may be appropriate for answering questions about long-term patterns of association (Lusseau et al. 2006), they can be problematic for answering questions about information transfer or disease transmission where the timing of contacts matters (Bansal et al. 2010). Static networks can be particularly problematic when contact data are collapsed over a time interval that is longer than the average duration of infectiousness. In such instances, the network structure will suggest a greater number of potential contacts between an infected node and its neighboring nodes than is actually possible during the infectious period (Cross et al. 2004). Several recent simulation studies have confirmed that when the true network structure is changing (i.e., dynamic), using a static network approach can misrepresent patterns of transmission and epidemic thresholds (Fefferman and Ng 2007; Volz and Meyers 2007, 2009). In systems where pathogens alter the contact behavior of infected hosts (Bouwman and Hawley 2010), static networks may be unable to identify pathogen-mediated shifts in network structure. Static networks can also be misleading in analyses of social behavior; for instance, data on agonistic interactions are sometimes aggregated into a matrix to produce a dominance hierarchy that includes some dyads that were never present at the same time, and many long-term studies include individuals that left the study early or entered the study late.

Statistical analysis of networks

The complex dependencies inherent in many contact and network datasets are not easily addressed by traditional statistical approaches. As a result, some ecological network analyses have been conducted using randomization tests (e.g., Mantel and partial Mantel tests; ESM 1; Whitehead 2008; Croft et al. 2011) that compare the properties of the observed network to a random null model of association between nodes. Often, Mantel tests are used to determine whether network structure is correlated with some other characteristic of dyads, such as their genetic relatedness or difference in age. We find many of these analyses unsatisfying because showing that individuals are not random is not as interesting as estimating the strength of the biological factors that drive the observed non-randomness. In addition, it is often unclear what the null random model should be (Cross et al. 2005; Whitehead 2008).

Other ecological network analyses involve first calculating a network metric and then statistically assessing the relationship between that metric and other data (e.g., degree centrality as a predictor of infection; ESM 1). This approach tends to ignore the estimation uncertainty associated with the network metrics as well as the bias associated with subsampling the network. Recently, exponential random graph models (ERGMs) have been developed to analyze network data (Snijders et al. 2006; Robins et al. 2007). Practitioners of this approach assume that the network data are one realization of a stochastic process and therefore estimate the probability of a contact (or edge) between individuals/nodes as a function of network parameters. However, because an edge is included in both the dependent and independent variables of the equation, an appropriate statistical estimation of ERGMs is more complicated than traditional generalized linear models. ERGMs have typically been used for static network analyses, but Snijders (2005) and others have begun to extend these approaches to dynamic networks. ERGM approaches are usually applied to networks with complete data, and many network estimates using ERGMS are highly biased by incomplete data (Huisman 2009). Consequently, we believe that the strength of proximity loggers lies outside of the

network paradigm. Here, we propose an alternative approach for analyzing contact data like those provided by proximity loggers that may be applicable to many field settings where the network is relatively weakly sampled.

Statistical analysis of contact rate

Our approach assesses the individual, dyadic, and environmental factors contributing to variation in contact rates among individuals while avoiding many of the problems associated with sampling networks by asking a different question than many network analyses—"What factors are associated with contact rate or the probability of contact between individuals *A* and *B*, given that they are located within the same group?" By focusing our analysis on within-group associations, we remove higher-order network dependencies that are not easily modeled with traditional statistical approaches. We focus here on characterizing the interaction rates within a group, but a full understanding of contact structure will require information on how individuals move among groups and how groups themselves interact. Defining what constitutes a group is sometimes not trivial, particularly when group membership changes frequently. We assume that group membership can be defined within some small time interval (e.g., hours to days).

Our approach utilizes generalized linear mixed models (GLMMs), which are increasingly applied to ecological datasets (Bolker et al. 2009). The so-called random effects in GLMM models are often used in the analysis of ecological data to account for the non-independence of multiple samples taken from the same individual ("repeated measures") or site ("subsampling"; Breslow and Clayton 1993; Gillies et al. 2006) and are often viewed as a statistical nuisance. In the analysis of interactions, however, individual and dyadic effects are of central interest, as is the variance among individuals and dyads. In cases where individuals, dyads, or periods are weakly sampled and data are unbalanced, the random-effects predictions are the best linear unbiased predictions with lower mean square errors than fixed effect estimates (Robinson 1991). Similar types of models have also been used in the psychology literature to analyze small groups and family dynamics (Kenny 1996; Kenny et al. 2002).

In many cases, it will be misleading to assess variation among individuals in contact rate by comparing the total numbers of contacts recorded by proximity loggers for each sampled individual because many populations are spatially structured such that some sampled individuals spend more time in the vicinity of other sampled individuals than others (and thus have greater opportunity for contacts to be recorded, regardless of true contact rate). To account for this, information on the spatial distribution of the sampled individuals is needed, particularly in systems where the group structure changes frequently. In our approach, zeros associated with no contacts between individuals of different groups are excluded. In addition, we insert zeros into the dataset whenever two marked individuals are known to be in the same group, but do not make contact (Fig. 1). These are important departures from an ERGM or SNA approach where these non-contacts are informative about the higher-order structure of who is in a group and how groups contact one another. Our focus, however, was on the contact rates within a group. Controlling for the distribution of sampled individuals could be done by direct observation, VHF, or GPS tracking. In our example, we use directly observed group membership information.

Proximity logger data include both the number of contacts and the duration of each contact. For some purposes like disease transmission, the total duration of contact during a time period may be more useful than either the number of contacts or the average duration of those contacts. The total duration of contact, however, is likely to be relatively complicated to statistically analyze because the distribution will be bimodal with peaks at zero for those dyads that did not contact one another and again at some average duration of contact, which will probably require zero-inflated (Lambert 1992; Hall 2000) or hurdle (Mullahy 1986; Gurmu 1998) modeling approaches. We focus on the number of contacts, but our approach is easily modified to instead investigate contact durations.

Let y_{lk} represent the number of contacts between dyad l for group observation k, where dyad l is the unique dyad for individuals i and j. Observation k may also be associated with group-level information about the location, time, habitat, and group size. Potential dyads that were never observed in the same group were excluded from the analysis. For each dyad, contacts are recorded twice, once on each logger in the pair. When loggers differ, we used the collars' data with the maximum number of contacts recorded for the pair (Fig. 1). We considered Poisson, overdispersed Poisson, and negative binomial data models, and in our example, the variance of the residuals from our best model had a roughly quadratic relationship with the mean (ESM 2), suggesting that a negative binomial formulation would be most appropriate (Ver Hoef and Boveng 2007). We used the Poisson–Gamma mixture model formulation of the negative binomial model such that

$$y_{lk} \sim \text{Poisson}(r_{lk}\lambda_{lk})$$

$$\lambda_{lk} = \exp(\beta_0 + \alpha_i + \alpha_j + \delta_l + \rho_k)$$

$$r_{lk} \sim \text{Gamma} \quad (\theta, \theta),$$

$$\alpha_i \sim \text{Normal}\left(0, \sigma_\alpha^2\right), \quad \alpha_j \sim \text{Normal}\left(0, \sigma_\alpha^2\right)$$

$$\delta_l \sim \text{Normal}\left(0, \sigma_\delta^2\right), \quad \rho_k \sim \text{Normal}\left(0, \sigma_\rho^2\right)$$

where β_0 is the global intercept, α_i and α_j are individual effects ("sociability"), δ_l are dyad effects (an interaction of individual i and j), and ρ_k are environmental effects (ESM 3). A gamma distribution with the same shape and scale parameter, θ, has a mean of 1 and thus only affects the variation in the predicted $\widehat{y_{lk}}$. Each observation period k represents a single observation of a group, and contacts are then summed for the 12 h before and after this observation. We refer to ρ_k as the environmental component of the variation in contact rate because it includes the predictable variation due to habitat, season, group composition, size, and density. In our example, multiple elk groups may be observed on a given day, resulting in multiple ρ estimates—one for each observed group. Elk group membership is relatively fluid; therefore, we do not often have multiple observations of exactly the same group over time. The width of time interval over which to sum contacts has important ramifications. Our choice of a 24-h period was primarily motivated by the frequent switching of individuals among groups, which would result in higher misclassification of group membership over longer time intervals.

This model allows for an individual effect (α_i) to capture the relative sociality of individuals as well as a dyadic interaction term (δ_l) that represents whether or not pairs of individuals interact more or less often than expected given the relative sociality of the individuals in the pair. One can build additional hierarchical levels into this model by incorporating variables that help predict the α_i, δ_l, or ρ_k effects. For example, to assess whether individuals of the same sex were more likely to make contact, we could assume that $\delta_l \sim \text{Normal}\left(\omega z_l, \sigma_\delta^2\right)$, where z_l is an indicator variable representing whether the pair was of the same sex or not. A model of particular importance to disease ecologists would assume $\rho_k \sim \text{Normal}\left(\gamma g_k, \sigma_\rho^2\right)$, where g is the standardized group size for observation k and γ indicates how the number of contacts between a pair of sampled individuals changes with group size. In this example, we are particularly interested in the estimates and relative magnitude of the variances σ_α^2, σ_δ^2, and σ_k^2 and the comparative fits of the models with or without individual, dyad, and environmental effects (Gelman 2005).

The above model shares a basic similarity with the diallel cross model of plant geneticists, whereby researchers are interested in the breeding value of two parents and each offspring is a data point (Kempthorne 1956). If contacts are asymmetric such that there is an obvious receiver and donor, then it is relatively straightforward to include these effects in a traditional regression (Whitehead 2008). When contacts are symmetric, without a biological interpretation of a receiver or donor effect, as they are for proximity logger data, then there is no obvious way to run the above model in the lme4 package of R (Bates et al. 2011) because there are two individual covariates (α_i and α_i) for each contact that are realizations of the same distribution of individuals effects. With effort, it can be analyzed with SAS PROC GLIMMIX® (SAS Institute Inc. 2008), but not very conveniently for even moderate-sized datasets. It is tempting to include each recorded contact twice in the dataset, once for each collar in which it was recorded and then include only one individual effect α_i for the recording collar or individual. However, this approach would bias the precision of the estimates of the other covariates due to pseudoreplicating each contact event (Hurlbert 1984). To circumvent these issues, we model these data using a Bayesian approach in WinBUGS (Lunn et al. 2000) where we can account for the two individuals involved in each contact (ESM 3). If one is willing to drop the individual effects and allow the dyadic effects to account for both the main effect of the individuals as well as their interaction, then simpler models could be run in most statistical packages. At present, however, a negative binomial mixed effect model is not supported within lme4 (Bates et al. 2011).

We used uninformative prior distributions on all parameters where possible. We assumed a diffuse normal prior for β_0 with a mean of 0 and a precision of 0.0001. We assigned the random effects α_{ij}, γ_i, and ρ_k normal prior distributions with a mean of 0 and a standard deviation with a hyperprior of *Uniform*(0, 3). We also ran several models with *Uniform*(0, 20) prior distributions for the standard deviations; the results were very similar. The prior distribution for θ was *Uniform*(0, 100). We also tested a prior distribution for exp(θ) as normal with a mean of 0 and a precision of 0.0001, and our posterior mean θ was nearly identical. We used the R2WinBUGS package to call WinBUGS version 1.4.3 (Lunn et al. 2000) from R version 2.13.2 (R Development Core Team 2011). All models were run for 20,000 iterations on four different Markov chains and the first half of each chain was discarded (Table 1). We assessed convergence using the Gelman–Rubin–Brooks statistic, where $\widehat{R}1.1$ for all parameters, which indicated that relatively little variation was associated with a specific MCMC chain (Gelman and Hill 2007).

We demonstrate our approach using a proximity logger dataset from elk (*Cervus canadensis*) in western Wyoming. This dataset will be analyzed more extensively elsewhere; here, we use it primarily to illustrate the general approach and thus will describe the dataset only briefly. We monitored roughly 60 elk per year from March 2009 to July 2011 across five different sites, placing approximately 30 collars in two regions each year. At each site, the proportion of individuals sampled is probably <5 % of the total number elk. We outfitted female elk with proximity loggers during

Table 1 Statistical model comparison and means of the posterior distributions of the individual, dyad, and environmental standard deviations using the elk proximity logger data. Numbers in parentheses are 95 % credible intervals

No.	Model	Deviance	DIC	pD[a]	σ_α[b]	σ_δ	σ_ρ
1	$\beta_0+\alpha_i+\alpha_j+\delta_l+\rho_k$	109,903	128,610	18,705	0.28 (0.24, 0.33)	0.28 (0.26, 0.31)	0.55 (0.48, 0.63)
2	$\beta_0+\alpha_i+\alpha_j+\rho_k$	109,736	128,615	18,879	0.28 (0.24, 0.32)	–	0.55 (0.48 0.64)
3	$\beta_0+\delta_l+\rho_k$	109,887	128,658	18,769		0.43 (0.41, 0.46)	0.57 (0.49, 0.66)
4	$\beta_0+\alpha_i+\alpha_j+\delta_l$	109,467	128,760	19,292	0.31 (0.26, 0.36)	0.27 (0.25, 0.29)	–

DIC deviance information criterion

[a] $pD=D_{\text{bar}}-D_{\text{hat}}$ (an approximation of the model complexity)

[b] Estimates of the standard deviation among individuals (α), dyads (δ), and environments (ρ)

captures (January through March), and they were programmed to drop off the elk in July. We calibrated each collar individually so that interactions were recorded at a distance of 3–4 m off the animal, which equated to roughly 2 m when the loggers were then tested on horses (Creech et al. 2012). The amount of separation time required between interactions before they were considered separate events was 90 s.

When we observed elk groups containing two or more proximity-collared individuals, we recorded the time, identity of collared individuals, and group size for each observation. In four of the regions, the elk were supplementally fed from December to March or April (Cross et al. 2007). In these regions, we used contact data from January to March when all the elk with loggers were known to be using the feed grounds. During the feeding season, all the sampled individuals were defined as being in the same group because most sampled individuals on the feed grounds contacted one another within a day. While not on the feed grounds, we delineated elk groups based upon relatively consistent internal spacing and individuals moving in roughly the same direction. The resulting dataset included 247 observations of groups (103 of those were while elk were being supplementally fed), which included 150 different individuals and 1,571 out of 11,175 possible dyads.

Results and discussion

As is typical for patterns of contact in most species, a small proportion of elk had very high contact rates. Thus, our contact data were highly right-skewed, but the negative binomial model provided a relatively good correspondence between the empirical data and the modeled predictions (Fig. 2). We fit models with and without different combinations of individual, dyad, and environmental effects; however, models without dyad, individual, or environmental effects tended to have higher DIC scores than the model that included all three (Table 1). This suggests that all effects were important enough to warrant the increased model complexity. The hyperparameter variance estimates, here shown as standard deviations, indicate that individual and

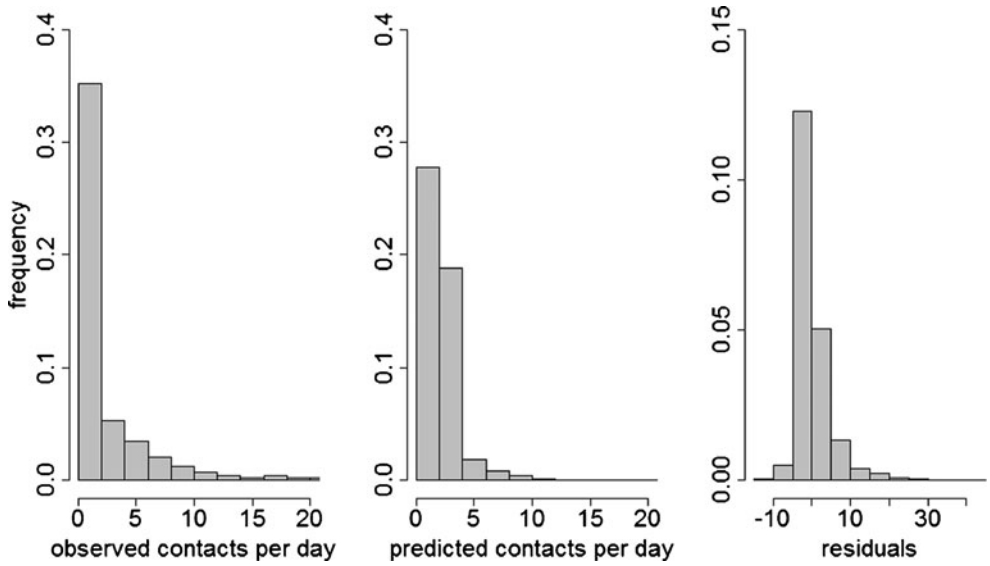

Fig. 2 The observed frequency distribution of the number of elk to elk contacts within a group in 24 h (**a**). The expected number of contacts (**b**) derived from the top-ranked model (lowest DIC score, Table 1) and the residuals (**c**)

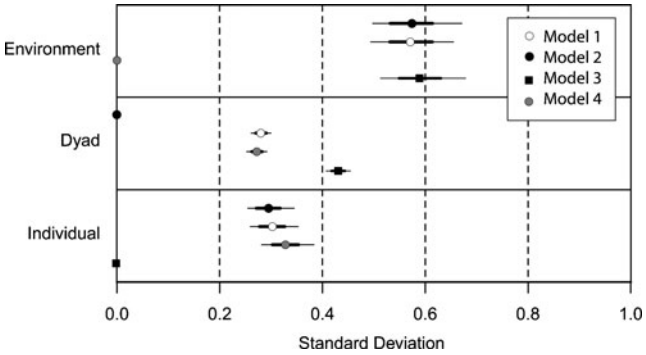

Fig. 3 Standard deviations of the individual, period, and environment random effects from the multilevel models 1–4 (Table 1). Point estimates (standard deviation scale) are the medians of the posterior distributions with 95 % (*wide*) and 68 % (*narrow*) intervals

dyad effects were roughly half as variable as the environmental effects (Fig. 3 and Table 1). Each pair of adult female elk, on average, made contacts with one another about twice per day if they were in the same group (exp(0.75)=2.1; Fig. 4). A pair in which one elk had an individual effect (α_i) 1 standard deviation higher than average would be expected to interact 2.8 times per day (exp(0.75+0.28)= 2.8; Table 1), while an observation period effect (ρ_k) 1 standard deviation higher than average would equate to 3.7 interactions per day for all those pairs present (exp(0.75+ 0.55)=3.7; Table 1). Therefore, in this study, super-spreading events are likely to be driven more by the environmental context than any particular individual or dyad. This may be a beneficial insight for managers because identifying and managing environments where many contacts occur may be logistically easier than identifying super-spreader individuals just prior to an epidemic. An important next step to these analyses will be to identify which covariates help predict the variability in the observation periods (e.g., habitat, season, group size). Low variation among dyads indicates that any effects of friends and enemies appear to be weak or, put another way, an individual's contacts do not appear to be highly concentrated among only a few other individuals.

Elk are a migratory species inhabiting different winter and summer ranges. Although we outfitted elk on winter range where groups are larger, we recorded contacts and observed groups as they migrated to their summer ranges, and the spatial distribution of loggers generally transitioned from few large groups each containing many collared individuals to many small groups each containing fewer collared individuals. As a result, the contact rates may appear to decline even though the contact rate per pair within a group is constant. For example, the total number of contacts at a site per day divided by the number of possible dyads at the site showed strong temporal trends (Fig. 4a). However, it is unclear whether this trend is due to changes in how dyads contact one another within a group or the spatial distribution of loggers. In our analyses, we control for the distribution of collars by limiting the data to include only those days and groups where group membership is known. We then inserted zeros for pairs without contacts that were known to be present in the group and exclude those pairs that were not present in the same group. The resulting temporal trend in within-group contact rate disappears, although the contact rate appears to become more variable over time (Fig. 4b and ESM 4). This indicates that the temporal trend in Fig. 4a is largely driven by the spatial distribution of loggers, whereby marked individuals are splitting up and not all loggers are present within the same group.

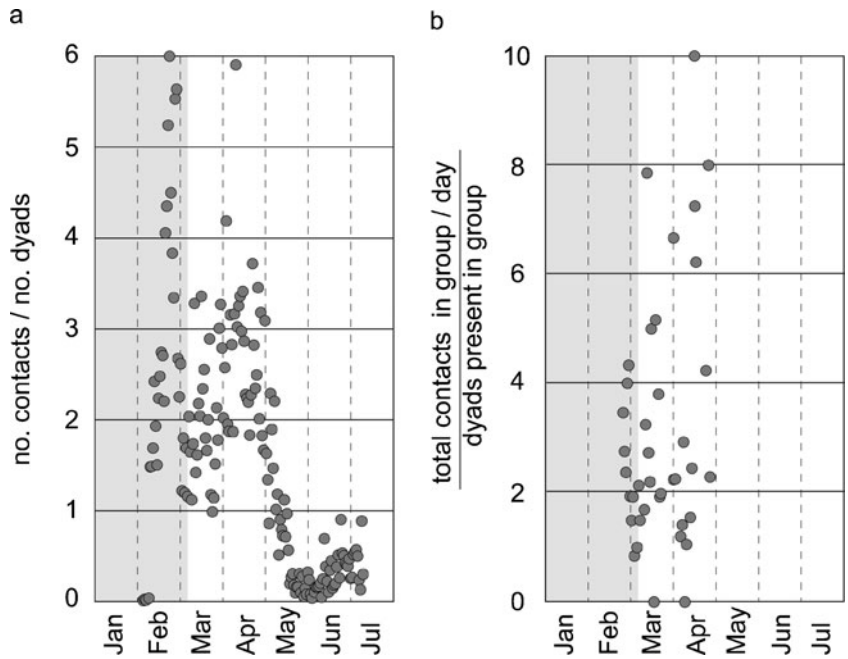

Fig. 4 Total number of contacts among all possible pairs of individuals at a site per day divided by the number of possible pairs at that site (**a**) and the total number of contacts within a group divided by the number of marked pairs in that group (**b**). The data in (**b**) is a subset of (**a**) because it includes only those days and groups for which group membership is known. *Shading* indicates the time period when elk were supplementally fed at this site (Muddy Creek 2010)

There are several important caveats associated with proximity logger data, and these transfer to our analysis as well. First, variability exists among the loggers in their ability to send and receive contact signals. This has important ramifications for the interpretation of the individual and dyadic effects, whereby a portion of that variability is due to logger performance differences as opposed to biological variation. Future analyses should work to use data collected on logger performance before and after the study to predict the variation in individual and dyad effects so that the residual variation in these estimates reflects biological variation among individuals and pairs of individuals. A second challenge to applying this approach is that proximity logger data are likely to be unbalanced (i.e., not all dyads are observed in all periods), and in many datasets, there will be relatively few dyads present in a group and some dyads will only rarely be observed together. In our case, although the median number of observations per dyad was 25, we had 84 dyads observed only one time. The less data available for a particular dyad, the more strongly that dyad's estimated effect is pulled toward the overall mean (Gelman and Hill 2007). As a result, one should be cautious of making inferences about particular dyads, individuals, or periods, especially when they are poorly sampled. However, there may be biological mechanisms driving the low number of observations for some dyads or environments; therefore, restricting analyses to just those cases that are well sampled may also induce bias. In simulated datasets, we found that the estimates of the population-level variability among individuals, dyads, and environments (σ_α^2, σ_δ^2, and σ_k^2) were relatively good as long as a large number of individuals, dyads, and periods were available (data not shown).

A final caveat is that our modeling approach assumes dyads are independent of one another within a group, although group observations may differ in their average contact rate. In other words, if individuals A and B contact often, and so do B and C, we assume that this does not imply a high contact rate between A and C. This assumption is unlikely to hold in cases where the spatial distribution of individuals is an important determinant of interaction within a group. For example, seals on a beach may be more spatially constrained to interact with their neighbors in the group compared to elk that can move relatively quickly from one side of the group to another. Exponential random graph models would be one approach that could account for these higher-order interactions, but then other issues arise about the application of ERGMs to sampled and dynamic networks. For adult female elk, we believe that much of the correlation among dyads is primarily attributable to group membership and that there is minimal hierarchical structure within a group. This should be more rigorously tested and may not be true in other cases.

Technological advances such as proximity loggers allow researchers to collect animal contact data with much greater resolution and efficiency than in the past, providing new opportunities but also ushering in new theoretical and statistical challenges. We have provided a potential method for analyzing interaction data in a multiple random-effects modeling framework that avoids many of the difficulties associated with networks, particularly sparsely sampled networks that are common in studies of animal contact rates. Despite several outstanding statistical issues, we hope that our approach is a useful stepping stone for future advances that will allow researchers to understand the factors affecting variation in contact rate, which is likely to create many new insights in multiple fields.

Acknowledgments This work was supported by the Wyoming Wildlife–Livestock Disease Partnership, National Science Foundation and National Institutes of Health Ecology of Infectious Disease Grant DEB-1067129, Wyoming Game and Fish Department, and the U.S. Geological Survey. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government. We thank B. Scurlock, J. Rogerson, E. Maichak, J. Henningsen, D. Damm, A. Williams, A. Barbknecht, and A. Roosa for their assistance in the field. D. Stinson assisted with aerial flights. Animals were under the Montana State University animal use and care protocol (#2010-02)

Ethical standards This research complies with the current laws of the country in which they were performed.

Conflict of interest The authors declare that they have no conflict of interest.

References

Bansal S, Grenfell BT, Meyers LA (2007) When individual behavior matters: homogeneous and network models in epidemiology. J R Soc Interface 4:879–891

Bansal S, Read J, Pourbohloul B, Meyers LA (2010) The dynamic nature of contact networks in infectious disease epidemiology. J Biol Dynam 4:478–489

Bates D, Maechler M, Bolker B (2011) lme4: linear mixed-effects models using S4 classes. R package version 0.999375-42 edition

Böhm M, Hutchings MR, White PCL (2010) 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

Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH, White JS (2009) Generalized linear mixed models: a practical guide for ecology and evolution. Trends Ecol Evol 24:127–135

Borgatti SP, Carley KM, Krackhardt D (2006) On the robustness of centrality measures under conditions of imperfect data. Soc Networks 28:124–136

Bouwman KM, Hawley DM (2010) Sickness behavior as an evolutionary trap? Male house finches preferentially feed near diseased conspecifics. Biol Lett 6:462–465

Bradley BJ, Doran-Sheehy DM, Lukas D, Boesch C, Vigilant L (2004) Dispersed male networks in western gorillas. Curr Biol 14:510–513

Breslow NE, Clayton DG (1993) Approximate inference in generalized linear mixed models. J Am Stat Assoc 88:9–25
Carrington PJ, Scott J, Wasserman S (eds) (2005) Models and methods in social network analysis. Cambridge University Press, Cambridge
Clay CA, Lehmer EM, Previtali A, St Jeor S, Dearing MD (2009) Contact heterogeneity in deer mice: implications for Sin Nombre virus transmission. Proc R Soc Lond B 276:1305–1312
Costenbader E, Valente TW (2008) The stability of centrality measures when networks are sampled. Soc Networks 25:283–307
Creech T, Cross PC, Scurlock BM, Maichak EJ, Rogerson JD, Henningsen JC, Creel S (2012) Effects of low-density feeding on elk-fetus contact rates on Wyoming feedgrounds. J Wildl Manage 76:877–886
Creel S, Creel N, Wildt DE, Monfort S (1992) Behavioural and endocrine mechanisms of reproductive suppression in Serengeti dwarf mongooses. Anim Behav 43:231–245
Croft DP, Madden JR, Franks DW, James R (2011) Hypothesis testing in animal social networks. Trends Ecol Evol 26:502–507
Cross PC, Lloyd-Smith JO, Bowers J, Hay C, Hofmeyr M, Getz WM (2004) Integrating association data and disease dynamics in a social ungulate: bovine tuberculosis in African buffalo in the Kruger National Park. Ann Zool Fenn 41:879–892
Cross PC, Lloyd-Smith JO, Getz WM (2005) Disentangling association patterns in fission–fusion societies using African buffalo as an example. Anim Behav 69:499–506
Cross PC, Edwards WH, Scurlock BM, Maichak EJ, Rogerson JD (2007) Effects of management and climate on elk brucellosis in the Greater Yellowstone Ecosystem. Ecol Appl 17:957–964
Fefferman NH, Ng KL (2007) How disease models in static networks can fail to approximate disease in dynamic networks. Phys Rev E 76:1–11
Frantz TL, Cataldo M, Carley KM (2009) Robustness of centrality measures under uncertainty: examining the role of network topology. Comp Math Organ Theory 15:303–328
Gelman A (2005) Analysis of variance—why it is more important than ever. Ann Stat 33:1–31
Gelman A, Hill J (2007) Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, Cambridge
Gillies CS, Hebblewhite M, Nielsen SE, Krawchuk MA, Aldridge CL, Frair JL, Saher DJ, Stevens CE, Jerde CL (2006) Application of random effects to the study of resource selection by animals. J Anim Ecol 75:887–898
Goodall J (1963) My life among wild chimpanzees. Natl Geogr 124:273–308
Gurmu S (1998) Generalized hurdle count data regression models. Econ Lett 58:263–268
Hall DB (2000) Zero-inflated poisson binomial regression with random effects: a case study. Biometrics 56:1030–1039
Hamede RK, Bashford J, McCallum H, Jones M (2009) Contact networks in a wild Tasmanian devil (*Sarcophilus harrisii*) population: using social network analysis to reveal season variability in social behaviour and its implications for transmission of devil facial tumour disease. Ecol Lett 12:1147–1157
Huisman M (2009) Imputation of missing network data: some simple procedures. J Soc Struct 10:1–29
Hurlbert SH (1984) Pseudoreplication and the design of ecological field experiments. Ecol Monogr 52:187–211
Ji W, White PCL, Clout MN (2005) Contact rates between possums revealed by proximity data loggers. J Appl Ecol 42:595–604
Kempthorne O (1956) The theory of the diallel cross. Genetics 41:451–459
Kenny DA (1996) The design and analysis of social-interaction research. Annu Rev Psychol 47:59–86
Kenny DA, Mannetti L, Pierro A, Livi S, Kashy DA (2002) The statistical analysis of data from small groups. J Pers Soc Psychol 83:126–137
Kenny DA, Kashy DA, Cook WL (2006) Dyadic data analysis. Guilford, New York
Krause J, Croft DP, James R (2007) Social network theory in the behavioural sciences: potential applications. Behav Ecol Sociobiol 62:15–27
Lambert D (1992) Zero-inflated Poisson regression with an application to defects in manufacturing. Technometrics 34:1–14
Lee SH, Kim P, Jeong H (2006) Statistical properties of sampled networks. Phys Rev E 73:016102
Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM (2005) Superspreading and the effect of individual variation on disease emergence. Nature 438:355–359
Lunn DJ, Thomas A, Best N, Spiegelhalter DJ (2000) WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. Stat Comput 10:325–337
Lusseau D, Newman MEJ (2004) Identifying the role that individual animals play in their social network. Proc R Soc Lond B 271: S477–S481
Lusseau D, Wilson B, Hammond PS, Grellier K, Durban JW, Parsons KM, Barton TR, Thompson PM (2006) Quantifying the influence of sociality on population structure in bottlenose dolphins. J Anim Ecol 75:14–24
Marschall N (2007) Methodological pitfalls in social network analysis: why current methods produce questionable results. VDM, Saarbrücken
Marsh MK, Hutchings MR, McLeod SR, White PCL (2010) Spatial and temporal heterogeneities in the contact behaviour of rabbits. Behav Ecol Sociobiol 65:183–195
Mullahy J (1986) Specification and testing of some modified count data models. Econometrics 3:341–365
Pereira ME (1988) Agonistic interactions of juvenile savanna baboons. Ethology 79:195–217
Prange S, Jordan T, Hunter C, Gehrt SD (2006) New radiocollars for the detection of proximity among individuals. Wildl Soc Bull 34:1333–1344
Prange S, Gehrt SD, Hauver S (2011) Frequency and duration of contacts between free-ranging raccoons: uncovering a hidden social system. J Mammal 92:1331–1342
R Development Core Team (2011) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
Robins G, Pattison P, Kalish Y, Lusher D (2007) An introduction to exponential random graph (p*) models for social networks. Soc Networks 29:173–191
Robinson GK (1991) That BLUP is a good thing: the estimation of random effects. Stat Sci 6:15–32
SAS Institute Inc. (2008) SAS/STAT® 9.2 user's guide. SAS Institute, Cary, NC
Snijders TAB (2005) Models for longitudinal network data. In: Carrington PJ, Scott J, Wasserman S (eds) Models and methods in social network analysis. Cambridge University Press, Cambridge, pp 215–247
Snijders TAB, Pattison PE, Robins GL, Handcock MS (2006) New specifications for exponential random graph models. Sociol Methodol 36:99–153
Stumpf MPH, Wiuf C, May RM (2005) Subnets of scale-free networks are not scale-free: sampling properties of networks. P Natl Acad Sci USA 102:4221–4224
Vander Wal E, Paquet PC, Andres JA (2012a) Influence of landscape and social interactions on transmission of disease in a social cervid. Mol Ecol 21:1271–1282
Vander Wal E, Yip H, McLoughlin PD (2012b) Sex-based differences in density-dependent sociality: an experiment with a gregarious ungulate. Ecology 93:206–212
Ver Hoef JM, Boveng P (2007) Quasi-Poisson vs. negative binomial regression: how should we model overdispersed count data? Ecology 88:2766–2772
Volz E, Meyers LA (2007) Susceptible-infected-recovered epidemics in dynamic contact networks. Proc R Soc Lond B 274:2925–2933

Volz E, Meyers LA (2009) Epidemic thresholds in dynamic contact networks. J R Soc Interface 6:233–241

Walrath R, Van Deelen TR, VerCauteren KC (2011) Efficacy of proximity loggers for detection of contacts between maternal pairs of white-tailed deer. Wildl Soc Bull 35:452–460

Wasserman S, Faust K (1994) Social network analysis: methods and applications. Cambridge University Press, Cambridge

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. The University of Chicago Press, Chicago

Wolf JBW, Mawdsley D, Trillmich F, James R (2007) Social structure in a colonial mammal: unravelling hidden structural layers and their foundations by network analysis. Anim Behav 74:1293–1302

Woolhouse MEJ, Dye C, Etard JF, Smith T, Charlwood JD, Garnett GP, Hagan P, Hii JLK, Ndhlovu PD, Quinnell RJ, Watts CH, Chandiwana SK, Anderson RM (1997) Heterogeneities in the transmission of infectious agents: implications for the design of control programs. P Natl Acad Sci USA 94:338–342