Published in final edited form as:

Soc Networks. 2012 October 1; 34(4): 691–700. doi:10.1016/j.socnet.2012.08.005.

The analysis of multiple ties in longitudinal egocentric network data: A case study on bidirectional relationships between trust and drug use

W. Scott Comulada, Dr.P.H.^{a,*}, Stephen Q. Muth, B.A.^b, and Carl A. Latkin, Ph.D.^c

^aDepartment of Psychiatry and Biobehavioral Sciences, University of California at Los Angeles, United States; scomulad@ucla.edu

^bQuintus-ential Solutions, Colorado, United States; quintus@earthlink.net

^cDepartment of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health Johns Hopkins University, United States; clatkin@jhsph.edu

Abstract

We extend multi-level models to examine single egocentric network ties to the joint analysis of paired dynamic ties. Two analytic challenges are addressed. First, inference needs to account for multiple layers of nesting: ties are nested within pairs, pairs are nested within time points, and time points are nested within egos. Second, the focus is on the relationship between two dynamic ties; specification of outcome and predictor may be difficult. Instead, we treat both ties as outcomes. Our approach is used to analyze trust and reported drug use between egos and alters over time in a Bayesian framework.

Keywords

Egocentric data; dyadic analysis; longitudinal bivariate regression; Bayesian inference; dynamic networks

1. Introduction

The decision of an individual, i.e. ego, to engage in drug use with a member of their personal network, i.e. alter, hinges on the interpersonal relationship with that alter. Among injecting drug users, men are more likely to share needles with partners who give them emotional support, women are more likely to share with partners who inject with them (Unger et al., 2006), and both men and women are more likely to share with members they trust (Cepeda et al., 2010). To date, these type of ego-alter relationships have largely been explored through cross-sectional data that lacks the information to establish temporal precedence and a more causal framework. To better understand network processes related to drug use, it is of interest to determine whether individuals engage in drug use with network members they trust or develop trust after a period of joint use, for example. In this article,

© 2012 Elsevier B.V. All rights reserved.

*Corresponding author: W. Scott Comulada, Semel Institute Center for Community Health, 10920 Wilshire Blvd Suite 350, Los Angeles, California 90024, U.S.A.; Telephone: +1(310)794-0938; Fax: +1(310)794-8297; scomulad@ucla.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

this question is addressed by extending the work of Lubbers et al. (2010) to examine the relationship between two social network measures in longitudinal egocentric data. They used a regression model that treats one measure as an outcome, i.e. a univariate outcome, and the other measure as a predictor. Instead we apply an analytic approach that allows both measures (e.g., ego-alter ties for trust and joint drug use) to be treated as outcomes, i.e. bivariate outcomes.

Univariate longitudinal models have been a recent and key development in the exploration of egocentric data that is typically comprised of small and disconnected networks. Traditional models for longitudinal social network data, including stochastic actor-based models (Snijders, 1996, 2001, 2008) and exponential random graph models (Morris et al., 2008; Robins et al., 2007; Snijders, 2001; Wang et al., 2004; Wasserman et al., 1996), have been applied to sociometric data that is comprised of large and connected networks, such as social networks in school classrooms. A motivation for the analysis of longitudinal egocentric data is the ability to inform temporal precedence by linking relational outcomes to predictors across different, i.e. lagged, time points versus analyses on cross-sectional egocentric data that are conducted at a single time point. For example, the longitudinal analyses of Lubbers et al. (2010) explored how alter degree centrality at the first time point predicted the frequency of contact between egos and alters two years later at a second time point. A benefit of their approach is the ability to model the natural hierarchy of the data. For example alters at the first level are nested within egos at the second level. The longitudinal model of Lubbers et al. (2010) is an extension of multi-level models (MLM) that have been applied to cross-sectional egocentric data (de Miguel and Tranmer, 2010; Hall, 2010; Unger et al., 2006; Valente 2010). In cross-sectional analyses, variation that is induced by correlations between nested alter observations is modeled through underlying ego-level constructs referred to as random effects (RE). Longitudinal analyses may introduce additional RE to account for correlations between repeated alter or ego-level observations over time.

The univariate longitudinal framework often encounters a chicken and egg problem. The a priori specification of predictor and outcome implies that temporal precedence between measures of interest is understood. For example, the models of Lubbers et al. (2010) related predictors at the first time point to outcomes at the second time. However, a longitudinal analysis often explores longitudinal relationships with the goal of drawing a conclusion about temporal precedence, as is the case in the current paper. We explore the dynamic relationship between ego-alter trust and joint drug use and which type of social tie precedes the other. Another example is provided by Bohnert et al. (2009). They examined whether ego or network-level drug use precedes the other in longitudinal egocentric data collected in drug-using neighborhoods in Baltimore, Maryland (SHIELD study). Directional hypotheses as required by a predictor-outcome specification would have been difficult as competing theories of social selection, such as homophily (Lazarsfeld and Merton, 1954), and network influence posit different mechanisms. Instead Bohnert et al. (2009) used a bivariate-outcome approach to simultaneously model bidirectional correlations between outcomes for ego drug use and a composite measure for network-level drug use at concurrent and lagged time points. A univariate approach would have required two models, a model with ego drug use as an outcome and network-level drug use as a predictor, and a vice versa. In addition to greater efficiency offered by the the bivariate versus univariate-outcome approach, the required specification of outcome and predictor of a univariate model is not trivial; results may differ if the roles of outcome and predictor are reversed (Weiss, 2005). Therefore, the examination of a single univariate-model may lead to the wrong conclusions.

In this article, we build upon the work of Bohnert et al. (2009) in two ways and hereafter refer to our modeling approach as a *bivariate MLM*. First, the research question posited by

Bohnert et al. (2009) targeted ego-level outcomes, an ego characteristic and a composite measure that aggregated alter-level information to the ego level. In our analyses, both outcomes represent ego-alter ties. Even when interest is at the ego level, an analysis that captures variation in the individual ego-alter ties is preferred over an analysis that aggregates individual ties to the ego level and does not capture tie variability. The magnitude of the estimated relationship between aggregate measures, often operationalized as a correlation coefficient in bivariate models, will generally be smaller relative to correlations between RE that are modeled through MLM and represent the original measures (Stone et al., 2007, pg. 100). Therefore, we examine the dynamic relationship between feels of trust and joint drug that was reported by egos for each alter through a bivariate MLM model that accounts for two layers of nesting. Similar to Bohnert et al. (2009), we account for nesting of repeated ego assessments over time. In addition, we also account for the nesting of ego-alter ties that are nested within egos.

A second way we expand upon the bivariate framework of Bohnert et al. (2009) is to implement Bayesian versus maximum likelihood inference in fitting bivariate MLM. In doing so, we overcome two related challenges that have hampered implementation of bivariate MLM: a lack of available computational algorithms and statistical software packages. Bayesian inference has been found to lead to fewer convergence problems for a variety of complex models compared to likelihood inference. For example, Caimo and Friel (2011) implemented exponential random graph models through Markov Chain Monte Carlo simulation (MCMC; Gamerman, 1997; Robert and Casella, 1999); Bayesian standard errors fell into a reasonable range, while likelihood standard errors were unusually large and indicated convergence problems. We illustrate bivariate MLM on two longitudinal egocentric data sets through WinBUGS, a free and downloadable statistical software package that implements Bayesian models through MCMC (Spiegelhalter, Thomas, Best, 2000). Through a logistic specification, bivariate MLM is applied to paired binary outcomes that represent ego-alter network ties for trust and joint drug use. While aiding computation, it should be noted that Bayesian inference relies on MCMC estimation that does not yield definitive convergence criteria, in contrast to typical likelihood algorithms. Instead MCMC estimation relies on diagnostic tests to achieve a reasonable level of certainty that convergence has occurred. We address this issue by illustrating diagnostic tests that are applied to our analyses.

There is another analytic issue that pertains to the general application of MLM to social network analysis. MLM gives more weight to measurements from larger networks through RE that are estimated for each network. The assumption is that measurements represent the population, regardless of network size. If smaller networks are less representative then MLM reduce the impact of reporting bias. Perhaps egos who report smaller networks want to terminate the interview early and give superficial responses. What is less clear is the impact if reporting bias is more likely in larger networks. Egos reporting larger networks may become fatigued and prone to give superficial tie responses or simply be less able to accurately report on all of their network members. While there has been general discussion on reporting bias in the social network literature (e.g., see Wasserman and Faust, 1994), an examination of the impact of reporting bias on statistical inference has been lacking. We begin to address this literature gap through simulation studies. Reporting errors are introduced into larger networks in our analysis data and bivariate MLM are rerun. Estimated associations between ties are compared between analyses on the original and simulated data.

Notation for our proposed bivariate MLM is given in Section 2. The Bayesian framework for bivariate MLM and diagnostic tests are presented in Section 3. Motivating examples for the bivariate MLM are presented in Section 4. Bivariate MLM are applied to paired ego-alter ties across two longitudinal egocentric data sets in Section 5.1. We conduct simulation

studies to explore the impact of reporting errors in larger networks in Section 5.2. Discussion follows in Section 6.

2. Bivariate multi-level models (MLM)

Bivariate MLM are applied to paired ego-alter network ties that are treated as outcome pairs and assessed within egocentric networks over time. The basic idea is to fit separate random effect (RE) regressions to each outcome. The focus of the analysis is on the correlation between RE from different outcomes that is estimated through the RE covariance structure. Bivariate MLM were originally proposed for continuous paired measures (Reinsel, 1982, 1984) and easily extended to binary ties; the correlation represents the association between the log-odds of each tie at the ego level.

Notation for the bivariate regression with correlated random effects follows. Let (Y_{1ijt}, Y_{2ijt}) be a set of bivariate responses for alter j within egocentric network i at time point $t = 1, 2, \dots, T$. Let $\beta_k = (\beta_{k1}, \dots, \beta_{kp})'$ be a vector of regression coefficients for response $k (= 1, 2)$ with covariate vector x_{kijt} . Let λ_i be a vector of RE for ego i with covariate vector z_{kijt} . RE are assumed to follow a multivariate normal distribution with a mean vector of zeros and covariance matrix Σ , $\lambda_i \sim N(0, \Sigma)$. The bivariate MLM is

$$\eta_{kijt} = x_{kijt}'\beta_k + z_{kijt}'\lambda_i, \quad (1)$$

where η_{kijt} is a link function for response Y_{kijt} . Our analyses are conducted on binary tie responses through bivariate logistic MLM where $\eta_{kijt} = \rho_{kijt} / (1 - \rho_{kijt})$. Conditional on RE, ρ_{kijt} is the probability that binary indicator Y_{kijt} equals one for the presence of the relation, $\rho_{kijt} = \Pr(Y_{kijt} = 1 \mid \text{RE})$. We note that it may also be warranted to include RE for each alter. Our models do not include alter-level RE, because most responses were only observed on the same alter at a single time point. The parameters of interest are the correlations between RE that are formulated from the off-diagonal covariance parameters, Σ_{ab} , $a \neq b$, and diagonal variance parameters, Σ_{aa} . As an example, correlation r_{12} between RE for the first tie at Wave 1 and the second tie at Wave 2 is expressed as $r_{12} = \Sigma_{12} / \text{square root}(\Sigma_{11} \Sigma_{22})$.

We consider two popular covariance structures for Σ to model the outcome trajectories over time. While trajectories are not the focus of our analysis, proper specification reduces error in modeling the outcomes. In turn, reduced error improves estimation of the correlation between RE. Repeated-measures ANOVA, hereafter simply referred to as *ANOVA*, models separate outcome means at each time point through β_k . ANOVA typically accommodates deviations in outcome mean levels through inclusion of RE λ_i for each unit i . We extend ANOVA to bivariate MLM by modeling deviations at each time point across each unit i ; RE are included for each time point. ANOVA does not make assumptions about the shape of the outcome and is particularly amenable to the analysis of data over a few time points where it may be difficult to model a time trend. ANOVA was implemented by Bohnert, Bradshaw, Latkin (2009) in their analyses of the SHIELD egocentric data across three waves. Similarly, we utilize ANOVA in our SHIELD analysis in Section 5. ANOVA becomes less practical with increasing numbers of waves. Instead we turn to another popular covariance structure, the random intercept and slope model (RIAS). Continuous time is modeled through two β_k coefficients, an intercept and slope for the mean outcome level at Wave 1 and over time respectively. Similarly, λ_i model random intercepts and slopes for each ego. RIAS is utilized for analysis of another egocentric data set (STEP study discussed in Section 4) across four waves in Section 5.

3. Bayesian framework

Bivariate logistic MLM are estimated in a Bayesian framework. Model parameters $\Theta = (\beta_k \lambda_i \Sigma)'$ are estimated through the posterior distribution $p(\Theta|Y)$ that is proportional to the product of the data likelihood $p(Y|\Theta)$ for outcome Y (i.e., binary ego-alter ties) and prior distribution of the model parameters $p(\Theta)$, $p(\Theta|Y) \propto p(Y|\Theta) p(\Theta)$. The data likelihood is formulated for a logistic regression model and expresses Y through a binomial sampling distribution. There are two approaches to selecting a prior distribution for each parameter in Θ . In the absence of model estimates from prior studies or from a theoretical basis, non-informative priors are typically chosen. For example, a normal distribution that is centered on zero and is specified with a variance that is large enough to cover the range of possible parameter values is often used as a prior distribution for regression coefficients. In Section 4.1, we illustrate how informative prior distributions can be selected by incorporating model estimates from the SHIELD study into prior distributions for Bayesian inference on egocentric measures from a subsequent and similarly conducted study (STEP study discussed in Section 4).

Estimation of Θ relies on MCMC simulation where random draws are taken from $p(\Theta|Y)$, or from a distribution that is equivalent to $p(\Theta|Y)$ up to a normalizing constant. Means (M) and standard deviations (SD) are calculated on the random draws for each parameter and are equivalent to point estimates and standard errors in likelihood inference. A linear combination of the parameters $L(\Theta)$, e.g. the difference between two RE correlations, can also be estimated by applying $L(\Theta)$ to each set of random draws from $p(\Theta|Y)$ and calculating M and SD on the linear combination of those draws. We also calculate 95% posterior credible intervals (PI) as the range of draws that fall within the 2.5% and 97.5% percentiles of draws from the posterior distribution. PI that do not include zero are considered statistically significant, relating to a p-value $< .05$ in classical likelihood inference.

3.1 Model Diagnostics

MCMC model diagnostics are used to establish a reasonable level of certainty that convergence has occurred in the estimation of model parameters in Θ . Under MCMC, convergence means that the sequence of random draws, known as a chain, has converged to draws from the posterior distribution. “Mixing” assesses how well the draws are sampled from the parameter space of the posterior distribution $p(\Theta|Y)$. If autocorrelation is high between draws, then mixing will be poor, and a longer chain will be needed for convergence. Adequate chain length can be assessed by testing whether different chains, initiated from different starting parameter values, converge to the same solution. Diagnostics can generally be grouped by whether they focus on mixing or convergence to a single solution. We discuss several of the popular diagnostic plots and tests that are readily available in ‘coda’, an R package that is available on the Comprehensive R Archive Network (<http://cran.r-project.org/>), and can be conducted on output from WinBUGS, the statistical software used to implement bivariate MLM. It is important to examine a number of diagnostic plots and tests as tests are sensitive to different types of convergence problems, e.g. mixing versus convergence to a single solution.

Plots—*Density plots* show the distribution of the sample draws at each iteration. Densities that deviate from a normal curve, e.g. with a long tail, suggest poor mixing. *Autocorrelation plots* graph correlation versus the number of lags between subsequent draws. Autocorrelation that remains high after numerous lags indicates poor mixing. *Trace plots*, i.e. *time series plots*, of multiple chains, initiated from different starting values, visually show whether or not the chains converge to the same solution. See Toft et al. (2007) for example of convergent and non-convergent chains.

Tests—The impact of autocorrelation is reflected in the *effective sample size*, defined as the number of independent draws that the given chain is equivalent to; a small sample size relative to the number of total iterations indicates poor mixing. *Raftery and Lewis* (1992) proposed a dependence factor which represents the degree of inflation the autocorrelation imposes on the required sample size for an adequate length chain; values large than five indicate strong autocorrelation. *Gelman and Rubin* (1992) proposed a diagnostic test of whether or not multiple chains converge to the same solution; statistics substantially greater than 1 indicate a lack of convergence. *Geweke* (1992) proposed a Z-score test for the equality of means between the first and last part of a chain (we use the coda default for the first 10% and last 50% of the chain); large Z-scores indicate a lack of convergence. *Heidelberger and Welch* (1983) proposed a test of the null hypothesis that the draws come from a stationary distribution. Small p-values indicate a lack of convergence.

4. Motivating examples

Bivariate MLM are illustrated through two egocentric data sets that were collected as part of social network-based interventions designed to reduce HIV transmission risk behaviors: the SHIELD study (Bohnert et al., 2009; Latkin et al., 2003a, b) and STEP study (Tobin et al., 2010). In order to evaluate the interventions, participants were tracked over time and queried on relations and shared behaviors with their personal network members, including joint drug use. The longitudinal and egocentric data structure provides an opportunity to investigate dynamic relationships between different types of ego-alter social ties. In this article, we address the research question that was raised earlier. Do egos engage in drug use with alters they trust or vice versa?

Both studies were conducted in a low-income African American drug using community in Baltimore, Maryland. Index participants, i.e. egos, were recruited from 1997 to 1999 in the SHIELD study and from 2004 to 2006 in the STEP study and randomized to an intervention or control condition. Index participants were encouraged to recruit social network members who were drug or sex partners as secondary participants. SHIELD analyses treated both index and secondary participants (60% of the analysis sample) as egos. STEP analyses excluded secondary participants. SHIELD participants were assessed across five waves, roughly nine months apart; analyses are conducted on social network measures that were assessed across three time points (Waves 1, 2, and 4). STEP participants were assessed across four waves (Waves 1–4) at six-month intervals. Repeated assessment on the same alters was possible, though most participants only reported on a particular alter at a single wave (69% of alters in SHIELD and 75% of alters in STEP). Different methods were used in each study to link observations on the same alters across waves. In the SHIELD study, linkage was based on a combination of alters' first names and demographics provided by egos and could not be carried out with absolute certainty. This potential linkage error does not diminish the utility of SHIELD data to illustrate our models, but warrants noting that the assumed repeated structure between ego-alter ties impacts alter-level regression coefficients which may be estimated with less precision. RE correlations, the focus of our analyses, are likely to be attenuated since RE estimates are influenced by model covariates. In the STEP study, egos were explicitly queried as to which alters were the same at each assessment.

Analyses are conducted on egos in the control conditions who were assessed on network measures at each wave (e.g. across all three and four possible waves in the SHIELD and STEP studies respectively). In general the retention rates exceeded 85% per wave. Alter observations were excluded at waves if egos did not report on alters at that wave or alter reports contained missing data. SHIELD analyses were conducted on 121 egos and 1500 alters. STEP analyses were conducted on 114 egos and 2001 alters. Outcome measures represent two types of binary ties for feelings of trust and joint drug use that egos reported

for each alter. Social networks were constructed through an adapted version of the Arizona Social Support Inventory (Barrera and Gottlieb, 1981). Egos were asked how much they trusted each alter on a scale of 1 (“don’t trust at all”) to 10 (“trust with my life”) and if they did drugs with that alter (39% in SHIELD and 35% in STEP). Similar to Bohnert et al. (2009), we dichotomized trust as 1 = 10 and 0 = 1–9. Almost half of the trust responses were rated as a 10 (51% in SHIELD and 48% in STEP).

Covariates encompassed domains of ego-level, network, and alter-level characteristics. We included ego gender (47% female in SHIELD; 46% in STEP), network size at the earliest wave that network characteristics were assessed (Wave 1 in most cases; median size = 9, range = 3 to 20 in SHIELD; median size = 8, range = 2 to 18 in STEP), and whether the network member was reported to be a friend (versus kin, co-worker, etc.; 43% in SHIELD and 50% in STEP).

4.1 Data priors

Uninformative priors are specified for SHIELD analyses. Priors on regression coefficients β_{kp} , $p = 1, \dots, P$, are set to be normal, $\beta_{kp} \sim N(0, 10)$. The STEP study was conducted in a similar manner to the SHIELD study. Therefore, we specify informative STEP priors on coefficients, $\beta_{kp} \sim N(\mu, \phi)$, where μ and ϕ is the parameter and variance estimate from the equivalent coefficient in the SHIELD analysis. Since time is modeled differently across studies, we use an ad hoc procedure to specify the STEP priors on the time coefficients. The μ are set to the smaller of the estimated wave coefficients that represent mean outcome differences between consecutive waves. The ϕ are set to the larger of the estimated variances between coefficients.

The prior on $H \times H$ random effect covariance matrix Σ is set to be inverse-Wishert, $\Sigma \sim IW(S, \nu)$, with scale matrix S and degrees of freedom ν ; $H = 6$ for SHIELD analyses and $H = 4$ for STEP analyses. The expectation of Σ under the inverse-Wishert distribution is specified as $E(\Sigma) = (\nu - H - 1)^{-1} S$ in WinBUGS. We set $\nu = H + 2$ so that $E(\Sigma) = S$, e.g. $\nu = 8$ for SHIELD analyses. It is important to note the parameterization of $E(\Sigma)$ in WinBUGS may differ from notation in some texts, e.g. Gelman et al., (1995, pp. 474–475), $E(\Sigma) = (\nu - H - 1)^{-1} S^{-1}$. We specify a diagonal matrix for the prior of S . Similar to Comulada and Weiss (2007), we use our knowledge of the range of the binary responses to construct reasonable estimates of the random effect variances, i.e. the diagonal elements of S . Probabilities must fall between .01 to .99, or between -4.6 to 4.6 on the logit scale, i.e. $\log(.01)/(1 + \log(.01)) \approx -4.6$. Since six standard deviations span 99% of a normal density, the standard deviation of the random effects is estimated to be $(4.6 - (-4.6))/6 \approx 1.5$. Lastly, the RE variance is $1.5^2 \approx 2.4$. We set $S_{11} = S_{22} = \dots S_{HH} = 2.4$. Off-diagonal elements are specified to be zero, assuming no association between RE correlations a priori.

5. Results

5.1 SHIELD and STEP analyses

Figure 1 shows plots for feelings of trust and joint drug use with alters that was reported by three randomly selected egos in the SHIELD data; similar patterns were seen for the STEP data. Separation between the majority of trust and drug use responses at concurrent waves suggests a negative association. Figure 1 also illustrates that most alter observations only occur at a single time point, which would make it difficult to estimate lagged correlations between individual trust and drug use responses. Rather we rely on bivariate logistic MLM to estimate lagged correlations between RE at the ego level.

Diagnostic tests utilize a 1000 iteration burn-in period, i.e. the period during which model parameter draws are not incorporated into the parameter estimates, followed by 10,000

iterations. Two chains are produced to conduct multiple-chain diagnostics, e.g. the Gelman and Rubin test. Diagnostics for plots and tests discussed in Section 4.3 are conducted across both data sets and are shown for the SHIELD data in Tables 1 and Figure 2. For the drug use outcome across both studies, the parameter for having a network member who is a friend exhibited some of the most favorable convergence characteristics. For example, in the SHIELD data set, the effective sample size = 5567.4, the Gelman & Rubin statistic = 1, The Geweke Z-score = 0.17, the Heidelberger & Welch p-value = .30, and the Raftery-Lewis dependence factor = 1.1. The density plot is normal, autocorrelation drops off quickly, in less than ten lags, and sample draws in the trace plot show a lot of overlap and good mixing. Conversely, The RE correlation between trust at Wave 2 and drug use at Wave 3 in the SHIELD study exhibited some of the worst convergence characteristics. Overall, the parameters are fairly well-behaved, though diagnostics suggest a larger number of iterations should be drawn to ensure adequate mixing. Across both data sets, the effective sample size for parameters ranges from 544.6 to 6040.2 out of 10,000 iterations, or roughly 1 out of 10 iterations, on average. This is also visually observed in autocorrelation plots; autocorrelation tends to drop off after a lag of about 10 iterations. Therefore, final models are based on 100,000 iterations and thinned by 10 samples, after a 1000 iteration burn-in; parameter estimates are based on 10,000 iterations.

Covariate estimates are presented in Table 2. Though not a focus of the analysis, it is interesting to note covariates that were influential across both studies. Egos were more likely to report trusting and less likely to report engaging in drug use with alters over time as indicated by the “Wave” covariates. Egos were less likely to trust and more likely to engage in drug use with alters classified as friends versus a family member or other type of associate.

Tables 3 and 4 show correlation estimates (r) from the models. As expected from Figure 1, both STEP and SHIELD analyses indicate a negative association between egos’ feelings of trust and joint ego-alter drug use with alters at concurrent waves. SHIELD analyses show negative correlations between RE at each time point ($r_{11} = -.46$; $r_{22} = -.32$; and $r_{44} = -.21$), though correlation is only significant at Wave 1. STEP analyses show negative correlations between RE at baseline ($r = -.50$).

Correlation results across waves are less conclusive. SHIELD analyses indicate joint drug use at Wave 1 to be inversely related to trust at Waves 2 ($r_{21} = -.40$) and 4 ($r_{41} = -.38$). However, these correlations do not significantly differ from correlations between trust at Wave 1 and drug use at Waves 2 ($r_{12} = -.34$) or 4 ($r_{14} = -.14$) which are also negative, e.g. $r_{41} - r_{14} = -.24$ (PI = $-.76$ to $.26$). STEP analyses indicate initial trust (Intercept) is associated with greater joint drug use over time (Slope; $r = .31$), but also indicate initial joint drug use to be associated with trust over time ($r = .41$). RE slopes for trust and joint drug use were negatively correlated ($r = -.35$).

As discussed in Section 4, STEP analyses use informative priors for the regression coefficients. As a sensitivity analysis, we rerun STEP analyses with non-informative priors, $\beta_{kp} \sim N(0, 10)$. Results are similar and lead to the same conclusions. For example, the correlation between RE slopes based on informative priors in Table 4 changed from $-.35$ to $-.33$.

5.2 Simulation study

Simulation studies examine the impact of incorrectly specified ties in larger egocentric networks on statistical inference through bivariate MLM. We start with our analysis data sets from the SHIELD and STEP studies as the gold standards to compare our simulated data sets against. Error is introduced into the joint drug use response; trust is not altered.

This may play out in general for the analysis of paired ties where one type of tie measures less socially acceptable behaviors and is therefore, more unreliable. Drug responses are randomly chosen to be incorrect with probability p across networks with more than q alters. Two types of errors are simulated: 1) responses are altered by changing responses from zero to one, or vice versa; and 2) responses are set to zero, regardless of the original value. Error 1) may be more attributable to an ego who is less familiar with their network members and error 2) may be more attributable to egos who give the same 'no' response to get through an interview or not report less acceptable behaviors. We simulate two data sets each for all six possible combinations with $p = .20, .50$, and $.80$, and $q = 8$ (29% and 28% of the networks across waves in the SHIELD and STEP studies respectively) and 10 (14% of the networks across waves in both the SHIELD and STEP studies). Bivariate logistic MLM are fit to the simulated data using the same modeling approach, including specified priors and the number of iterations that are implemented in Section 5.1.

Figures 3a and 3b show estimated correlations for simulations ($q = 10$) on the SHIELD and STEP studies respectively. Horizontal lines represent correlations from bivariate logistic MLM applied to the original data sets (Tables 3 and 4). Correlations from simulated data sets are represented by points. Dots for the correlation between the trust and drug use slopes that correspond to error 2) tend to be closer to the horizontal line, indicating that error 2) has the least impact on the correlation relative to error 1). Both error types result in a noticeable loss of statistical significance. This is especially true for the estimation of correlations on RE for lagged versus concurrent observations, which unfortunately may be of greater interest in a longitudinal study. For example, statistical significance is lost across all simulated data sets for the correlation between the trust intercept and drug use slope (Figure 3b). As expected, results for $q = 8$ result in an even greater loss of statistical significance (not shown, but available upon request from the authors).

6. Discussion

The growing prominence of longitudinal social network studies in the form of both network-based intervention trials (e.g., Heckathorn, 1999) and observational studies (e.g., Christakis and Fowler, 2008) over the past several decades have called for new analytic tools to better understand complex relationships between multiple dynamic network ties. Towards this goal we proposed a bivariate-outcome framework for the analysis of longitudinal egocentric network data that is also gaining prominence in other fields of research to explore dynamic relationships (e.g., Audrain-McGovern et al., 2003; Comulada et al., 2010; Huang et al., 2010). Similar to cross-sectional egocentric studies that utilized univariate-outcome models and examined the relationship between ego-alter social cohesion, e.g. trust (Cepeda et al., 2011) and emotional support (Unger et al., 2006), and needle sharing, we found significant relationships between egos' feelings of trust and joint drug use with alters at the first wave. Needle sharing studies found positive associations in contrast to negative associations reported by egos for general drug use in our study, which may reflect the different domains of substance use that were examined.

A main contribution of our bivariate approach is the ability to jointly examine multiple relationships between social network measures across waves and begin to establish temporal precedence. For example, we examined how egos' feelings of trust towards alters at Wave 1 was related to joint ego-alter drug use at Wave 4 (r_{14}), and vice versa (r_{41}). If r_{14} was found to be significant, but not r_{41} , it would suggest that trust, or a lack of trust depending on the direction of association, precedes joint drug use. Unfortunately, in our study, both r_{14} and r_{41} were significant and our results were inconclusive. Moreover, the direction of the relationship between trust and joint drug use differed across studies. SHIELD analyses found negative correlations between trust at earlier waves and joint drug use at later waves,

and vice versa. STEP analyses found a positive correlation between trust at the first wave and the trajectory of joint drug use over time (i.e., the joint drug use slope), and vice versa.

The univariate-outcome approach would have required multiple models to fully test relationships between trust and joint drug use across waves. The temptation would have been to draw a conclusion after the first model were fit, for example, that trust precedes joint drug use. Furthermore, the bivariate model can easily estimate and test the statistical significance between multiple relationships across waves which are more difficult to construct through multiple univariate-outcome models. For example, SHIELD analyses estimated the difference between the correlation of trust at Wave 4 and joint drug use at Wave 1 (r_{41}) and r_{14} , $r_{41} - r_{14} = -.24$. The PI included zero (PI = $-.76$ to $.26$) and indicated that the correlations for the bidirectional relationships did not significantly differ.

Our analyses focused on correlations between two different types of ego-alter ties (feelings of trust and joint drug use) at the ego level through RE that acted as underlying ego-level constructs. Ego-level RE can also provide insight into relations between different types of alter-alter ties and between ego-alter and alter-alter ties. In particular, a better understanding of relations between ego-alter and alter-alter ties at the ego-level could elucidate how ego choices influence network structure and how network structure influences egos' choices (Lubbers et al., 2010), perhaps in the exploration of ego and alter drug-related influences. This would be related to work by Bohnert et al. (2009) that explored homophily versus network structure influence related to drug use. If alter-level RE are included for repeated alter observations over time, dynamic bivariate relationships can also be modeled between social ties or other network characteristics at the alter-level. An application would be in the exploration of the relationship between alter network characteristics, e.g. centrality, and alter behavior, e.g. drug use. We did not explore alter network or individual characteristics in our data because most of the alters were not reported on across multiple time points.

Our approach retains a limitation that is common to social network analysis. Egocentric networks are not completely assessed and contain missing or incorrectly specified responses on alters that are assessed. Heckathorn (1997, 2002) presents sampling and modeling approaches to minimize the impact of incomplete networks on inference. If it can be assumed that the missing data mechanism is non-informative, i.e. does not depend on unobserved variables, missing data can be imputed (Rubin, 1987). A benefit of the Bayesian approach over likelihood methods is the ability to treat missing data points as unknown model parameters to be estimated with the other model parameters. A single step is involved on the part of the analyst. In contrast, likelihood methods require two steps: multiple imputation and then model fitting to the imputed data sets. However, both approaches are less practical for larger fractions of missing data. Not surprisingly, our simulation results demonstrated that associations between ties lose statistical significance as error rates increase in larger egocentric networks. To some degree, this response bias can be evaluated by checking for associations between network size and the network tie of interest. Further study is needed to explore the full impact on inference and research scenarios where these error rates are most likely to be the most prevalent.

Another limitation of the bivariate MLM is the implementation in available statistical software. As a first step, we attempted to fit bivariate MLM to the SHIELD and STEP data through likelihood and pseudo-likelihood inference as implemented through the NLMIXED and GLIMMIX procedures in SAS (SAS Institute Inc., Cary, NC), a popular and commercial software package. Not surprisingly, we were unsuccessful. While we had success with Bayesian inference, the additional programming steps, uncertainty around convergence and additional computing time that MCMC requires is understandably daunting

to some researchers. Our final models based on 100,000 iterations took several hours to run on a computer with an Intel Xeon 3Ghz CPU with 3.25 GB of RAM.

Despite challenges, MLM analyses that focus on ego-level characteristics, as was done in this article, can attenuate some of the modeling issues discussed above. For many network characteristics the boundaries are fuzzy, such as the distinction between a friend or drug partner. Inference on alter-level characteristics may be difficult. By focusing on relationships between ego-level RE, some of the measurement error at the alter level can be accounted for and we can at least arrive at a clearer understanding of ego-level relations. It is perhaps the ego level that is the most important in egocentric data that is determined by the ego, interviewer, and the interaction between them. Since perceived alter behaviors may be more strongly related to ego behavior than actual alter behavior (Valente, 2005), inference at the ego level is often the proper focal point, especially in behavioral interventions that target egos' behaviors.

Acknowledgments

This study was funded by the National Institutes of Health (K01MH089270, R01DA13142, and R01DA016555) and the Open Society Institute-Baltimore Drug Addiction Treatment Program (20007265). We thank Phillip Bonacich for helpful comments regarding the manuscript.

References

- Audrain-McGovern J, Rodriguez D, Moss HB. Smoking progression and physical activity. *Cancer Epidemiology, Biomarkers and Prevention*. 2003; 12:1121–1129.
- Barrera, M., Jr; Gottlieb, BH. Social support in the adjustment of pregnant adolescents: Assessment issues. In: Gottlieb, BH., editor. *Social networks and social support*. Beverly Hills, CA: Sage Publications; 1981. p. 69-96.
- Bohnert ASB, Bradshaw CP, Latkin CA. A social network perspective on heroin and cocaine use among adults: Evidence of bidirectional influences. *Addiction*. 2009; 104:1210–1218. [PubMed: 19563564]
- Caimo A, Friel N. Bayesian inference for exponential random graph models. *Social Networks*. 2011; 33:41–55.
- Cepeda JA, Odinokova VA, Heimer R, Grau LE, Lyubimova A, Safiullina L, Levina OS, Niccolai LM. Drug network characteristics and HIV risk among injection drug users in Russia: The roles of trust, size, and stability. *AIDS and Behavior*. 2011; 15:1003–1010. [PubMed: 20872063]
- Christakis NA, Fowler JH. The collective dynamics of smoking in a large social network. *The New England Journal of Medicine*. 2008; 358:2249–2258. [PubMed: 18499567]
- Comulada WS, Weiss RE. On models for binomial data with random numbers of trials. *Biometrics*. 2007; 63:610–617. [PubMed: 17688514]
- Comulada WS, Rotheram-Borus MJ, Pequegnat W, Weiss RE, Desmond K, Arnold E, Remien RH, Morin SF, Weinhardt L, Johnson MO, Chesney MA. Relationships over time between mental health symptoms and transmission risk among persons living with HIV. *Psychology of Addictive Behaviors*. 2010; 24:109–118. [PubMed: 20307117]
- de Miguel Luken V, Tranmer M. Personal support networks of immigrants to Spain: A multilevel analysis. *Social Networks*. 2010; 32:253–262.
- Gelman, A.; Carlin, JB.; Stern, HS.; Rubin, DB. *Bayesian Data Analysis*. 1st edition. Chapman and Hall; 1995.
- Gamerman, D. *Markov Chain Monte Carlo*. 1st edition. Chapman and Hall; 1997.
- Hall JA. Parents' networks: Egocentric networks and unique and shared sources of social support. *Connections*. 2010; 30:41–49.
- Heckathorn DD. Respondent driven sampling: A new approach to the study of hidden populations. *Social Problems*. 1997; 44:174–199.

- Heckathorn DD. Respondent-Driven Sampling II: Deriving valid population estimates from chain-referral samples of hidden populations. *Social Problems*. 2002; 49:11–34.
- Heckathorn DD, Broadhead RS, Anthony DL. AIDS and social networks: HIV prevention through network mobilization. *Sociological Focus*. 1999; 32:159–179.
- Huang DYC, Evans E, Hara M, Weiss RE, Hser Y-I. Employment trajectories: exploring gender differences and impacts of drug use. *Journal of Vocational Behavior*. 2010; 79:277–289. [PubMed: 21765533]
- Latkin CA, Hua W, Forman VL. The relationship between social network characteristics and exchanging sex for drugs or money among drug users in Baltimore, MD, USA. *International Journal of STD & AIDS*. 2003a; 14:770–775. [PubMed: 14624742]
- Latkin CA, Sherman S, Knowlton A. HIV prevention among drug users: outcome of a network-oriented peer outreach intervention. *Health Psychology*. 2003b; 22:332–339. [PubMed: 12940388]
- LaZarsfeld, P.; Merton, RK. Friendship as a Social Process: A Substantive and Methodological Analysis. In: Berger, M.; Abel, T.; Page, CH., editors. *Freedom and Control in Modern Society*. New York: Van Nostrand; 1954. p. 18-66.
- Lubbers MJ, Molina JL, Lerner J, Brandes U, Avila J, McCarty C. Longitudinal analysis of personal networks. The case of Argentinean migrants in Spain. *Social Networks*. 2010; 32:91–104.
- Morris M, Handcock MS, Hunter DR. Specification of exponential-family random graph models: terms and computational aspects. *Journal of Statistical Software*. 2008; 24:1–24. [PubMed: 18612375]
- Raftery AE, Lewis SM. One long run with diagnostics: Implementation strategies for Markov chain Monte Carlo. *Statistical Science*. 1992; 7:493–497.
- Reinsel G. Multivariate repeated-measurement of growth curve models with multivariate random-effects covariance structure. *Journal of the American Statistical Association*. 1982; 77:190–195.
- Reinsel G. Estimation and prediction in a multivariate random effects generalized linear model. *Journal of the American Statistical Association*. 1984; 79:406–414.
- Robert, C.; Casella, G. *Monte Carlo Statistical Methods*. 1st edition. Springer; 1999.
- Robins G, Pattison P, Kalish Y, Lusher D. An introduction to exponential random graph (p^*) models for social networks. *Social Networks*. 2007; 29:173–191.
- Rubin, DB. *Multiple Imputation for Nonresponse in Surveys*. New York, NY: Wiley; 1987.
- Snijders TAB. Stochastic actor-oriented dynamic network analysis. *Journal of Mathematical Sociology*. 1996; 21:149–172.
- Snijders, TAB. The statistical evaluation of social network dynamics. In: Sobel, M.; Becker, M., editors. *Sociological Methodology*. Boston and London: Basil Blackwell; 2001. p. 361-395.
- Snijders, TAB.; Steglich, CEG.; Schweinberger, M.; Huisman, M. *Manual for SIENA version 3.2*. Groningen: ICS, University of Groningen; Oxford: Department of Statistics, University of Oxford; 2008. Available from: <http://stat.gamma.rug.nl/snijders/siena.html>.
- Spiegelhalter, DJ.; Thomas, A.; Best, NG. *WinBUGS Version 1.4 User Manual*. MRC Biostatistics Unit; 2000. <http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml>.
- Stone, AA.; Shiffman, S.; Atienza, A.; Nebeling, L. *Self Reports in Health Research*. Oxford University Press; 2007. *The Science of Real-time Data Capture*.
- Tobin KE, Kuramoto SJ, Davey-Rothwell MA, Latkin CA. The STEP into Action study: a peer-based, personal risk network-focused HIV prevention intervention with injection drug users in Baltimore, Maryland. *Addiction*. 2010; 106:366–375. [PubMed: 21054614]
- Toft N, Innocent GT, Gettinby G, Reid SWJ. Assessing the convergence of Markov Chain Monte Carlo methods: an example from evaluation of diagnostic tests in absence of a gold standard. *Preventative Veterinary Medicine*. 2007; 79:244–256.
- Unger JB, Kipke MD, De Rosa CJ, Hyde J, Ritt-Olson A, Montgomery S. Needle-sharing among young IV drug users and their social network members: The influence of the injection partner's characteristics on HIV risk behavior. *Addictive Behaviors*. 2006; 31:1607–1618. [PubMed: 16459023]
- Valente, TW. *Social Networks and Health: Models, Methods, and Applications*. Vol. chapter 4. New York: Oxford University Press; 2010.

- Valente, TW. Network models and methods for studying the diffusion of innovations. In: Carrington, PJ.; Scott, J.; Wasserman, S., editors. *Models and Methods in Social Network Analysis*. New York, NY: Cambridge University Press; 2005. p. 106
- Wang, P.; Robins, G.; Pattison, P. Pnet: a program for the simulation and estimation of exponential random graph models. University of Melbourne; 2004.
- Wasserman, S.; Faust, K. *Social Network Analysis: Methods and Applications*. Vol. chapter 2. Cambridge University Press; 1994.
- Wasserman S, Pattison P. Logit models and logistic regressions for social networks: I. an introduction to Markov graphs and p^* . *Psychometrika*. 1996; 61:401–425.
- Weiss, RE. *Modeling Longitudinal Data*. New York: New York: Springer-Verlag; 2005. p. 380

- Multi-level models are applied to pairs of longitudinal egocentric network ties.
- Tie relationships are examined between concurrent and lagged time points.
- Models and diagnostic tests are presented in a Bayesian framework.

\$watermark-text

\$watermark-text

\$watermark-text

Figure 1.
SHIELD dataset: Profile plots of ego-alter ties for trust (●—) and drug use outcomes (◊---) from 3 randomly selected egos. A small constant (.1) was added to trust outcome values.

Drug use outcome: friend parameter

Random effect correlation between trust (Wave 2) and drug use (Wave 3)

Figure 2.
SHIELD dataset: MCMC density and autocorrelation plots of a single chain, and trace plots of last 200 iterations for two chains based on 10,000 iterations with 1,000 iteration burn-in.

a

b

Figure 3.

a. SHIELD dataset: Comparison of random effect correlations for trust (T) and drug use (D) outcomes across waves (1, 2, and 4) between original data (represented by horizontal line) and data with simulated drug response errors for network sizes > 10. Data point clusters (4 points each) from left to right correspond to 20% (squares), 50% (circles), and 80% error rates (triangles). Within each cluster, incorrect response = 1 - original response for first two simulated data sets and response is set to zero, regardless of the original response, in last 2 simulated data sets. Significant correlations are indicated by * in original data and solid points in simulated data.

b. STEP dataset: Comparison of random intercept and slope correlations for trust (T) and drug use (D) outcomes across waves between original data (represented by horizontal line) and data with simulated drug response errors for network sizes > 10 . Data point clusters (4 points each) from left to right correspond to 20% (squares), 50% (circles), and 80% error rates (triangles). Within each cluster, incorrect response = $1 - \text{original response}$ for first two simulated data sets and response is set to zero, regardless of the original response, in last 2 simulated data sets. Significant correlations are indicated by * in original data and solid points in simulated data.

\$watermark-text

\$watermark-text

\$watermark-text

Table 1

SHIELD dataset: MCMC diagnostics for parameters from bivariate logistic multi-level model based on 10,000 iterations with 1,000 iteration burn-in.

Parameter	Effective sample size	Gelman and Rubin	Geweke (Z-score)	Heidelberger and Welch (p-value)	Raftery-Lewis (dep. factor)
Trust					
Intercept	1050.0	1	0.69	0.8	3.2
Wave 2 indicator	2207.8	1	1.28	0.14	2.6
Wave 4 indicator	1603.0	1	-0.15	0.39	2.7
Index female gender	830.7	1	1.30	0.51	3.5
Network size	856.4	1	-0.77	0.69	5.3
Network member is friend	3370.3	1	-1.37	0.06	1.2
Random effect variances					
Wave 1	1675.5	1.01	-0.0083	0.52	3.3
Wave 2	603.1	1	0.36	0.75	4.0
Wave 4	544.6	1	0.95	0.04	5.8
Drug use					
Intercept	2264.3	1	-0.41	0.79	2.5
Wave 2 indicator	3160.8	1	1.57	0.38	1.4
Wave 4 indicator	2553.6	1	-0.02	0.64	2.6
Index female gender	2028.8	1	-0.28	0.63	2.6
Network size	1854.0	1	-0.15	0.91	1.6
Network member is friend	5567.4	1	0.17	0.30	1.1
Random effect variances					
Wave 1	1504.1	1	0.52	0.53	3.1
Wave 2	856.0	1	-0.93	0.79	3.7
Wave 4	787.4	1	1.37	0.21	6.2
Correlations between random effects for trust at wave i and drug use at wave j (ij)					
(11)	1637.4	1	-0.29	0.78	3.5
(12)	920.4	1	0.30	0.98	3.1
(13)	680.6	1.01	-1.06	0.17	4.5

\$watermark-text

\$watermark-text

\$watermark-text

Parameter	Effective sample size	Gelman and Rubin	Geweke (Z-score)	Heidelberger and Welch (p-value)	Raftery-Lewis (dep. factor)
(21)	969.9	1	0.27	0.43	3.4
(22)	905.7	1	0.24	1.00	3.2
(23)	542.9	1.02	-1.10	0.13	6.4
(31)	1019.1	1	0.043	0.68	3.1
(32)	543.2	1	0.39	0.80	3.2
(33)	705.8	1	-2.16	0.052	3.1

\$watermark-text

\$watermark-text

\$watermark-text

Table 2

Parameter estimates from fixed effects for bivariate logistic model represented by posterior means (M) and standard deviations (SD)

Parameter	SHIELD data		STEP data	
	M	SD	M	SD
Trust				
Intercept	.86	.38 *	1.29	.36 *
Wave (continuous)			.19	.081 *
Wave 2 indicator	.46	.20 *		
Wave 4 indicator	.23	.22		
Index female gender	-.073	.27	.17	.27
Network size	.0070	.038	-.039	.029
Network member is friend	-2.53	.14 *	-2.97	.12 *
Shared drug use				
Intercept	-1.00	.26 *	-.84	.24 *
Wave (continuous)			-.44	.066 *
Wave 2 indicator	-.59	.17 *		
Wave 4 indicator	-.44	.18 *		
Index female gender	-.078	.17	.28	.16
Network size	.0078	.024	-.014	.020
Network member is friend	1.85	.12 *	2.06	.10 *

* posterior credible interval does not include zero, indicating significance.

Table 3

SHIELD dataset: Correlations between random effects across three waves as estimated by bivariate logistic multi-level model

		Drug use		
	Wave	1	2	4
Trust	1	-.46 [*]	-.34	-.14
	2	-.40 [*]	-.32	-.13
	4	-.38 [*]	-.25	-.21

* posterior credible interval does not include zero, indicating significance.

Table 4

STEP dataset: Correlations between random intercepts and slopes as estimated by bivariate logistic multi-level model

		Drug use Intercept	Slope
Trust	Intercept	-.50 *	.31 *
	Slope	.41 *	-.35 *

* posterior credible interval does not include zero, indicating significance.