Regression of exchangeable relational arrays

Marrs, Frank Wayne III
Fosdick, Bailey
McCormick, Tyler

To be published in: Biometrika

DOI to publisher's version: 10.1093/biomet/asac031

Regression of exchangeable relational arrays

Frank W. Marrs[1], Bailey K. Fosdick[2], and Tyler H. McCormick[3]

LA-UR-20-30061

[1]Statistical Sciences, Los Alamos National Laboratory,
fmarrs3@lanl.gov

[2]Department of Statistics, Colorado State University

[3]Departments of Statistics and Sociology, University of Washington

June 23, 2022

Abstract

Relational arrays represent measures of association between pairs of actors, often in varied contexts or over time. Trade flows between countries, financial transactions between individuals, contact frequencies between school children in classrooms, and dynamic protein-protein interactions are all examples of relational arrays. Elements of a relational array are often modeled as a linear function of observable covariates. Uncertainty estimates for regression coefficient estimators – and ideally the coefficient estimators themselves – must account for dependence between elements of the array (e.g. relations involving the same actor) and existing estimators of standard errors that recognize such relational dependence rely on estimating extremely complex, heterogeneous structure across actors. This paper develops a new class of parsimonious coefficient and standard error estimators for regressions of relational arrays. We leverage an exchangeability assumption to derive standard error estimators that pool information across actors and are substantially more accurate than existing estimators in a variety of settings. This exchangeability assumption is pervasive in network and array models in the statistics literature, but not previously considered when adjusting for dependence in a regression setting with relational data. We demonstrate improvements in inference theoretically, via a simulation study, and by analysis of a data set involving international trade.

Keywords: array data; weighted network; dependent data; generalized least squares.

1 Introduction

Entries in relational arrays quantify pairwise interactions between actors that may be of multiple types or observed over time. Examples include annual flows of migrants between countries (Aleskerov et al., 2017) and interactions between students over the course of a semester (Han

et al., 2016). In economics, relational arrays are used to describe monetary transfers between individuals as part of informal insurance markets (see Bardham, 1984; Fafchamps, 2006; Foster and Rosenzweig, 2001; Attanasio et al., 2012; Banerjee et al., 2013)). Other examples of data that are naturally represented as relational arrays include gene expressions (Zhang and Horvath, 2005) and international relations (Fagiolo et al., 2008).

A relational array $Y = (y_{ijr})$ where $(i, j = 1, \ldots, n, i \neq j, r = 1, \ldots, R)$, is composed of a series of R matrices of size $(n \times n)$, each of which describes the directed pairwise relationships among n actors of type r, e.g. time period r or relation context r. The diagonal elements of each matrix, for example y_{iir}, are assumed to be undefined, as we do not consider, e.g., international relations of a country with itself. The relationship from actor i to actor j may differ than that from j to i, $y_{ijr} \neq y_{jir}$, however, the methods we propose extend to the symmetric relation case (see Section S3).

The primary goal in our setting is inference for linear regressions exploring the effects of exogenous covariates on the values in the relational array, expressed as

$$y_{ijr} = \beta^T x_{ijr} + \xi_{ijr}, \quad (i, j = 1, \ldots, m;\ i \neq j;\ r = 1, \ldots, R), \tag{1}$$

where y_{ijr} is a continuous directed measure of the rth relation from actor i to actor j, x_{ijr} is a $(p \times 1)$ vector of covariates, and ξ_{ijr} is an unobserved, scalar random error. For example, considering informal insurance markets, Fafchamps and Gubert (2007) examine how covariates such as geographical proximity and kinship relate to risk sharing relations after economic shocks.

A core challenge in making inference on β arises from the innate dependencies among error relations involving the same actor. For example, dependence often exists between trade relations involving the same country or between economic transfers originating from the same individual. This dependence may arise due to variation unaccounted for in the covariates, for example, from differences in production levels between nations or from individual differences in risk aversion. Standard regression techniques may lead to poor estimates of β and/or incorrect conclusions regarding the significance of the estimate of β. Approaches to account for error dependence in relational arrays have appeared in the statistics, biostatistics, and econometrics literatures and can be characterized into two broad classes.

The first set of approaches impose a parametric model on the errors. Specifically, they either use latent variables to model the array measurements as conditionally independent given the latent structure (see Holland et al., 1983; Wang and Wong, 1987; Hoff et al., 2002; Li and Loken, 2002; Hoff, 2005) or model the error covariance structure directly subject to a set of simplifying assumptions (see Hoff, 2011; Fosdick and Hoff, 2014; Hoff, 2015). While these methods allow for (possibly) improved estimation of β and appropriate standard error estimators in the presence of relational dependence, the accuracy of inference on β depends on the extent to which the true error structure is consistent with the specified parametric model. In addition, many of these models are estimated in a Bayesian paradigm using Markov chain Monte Carlo approaches, which are commonly computationally expensive to estimate.

The second set of approaches to accounting for relational dependence relies heavily on empirical estimates of the error structure based on the regression residuals, first proposed by Fafchamps and Gubert (2007) and based on the spatial dependence work of Conley (1999). This framework is model agnostic, making as few assumptions as possible about the data generating process. One empirical approach estimates the regression coefficients using ordinary least squares, and then utilizes a sandwich covariance estimator – which is robust to a wide array of error structures – for the

standard errors of the regression coefficients (Fafchamps and Gubert, 2007; Aronow et al., 2015). In finite samples, this estimator is hindered by the need to estimate a large number of covariance parameters with limited observations (King and Roberts (2015) for a discussion in other contexts), and is the reason why Wakefield (2013) suggests such estimators be labeled empirical rather than robust. We observe that standard errors from this empirical framework are often highly variable and are anticonservative.

In this work, we introduce an empirical estimation approach for relational arrays that incorporates an exchangeability assumption. This assumption is implicit in many of the model-based approaches discussed previously and is a hallmark of Bayesian hierarchical models within, and outside, the relational context (Orbanz and Roy (2015)). Our key contribution is to define the covariance matrix (and a corresponding estimator) of the relational error array under exchangeability. Use of our parsimonious estimator produces superior estimates of β and its standard errors relative to existing approaches, and our estimator is easier to compute than existing Bayesian model-based and exchangeable bootstrapping (Menzel, 2017; Green and Shalizi, 2017) approaches. Reproduction code is available at `https://github.com/fmarrs3/netreg_public` and methods are implemented in `R` package `netregR`.

2 Inference in relational regression

2.1 Estimation of regression coefficients

We employ a least squares framework to perform inference on β in the relational regression model in (1) Aitkin (1935). An unbiased estimator for β is the ordinary least squares estimator, $\widehat{\beta} = \left(X^T X\right)^{-1} X^T y$, where X is an $\left(Rn(n-1) \times p\right)$ matrix of $(p \times 1)$ covariate vectors (x_{ijr}) and y is a vectorized representation of (y_{ijr}). The least squares estimator is the best linear unbiased estimator for β when the covariance matrix $\Omega = \text{var}(y \mid X)$ is proportional to the identity matrix. Dependence is expected in relational data, e.g. between relations (i, j, r) and (i, k, r) which share actor i. If Ω were known, the best linear unbiased estimator for β is the generalized least squares estimator of Aitkin (1935),

$$\widehat{\beta}_{GLS} = (X^T \Omega^{-1} X)^{-1} X^T \Omega^{-1} y. \tag{2}$$

In practice, Ω is unknown and must be estimated. Given estimator $\widehat{\Omega}$, alternating estimation of Ω with (2) (replacing Ω with $\widehat{\Omega}$ at each iteration) is termed feasible generalized least squares. When $\widehat{\Omega}$ is consistent, feasible generalized least squares is asymptotically efficient for β (Greene, 2003; Hansen, 2015).

Regardless of whether the ordinary least squares estimator or (2) is used to estimate β, uncertainty estimates are required for inference. A common approach is to approximate the distribution of the β estimate as a multivariate normal random variable and construct confidence intervals using an estimator of its variance: in the ordinary least squares setting,

$$\text{var}\left(\widehat{\beta} \mid X\right) = (X^T X)^{-1} X^T \Omega X (X^T X)^{-1}, \tag{3}$$

and in the generalized least squares setting,

$$\text{var}\left(\widehat{\beta}_{GLS} \mid X\right) = (X^T \widetilde{\Omega}^{-1} X)^{-1} X^T \widetilde{\Omega}^{-1} \Omega \widetilde{\Omega}^{-1} X (X^T \widetilde{\Omega}^{-1} X)^{-1}, \tag{4}$$

where $\widetilde{\Omega}$ is the final estimate of Ω from the generalized least squares procedure. Variance estimators are often constructed by substituting an estimator for Ω in (3) and (4), and are commonly termed sandwich estimators (Huber, 1967; White, 1980). Thus, inference for β requires an estimator for Ω, regardless of how β is estimated, and properties of the estimator of var$(\widehat{\beta} \mid X)$ depend strongly on the estimator of Ω.

2.2 Dyadic clustering estimator

Fafchamps and Gubert (2007), Cameron et al. (2011), Aronow et al. (2015), and Tabord-Meehan (2018) propose and describe the properties of a flexible standard error estimator for relational regression which makes the sole assumption that two relations (i, j, r) and (k, l, s) are independent if (i, j) and (k, l) do not share an actor. This assumption implies that cov$(y_{ijr}, y_{kls} \mid X) =$ cov$(\xi_{ijr}, \xi_{kls} \mid X) = 0$ for non-overlapping relation pairs, but places no restrictions on the covariance elements for pairs of relations that share an actor. Let Ω_{DC} denote the covariance matrix of ξ, subject to this non-overlapping pair independence assumption. Fafchamps and Gubert (2007) propose estimating each nonzero entry of Ω_{DC} with a product of residuals, i.e. using $e_{ijr}e_{iks}$ to estimate cov(ξ_{ijr}, ξ_{iks}), where $e_{ijr} = y_{ijr} - \widehat{\beta}^T x_{ijr}$. The estimator $\widehat{\Omega}_{DC}$ can be seen as that which takes the empirical covariance of the residuals defined by ee^T, where e is a vector of the set of residuals (e_{ijr}), and introduces zeros to enforce the non-overlapping pair independence assumption. Fafchamps and Gubert (2007) propose a sandwich variance estimator for var$(\widehat{\beta} \mid X)$ in (3) based on $\widehat{\Omega}_{DC}$,

$$\widehat{V}_{DC} = (X^T X)^{-1} X^T \widehat{\Omega}_{DC} X (X^T X)^{-1}. \tag{5}$$

We refer to $\widehat{V}_{DC}$ as the dyadic clustering estimator as it owes its derivation to the extensive literature on cluster-robust standard error estimators.

The dyadic clustering estimator in (5) has the attractive properties that it is asymptotically consistent under wide range of error dependence structures and is fast to compute. However, $\widehat{\Omega}_{DC}$ estimates $O(R^2 n^3)$ nonzero covariance elements separately based on $O(R^2 n^2)$ dependent observations, and thus, $\widehat{V}_{DC}$ is inherently quite variable. Only when there is extreme heterogeneity in the true covariance structure is the dyadic clustering method optimal and it will suffer a loss of efficiency otherwise.

3 Standard errors under exchangeability

3.1 Exchangeability in relational models

A common modeling assumption for relational and array structured errors is exchangeability. Defined by de Finetti for a univariate sequence of random variables, exchangeability was generalized to array data by Hoover (1979) and Aldous (1981). The errors in a relational data model are jointly exchangeable if the probability distribution of the error array $\Xi = (\xi_{ijr})$ is invariant under any simultaneous permutation of the rows and columns, and secondary permutation of the third dimension. Mathematically, this means pr$(\Xi) =$ pr$\{\Pi(\Xi)\}$, where $\Pi(\Xi) = \{\xi_{\pi(i)\pi(j)\nu(r)}\}$ is the error array with its indices reordered according to permutation operators π and ν. Intuitively, exchangeability in the regression context means the observed covariates are sufficiently informative such

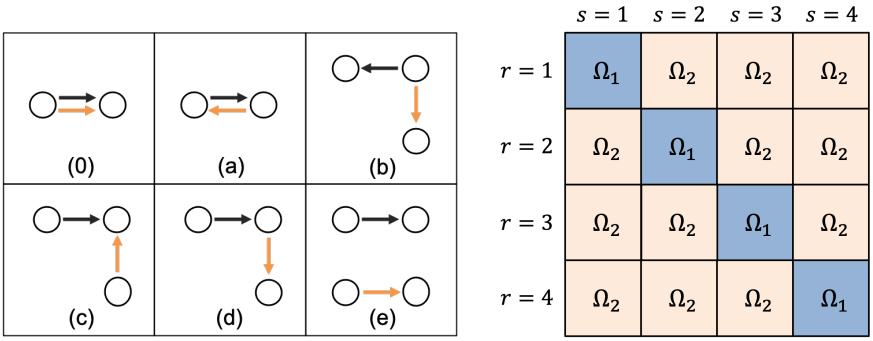

Figure 1: Left: Six distinguishable configurations of relation pairs (black and orange arrows) in an exchangeable relational model with unlabeled actors, corresponding to twelve covariance values (six each for $s = r$ and $s \neq r$). Right: Block structure in Ω_E for $R = 4$, where blocks Ω_1 and Ω_2 correspond to $s = r$ and $s \neq r$, respectively.

that the labels of the rows and columns in the error array are uninformative. Similarly the ordering of the third dimension of the error array is uninformative to its distribution. This assumption may be appropriate when the third dimension of the array represents different contexts of observations (such as economic trade sectors) that have no inherent ordering, or when the third dimension represents time periods, but the bulk of the temporal variation is accounted for in the covariates. Many of the conditionally independent parametric latent variable models cited in the Introduction have this joint exchangeability property (Hoff, 2008; Bickel and Chen, 2009).

3.2 Impact of exchangeability on covariance structure

Li and Loken (2002) and Hoff (2005) describe several particular random effects models for $R = 1$. The corresponding error covariance matrices have different entries depending on the model, yet all covariance matrices have at most six unique entries. A key contribution of this paper is to formalize and extend this observation, showing that any jointly exchangeable model for relational array Ξ results in an Ω of the same form, with at most six unique terms when $R = 1$ and at most twelve unique terms when $R > 1$.

Proposition 1. *If a probability model for a directed relational array Ξ is jointly exchangeable and has finite second moments, then the covariance matrix of Ξ contains at most twelve unique values.*

The twelve (possibly) unique entries in Ω correspond to the twelve distinguishable configurations of relation pairs (i, j, r) and (k, l, s) with unlabeled actors. The twelve configurations can be separated into two sets of six identical configurations of relations (i, j) and (k, l) with unlabeled actors, as depicted in Figure 1, where each set corresponds to $r = s$ and $r \neq s$. A proof is provided in Section S4.

3.3 Covariance matrices of exchangeable relational arrays

Similar to the dyadic clustering estimator, we assume non-overlapping relation pairs are independent, such that $\text{cov}(\xi_{kls}, \xi_{ijr}) = 0$, for any s and r when (i, j, k, l) are distinct. This assumption sets two of the twelve parameters in Ω to zero. We introduce a new class of covariance matrices which contain ten possibly nonzero entries, $\phi_0^{(\eta)}, \phi_a^{(\eta)}, \phi_b^{(\eta)}, \phi_c^{(\eta)}, \phi_d^{(\eta)}, (\eta = 1, 2)$, associated with the left panel in Figure 1. The separation of covariances by $r = s$ and $r \neq s$ implies that Ω consists of blocks of Ω_1 and Ω_2, each consisting of five nonzero terms for $r = s$ ($\eta = 1$) and $r \neq s$ ($\eta = 2$), respectively (Figure 1, right panel). We define an exchangeable covariance matrix as any covariance matrix of this form and denote it Ω_E.

4 Exchangeable estimator definition and evaluation

4.1 Exchangeable covariance estimator

Consider relational regression models with Ω of the exchangeable form Ω_E. The proposed exchangeable estimator of $\text{var}(\widehat{\beta} \mid X)$ is then

$$\widehat{V}_E = (X^T X)^{-1} X^T \widehat{\Omega}_E X (X^T X)^{-1}, \quad \widehat{\Omega}_E = \sum_{\eta=1}^{2} \sum_{u=0}^{d} \widehat{\phi}_u^{(\eta)} \mathcal{S}_u^{(\eta)}, \tag{6}$$

where $\mathcal{S}_u^{(\eta)}$ denotes the $\{Rn(n-1) \times Rn(n-1)\}$ binary matrix with 1's in the entries corresponding to relation pairs of type $(u = 0, a, b, c, d; \eta = 1, 2)$ as defined in Figure 1. We propose estimating the ten parameters in Ω by averaging the residual products that share the same index configurations, corresponding to (0)-(d) in Figure 1. For example, the estimate of $\text{cov}(\xi_{kls}, \xi_{ijr})$, corresponding to $u = b$ and $\eta = 2$, is

$$\widehat{\phi}_b^{(2)} = \binom{R}{2}^{-1} \frac{1}{n(n-1)(n-2)} \sum_{r \neq s} \sum_{i} \sum_{j \neq i} e_{ijr} \Big(\sum_{k \neq j} e_{iks} - e_{ijs} \Big). \tag{7}$$

The remaining nine estimators for $(s = 0, a, \ldots, e; \eta = 1, 2)$, are defined analogously, and $\widehat{\Omega}_E$ may be interpreted as the projection of $\widehat{\Omega}_{DC}$ into the vector space over symmetric matrices of the form of Ω_E. We provide theoretical details for the proposed exchangeable estimator, including asymtotic normality of least squares under exchangeability and consistency of $\widehat{V}_E$ in Sections S7- S9.

4.2 Comparison of exchangeable estimator with dyadic clustering

It is intuitive that the moment-based exchangeable estimator is consistent, and more efficient than the dyadic clustering estimator whenever the exchangeability assumption is satisfied. One might expect the highly parametrized dyadic clustering estimator to trade off high variance for reduced bias. However, we derive the result that the dyadic clustering estimator is biased downwards, and that this bias is larger than twice the bias of the exchangeable estimator. One concludes that one tradeoff for the robustness of the dyadic clustering estimator is anticonservatism. The proof of Theorem 2 is provided in Section S6. We provide additional theoretical details for the proposed exchangeable estimator (comparison of mean-square error with dyadic clustering) in Section S10.

Theorem 2. *Consider error vector ξ and normally distributed covariate vector x with mean zero, exchangeable covariance matrix, and bounded fourth moments, where*

$$y_{ij} = \beta_1 + x_{ij}\beta_2 + \xi_{ij}.$$

Then, the dyadic clustering estimator for var$(\widehat{\beta}_2)$ is biased downwards,

$$n^2 \textit{Bias}(\widehat{V}_{DC}) + O(n^{-1/2}) \leq -2n^2 |\textit{Bias}(\widehat{V}_E)| + O(n^{-1/2}) \leq 0,$$

noting that both $n^2 \textit{Bias}(\widehat{V}_{DC})$ and $|n^2 \textit{Bias}(\widehat{V}_E)|$ are $O(1)$.

We conducted a simulation study to compare the bias and 95% confidence interval coverage when using the exchangeable and dyadic clustering estimators. We simulated from a model with three covariates (one each of binary, positive real, and real-valued) with exchangeable and non-exchangeable error models, for $R = 1$ and $n = 20, 40, 80, 160, 320$ (see Section S11 for details). Figure 2 shows the estimated mean coverage and middle 95% of coverages across various X realizations. In all settings, the estimated mean coverage of the exchangeable estimator is closer to the nominal 0.95 level than the dyadic clustering estimator. This difference is most pronounced for the binary covariate, where there is reduced signal-to-noise relative to the other covariates. The average bias of the dyadic clustering estimator is typically more than four times that of the exchangeable estimator under the exchangeable error model, confirming Theorem 2 and driving the poorer coverage performance (see Section S11.2 for details).

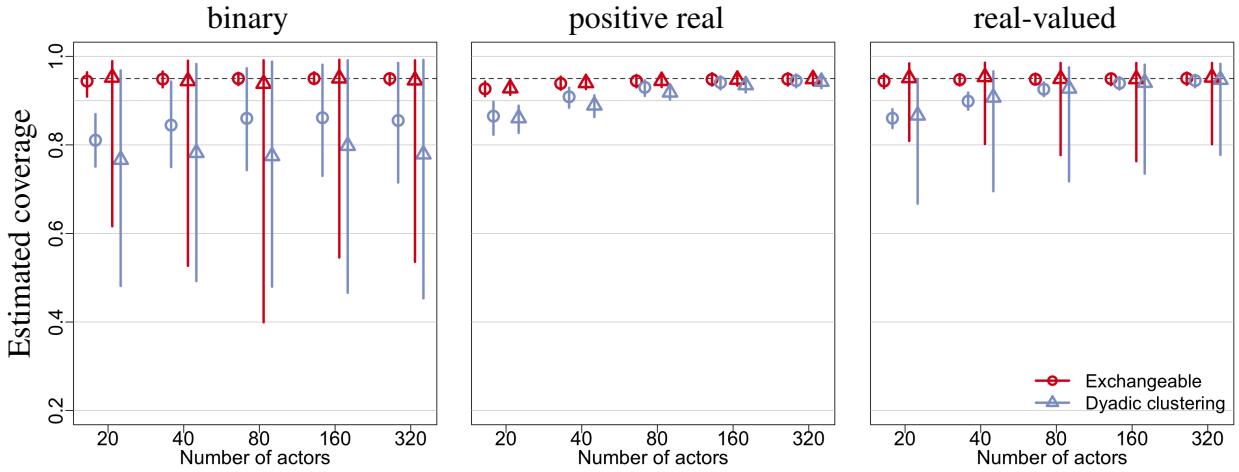

Figure 2: Estimated probability that true coefficient is in 95% confidence interval for each of three covariates (binary, positive, and real-valued) when the errors are generated from exchangeable (circles) and non-exchangeable (triangles) models. Points denote mean estimated coverage and lines represent the middle 95% of coverages for exchangeable (black) and dyadic clustering (gray) estimators.

5 Patterns in international trade

We demonstrate the implications of using our exchangeable estimator in a study of international trade among 58 countries over $R = T = 20$ years. These data were previously analyzed and

made available by Westveld and Hoff (2011). Following Westveld and Hoff (2011), Ward and Hoff (2007), and Tinbergen (1962), we use a modified gravity mean model to represent log yearly trade between each pair of countries as linear function of seven covariates in years 1981-2000. Westveld and Hoff (2011) propose a model – which we refer to as the mixed effects model – which explicitly decomposes the regression error term ϵ_{ijt} for each time period and pair of actors into time-dependent sender and receiver effects, resulting in 13 error covariance parameters which are estimated using a latent variable representation and Bayesian Markov chain Monte Carlo methodology. We propose estimating the gravity mean model using feasible generalized least squares, assuming the errors are jointly exchangeable. As noted in Section 4.1, the proposed approach estimates ten error covariance parameters. We provide a method for efficient inversion of Ω_E in this setting in Section S13.

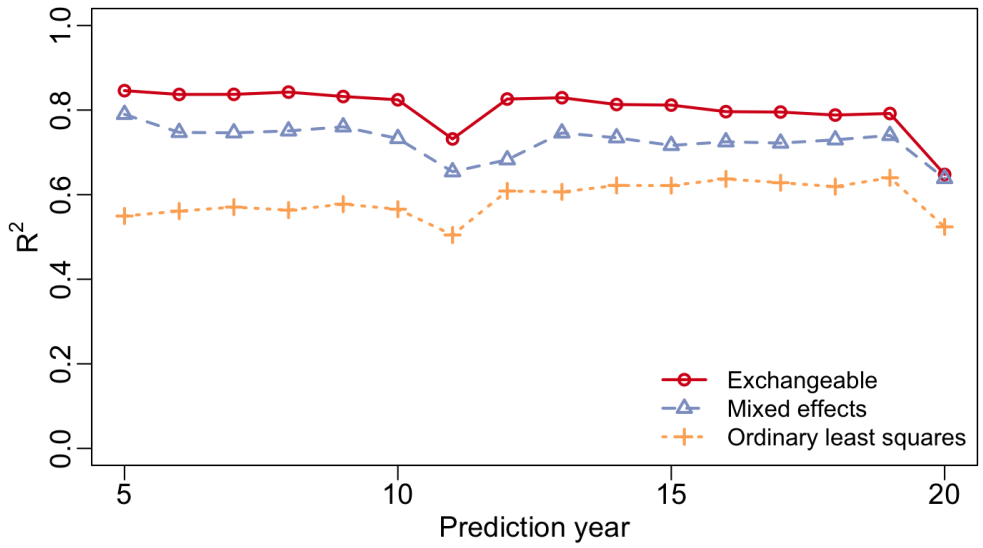

Figure 3: R^2 when predicting one-year-ahead trade flows using exchangeable, mixed effects, and ordinary least squares approaches.

We compared the exchangeable and mixed effects approaches (and ordinary least squares as a baseline) in an out-of-sample prediction study. Here we estimated the regression coefficients using the first K years of trade data for $K = 4, ..., 19$ and used the estimates to predict trade values in the following year. Figure 3 provides the coefficient of determination, R^2, for the three procedures when predicting trade flows in years 5 through 20. There is a median increase in R^2 of about 10% (30%) when using the proposed exchangeable approach relative to the mixed effects approach (and ordinary least squares, respectively). The proposed approach performs better than the other approaches for all time periods, although the gap in performance decreases as T increases. These results suggest the more parsimonious exchangeable approach represents the data better than the mixed effects model, and yet, the exchangeable approach runs in a small fraction of the time of the mixed effects approach. See Section S14 for additional details.

We compared the coefficients of the ordinary least squares, mixed effects, and exchangeable approaches. The R^2 between the ordinary least squares and mixed effects coefficients is about 0.46, while the exchangeable and mixed effects coefficients have an R^2 of 0.78. About 40% of

the ordinary least squares coefficients (using dyadic clustering standard errors) were significantly different from the mixed effects coefficients, while only 1% of the exchangeable coefficients were significantly different from the the mixed effects coefficients. Finally, the ordinary least squares coefficients have standard errors that are, on average, over 1.5 times the standard errors of the exchangeable coefficients (with exchangeable standard errors). See Section S14.3 for more details. Together with the one-year-ahead prediction results, the coefficient and standard error comparisons suggest that the proposed approach can revise ordinary least squares coefficients in the direction of a higher fidelity model, giving more precise estimates of the coefficients, while requiring few modeling decisions and with limited runtime penalty. Based on the success of the exchangeable approach in the trade data analysis and simulation study, we recommend that researchers use the feasible generalized least squares estimator of the coefficient vector β as demonstrated here (unless an unbiased estimator of the coefficient vector is specifically desired).

6 Discussion

The proposed exchangeable estimator leverages exchangeability for maximal symmetry (and thus maximal parsimony) in the covariance matrix of relational array Y. The exchangeability assumption may not be appropriate when the true error covariances are substantially heterogeneous. We propose using a permutation test based on the dyadic clustering estimator for testing the hypothesis of exchangeable errors. The procedure consists of generating a null distribution of $\widehat{V}_{DC}$ in (5) by randomly permuting the residual array in a manner consistent with exchangeability. If the observed estimator is extreme relative to the null distribution, this suggests the errors are non-exchangeable. Details and simulations are available in Section S15.

Supplementary material

The supplementary material contains details about the estimator and analyses, proofs, further theoretical details, and a proposed test for exchangeability.

References

Aitkin, A. (1935). On least squares and linear combination of observations. *Proc. R. Soc. Edinb.*, 55:42–48.

Aldous, D. J. (1981). Representations for partially exchangeable arrays of random variables. *J. Multivariate Anal.*, 11(4):581–598.

Aleskerov, F., Meshcheryakova, N., Rezyapova, A., and Shvydun, S. (2017). Network analysis of international migration. In Kalyagin, V. A., Nikolaev, A. I., Pardalos, P. M., and Prokopyev, O. A., editors, *Models, Algorithms, and Technologies for Network Analysis*, pages 177–185, Cham. Springer International Publishing.

Aronow, P. M., Samii, C., and Assenova, V. A. (2015). Cluster–robust variance estimation for dyadic data. *Political Anal.*, 23(4):564–577.

Attanasio, O., Barr, A., Cardenas, J. C., Genicot, G., and Meghir, C. (2012). Risk pooling, risk preferences, and social networks. *Am. Econ. J.*, 4(2):134–167.

Austern, M. and Orbanz, P. (2018). Limit theorems for invariant distributions. *arXiv:1806.10661*.

Banerjee, A., Chandrasekhar, A. G., Duflo, E., and Jackson, M. O. (2013). The diffusion of microfinance. *Science*, 341:363–373.

Bardham, P. (1984). *Land, labor and rural poverty*. Oxford University Press.

Bickel, P. J. and Chen, A. (2009). A nonparametric view of network models and Newman–Girvan and other modularities. *Proc. Natl. Acad. Sci. USA*, 106(50):21068–21073.

Bolthausen, E. (1982). On the central limit theorem for stationary mixing random fields. *Ann. Probab.*, 10(4):1047–1050.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2011). Robust inference with multiway clustering. *J. Bus. Econom. Statist.*, 29(2).

Cameron, A. C. and Miller, D. L. (2014). Robust inference for dyadic data. *Working paper, University of California-Davis*.

Conley, T. G. (1999). GMM estimation with cross sectional dependence. *J. Econometrics*, 92(1):1–45.

Cramér, H. and Wold, H. (1936). Some theorems on distribution functions. *J. Lond. Math. Soc.*, 1(4):290–294.

Fafchamps, M. (2006). Development and social capital. *J. Develop. Studies*, 42(7):1180–1198.

Fafchamps, M. and Gubert, F. (2007). The formation of risk sharing networks. *J. Dev. Econ.*, 83(2):326–350.

Fagiolo, G., Reyes, J., and Schiavo, S. (2008). On the topological properties of the world trade web: A weighted network analysis. *Phys. A*, 387(15):3868–3873.

Fortini, S., Ladelli, L., and Regazzini, E. (2012). Central limit theorem with exchangeable summands and mixtures of stable laws as limits. *arXiv:1204.4357*.

Fosdick, B. K. and Hoff, P. D. (2014). Separable factor analysis with applications to mortality data. *Ann. Appl. Stat.*, 8(1):120–147.

Foster, A. D. and Rosenzweig, M. R. (2001). Imperfect commitment, altruism, and the family: Evidence from transfer behavior in low-income rural areas. *Rev. Econ. Stat.*, 83(3):389–407.

Green, A. and Shalizi, C. R. (2017). Bootstrapping exchangeable random graphs. *arXiv:1711.00813*.

Greene, W. H. (2003). *Econometric analysis*. Prentice Hall, 5 edition.

Guyon, X. (1995). *Random fields on a network: Modeling, statistics, and applications*. Springer Science & Business Media.

Han, G., McCubbins, O., and Paulsen, T. (2016). Using social network analysis to measure student collaboration in an undergraduate capstone course. *NACTA Journal*, 60(2):176.

Hansen, B. E. (2015). *Econometrics*. Online at `http://www.ssc.wisc.edu/~bhansen/econometrics/`.

Hoff, P. D. (2005). Bilinear mixed-effects models for dyadic data. *J. Amer. Statist. Assoc.*, 100(469):286–295.

Hoff, P. D. (2008). Modeling homophily and stochastic equivalence in symmetric relational data. In *Adv. Neural Inf. Process Syst.*, pages 657–664.

Hoff, P. D. (2011). Separable covariance arrays via the Tucker product, with applications to multivariate relational data. *Bayesian Anal.*, 6(2):179–196.

Hoff, P. D. (2015). Multilinear tensor regression for longitudinal relational data. *Ann. Appl. Stat.*, 9(3):1169–1193.

Hoff, P. D., Raftery, A. E., and Handcock, M. S. (2002). Latent space approaches to social network analysis. *J. Amer. Statist. Assoc.*, 97(460):1090–1098.

Holland, P. W., Laskey, K. B., and Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Soc. Networks*, 5(2):109–137.

Hoover, D. N. (1979). Relations on probability spaces and arrays of random variables. *Preprint, Inst. for Adv. Study, Princeton, NJ*, 2.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proc. Fifth Berkeley Symp. on Math. Statist. and Prob.*, volume 1, pages 221–233.

Kallenberg, O. (2006). *Probabilistic Symmetries and Invariance Principles*. Springer Science & Business Media.

King, G. and Roberts, M. E. (2015). How robust standard errors expose methodological problems they do not fix, and what to do about it. *Polit. Anal.*, 23(2):159–179.

Li, H. and Loken, E. (2002). A unified theory of statistical analysis and inference for variance component models for dyadic data. *Statist. Sinica*, pages 519–535.

Lumley, T. and Hamblett, N. M. (2003). Asymptotics for marginal generalized linear models with sparse correlations. *UW Biostatistics Working Paper Series*.

Menzel, K. (2017). Bootstrap with clustering in two or more dimensions. *arXiv:1703.03043*.

Orbanz, P. and Roy, D. M. (2015). Bayesian models of graphs, arrays and other exchangeable random structures. *IEEE Trans. Pattern Anal. Mach. Intell.*, 37(2):437–461.

Tabord-Meehan, M. (2018). Inference with dyadic data: Asymptotic behavior of the dyadic-robust t-statistic. *Journal of Business & Economic Statistics*, pages 1–10.

Tinbergen, J. (1962). *Shaping the world economy: Suggestions for an international economic policy*. Twentieth Century Fund, New York.

Wakefield, J. (2013). *Bayesian and frequentist regression methods*. Springer Science & Business Media.

Wang, Y. J. and Wong, G. Y. (1987). Stochastic blockmodels for directed graphs. *J. Amer. Statist. Assoc.*, 82(397):8–19.

Ward, M. D. and Hoff, P. D. (2007). Persistent patterns of international commerce. *J. Peace Res.*, 44(2):157–175.

Westveld, A. H. and Hoff, P. D. (2011). A mixed effects model for longitudinal relational and network data, with applications to international trade and conflict. *Ann. Appl. Stat.*, pages 843–872.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, pages 817–838.

Zhang, B. and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.*, 4(1).

Supplementary material for "Regression of exchangeable relational arrays"

.

Contents Page No.

S1 Patterns of exchangeable covariance matrices

Figure S1 shows the structure of Ω_E for a relational matrix with four actors $\{A, B, C, D\}$ and $R = 1$.

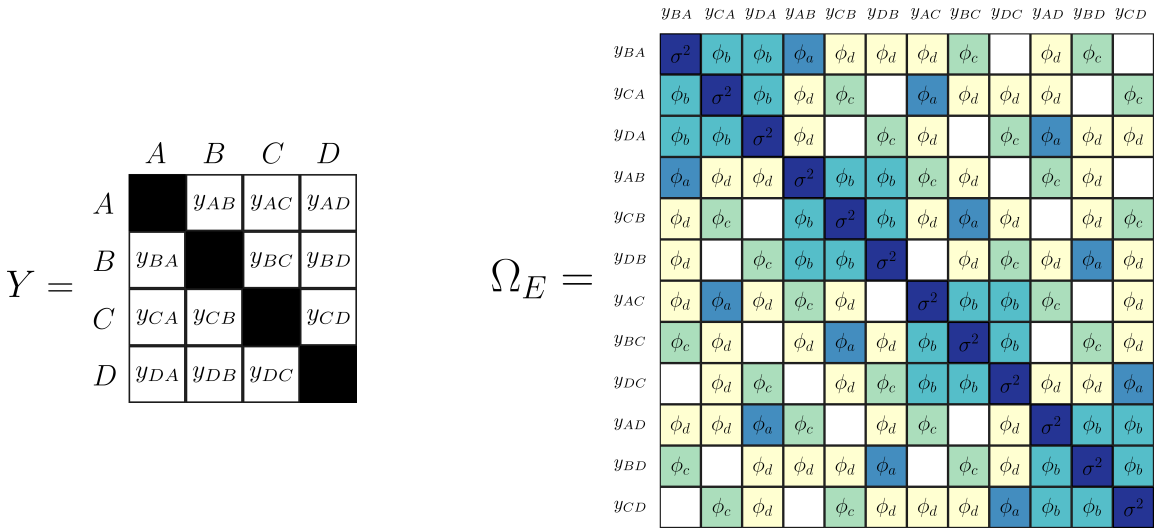

Figure S1: Consider a matrix Y containing the relations among four actors $\{A, B, C, D\}$ shown on the left. Since the relation between an actor and itself is undefined, the diagonal entries (blacked out in the picture) are not regarded as part of Y. Assuming joint exchangeability of the actors and that relations involving non-overlapping sets of actors are independent, the covariance matrix Ω_E contains five unique values.

S2 Exchangeable estimator details

The text gives an example for the estimator of $\phi_b^{(2)}$. We explicitly define the empirical mean estimates used in the exchangeable estimator in (6) below:

$$\widehat{\phi}_0^{(1)} = \frac{1}{Rn(n-1)} \sum_r \sum_i \sum_{j \neq i} e_{ijr}^2,$$

$$\widehat{\phi}_a^{(1)} = \frac{1}{Rn(n-1)} \sum_r \sum_i \sum_{j \neq i} e_{ijr} e_{jir},$$

$$\widehat{\phi}_b^{(1)} = \frac{1}{Rn(n-1)(n-2)} \sum_r \sum_i \sum_{j \neq i} e_{ijr} \Big(\sum_{k \neq i} e_{ikr} - e_{ijr} \Big),$$

$$\widehat{\phi}_c^{(1)} = \frac{1}{Rn(n-1)(n-2)} \sum_r \sum_i \sum_{j \neq i} e_{ijr} \Big(\sum_{k \neq j} e_{kjr} - e_{ijr} \Big),$$

$$\widehat{\phi}_d^{(1)} = \frac{1}{2Rn(n-1)(n-2)} \sum_r \sum_i \sum_{j \neq i} e_{ijr} \Big(\sum_{k \neq i} e_{kir} + \sum_{k \neq j} e_{jkr} - 2e_{jir} \Big),$$

$$\widehat{\phi}_0^{(2)} = \binom{R}{2}^{-1} \frac{1}{n(n-1)} \sum_{r \neq s} \sum_i \sum_{j \neq i} e_{ijr} e_{ijs},$$

$$\widehat{\phi}_a^{(2)} = \binom{R}{2}^{-1} \frac{1}{n(n-1)} \sum_{r \neq s} \sum_i \sum_{j \neq i} e_{ijr} e_{jis},$$

$$\widehat{\phi}_b^{(2)} = \binom{R}{2}^{-1} \frac{1}{n(n-1)(n-2)} \sum_{r \neq s} \sum_i \sum_{j \neq i} e_{ijr} \Big(\sum_{k \neq i} e_{iks} - e_{ijs} \Big),$$

$$\widehat{\phi}_c^{(2)} = \binom{R}{2}^{-1} \frac{1}{n(n-1)(n-2)} \sum_{r \neq s} \sum_i \sum_{j \neq i} e_{ijr} \Big(\sum_{k \neq j} e_{kjs} - e_{ijs} \Big),$$

$$\widehat{\phi}_d^{(2)} = \binom{R}{2}^{-1} \frac{1}{2n(n-1)(n-2)} \sum_{r \neq s} \sum_i \sum_{j \neq i} e_{ijr} \Big(\sum_{k \neq i} e_{kis} + \sum_{k \neq j} e_{jks} - 2e_{jis} \Big).$$

Recall from the text that $\widehat{\phi}_0^{(1)}$ is an estimator of var(ξ_{ijr}), $\widehat{\phi}_a$ is an estimator of cov(ξ_{ijr}, ξ_{jir}), $\widehat{\phi}_b^{(1)}$ is an estimator of cov(ξ_{ijr}, ξ_{kjr}), $\widehat{\phi}_c^{(1)}$ is an estimator of cov(ξ_{ijr}, ξ_{ikr}), and $\widehat{\phi}_d^{(1)}$ is an estimator of cov(ξ_{ijr}, ξ_{jkr}) (and equivalently an estimator of cov(ξ_{ijr}, ξ_{kir})). The estimator $\widehat{\phi}_i^{(2)}$, for $i = 0, a, b, c, d$, is the analogous estimator to $\widehat{\phi}_i^{(1)}$, only when $r \neq s$.

S3 Undirected arrays

This section specializes the results presented in the manuscript to undirected relational data. Consider the case when $R = 1$ and suppose the relational data contains the relations among n actors. The covariance of the errors Ω contains six possibly unique elements

$$\text{var}(\xi_{ijr}) = \theta_0^{(1)}, \quad \text{cov}(\xi_{ijr}, \xi_{kir}) = \theta_a^{(1)}, \quad \text{cov}(\xi_{ijr}, \xi_{klr}) = 0,$$
$$\text{cov}(\xi_{ijr}, \xi_{ijs}) = \theta_0^{(2)}, \quad \text{cov}(\xi_{ijr}, \xi_{kis}) = \theta_a^{(2)}, \quad \text{cov}(\xi_{ijr}, \xi_{kls}) = 0, \quad r \neq s.$$

As in the directed case, we assume covariances corresponding to relations which share no common actor are zero. We again propose estimating the remaining nonzero covariances using the corresponding means of residuals.

S4 Proof of Proposition 1

Proof. Consider a probability model for a directed relational array Ξ that satisfies the joint exchangeability and second moment criteria defined in the proposition. Now consider arbitrary permutations $\pi(\cdot)$ of the actor set $(1, ..., n)$ and $\nu(\cdot)$ of the actor set $(1, ..., R)$. Exchangeability implies the that the probability distribution of ξ_{ijr} is the same as $\xi_{\pi(i)\pi(j)\nu(r)}$. Thus, every entry in Ξ is marginally identically distributed, and

$$\text{var}(\xi_{ijr}) = \text{var}\{\xi_{\pi(i)\pi(j)\nu(r)}\}.$$

This is one of the entries in Ω, the covariance matrix of Ξ.

We now move onto the covariances in Ω; it remains to show that there are 11 unique entries beyond the variance. This result follows from two facts. The first fact is that exchangeability implies that the bivariate distribution of the pair (ξ_{ijr}, ξ_{klr}) must be the same as bivariate distribution of $\{\xi_{\pi(i)\pi(j)r}, \xi_{\pi(k)\pi(l)r}\}$, and thus,

$$\text{cov}(\xi_{ijr}, \xi_{klr}) = \text{cov}\{\xi_{\pi(i)\pi(j)r}, \xi_{\pi(k)\pi(l)r}\}.$$

The second fact is that permutations preserve equality and inequality, meaning that if $i = j$, then $\pi(i) = \pi(j)$, and if $i \neq j$, then $\pi(i) \neq \pi(j)$, for any permutation π. We will enumerate the 11 patterns of indices that have the same bivariate distributions, as implied by the equality-inequality preservation of permutations, being mindful of the fact that the first two indices in ξ_{ijr} refer to the same index set $(1, \ldots, n)$.

First take the case when $i = k$ and $j = l$. When $r = s$, we have the variance previously discussed. When $r \neq s$, we obtain the second set of equivalent bivariate distributions (after the variance):

$$(\xi_{ijr}, \xi_{ijs}), \;\; (i = 1, \ldots n; j = 1, \ldots, n; i \neq j; r = 1, \ldots, R; s = 1, \ldots, R; r \neq s),$$

where it is understood that $i \neq j$ due to the network setting. Moving on, consider the case when $i = k$ and $j \neq l$, noting that, by assumption, $i \neq j$ and $k \neq l$. Then, we obtain two sets with equivalent bivariate distributions, pertaining to $r = s$ and $r \neq s$:

$$(\xi_{ijr}, \xi_{ilr}), \;\; (i = 1, \ldots n; j = 1, \ldots, n; l = 1, \ldots, n; l \neq j; r = 1, \ldots, R),$$
$$(\xi_{ijr}, \xi_{ils}), \;\; (i = 1, \ldots n; j = 1, \ldots, n; l = 1, \ldots, n; l \neq j; r = 1, \ldots, R; s = 1, \ldots, R; r \neq s).$$

Thus far, we have enumerated four sets of equivalent bivariate distributions under exchangeability. The remaining eight may be determined by cycling through combinations of indices in (i, j) equal to, and not equal to, indices in (k, l), for both $r = s$ and $r \neq s$. Table S1 summarizes the twelve index patterns that result; the first and third row in the table have already been discussed. The second row, first column, states that pairs of entries in Ξ with the index pattern (ξ_{ijr}, ξ_{jir}) form a set with the same bivariate distribution.

Table S1: Twelve index patterns that form a set of equivalent bivariate distributions with the entry in Ξ corresponding to indices (i, j, r). The row labels correspond to those in Figure 1 and the column labels indicate index sets that share the same third dimension "slice" of Ξ with index set (i, j, r).

	$s = r$	$s \neq r$
(0)	(i, j, r)	(i, j, s)
(a)	(j, i, r)	(j, i, s)
(b)	(i, l, r)	(i, l, s)
(c)	(k, j, r)	(k, j, s)
(d)	(k, i, r)	(k, i, s)
(e)	(k, l, r)	(k, l, s)

Finally, one might at first think that we have missed the bivariate distribution of (ξ_{ijr}, ξ_{jlr}). However, we note that the bivariate distribution of random variables (A, B) is equal to that of (B, A), and thus, the bivariate distribution of (ξ_{ijr}, ξ_{jlr}) is captured in the first column of row (d) of Table S1. Since we have enumerated all entries in Ω, and there are twelve possibly unique sets with equivalent covariances (or variances), the result is shown for $R > 1$. When $R = 1$, it is sufficient to consider only the first column of Table S1, which establishes that there are at most six unique entries in Ω for $R = 1$. □

S5 Eigenvalues of exchangeable covariance matrix

S5.1 Positive definite $\widehat{\Omega}_E$

Since the entries in the exchangeable covariance matrix estimator $\widehat{\Omega}_E$ are empirical averages, it is possible the estimate $\widehat{\Omega}_E$ is not positive definite. Here we briefly investigate the constraints on the parameters that guarantee the resulting covariance matrix is positive definite when $R = 1$. Note that for computing the sandwich estimator variance of $\widehat{\beta}$ and making inference on $\widehat{\beta}$, positive definiteness of $\widehat{\Omega}_E$ is not necessary. However, if generalized least squares is employed to estimate β, the inverse of the covariance matrix estimator is required, and hence it is desirable that $\widehat{\Omega}_E$ be positive definite.

S5.2 Undirected relational data

We first focus on the undirected case, where the exchangeable covariance matrix contains two distinct nonzero entries: a variance σ^2 and a parameter ϕ in the off-diagonal representing the correlation between any pairs of relations that share an actor. Below we consider the correlation matrix, rather than the covariance matrix, which contains only a single nonzero correlation value. We denote this value by $a = \theta_a^{(1)}/\theta_0^{(1)}$.

Based on a thorough empirical investigation, we conjecture that the exchangeable correlation matrix corresponding to an undirected set of relations among n actors, which has nonzero value a in select off-diagonal entries, has exactly three eigenvalues as given below.

Eigenvalue	Multiplicity
$1 + 2(n-2)a$	1
$1 - 2a$	$\frac{1}{2}n(n-3)$
$1 + (n-4)a$	$n-1$

The correlation matrix is positive definite if and only if all eigenvalues are positive. Thus, if $a \in (1/(2(n-2)), 1/2)$, the correlation matrix is positive definite. Notice that the upper bound on a does not vary with n. Using the relation between a and $(\theta_0^{(1)}, \theta_a^{(1)})$, this constraint can be re-expressed as a constraint on the covariance parameters.

S5.3 Directed relational data

Since matrices of the form of Ω_E form a five-dimensional linear subspace over square symmetric matrices, it is intuitive that the eigenspace of such matrices is limited. We find empirically that square symmetric matrices of the form of Ω_E have five unique eigenvalues (when $R = 1$). As in the undirected case, we again focus on the exchangeable correlation matrix which contains four nonzero off-diagonal elements $\{a, b, c, d\}$ corresponding, respectively, in placement to $\{\phi_a^{(1)}, \phi_b^{(1)}, \phi_c^{(1)}, \phi_d^{(1)}\}$ in the exchange covariance matrix Ω_E. Note $a = \phi_a^{(1)}/\phi_0^{(1)}$, $b = \phi_b^{(1)}/\phi_0^{(1)}$, and so on. Based on empirical studies, we conjecture the eigenvalues for the exchangeable correlation matrix associated with a directed set of relations among n actors has exactly five eigenvalues as given below.

Eigenvalue(s)	Multiplicity
$1 + a + (n-2)(b+c) + 2(n-2)d$	1
$1 + a - (b + c + 2d)$	$(n-1)(n-2)/2 - 1$
$1 - (a + b + c) + 2d$	$(n-1)(n-2)/2$
$\{(n-3)(b+c) - 2d + 2\}/2 \pm (\alpha + \beta)^{1/2}/2$	$n-1$

where $\alpha = (c^2 + b^2)(n^2 - 2n + 1) + 4d^2(n^2 - 6n + 9) + 2bc(1 - n^2 + 2n)$ and $\beta = ad(8n - 24) + (b + c)d(12 - 4n) + 4a\{a - (b + c)\}$. As in the undirected case, these constraints can be re-expressed as constraints on the original five covariance parameters.

S6 Proof of Theorem 2: bias of the exchangeable estimator

S6.1 Notation and proof preliminaries

In this section we prove that the dyadic clustering estimator is biased downwards, and that this bias is greater in absolute value than the bias of the exchangeable estimator, in the case when $R = 1$.

First, we make some definitions. Recall that the dyadic clustering estimator is based on the outer product of residuals, with entries corresponding to pair of relations that do not share an actor set to zero. This operation may be viewed as a projection onto a vector space of square, symmetric matrices of dimension $n(n-1)$ with zeros in the appropriate entries. The dimension of this vector space is $O(n^3)$. We define the vector space below:

Definition S3. *Vector space $\mathcal{D}$: Let $M_{jk,lm}$ be the square matrix of dimension $n(n-1)$ with a '1' in the entry corresponding to relations jk and lm, and zeros elsewhere. Then,*

$$\mathcal{D} = \textit{span}\left\{M_{jk,lm} : (j,k) \cap (l,m) = \varnothing\right\}.$$

We define the orthogonal projection of matrix A onto $\mathcal{D}$ as $P_{\mathcal{D}}(A)$.

A similar definition to Definition S3 may be made concerning the space of covariance matrices corresponding to jointly exchangeable random variables, as defined below.

Definition S4. *Vector space $\mathcal{E}$: Let $\mathcal{S}_i$ be an indicator matrix of entries in square, symmetric matrices of dimension $n(n-1)$ corresponding to pairs of relations that share an actor in the ith manner, for $i = 1, \ldots, 5$. Then,*

$$\mathcal{E} = \textit{span}\left(\mathcal{S}_i : i = 1, \ldots, 5\right).$$

We define the orthogonal projection of matrix A onto $\mathcal{E}$ as $P_{\mathcal{E}}(A)$.

It will be useful to write the dyadic clustering and exchangeable estimators as functions of the projections in Definitions S3 and S4, supposing that X is orthogonal, that is, $X^T X = f(n)I$, where $f(n)$ is a function of the number of actors n. The dyadic clustering estimator of var$(\widehat{\beta}_k)$ is then

$$\begin{aligned}(\widehat{V}_{DC})_{kk} &= f(n)^{-2} X_k^T P_{\mathcal{D}} \left\{(y - X\widehat{\beta})(y - X\widehat{\beta})^T\right\} X_k, \\ &= f(n)^{-2} \text{tr}\left\{(y - X\widehat{\beta})(y - X\widehat{\beta})^T P_{\mathcal{D}}\left(X_k X_k^T\right)\right\},\end{aligned} \tag{S1}$$

where X_k is column k of covariate matrix X. The second equality in (S1) results from the application of a property of inner products, which we apply to the inner product over matrices, where, for any projection operator $P(.)$, we have that $\text{tr}\{AP(B)\} = \text{tr}\{P(A)B\} = \text{tr}\{P(A)P(B)\}$. An analogous definition to (S1) holds for the exchangeable estimator of var$(\widehat{\beta}_k)$ when replacing the projections onto $\mathcal{D}$ in (S1) with projections onto $\mathcal{E}$.

Finally, to prove Theorem 2, we need some relationships between the error covariance parameters $(\phi_3^{(1)}, \phi_4^{(1)}, \phi_5^{(1)})$ when Ω is positive definite. We summarize these relationships in the inequalities in the following proposition. We drop the superscript '(1)' to lighten notation.

Proposition S5. *Let $\Omega \in \mathcal{E}$ be an exchangeable correlation matrix, with $\Omega = I + \sum_{i=2}^{5} \phi_i \mathcal{S}_i$. Then,*

$$\begin{aligned} 0 \leq \phi_3 + \phi_4 &\leq 1, \\ \phi_3^2 + \phi_5^2 &\leq \phi_3, \\ \phi_4^2 + \phi_5^2 &\leq \phi_4. \\ \phi_3^2 + \phi_4^2 + 2\phi_5^2 &\leq 1, \end{aligned}$$

Proof. The inequalities result from the requirement that eigenvalues of Ω are positive for all n. Taking n arbitrarily large, the eigenvalues of Ω are

$$\begin{aligned} n(\phi_3 + \phi_4 + 2\phi_5) &\geq 0, \\ 1 + \phi_2 - \phi_3 - \phi_4 - 2\phi_5 &\geq 0, \\ 1 - \phi_2 - \phi_3 - \phi_4 + 2\phi_5 &\geq 0, \\ n\left[\phi_3 + \phi_4 \pm \left\{(\phi_3 - \phi_4)^2 + 4\phi_5^2\right\}^{1/2}\right] &\geq 0. \end{aligned}$$

The second and third eigenvalues give that $\phi_3 + \phi_4 \leq 1$. The fourth and fifth eigenvalues imply $\phi_3 + \phi_4 \geq 0$.

Using the fourth and fifth eigenvalue and rearranging,

$$\begin{aligned} 4\phi_5^2 + (\phi_3 - \phi_4)^2 &\leq (\phi_3 + \phi_4)^2 \\ 4\phi_5^2 &\leq 4\phi_3\phi_4. \end{aligned} \tag{S2}$$

Then, we have that

$$\begin{aligned} \phi_3^2 + \phi_5^2 &\leq \phi_3^2 + \phi_3\phi_4, \\ &\leq \phi_3^2 + \phi_3(1 - \phi_3) \\ &\leq \phi_3, \end{aligned} \tag{S3}$$

which establishes the second result. The third result follows from an analogous argument to (S3) for $\phi_4^2 + \phi_5^2$. The final result follows from summing the previous two, such that

$$\phi_3^2 + \phi_4^2 + 2\phi_5^2 \leq \phi_3 + \phi_4 \leq 1.$$

□

S6.2 Proof of Theorem 2

In this proof, we start by establishing the expression for the true variance of $\widehat{\beta}_2$. Then, we write the bias of the dyadic clustering and exchangeable estimators as a function of the variance parameters of x and ξ. Then, we write the bias of the dyadic clustering estimator as a function of the bias of the exchangeable estimator, and then bound the absolute value of the bias of the exchangeable estimator, which gives the result. Throughout, we assume that the bias of the exchangeable and dyadic clustering estimators, scaled by n^2, are $O(1)$. We verify this assumption at the end of the proof. Without loss of generality, we prove for the scaled problem where $E(xx^T) = \Psi = \sum_{i=1}^5 \psi_i \mathcal{S}_i$ with $\psi_1 = 1$.

We begin by expressing the true variance of $\widehat{\beta}_2$. We require that $n^{1/2}(\widehat{\beta}_2 - \beta_2)$ converges in distribution to a normal random variable with finite variance, which may be expected from Proposition 3.2 in Tabord-Meehan (2018). By assumption, $\widehat{\beta}_2 = (x^T x)^{-1} x^T y$. The true variance of $\widehat{\beta}_2$ is then

$$\begin{aligned} n\text{var}(\widehat{\beta}_2) &= n(x^T x)^{-2} x^T \Omega x, \\ &= O(n^{-3}) x^T \Omega x + O(n^{-1}) = O(1), \end{aligned}$$

where we use that $x^T x = n^2 + O(n^{-1})$ by assumption. Using (S1), it will be useful to rewrite the true variance

$$\begin{aligned} n\text{var}(\widehat{\beta}_2) &= n(x^T x)^{-2} \text{tr}\left\{\Omega P_{\mathcal{D}}\left(xx^T\right)\right\}, \\ &= n(x^T x)^{-2} \text{tr}\left\{\Omega P_{\mathcal{E}}\left(xx^T\right)\right\}, \end{aligned}$$

where we use the fact that Ω is a member of both spaces $\mathcal{D}$ and $\mathcal{E}$ by assumption.

We now write the expectation of the dyadic clustering estimator, taking only the leading terms. First, the expectation

$$E\{(y - X\widehat{\beta})(y - X\widehat{\beta})^T \mid X\} = (I - P_X)\Omega(I - P_X),$$

where P_X is the projection onto the column space of the covariate matrix X. By assumption, we have that

$$\begin{aligned} P_X &= \{n(n-1)\}^{-1} 11^T + (x^T x)^{-1} xx^T, \\ &= \{n(n-1)\}^{-1} \left\{11^T + (1 + O(n^{-1}))xx^T\right\}, \\ &= \{n(n-1)\}^{-1} \left(11^T + xx^T\right) + O(n^{-1}), \end{aligned}$$

where 1 is a vector of 1's of appropriate length (in this case, length $n(n-1)$). Using the expression of the dyadic clustering estimator in (S1), its expectation is

$$\begin{aligned} n^2 E\{(\widehat{V}_{DC})_{22}\} &= n^2 (x^T x)^{-2} \text{tr}\Big\{\Big(I - P_X\Big)\Omega\Big(I - P_X\Big) P_{\mathcal{D}}\left(xx^T\right)\Big\}, \\ &= n^{-2}\text{tr}\Big\{\Big(I - P_X\Big)\Omega\Big(I - P_X\Big) P_{\mathcal{D}}\left(xx^T\right)\Big\} + O(n^{-1}), \\ &= n^2 \text{var}(\widehat{\beta}_2) + n^{-2}\text{tr}\Big\{P_X \Omega P_X P_{\mathcal{D}}\left(xx^T\right)\Big\} - 2n^{-2}\text{tr}\Big\{\Omega P_X P_{\mathcal{D}}\left(xx^T\right)\Big\} + O(n^{-1}), \\ &= n^2 \text{var}(\widehat{\beta}_2) + n^{-6}\text{tr}\Big\{11^T \Omega xx^T P_{\mathcal{D}}\left(xx^T\right)\Big\} - 2n^{-4}\text{tr}\Big\{\Omega 11^T P_{\mathcal{D}}\left(xx^T\right)\Big\} \ldots \qquad \text{(S4)} \\ &\quad + n^{-6}\text{tr}\Big\{xx^T \Omega xx^T P_{\mathcal{D}}\left(xx^T\right)\Big\} - 2n^{-4}\text{tr}\Big\{\Omega xx^T P_{\mathcal{D}}\left(xx^T\right)\Big\} \ldots \\ &\quad + n^{-6}\text{tr}\Big\{11^T \Omega 11^T P_{\mathcal{D}}\left(xx^T\right)\Big\} + O(n^{-1}). \end{aligned}$$

The second term in (S4) is zero in probability, as the row sums of Ω are all the same, and are $O(n)$. Thus,

$$1^T \Omega x = O(n) 1^T x.$$

Then, the second term is

$$n^{-6}\text{tr}\Big\{11^T \Omega xx^T P_{\mathcal{D}}\left(xx^T\right)\Big\} = (n^{-2} 1^T x)(n^{-3} x^T P_{\mathcal{D}}\left(xx^T\right) 1) = O(n^{-1}), \qquad \text{(S5)}$$

since $n^{-2}1^T x = O(n-1)$ by assumption, which implies that the second term vanishes relative to the rest.

The third term in (S4) may be written

$$-2n^{-4}\text{tr}\Big\{\Omega 11^T P_{\mathcal{D}}\left(xx^T\right)\Big\} = -2n^{-4} c_n 1^T P_{\mathcal{D}}\left(xx^T\right) 1,$$

where $c_n = O(n)$ is the row sum of Ω, which again is the same for all rows. Further, using the Cauchy-Scharz inequality,

$$1^T P_{\mathcal{D}}\left(xx^T\right) 1 = \sum_{jk,lm \in \Theta_0} x_{jk} x_{lm} \leq \left(\sum_{jk,lm \in \Theta_0} x_{jk}^2 x_{lm}^2\right)^{1/2} = ||P_{\mathcal{D}}\left(xx^T\right)||_F.$$

Substituting, the third term is

$$-2n^{-3}\text{tr}\Big\{\Omega 11^T P_{\mathcal{D}}\left(xx^T\right)\Big\} = n^{-3}||P_{\mathcal{D}}\left(xx^T\right)||_F.$$

We now bound the order of $||P_{\mathcal{D}}\left(xx^T\right)||_F$. By assumption, $x^T x = n(n-1) + O(n^{-1})$, and thus

$$||P_{\mathcal{D}}\left(xx^T\right)||_F^2 = \sum_{jk,lm\in\Theta_0} x_{jk}^2 x_{lm}^2 = O(n^3).$$

Using this bound, the third term is

$$-2n^{-4}\text{tr}\Big\{\Omega 11^T P_{\mathcal{D}}\left(xx^T\right)\Big\} = O(n^{-3/2}), \tag{S6}$$

and the third term vanishes relative to the other terms in (S4). Further, the sixth term in (S4) is of the same order,

$$n^{-6}\text{tr}\Big\{11^T\Omega 11^T P_{\mathcal{D}}\left(xx^T\right)\Big\} = n^4 c_n \text{tr}\Big\{11^T P_{\mathcal{D}}\left(xx^T\right)\Big\} = O(n^{-3/2}), \tag{S7}$$

and may also be neglected.

Using (S5), (S6), and (S7), the expression for the bias of the dyadic clustering estimator based on (S4) becomes

$$n^2\text{Bias}\left\{(\widehat{V}_{DC})_{22}\right\} = n^{-6}\text{tr}\Big\{xx^T\Omega xx^T P_{\mathcal{D}}\left(xx^T\right)\Big\} - 2n^{-4}\text{tr}\Big\{\Omega xx^T P_{\mathcal{D}}\left(xx^T\right)\Big\} + O(n^{-1}). \tag{S8}$$

Since, by the properties of projections, $||P_{\mathcal{E}}\left(xx^T\right)||_F \leq ||P_{\mathcal{D}}\left(xx^T\right)||_F$ since $\mathcal{E} \subset \mathcal{D}$, we may write the bias of the exchangeable estimator by analogy:

$$n^2\text{Bias}\left\{(\widehat{V}_{E})_{22}\right\} = n^{-6}\text{tr}\Big\{xx^T\Omega xx^T P_{\mathcal{E}}\left(xx^T\right)\Big\} - 2n^{-4}\text{tr}\Big\{\Omega xx^T P_{\mathcal{E}}\left(xx^T\right)\Big\} + O(n^{-1}). \tag{S9}$$

We now explicitly determine the value of the exchangeable bias in (S9). By assumption,

$$|\Theta_i|^{-1} x^T \mathcal{S}_i x = \psi_i + O(n^{-1/2}), \quad (i = 1, \ldots, 5),$$

recalling that $E(xx^T) = \Psi = \sum_{i=1}^5 \psi_i \mathcal{S}_i$. Then, the projection of xx^T onto the exchangeable space is

$$P_{\mathcal{E}}\left(xx^T\right) = \sum_{i=1}^5 \psi_i \mathcal{S}_i + O(n^{-1/2}), \tag{S10}$$

and the scaled true variance is

$$n\text{var}(\widehat{\beta}_2) = n^{-3} x^T \Omega x = \sum_{i=1}^5 \frac{|\Theta_i|}{n^3}\psi_i\phi_i + O(n^{-1/2}). \tag{S11}$$

Using (S10) and (S11), the first term in the exchangeable bias in (S9) is

$$n^{-6}\text{tr}\Big\{xx^T\Omega xx^TP_{\mathcal{E}}\left(xx^T\right)\Big\} = \left(n^{-3}x^T\Omega x\right)\left\{n^{-3}x^TP_{\mathcal{E}}\left(xx^T\right)x\right\},$$
$$= \left(\sum_{i=1}^{5}\frac{|\Theta_i|}{n^3}\psi_i\phi_i\right)\left(\sum_{j=1}^{5}\frac{|\Theta_j|}{n^3}\psi_j^2\right) + O(n^{-1/2}).$$

The second term in the exchangeable bias in (S9) is

$$-2n^{-4}\text{tr}\Big\{\Omega xx^TP_{\mathcal{E}}\left(xx^T\right)\Big\} = -2n^{-4}\sum_{i=1}^{5}\sum_{j=1}^{5}\phi_i\psi_jx^T\mathcal{S}_i\mathcal{S}_jx.$$

Now we must determine $n^{-4}x^T\mathcal{S}_i\mathcal{S}_jx$ for each i and j. Recalling that $\mathcal{S}_i$ is a matrix that indicates pairs of relations jk and lm that share an actor in the i^{th} manner,

$$x^T\mathcal{S}_i\mathcal{S}_jx = \sum_{(rs,tu)\in\Theta_i}\sum_{(rs,ab)\in\Theta_j}x_{tu}x_{ab}. \tag{S12}$$

Then, the sums (S12) may be determined by examining the relationships between the actors in relations tu and ab. First, for any i and j, tu must share at least one actor with ab, as the expectation $E(x_{tu}x_{ab})$ is zero otherwise. This result implies that every sum in (S12) is at most $O(n^4)$. As an example, we consider $i = 3$ and $j = 3$. In this case, tu is of the form ru and ab is of the form rb. Whenever $b \neq u$, the expectation $E(x_{ru}x_{rb}) = \psi_3$. When $b = u$, the expectation $E(x_{ru}x_{rb}) = \psi_1$, although there are an order of magnitude fewer of these terms. Finally, similar to the expectation argument, the variance of the sums in (S12) consists of at most $O(n^6)$ finite terms, and thus tends to zero. Then, the sum tends to the limit of its expectation. Summarizing, we have that

$$n^{-4}\sum_{(rs,tu)\in\Theta_3}\sum_{(rs,ab)\in\Theta_3}x_{tu}x_{ab} = n^{-4}\sum_{(rs,tu)\in\Theta_3}\sum_{(rs,ab)\in\Theta_3}E(x_{tu}x_{ab}) + O(n^{-1/2}),$$
$$= \frac{n}{2}n^{-4}|\Theta_3|\psi_3 + n^{-3}|\Theta_1|\psi_1 + O(n^{-1/2}),$$
$$= \psi_3 + O(n^{-1/2}).$$

Similar counting operations give the following result, for $i \leq j$,

$$n^{-4}x^T\mathcal{S}_i\mathcal{S}_jx = \begin{cases}\psi_3 + O(n^{-1/2}), & i = j = 3\\ \psi_4 + O(n^{-1/2}), & i = j = 4\\ \psi_3 + \psi_4 + O(n^{-1/2}), & i = j = 5\\ \psi_5 + O(n^{-1/2}), & i = 3,\ j = 5\\ \psi_5 + O(n^{-1/2}), & i = 4,\ j = 5\\ O(n^{-1/2}) & \text{otherwise.}\end{cases}$$

Substituting, we have the following expression for the bias of the exchangeable estimator

$$n^2\text{Bias}\left\{(\widehat{V}_E)_{22}\right\} = (\phi_3\psi_3 + \phi_4\psi_4 + 2\phi_5\psi_5)(\psi_3^2 + \psi_4^2 + 2\psi_5^2)\ldots \tag{S13}$$
$$- 2\left\{\phi_3(\psi_3^2 + \psi_5^2) + \phi_4(\psi_4^2 + \psi_5^2) + 2\phi_5\psi_5(\psi_3 + \psi_4)\right\} + O(n^{-1/2}).$$

We now return to the bias of the dyadic clustering estimator in (S8). The first term in the bias of the dyadic clustering estimator is

$$n^{-6}\text{tr}\Big\{xx^T\Omega xx^T P_{\mathcal{D}}\left(xx^T\right)\Big\} = \left(\sum_{i=1}^{5}\frac{|\Theta_i|}{n^3}\psi_i\phi_i\right)\left\{n^{-3}x^T P_{\mathcal{D}}\left(xx^T\right)x\right\} + O(n^{-1/2}),$$
$$= \left(\sum_{i=1}^{5}\frac{|\Theta_i|}{n^3}\psi_i\phi_i\right)\left(n^{-3}\sum_{i=1}^{5}\sum_{jk,lm\in\Theta_i}x_{jk}^2x_{lm}^2\right) + O(n^{-1/2}). \quad \text{(S14)}$$

As with the exchangeable estimator, the sum in (S14) converges to the limit of its expectation, as the variance tends to zero. Then, using Iserlis' theorem,

$$n^{-3}\sum_{i=1}^{5}\sum_{jk,lm\in\Theta_i}x_{jk}^2x_{lm}^2 = n^{-3}\sum_{i=1}^{5}\sum_{jk,lm\in\Theta_i}E(x_{jk}^2x_{lm}^2) + O(n^{-1/2}),$$
$$= n^{-3}\sum_{i=1}^{5}|\Theta_i|(1+2\psi_i^2) + O(n^{-1/2}),$$

using the assumption that $\psi_1 = 1$. Then, the first term in the bias of the dyadic clustering estimator in (S8) is

$$n^{-6}\text{tr}\Big\{xx^T\Omega xx^T P_{\mathcal{D}}\left(xx^T\right)\Big\} = \left(\sum_{i=1}^{5}\frac{|\Theta_i|}{n^3}\psi_i\phi_i\right)\left(4+2\sum_{j=1}^{5}\frac{|\Theta_j|}{n^3}\psi_j^2\right) + O(n^{-1/2}).$$

We now turn to the second term in the bias of the dyadic clustering estimator in (S8).

$$-2n^{-4}x^T P_{\mathcal{D}}\left(xx^T\right)\Omega x = -2n^{-4}\sum_{i=1}^{5}\sum_{j=1}^{5}\phi_i x^T\mathcal{S}_i\left(xx^T\circ\mathcal{S}_j\right)x, \quad \text{(S15)}$$

where '$\circ$' represents element-wise multiplication. As with the exchangeable estimator, the sum in (S15) has variance that tends to zero, and converges to its expectation. Writing in sum form and using Isserlis' theorem,

$$n^{-4}x^T\mathcal{S}_i\left(xx^T\circ\mathcal{S}_j\right)x = n^{-4}\sum_{rs,tu\in\Theta_i}\sum_{rs,ab\in\Theta_j}E(x_{rs}x_{tu}x_{ab}^2) + O(n^{-1/2}),$$
$$= n^{-4}\sum_{rs,tu\in\Theta_i}\sum_{rs,ab\in\Theta_j}E(x_{rs}x_{tu}) + 2E(x_{ab}x_{rs})E(x_{ab}x_{tu}) + O(n^{-1/2}),$$
$$= n^{-3}|\Theta_j|\psi_i + 2n^{-4}\sum_{rs,tu\in\Theta_i}\sum_{rs,ab\in\Theta_j}E(x_{ab}x_{rs})E(x_{ab}x_{tu}) + O(n^{-1/2}),$$
$$= n^{-3}|\Theta_j|\psi_i + 2n^{-4}\sum_{rs,tu\in\Theta_i}\sum_{rs,ab\in\Theta_j}\psi_j E(x_{ab}x_{tu}) + O(n^{-1/2}),$$
$$= n^{-3}|\Theta_j|\psi_i + 2n^{-4}\psi_j x^T\mathcal{S}_i\mathcal{S}_j x + O(n^{-1/2}).$$

Substituting into (S15), the second term in the bias of the dyadic clustering estimator is

$$-2n^{-4}x^T P_{\mathcal{D}}\left(xx^T\right)\Omega = -8\left(\sum_{i=1}^{5}\frac{|\Theta_i|}{n^3}\psi_i\phi_i\right) - 4n^{-4}\sum_{i=1}^{5}\sum_{j=1}^{5}\phi_i\psi_j x^T\mathcal{S}_i\mathcal{S}_j x + O(n^{-1/2}). \quad \text{(S16)}$$

Then, combining (S14) and (S16), we obtain the following expression for the bias of the dyadic clustering estimator,

$$n^2\text{Bias}\left\{(\widehat{V}_{DC})_{22}\right\} = \left(\sum_{i=1}^{5}\frac{|\Theta_i|}{n^3}\psi_i\phi_i\right)\left(-4+2\sum_{j=1}^{5}\frac{|\Theta_j|}{n^3}\psi_j^2\right) - 4n^{-4}\sum_{i=1}^{5}\sum_{j=1}^{5}\phi_i\psi_j x^T\mathcal{S}_i\mathcal{S}_j x + O(n^{-1/2}),$$

wherein we recognize the bias of the exchangeable estimator, and

$$n^2\text{Bias}\left\{(\widehat{V}_{DC})_{22}\right\} = 2n^2\text{Bias}\left\{(\widehat{V}_{E})_{22}\right\} - 4\left(\sum_{i=1}^{5}\frac{|\Theta_i|}{n^3}\psi_i\phi_i\right) + O(n^{-1/2}). \tag{S17}$$

Since we have Ω positive definite and

$$0 \geq -4n^{-3}x^T\Omega x = -4\left(\sum_{i=1}^{5}\frac{|\Theta_i|}{n^3}\psi_i\phi_i\right) + O(n^{-1/2}).$$

Thus, we have shown that the bias of the dyadic clustering estimator is less than twice the bias of the exchangeable estimator. To establish that this same property holds for the absolute values of the biases, it remains to show that $n^2\text{Bias}\{(\widehat{V}_E)_{22}\} \leq x^T\Omega x + O(n^{-1/2})$, that is, that the right hand side of (S17) is negative and bounded away from zero. Combining terms in (S13), the bias of the exchangeable estimator is

$$\begin{aligned} n^2\text{Bias}\left\{(\widehat{V}_{E})_{22}\right\} = {} & \phi_3\psi_3\{\psi_3^2+\psi_4^2+2\psi_5^2-2\psi_3(1+\psi_5^2/\psi_3^2)\}\ldots \\ & \phi_4\psi_4\{\psi_3^2+\psi_4^2+2\psi_5^2-2\psi_4(1+\psi_5^2/\psi_4^2)\}\ldots \\ & 2\phi_5\psi_5\{\psi_3^2+\psi_4^2+2\psi_5^2-2(\psi_3+\psi_4)\}+O(n^{-1/2}). \end{aligned}$$

Now, using Proposition S5, we have that

$$\begin{aligned} \psi_3^2+\psi_4^2+2\psi_5^2-2\psi_3(1+\psi_5^2/\psi_3^2) &\leq \psi_3^2+\psi_4^2+2\psi_5^2 \leq 1, \\ \psi_3^2+\psi_4^2+2\psi_5^2-2\psi_4(1+\psi_5^2/\psi_4^2) &\leq \psi_3^2+\psi_4^2+2\psi_5^2 \leq 1, \\ \psi_3^2+\psi_4^2+2\psi_5^2-2(\psi_3+\psi_4) &\leq \psi_3^2+\psi_4^2+2\psi_5^2 \leq 1, \end{aligned} \tag{S18}$$

and thus,

$$\begin{aligned} n^2\text{Bias}\left\{(\widehat{V}_{E})_{22}\right\} &\leq (\phi_3\psi_3+\phi_4\psi_4+2\phi_5\psi_5)+O(n^{-1/2}), \\ &\leq n^{-3}x^T\Omega x + O(n^{-1/2}). \end{aligned} \tag{S19}$$

To establish a lower bound on the bias of the exchangeable estimator, using (S2) and (S3) in Proposition S5,

$$\begin{aligned} \psi_3\psi_4 &\geq \psi_5^2, \\ \psi_3+\psi_4 &\geq \psi_3+\psi_5^2/\psi_3, \\ -(\psi_3+\psi_4) &\leq -(\psi_3+\psi_5^2/\psi_3), \end{aligned} \tag{S20}$$

and a similar argument holds for $-\psi_4(1+\psi_5^2/\psi_4^2)$. Then, recalling from Proposition S5 that

$$2\psi_5^2 \geq \frac{1}{2}\psi_3^2 + \frac{1}{2}\psi_4^2 + \psi_3\psi_4,$$

and combining with (S20), we obtain the following bound on the key expressions in (S18),

$$\begin{aligned}\psi_3^2 + \psi_4^2 + 2\psi_5^2 - 2\psi_3(1+\psi_5^2/\psi_3^2) &\geq \psi_3^2 + \psi_4^2 + 2\psi_5^2 - 2(\psi_3+\psi_4), \\ &\geq (\psi_3+\psi_4)^2 - 2(\psi_3+\psi_4) + \frac{1}{2}\psi_3^2 + \frac{1}{2}\psi_4^2 - \psi_3\psi_4, \\ &\geq -1.\end{aligned} \tag{S21}$$

To obtain the final inequality in (S21), we use the fact that $0 \leq \phi_3 + \phi_4 \leq 1$ to find that

$$(\psi_3+\psi_4)^2 - 2(\psi_3+\psi_4) \geq -1,$$

and that

$$\frac{1}{2}\psi_3^2 + \frac{1}{2}\psi_4^2 - \psi_3\psi_4 \geq 0.$$

A similar argument to that in (S21) can be obtained starting with $\psi_3^2+\psi_4^2+2\psi_5^2-2\psi_4(1+\psi_5^2/\psi_4^2)$. Then, combining (S21) with the complete expression for the bias in (S18) and the upper bound in (S19),

$$n^2\left|\text{Bias}\left\{(\widehat{V}_E)_{22}\right\}\right| \leq n^{-3}x^T\Omega x + O(n^{-1/2}). \tag{S22}$$

Substituting the bound on the absolute value of the exchangeable bias (S22) into (S17), we have that

$$n^2\text{Bias}\left\{(\widehat{V}_{DC})_{22}\right\} \leq -2n^{-3}x^T\Omega x + O(n^{-1/2}) \leq -2n^2\left|\text{Bias}\left\{(\widehat{V}_E)_{22}\right\}\right| + O(n^{-1/2}), \tag{S23}$$

which establishes the result. To complete the proof, we note that by the explicit expression for $n^2\text{Bias}\{(\widehat{V}_E)_{22}\}$ in (S13), the exchangeable bias scaled by n^2 is $O(1)$. Similarly, $n^2\text{Bias}\{(\widehat{V}_{DC})_{22}\}$ is $O(1)$ by (S17). □

S6.3 A corollary to Theorem 2

In the following corollary we expand the conditions of the bias theorem slightly, such that the error array is not strictly in the exchangeable, similar to Corollary S12.

Corollary S6. *Theorem 2 holds under the relaxation exchangeability of ξ, where the first two entries in the covariance matrix of ξ are allowed to be heterogenous, that is, $E(\xi_{jk}^2) = \sigma_{jk}^2$ and $E(\xi_{jk}\xi_{kj}) = \theta_{jk}$, where $|\theta_{jk}| < \sigma_{jk}^2 < C < \infty$. Similarly, the conditions hold when $E(x_{jk}x_{jk}^T) = A_{jk}$ and $E(x_{jk}x_{kj}^T) = D_{jk}$, where the entries in matrices A_{jk} and D_{jk} are bounded.*

The result follows directly from noting that the proof of Theorem 2 depends only on the values of $E(x_{jk}x_{lm})$ and $E(\xi_{jk}\xi_{lm})$ for $(jk,lm) \in \Theta_i$ for $i \in \{3,4,5\}$, at least up to a vanishing constant. Thus, so long as the expectations $E(x_{jk}x_{lm})$ and $E(\xi_{jk}\xi_{lm})$ for $(jk,lm) \in \Theta_i$ for $i \in \{1,2\}$ are bounded, Theorem 2 holds. Corollary S6 states that even under heteroskedasticity in the covariance matrix of the error vector ξ, the exchangeable estimator still outperforms the dyadic clustering estimator in bias. Only when the entries in Ω corresponding to configurations (b), (c), and (d) are heterogeneous does the theory fail to hold.

S7 Theoretical properties of the exchangeable estimator

In this section, we provide additional theoretical results for the proposed exchangeable estimator. Namely, we show that the proposed exchangeable estimator is consistent and provides an improvement in mean-square error over the dyadic clustering estimator, when the assumed model is correct. First, we provide conditions under which the theory holds, and prove asymptotic normality of ordinary least squares (which is necessary for the results that follow). For brevity, we focus on the ordinary least squares coefficient estimator and the case of a singe time period or array observation, where $R = 1$.

S7.1 Conditions for theory

We define the conditions under which we establish the theory, starting with a formal definition of the class of exchangeable covariance matrices.

Definition S7. *An exchangeable covariance matrix is defined as $\Omega_E = E[\xi\xi^T]$ arising from mean-zero random vector $\xi = \mathrm{vec}(\Xi)$, where Ξ is a jointly exchangeable random matrix with entries ξ_{ij} independent ξ_{kl} whenever $\{i,j\} \cap \{k,l\} \neq \varnothing$.*

For the theoretical assessment of the exchangeable estimator, we take X random, but still evaluate $\widehat{\beta}$, $\widehat{V}_{DC}$, and $\widehat{V}_E$ conditional on X. We assume the rows of the matrix X are jointly exchangeable, meaning that a reordering of the rows $\{x_{jk}\}_{j,k=1}^n$ to $\{x_{\pi(j)\pi(k)}\}_{j,k=1}^n$ leaves the distribution of matrix X invariant for any permutation $\pi(.)$. As with the dependence in the errors, we assume that two rows of X that correspond to relations which do not share an actor are independent, that is row x_{ij}^T is independent row x_{kl}^T whenever $\{i,j\} \cap \{k,l\} = \varnothing$. This dependence in the rows of matrix X (along with some assumptions on the finiteness of its moments) implies the following:

$$E(x_{jk}x_{lm}^T) = M_i, \quad (jk,lm) \in \Theta_i, \quad (i = 0,a,b,c,d), \tag{S24}$$

where Θ_i is the set of pairs of relations (jk,lm) that share a member in the ith manner and '0' refers to self-relation (i.e. variance).

The theoretical setting is as follows:

Condition S8. *Define the following data generating process:*

(A1) The true data generating model is $y = X\beta + \xi$, where the errors ξ are mean-zero with exchangeable covariance matrix as in Definition S7.

(A2) At least one of $\{\phi_b, \phi_c, \phi_d\}$ is nonzero.

In addition, consider the following regularity conditions:

(B1) The covariate matrix X has rows that are jointly exchangeable with at least one of $\{M_i\}_{i\in\{b,c,d\}}$ in (S24) *nonzero, and where row x_{ij}^T is independent row x_{kl}^T whenever $\{i,j\} \cap \{k,l\} \neq \varnothing$.*

(B2) The fourth moments of the errors and the eight moments of the covariates are bounded: $E(|\xi_{jk}|^4) < C < \infty$ and $\max_{l\in\{1,2,\ldots,p\}} E(|x_{jk}^{(l)}|^8) < C' < \infty$ where $x_{jk} = [x_{jk}^{(1)}, \ldots, x_{jk}^{(p)}]^T$.

(B3) The errors Ξ and covariates X are independent.

(B4) X is full rank.

S7.2 Additional theoretical results

The following theorem establishes asymptotic normality of $\widehat{\beta}$, and supports the normal approximation to $\widehat{\beta}$ to produce confidence intervals.

Theorem S9. *Under Conditions S8,*

$$n^{1/2}(\widehat{\beta} - \beta) \to_d \mathrm{N}(0, V_0\big),$$
$$V_0 = E(x_{jk}x_{jk}^T)^{-1}\big\{\phi_b E(x_{jk}x_{jl}^T) + \phi_c E(x_{jk}x_{lk}^T) + 2\phi_d E(x_{jk}x_{lk}^T)\big\}E(x_{jk}x_{jk}^T)^{-1}, \quad (j \neq k \neq l),$$

where '$\to_d$' denotes convergence in distribution.

Theorem S9 is similar to Proposition 3.2 in Tabord-Meehan (2018), although with unbounded errors. It may be seen as an extension of central limit theorems for sums over infinitely exchangeable arrays (Fortini et al., 2012) to regression scenarios, and belongs to a broader literature on distributions of random processes that satisfy a symmetry property (Kallenberg, 2006; Austern and Orbanz, 2018).

We establish consistency of the proposed estimator, and its improvement in performance over the dyadic clustering estimator, in the following three theorems.

Theorem S10. *Under Conditions S8, the exchangeable covariance estimator is consistent in the sense that*

$$n\widehat{V}_E - n\mathit{var}(\widehat{\beta}) \to_p 0, \quad n \to \infty,$$

where '$\to_p$' denotes convergence in probability.

Theorem S11. *Under Conditions S8, the mean-square error of the exchangeable estimator is less than that of the dyadic clustering estimator with probability approaching 1, that is*

$$\mathit{pr}\left[E\left\{\left(\widehat{V}_E - \mathit{var}(\widehat{\beta} \mid X)\right)^2 \mid X\right\} \leq E\left\{\left(\widehat{V}_{DC} - \mathit{var}(\widehat{\beta} \mid X)\right)^2 \mid X\right\}\right] \to 1.$$

S7.3 Corollary to theory

The proofs of Theorem S9 through Theorem S11 depend only on the uniformity of the covariance entries in Ω corresponding to ϕ_b, ϕ_c, and ϕ_d. Then, we can relax the assumption of exchangeability of the error vector ξ to allow some heterogeneity, namely heteroskedasticity, under which the theory still holds.

Corollary S12. *Theorem S9 through Theorem S11 hold under the relaxation of (A2) and (B1), where the first two entries in the covariance matrix of ξ are allowed to be heterogenous, that is, $E(\xi_{jk}^2) = \sigma_{jk}^2$ and $E(\xi_{jk}\xi_{kj}) = \theta_{jk}$, where $|\theta_{jk}| < \sigma_{jk}^2 < C < \infty$. Similarly, the conditions hold when $E(x_{jk}x_{jk}^T) = A_{jk}$ and $E(x_{jk}x_{kj}^T) = D_{jk}$, where the entries in matrices A_{jk} and D_{jk} are bounded.*

S8 Proof of asymptotic normality of ordinary least squares

S8.1 Notation and outline of proof

For this proof, and throughout the remainder of the supplementary material, we adopt slightly different notation to simplify the representation of the exchangeable covariance estimator. Recall that the exchangeable covariance estimator for the OLS estimating equations is

$$\widehat{V}_E = (X^T X)^{-1} X^T \widehat{\Omega}_E X (X^T X)^{-1},$$

where $\widehat{\Omega}_E$ is the exchangeable estimate of the error covariance matrix, consisting of five averages of residual products. Here we express $\widehat{\Omega}_E$ as

$$\widehat{\Omega}_E = \sum_{i=1}^{5} \widehat{\phi}_i \mathcal{S}_i, \quad \widehat{\phi}_i = \frac{\sum_{(jk,\ell m)\in\Theta_i} e_{jk} e_{\ell m}}{|\Theta_i|}, \quad (i = 1, \ldots, 5). \tag{S25}$$

This amounts to mapping $\sigma^2 \mapsto \phi_1$, $\phi_a \mapsto \phi_2$, ..., $\phi_d \mapsto \phi_5$, and re-indexing the $\mathcal{S}$ and M matrices accordingly. Additionally, when we consider sequences of jointly exchangeable random variables $\{W_{ij}\}_{i,j=1}^n$, it is understood that the sequence arises from a relational array such that entries with $i = j$ are undefined. Thus, sums over the sequence are of $n(n-1)$ terms and we define $\sum_{ij} W_{ij} = \sum_{i\neq j} W_{ij}$.

We work in the asymptotic regime where actors are added incrementally to the relational data set, i.e. n is continually increasing. To establish asymptotic normality of $\widehat{\beta}$, we wish to show

$$n^{1/2}(\widehat{\beta} - \beta) \to_d \mathrm{N}\big(0, M_1^{-1}\big[\phi_3 M_3 + \phi_4 M_4 + 2\phi_5 M_5\big] M_1^{-1}\big),$$

where $\{M_i\}_{i\in\{1,3,4,5\}}$ are as in (S24) and '$\to_d$' denotes element-wise convergence in distribution.

The motivation for the proof argument follows from the expression

$$n^{1/2}(\widehat{\beta} - \beta) = \left(\frac{\sum_{jk} x_{jk} x_{jk}^T}{n(n-1)}\right)^{-1} \frac{n^{1/2} \sum_{jk} x_{jk} \xi_{jk}}{n(n-1)}. \tag{S26}$$

We state that $\left(\frac{\sum_{jk} x_{jk} x_{jk}}{n(n-1)}\right)^{-1}$ converges in probability to M_1^{-1}, and then show asymptotic normality of the second multiplicative term in (S26).

In analyzing $\{x_{ij}\xi_{ij}\}_{i,j=1}^n$, by condition (B1), the joint exchangeability and independence of non-overlapping pairs of the sequence $\{\xi_{ij}\}_{i,j=1}^n$ extends to the component sequences in the vectors $\{x_{ij}\xi_{ij}\}_{i,j=1}^n$. Thus, to prove asymptotic normality of $\widehat{\beta}$, we first prove a theorem stating that the average of a mean-zero sequence of jointly exchangeable random variables is asymptotically normal. Specifically, for $\{W_{ij}\}_{i,j=1}^n$ mean zero and jointly exchangeable, we show

$$k_n \frac{\sum_{ij} W_{ij}}{\sigma} \to_d \mathrm{N}(0, 1) \tag{S27}$$

for some normalizing constant σ and fixed sequence $k_n \to 0$ as $n \to \infty$.

To prove (S27), we rely on a result from Bolthausen (1982), as well as a supporting lemma which we present here. Below we outline the significance of these results in the proof.

- Lemma S13 (Bolthausen (1982)): Provides a sufficient condition for asymptotic normality of a sequence of measures based on the standard normal characteristic function.
- Lemma S14: Provides a bound for a variance that surfaces in the proof of asymptotic normality in (S27).

From (S27), we immediately have the marginal asymptotic normality of the sample mean of each of the vector components in the sequence $\{x_{ij}\xi_{ij}\}_{i,j=1}^{n}$. To establish joint asymptotic normality, we employ the Cramér-Wold device (Cramér and Wold, 1936), where asymptotic normality of $\{v^T x_{ij}\xi_{ij}\}_{i,j=1}^{n}$, for all $v \in \mathbb{R}^p$ with $||v|| = 1$, establishes joint normality. To achieve the asymptotic normality of this inner product, we simply recognize that this inner product is itself the mean of an exchangeable sequence of random variables. Joint asymptotic normality of the mean of the sequence of vectors $\{x_{ij}\xi_{ij}\}_{i,j=1}^{n}$ establishes joint asymptotic normality of $\widehat{\beta}$ via (S26).

S8.2 Lemmas and theorem in support of Theorem S9

The following is Lemma 2 in Bolthausen (1982) and provides a sufficient condition for asymptotic normality. We abuse notation slightly, letting i be the imaginary unit where appropriate.

Lemma S13 (Bolthausen (1982))**.** *Let ν_n be a sequence of probability distributions over $\mathbb{R}$ which satisfies*

1. $\sup_n \int x^2 d\nu_n(x) < \infty$*, and*
2. *for all* $\lambda \in \mathbb{R}$, $\lim_n \int (i\lambda - x)\mathrm{e}^{i\lambda x} d\nu_n(x) = 0$.

Then, $\nu_n \to_d \mathrm{N}(0,1)$.

To provide intuition for Lemma S13, the integral in condition (2) is identically zero when ν_n is the standard normal distribution.

The next lemma provides a sufficient condition on the dependence structure in $\{W_{ij}\}_{i,j=1}^{n}$ necessary for the proof of asymptotic normality in (S27). Reecall that terms in $\{W_{ij}\}_{i,j=1}^{n}$ with $i = j$ are undefined.

Lemma S14. *Let $\{W_{ij}\}_{i,j=1}^{n}$ be a sequence of jointly exchangeable random variables as in Definition S7 with $||W_{ij}||_4 < L < \infty$, where $||W_{ij}||_p := E\left(|W_{ij}|^p\right)^{1/p}$ for $p > 0$. Then,*

$$\frac{1}{n^6} var\left(\sum_{ij}\sum_{kl\in\Theta_{ij}} W_{ij}W_{kl}\right) < \frac{CL^4}{n} \to 0 \text{ as } n \to \infty,$$

for some $C < \infty$, where Θ_{ij} is the set of ordered pairs (k,l) that share an index with (i,j).

Proof. By definition we write

$$\frac{1}{n^6}\mathrm{var}\left(\sum_{ij}\sum_{kl\in\Theta_{ij}} W_{ij}W_{kl}\right) = \frac{1}{n^6}\sum_{ij}\sum_{kl\in\Theta_{ij}}\sum_{rs}\sum_{tu\in\Theta_{rs}} \mathrm{cov}(W_{ij}W_{kl}, W_{rs}W_{tu}). \tag{S28}$$

Each covariance of (S28) is bounded by L^4. To bound the variance, we will show the number of nonzero entries in the sum is $O(n^5)$. For $\text{cov}(W_{ij}W_{kl}, W_{rs}W_{tu}) \neq 0$, there must be overlap between the index sets $\{i,j,k,l\}$ and $\{r,s,t,u\}$. Further, the sum in (S28) is taken over index sets that themselves contain overlap, i.e. $\{i,j\} \cap \{k,l\} \neq \varnothing$ and $\{r,s\} \cap \{t,u\} \neq \varnothing$. For example, the index sets $\{i,j,i,l\}$ and $\{i,s,i,u\}$ have nonzero covariance in (S28). Since there are 5 unique indices in the union of the sets $\{i,j,i,l\}$ and $\{i,s,i,u\}$, there are $O(n^5)$ such index set pairs of this form in total. There are 96 pairs of index sets that result in nonzero covariance $\text{cov}(W_{ij}W_{kl}, W_{rs}W_{tu})$. For example, another such pair of index sets is $\{i,j,i,l\}$ and $\{i,j,i,j\}$. Each of these 96 pairs of index sets is $O(n^5)$. Thus, the sum of covariances in (S28) is over $O(n^5)$ bounded elements. □

It is worth noting that we repeat the counting argument in the proof of Lemma S14 in many of the following proofs, including those in later sections. Now that we have Lemma S13 and S14, we prove that a general sequence of mean-zero exchangeable random variables is asymptotically normal.

Theorem S15. *Let $\{W_{ij}\}_{i,j=1}^n$ be a mean-zero sequence of jointly exchangeable random variables with at least one of $\{\phi_3, \phi_4, \phi_5\}$ nonzero. If $||W_{ij}||_4 < L < \infty$, then*

$$\frac{n^{1/2}\sum_{ij} W_{ij}}{n(n-1)} \to_d \text{N}(0, \phi_3 + \phi_4 + 2\phi_5) \text{ as } n \to \infty. \tag{S29}$$

Proof. We first show that $\phi_3 + \phi_4 + 2\phi_5$ is the correct limiting variance. Writing the variance of the expression on the left hand side of (S29) explicitly and recalling that entries such that $i = j$ are undefined, we see

$$\begin{aligned}
\text{var}\left(\frac{n^{1/2}}{n(n-1)}\sum_{ij} W_{ij}\right) &= \frac{n}{n^2(n-1)^2}\sum_{ij}\sum_{kl\in\Theta_{ij}} \text{cov}\left(W_{ij}, W_{kl}\right) \\
&= \frac{n^2(n-1)\left(\phi_1+\phi_2\right) + n^2(n-1)(n-2)\left(\phi_3+\phi_4+2\phi_5\right)}{n^2(n-1)^2} \\
&\to \phi_3 + \phi_4 + 2\phi_5 \text{ as } n \to \infty,
\end{aligned}$$

by the properties of joint exchangeability of $\{W_{ij}\}_{i,j=1}^n$ as described in Section 3.3. This variance is finite and nonzero by assumption. To prove (S29), it is sufficient to show

$$\bar{S}_n := \frac{\sum_{ij} W_{ij}}{n^{3/2}(\phi_3 + \phi_4 + 2\phi_5)^{1/2}} \to_d N(0,1). \tag{S30}$$

Define the limiting variance as $\sigma_n^2 = n^3(\phi_3 + \phi_4 + 2\phi_5)$ and the sum $S_n = \sum_{ij} W_{ij}$.

To establish (S30), we employ Lemma S13, where ν_n is the probability measure corresponding to $\bar{S}_n$ for all n. The first condition of Lemma S13 is satisfied since

$$E((\bar{S}_n)^2) = \frac{\text{var}\left(\sum_{ij} W_{ij}\right)}{n^3(\phi_3 + \phi_4 + 2\phi_5)} < CL^2$$

for $C < \infty$ and all n. Thus, to prove (S30), it is sufficient to show the second condition of Lemma S13: for all $\lambda \in \mathbb{R}$, as $n \to \infty$,

$$E\left((i\lambda - \bar{S}_n)\mathrm{e}^{i\lambda \bar{S}_n}\right) \to 0. \tag{S31}$$

We decompose the term in the expectation as in Guyon (1995) and Lumley and Hamblett (2003):

$$\begin{aligned}
(i\lambda - \bar{S}_n)\mathrm{e}^{i\lambda \bar{S}_n} &= A_1 - A_2 - A_3, \\
A_1 &= i\lambda \mathrm{e}^{i\lambda \bar{S}_n}\left(1 - \frac{1}{\sigma_n^2}\sum_{ij} W_{ij} S_{ij,n}\right), \\
A_2 &= \frac{\mathrm{e}^{i\lambda \bar{S}_n}}{\sigma_n}\sum_{ij} W_{ij}\left(1 - i\lambda \bar{S}_{ij,n} - \mathrm{e}^{-i\lambda \bar{S}_{ij,n}}\right), \\
A_3 &= \frac{1}{\sigma_n}\sum_{ij} W_{ij}\mathrm{e}^{i\lambda(\bar{S}_n - \bar{S}_{ij,n})}, \\
S_{ij,n} &= \sum_{kl \in \Theta_{ij}} W_{kl},
\end{aligned}$$

where $\bar{S}_{ij,n} = S_{ij,n}/\sigma_n$. To satisfy (S31) it remains to be shown that $\lim_{n\to\infty} E(A_m) = 0$ for each $m \in \{1, 2, 3\}$.

Beginning with A_1, $|\mathrm{e}^{i\lambda \bar{S}_n}| \leq 1$. Using this fact and Lemma S14,

$$\begin{aligned}
0 \leq E(|A_1|)^2 \leq E(|A_1^2|) &\leq \lambda^2 E\left(\left|1 - \frac{1}{\sigma_n^2}\sum_{ij} W_{ij} S_{ij,n}\right|^2\right) \\
&= \frac{\lambda^2}{\sigma_n^4}\mathrm{var}\left(\sum_{ij}\sum_{kl\in\Theta_{ij}} W_{ij}W_{kl}\right) + \lambda^2\left\{1 - \frac{\mathrm{var}\left(\sum_{ij} W_{ij}\right)}{\sigma_n^2}\right\}^2 \\
&\leq \lambda^2 \frac{CL^4}{n} + \lambda^2\left(1 - \frac{\sigma_n^2 + O(n^{-1})}{\sigma_n^2}\right)^2 \\
&= \lambda^2\left(\frac{CL^4}{n} + \frac{O\left(n^{-2}\right)}{\sigma_n^2}\right) \to 0
\end{aligned}$$

for all real-valued λ. $E(|A_1|)^2$ limiting to zero implies $E(|A_1|)$ limits to zero, and hence $E(A_1)$ limits to zero.

Now, for A_2, by Taylor expansion of $\mathrm{e}^{-i\lambda \bar{S}_{ij,n}}$, we can write

$$\left|1 - i\lambda \bar{S}_{ij,n} - \mathrm{e}^{-i\lambda \bar{S}_{ij,n}}\right| \leq c\lambda^2\left(\bar{S}_{ij,n}\right)^2,$$

for some $0 < c < \infty$ and all n, λ. Using this bound and the fact that $|\Theta_{ij}| = 4n - 6$, we evaluate

$E(|A_2|)$ directly below:

$$\begin{aligned} E(|A_2|) &\leq \frac{1}{\sigma_n} E\left(\sum_{ij} |W_{ij}| \left| 1 - i\lambda \bar{S}_{ij,n} - \mathrm{e}^{-i\lambda \bar{S}_{ij,n}} \right| \right), \\ &\leq \frac{c\lambda^2}{\sigma_n^3} \sum_{ij} E\left(|W_{ij}| \, S_{ij,n}^2 \right), \\ &\leq \frac{c\lambda^2}{\sigma_n^3} n(n-1)(4n-6)^2 L^3 \to 0, \end{aligned}$$

for all real λ. As $E(|A_2|)$ limits to zero, so does $E(A_2)$.

Finally, for A_3, the expression $S_{ij,n}$ sums all terms in the sequence $\{W_{ij}\}_{i,j=1}^n$ that depend upon W_{ij}, including W_{ij} itself. Thus, W_{ij} and $\bar{S}_n - \bar{S}_{ij,n}$ are independent. It follows immediately that

$$E\left(\frac{1}{\sigma_n} \sum_{ij} W_{ij} \mathrm{e}^{i\lambda(\bar{S}_n - \bar{S}_{ij,n})} \right) = \frac{1}{\sigma_n} \sum_{ij} E\big(W_{ij}\big) E\big(\mathrm{e}^{i\lambda(\bar{S}_n - \bar{S}_{ij,n})}\big) = 0,$$

since $E\left(W_{ij}\right) = 0$ for all ordered pairs (i, j).

Hence, $\lim_{n\to\infty} E(A_m) = 0$ for each $m \in \{1, 2, 3\}$ and we have the convergence in (S31), implying $\bar{S}_n \to_d \mathrm{N}(0, 1)$ by Lemma S13, which gives the desired result in (S29). □

S8.3 Proof of Theorem S9

We begin by writing

$$n^{1/2}(\widehat{\beta} - \beta) = \left(\frac{\sum_{jk} x_{jk} x_{jk}^T}{n(n-1)} \right)^{-1} \frac{n^{1/2} \sum_{jk} x_{jk} \xi_{jk}}{n(n-1)}, \tag{S32}$$

again emphasizing that entries in the sum with $j = k$ are undefined and omitted. Addressing the first multiplicative term in (S32), we recall that the inverse map is continuous. Then, by (S24) and the continuous mapping theorem, we have

$$\left(\frac{\sum_{jk} x_{jk} x_{jk}^T}{n(n-1)} \right)^{-1} \to_p M_1^{-1}. \tag{S33}$$

We now analyze the second multiplicative term in (S32). Showing asymptotic normality of this term is sufficient to show asymptotic normality of the expression on the left hand side of (S32). Recall $x_{jk}^T = [x_{jk}^{(1)}, x_{jk}^{(2)}, \ldots, x_{jk}^{(p)}]$. We wish to show that the sum of vectors

$$U_n := \frac{n^{1/2}}{n(n-1)} \sum_{jk} x_{jk} \xi_{jk} \to_d \mathrm{N}(0, \Sigma), \tag{S34}$$

for some limiting variance Σ. By the Cramér-Wold device (Cramér and Wold, 1936), U_n is asymptotically normal with asymptotic variance Σ if and only if $v^T U_n$ is asymptotically normal with

asymptotic variance $v^T \Sigma v$ for every vector $v \in \mathbb{R}^p$ such that $||v|| = 1$. Clearly,

$$v^T U_n := \frac{n^{1/2}}{n(n-1)} \sum_{jk} \widetilde{x}_{jk} \xi_{jk},$$

where we define $\widetilde{x}_{jk} = v^T x_{jk}$. We wish to apply Theorem S15 to the sequence $\{\widetilde{x}_{jk}\xi_{jk}\}_{j,k=1}^n$. First, the condition of finite moments in (B2) of Theorem S9 and $||v|| = 1$ implies that $||\widetilde{x}_{jk}\xi_{jk}||_4 < L$ for some finite $L < \infty$. Secondly, by the independence of X and Ξ in (B3) of Theorem S9, the sequence $\{\widetilde{x}_{jk}\xi_{jk}\}_{j,k=1}^n$ is a mean-zero exchangeable sequence of scalar random variables. Taking the variance directly,

$$\text{var}\left(\sum_{jk} \widetilde{x}_{jk}\xi_{jk}\right) = n^3 v^T \{\phi_1 M_1 O(n^{-1}) + \phi_2 M_2 O(n^{-1}) + \phi_3 M_3 + \phi_4 M_4 + 2\phi_5 M_5\} v. \quad \text{(S35)}$$

Then, we apply Theorem S15 with $\sigma_n^2 = \text{var}(\sum_{jk} \widetilde{x}_{jk}\xi_{jk})$ in (S35), which gives that

$$v^T U_n \to_d \text{N}(0, v^T[\phi_3 M_3 + \phi_4 M_4 + 2\phi_5 M_5]v).$$

Thus, by the Cramér-Wold device, we get the desired joint asymptotic normality

$$\frac{n^{1/2} \sum_{jk} x_{jk}\xi_{jk}}{n(n-1)} \to_d \text{N}\big(0, \phi_3 M_3 + \phi_4 M_4 + 2\phi_5 M_5\big). \quad \text{(S36)}$$

Combining the convergence in probability in (S33) and the asymptotic normality of (S36), we obtain the desired result. □

S9 Proof of consistency of the exchangeable estimator

S9.1 Notation and outline of proof

For the proof of the consistency of the exchangeable estimator $\widehat{V}_E$, we adopt the same change in notation as in Section S8, defined in (S25). We deviate slightly in that we denote Θ_i to denote dyadic pairs (j,k) and (l,m) that share a member in the ith manner. For example, for $i = 3$ we must have $j = l$ and $m \neq k$. We use the same assumptions as in Theorem S9.

This proof is outlined as follows. We initially prove that the exchangeable estimator $\widehat{V}_E$ is consistent if the exchangeable parameter estimates $\{\widehat{\phi}_i : i = 1, \ldots, 5\}$ are consistent for the true parameters. We then prove consistency of $\{\widehat{\phi}_i\}$ in two steps: (a) we show parameter estimates $\{\widetilde{\phi}_i\}$ based on the unobserved true errors Ξ are consistent and then (b) we show that the parameter estimates $\{\widehat{\phi}_i\}$ are asymptotically equivalent to $\{\widetilde{\phi}_i\}$. We require the consistency of $\widehat{\beta}$ result (implied by Theorem S9) for this last step.

S9.2 Proof of Theorem S10: consistency of the exchangeable estimator

First, from Theorem S9, the order of convergence of $\widehat{\beta}$ is $n^{1/2}$. Thus, we choose the rate n as our asymptotic regime for consistency of $\widehat{V}_E$. We wish to show that

$$n\widehat{V}_E - n\text{var}(\widehat{\beta}) \to_p 0. \quad \text{(S37)}$$

First we show that to prove consistency of $\widehat{V}_E$, it is sufficient to prove the consistency of the parameter estimates $\{\widehat{\phi}_i\}$ for the true parameters. We begin by writing the difference of variances $n\widehat{V}_E - n\text{var}(\widehat{\beta})$ in (S37), as

$$\begin{aligned}
&n(X^TX)^{-1}X^T\big(\widehat{\Omega}_E - \Omega_E\big)X(X^TX)^{-1} \\
&= \frac{n}{n^2(n-1)^2}\left\{\frac{X^TX}{n(n-1)}\right\}^{-1}\left\{\frac{X^T\sum_{i=1}^5|\Theta_i|\left(\widehat{\phi}_i-\phi_i\right)\mathcal{S}_iX}{|\Theta_i|}\right\}\left\{\frac{X^TX}{n(n-1)}\right\}^{-1} \\
&= \sum_{i=1}^5\frac{|\Theta_i|}{n(n-1)^2}\left(\widehat{\phi}_i-\phi_i\right)\left\{\frac{X^TX}{n(n-1)}\right\}^{-1}\left(\frac{\sum_{(jk,\ell m)\in\Theta_i}x_{jk}x_{\ell m}^T}{|\Theta_i|}\right)\left\{\frac{X^TX}{n(n-1)}\right\}^{-1} \\
&= \sum_{i=1}^5 c_{i,n}\left(\widehat{\phi}_i-\phi_i\right)h_{i,n}(X), \qquad \text{(S38)}
\end{aligned}$$

where $c_{i,n} = |\Theta_i|/n(n-1)^2$ and $h_{i,n}(X)$ contains the remaining terms which are functions of X. By the counting argument used to show Lemma S14, each $|\Theta_i|$ is at most $O(n^3)$, so each $c_{i,n} \to d_i$ for some finite constant d_i. Namely, $d_i = 0$ for $i \in \{1,2\}$, $d_i = 1$ when $i \in \{3,4\}$, and $d_i = 2$ for $i = 5$. To obtain the result in (S37), it is sufficient then to show $\widehat{\phi}_i - \phi_i \to_p 0$ and $h_{i,n}(X)$ converges in probability to some constant for all i. The latter comes easily, that is, by assumption and Slutsky's theorem,

$$h_{i,n}(X) \to_p M_1^{-1}M_iM_1^{-1}, \;\; i \in \{1,\ldots,5\}.$$

The continuous mapping theorem allows us to take the probability limit of $X^TX/(n(n-1))$ before inversion, as the inversion map is continuous.

We now consider consistency of the parameter estimates $\widehat{\phi}_i$. First, define error averages $\{\widetilde{\phi}_i : i = 1,\ldots,5\}$ analogous to the parameter estimates, such that for each i,

$$\widetilde{\phi}_i = \frac{1}{|\Theta_i|}\sum_{(jk,\ell m)\in\Theta_i}\xi_{jk}\xi_{\ell m}.$$

We will show $\widetilde{\phi}_i - \phi_i$ converges in probability to zero, and then do the same for $\widehat{\phi}_i - \widetilde{\phi}_i$. This is sufficient for showing $\widehat{\phi}_i - \phi_i \to_p 0$ as $\widehat{\phi}_i - \phi_i = (\widehat{\phi}_i - \widetilde{\phi}_i) + (\widetilde{\phi}_i - \phi_i)$.

To show convergence in probability of $\widetilde{\phi}_i - \phi_i$ to zero, we use the argument that the bias and variance both tend to zero. By assumption (A1), $E[\xi_{jk}\xi_{\ell m}] = \phi_i$ for every relation pair $(jk,\ell m) \in \Theta_i$. Thus, $E[\widetilde{\phi}_i - \phi_i] = 0$ for all n and $i \in \{1,\ldots,5\}$. We now turn to the variance:

$$\text{var}\left(\widetilde{\phi}_i\right) = \frac{1}{|\Theta_i|^2}\sum_{(jk,\ell m)\in\Theta_i}\sum_{(rs,tu)\in\Theta_i}\text{cov}\big(\xi_{jk}\xi_{\ell m},\xi_{rs}\xi_{tu}\big).$$

We again make a counting argument similar to that in Lemma (S14). By condition (B2), each of the $|\Theta_i|^2$ covariances in the sum above are bounded. The covariance between $\xi_{jk}\xi_{\ell m}$ and $\xi_{rs}\xi_{tu}$ is nonzero only if there is overlap between their two index sets. This reduces the number of nonzero covariances from the maximum possible $|\Theta_i|^2$ by a factor of at least n. Again, consider the case

of $i = 3$ where $|\Theta_3| = O(n^3)$. Each pair of relations in Θ_3 must be of the form (jk, jm), and thus the second set of indices must be of the form (js, ju), for example, for the covariance to be nonzero. The set of indices $\{j, k, j, m, j, s, j, u\}$ is of order $O(n^5) = |\Theta_3|^2 n^{-1}$. There are other forms of relation pairs in the second sum that give rise to nonzero covariance, such as (ks, ku) and so on. However, there are nine such forms, each of which is $O(n^5)$. Thus, the number of nonzero covariances is $O(n^5)$, and hence, we have

$$\text{var}\left(\widetilde{\phi}_i\right) = \frac{|\Theta_i|^2 O(n^{-1})}{|\Theta_i|^2} \to 0. \tag{S39}$$

This same argument holds for all i, and thus, we have the desired consistency: $\widetilde{\phi}_i - \phi_i \to_p 0$ for $i = 1, \ldots, 5$.

We now show that $\widehat{\phi}_i - \widetilde{\phi}_i$ converges in probability to zero. We first write the expression in terms of the estimated coefficients $\widehat{\beta}$:

$$\begin{aligned}\widehat{\phi}_i \; - \; \widetilde{\phi}_i &= \frac{\sum_{(jk,\ell m)\in\Theta_i} e_{jk}e_{\ell m} - \xi_{jk}\xi_{\ell m}}{|\Theta_i|} \\ &= \frac{1}{|\Theta_i|} \sum_{(jk,\ell m)\in\Theta_i} \left\{(\beta - \widehat{\beta})^T (x_{jk}x_{\ell m}^T)(\beta - \widehat{\beta}) - (\beta - \widehat{\beta})^T (\xi_{jk}x_{\ell m} + \xi_{\ell m}x_{jk})\right\}. \end{aligned} \tag{S40}$$

By Theorem S9, $\widehat{\beta} - \beta$ converges to zero in probability. By Slutsky's theorem, if the terms in (S40) involving elements of X and ξ converge in probability to any constant, then $\widehat{\phi}_i - \widetilde{\phi}_i$ converges in probability to zero. By (B1) and (S24) we have the convergence in probability of the term involving $x_{jk}x_{\ell m}^T$. Furthermore, by condition (B3), we have that $E(\xi_{jk}x_{\ell m}) = E(\xi_{\ell m}x_{jk}) = 0$. It remains to be shown that the variance of the error-covariate averages tend to zero. Consider the variance of the first error-covariate averages:

$$\begin{aligned}\text{var}\left(\frac{1}{|\Theta_i|}\sum_{(jk,\ell m)\in\Theta_i} \xi_{jk}x_{\ell m}\right) &= \frac{1}{|\Theta_i|^2} \sum_{(jk,\ell m)\in\Theta_i} \sum_{(rs,tu)\in\Theta_i} \text{cov}\left(\xi_{jk}x_{\ell m}, \xi_{rs}x_{tu}\right), \\ &= \frac{1}{|\Theta_i|^2} \sum_{(jk,\ell m)\in\Theta_i} \sum_{(rs,tu)\in\Theta_i} E\left(x_{\ell m}x_{tu}^T\right) \text{cov}\left(\xi_{jk}, \xi_{rs}\right). \end{aligned} \tag{S41}$$

In writing (S41), we use condition (B3) and simplify by conditioning on X and using the law of total variance. By the same counting arguments used to establish (S39), there are $|\Theta_i|^2 O(n^{-1})$ nonzero bounded covariances in (S41). Thus, we have

$$\text{var}\left(\frac{1}{|\Theta_i|}\sum_{(jk,\ell m)\in\Theta_i} \xi_{jk}x_{\ell m}\right) = \frac{|\Theta_i|^2 O(n^{-1})}{|\Theta_i|^2} \to 0.$$

Since the expectation and variance both tend to zero, we have

$$\frac{1}{|\Theta_i|}\sum_{(jk,\ell m)\in\Theta_i} \xi_{jk}x_{\ell m} \to_p 0.$$

The same argument applies to the second error-covariate term in (S40). Thus, we have shown that consistency of $\widehat{\beta}$ implies

$$\widehat{\phi}_i - \widetilde{\phi}_i \to_p 0.$$

□

S10 Proof of Theorem S11: mean-square error of the exchangeable and dyadic clustering estimators

In this section, we prove that the mean-square error of the estimator of var$(\widehat{\beta})$, conditional on X, is lower when using the exchangeable estimator than that when using the dyadic clustering estimator with high probability in X, assuming that the error structure is exchangeable. Before proving the theorem, we provide a lemma that states that the mean-square error of the each estimator is asymptotically equivalent to the mean-square error of each estimator based on the true errors, which vastly simplifies the proof of the mean-square error theorem. Even so, we must consider higher order moments of ξ than the covariances $\text{cov}\,(\xi_{jk}, \xi_{lm})$. So, we also provide a lemma in which we define the covariance of any pair of product of error relations $\text{cov}\,(\xi_{jk}\xi_{lm}, \xi_{rs}\xi_{tu})$ and define the limiting values of the covariance of the error averages, $n\text{cov}(\widetilde{\phi}_v, \widetilde{\phi}_w)$, for every pair $(v, w) \in \{1, \ldots, 5\} \times \{1, \ldots, 5\}$.

In this Section, we use the notation $O(n^a)$ and $\Theta(n^a)$, for some $a \in \mathbb{R}$, to denote the convergence a sequence of numbers to a constant (possibly zero) and a nonzero constant, respectively, as n grows to infinity. In other words, $X_n = O(n^a)$ means that the sequence $n^{-a}X_n$ converges to a constant that may be zero. The notation $X_n = \Theta(n^a)$ means that the sequence $n^{-a}X_n$ converges to a nonzero constant. Lastly, it follows that $X_n = O(n^{a-\epsilon})$ means that the sequence $n^{-a}X_n$ converges to zero.

We use similar notation for convergence of sequences of random variables. The notation $X_n = O_p(n^a)$ for some $a \in \mathbb{R}$ means that the sequence $n^{-a}X_n$ converges in distribution to a random variable (possibly a constant). The notation $X_n = o_p(n^a)$ for some $a \in \mathbb{R}$ means that the sequence $n^{-a}X_n$ converges in probability to zero. Finally, we define $X_n = \Theta_p(n^a)$ to mean that $n^{-a}X_n$ converges in distribution to a random variable with distribution that is not a point mass at zero, and thus possibly a nonzero constant (as will always be the case in this section).

S10.1 Lemmas in support of Theorem S11

The first lemma describes the covariances of parameter estimates based on the errors, which arise in the proof of the mean-square error theorem. Of interest are the covariances $\text{cov}(\widetilde{\phi}_v, \widetilde{\phi}_w)$ for $(v, w) \in \{3, 4, 5\} \times \{3, 4, 5\}$, as there are $\Theta(n)$ times as many of these covariances in $\widehat{V}_E$ as those covariances where at least one of v or w is in $\{1, 2\}$. However, we provide limiting values of all covariances for completeness. The proof of this lemma follows from recognizing that $\widetilde{\phi}_v$ is a sample average and from defining all possible covariances that make up $\text{cov}(\widetilde{\phi}_v, \widetilde{\phi}_w)$ and their multiplicities.

Lemma S16. *If ξ is a mean zero random vector with positive definite covariance matrix in the exchangeable class, $\Omega = \sum_{i=1}^{5} \phi_i \mathcal{S}_i$, and $E[\xi_{jk}^4] < L < \infty$, then the covariance $n\text{cov}\left(\widetilde{\phi}_v, \widetilde{\phi}_w\right)$ for $(v, w) \in \{1, \ldots, 5\} \times \{1, \ldots, 5\}$ converges to*

$$n\text{cov}\left(\widetilde{\phi}_v, \widetilde{\phi}_w\right) \to \begin{cases} \sum_{i=1}^{4} \alpha_i \beta_v \beta_w C\left(v, w\right)_i & (v, w) \in \{3, 4, 5\} \times \{3, 4, 5\} \\ \sum_{j=1}^{3} \gamma_j F(v, w)_j & (v, w) \in \{1, 2\} \times \{1, 2\}, \\ \sum_{k=1}^{4} \gamma_k D(v, w)_k & o.w., \end{cases} \tag{S42}$$

$$\begin{aligned} \textit{where } \alpha_i &:= 1 + 1_{[i>1]} + 1_{[i=4]}, \quad (i = 1, \ldots, 4), \\ \beta_i &:= 1 + 1_{[i=5]}, \quad i = (1, \ldots, 5), \\ \gamma_i &:= 1 + 1_{[i>2]}, \quad i = (1, \ldots, 4), \end{aligned}$$

and $C\left(v, w\right)_i$, $D\left(v, w\right)_i$, and $F\left(v, w\right)_i$ are unknown finite constants equal to $\text{cov}\left(\xi_{jk}\xi_{lm}, \xi_{rs}\xi_{tu}\right)$ for various configurations of the sets $\{j, k, l, m\}$ and $\{r, s, t, u\}$.

Proof. By definition,

$$n\text{cov}\left(\widetilde{\phi}_v, \widetilde{\phi}_w\right) = n|\Theta_v|^{-1}|\Theta_w|^{-1} \sum_{(jk,lm)\in\Theta_v} \sum_{(rs,tu)\in\Theta_w} \text{cov}\left(\xi_{jk}\xi_{lm}, \xi_{rs}\xi_{tu}\right). \tag{S43}$$

The sum is over $|\Theta_v||\Theta_w|$ terms. Whenever $\{j, k, l, m\} \cap \{r, s, t, u\} = \varnothing$, the covariance is zero. This removes a power of n from the sum in (S43), such that the sum is over $O(|\Theta_v||\Theta_w|n^{-1})$ possibly nonzero covariances. The scaled sum in (S43) converges – provided that the number of values that $\text{cov}\left(\xi_{jk}\xi_{lm}, \xi_{rs}\xi_{tu}\right)$ can take is finite – as each covariance is finite by assumption and the sequence of covariances is homogeneous as n grows by exchangeability. In the remainder of the proof, we enumerate and define the covariances $\text{cov}\left(\xi_{jk}\xi_{lm}, \xi_{rs}\xi_{tu}\right)$ in (S43) for particular pairs $(v, w) \in \{1, \ldots, 5\} \times \{1, \ldots, 5\}$, showing that the number of values that $\text{cov}\left(\xi_{jk}\xi_{lm}, \xi_{rs}\xi_{tu}\right)$ can take is finite. This is sufficient to establish convergence.

We begin by analyzing the case of interest, that is when both v and w are members of $\{3, 4, 5\}$. As an example, we focus on $v = 3$ and $w = 4$, where the first product of error relations corresponds to the same-sender covariance (b) in Figure 1 and the second corresponds to the same-receiver covariance (c) in Figure 1. In this case, both $|\Theta_3| = |\Theta_4| = \Theta(n^3)$. When $v = 3$ and $w = 4$, the covariance in (S43) becomes

$$n\text{cov}\left(\widetilde{\phi}_3, \widetilde{\phi}_4\right) =_a n^{-5} \sum_{jk} \sum_{l\notin\{j,k\}} \sum_{rs} \sum_{t\notin\{r,s\}} \text{cov}\left(\xi_{jk}\xi_{jl}, \xi_{sr}\xi_{tr}\right), \tag{S44}$$

where '$=_a$' denotes equality in the limit as n grows to infinity.

Only pairs of relation products that share a single actor will remain in the limit, as there are an order of n fewer covariances resulting from pairs of relation products that share two actors. One such pair of relation products that share a single actor correspond to the case when $s = j$, i.e. $\text{cov}\left(\xi_{jk}\xi_{jl}, \xi_{jr}\xi_{tr}\right)$, of which there are $\Theta(n^5)$ in the sum in (S44). There are $\Theta(n^4)$ covariances corresponding to the case when $s = j$ and $r = k$, i.e. $\text{cov}\left(\xi_{jk}\xi_{jl}, \xi_{jk}\xi_{tk}\right)$. The values of all covariances in (S44) are finite by assumption and not equal in general. However, by exchangeability, covariances resulting from pairs of relations that share an actor in the same way are equal.

For example, the covariance corresponding to $s = j$ is the same regardless of the node labeling, that is $\mathrm{cov}\left(\xi_{jk}\xi_{jl}, \xi_{jr}\xi_{tr}\right) = \mathrm{cov}\left(\xi_{ab}\xi_{ac}, \xi_{ad}\xi_{ed}\right)$ for any set $\{a, b, c, d, e\} \subset \{1, \ldots, n\}$ with $|\{a, b, c, d, e\}| = 5$.

There are nine ways that we may have $|\{j, k, l\} \cap \{r, s, t\}| = 1$ in (S44), i.e. there are nine ways that exactly one of $\{j, k, l\}$ equals exactly one of $\{r, s, t\}$. However, these reduce into four unique covariance values for each pair $(v, w) \in \{3, 4, 5\} \times \{3, 4, 5\}$. As an example, when $t = j$ the covariance is the same as that when $s = j$, that is $\mathrm{cov}\left(\xi_{jk}\xi_{jl}, \xi_{jr}\xi_{tr}\right) = \mathrm{cov}\left(\xi_{jk}\xi_{jl}, \xi_{sr}\xi_{jr}\right)$. Now we define these four covariance values and their multiplicities out of the nine possible ways that exactly one of $\{j, k, l\}$ equals exactly one of $\{r, s, t\}$:

- When $r = j$, we define the covariance $C(3, 4)_1 = \mathrm{cov}\left(\xi_{jk}\xi_{jl}, \xi_{sj}\xi_{tj}\right)$, of which there is one out of nine possible;
- When $s = j$, the covariance is the same as when $t = j$ (multiplicity two), and we define this covariance $C(3, 4)_2 = \mathrm{cov}\left(\xi_{jk}\xi_{jl}, \xi_{jr}\xi_{tr}\right)$;
- When $r = k$, the covariance is the same as when $r = l$ (multiplicity two), and we define this covariance $C(3, 4)_3 = \mathrm{cov}\left(\xi_{jk}\xi_{jl}, \xi_{sk}\xi_{tk}\right)$;
- We define the covariance when $s = k$ to be $C(3, 4)_4 = \mathrm{cov}\left(\xi_{jk}\xi_{jl}, \xi_{kr}\xi_{tr}\right)$, of which there are four, the remaining terms of which correspond to $t = k$, $s = l$, and $t = l$.

Now, noting that there are $n^5 + \Theta(n^4)$ covariances in the sum (S44) corresponding to each of the nine possible ways that exactly one of $\{j, k, l\}$ equals exactly one of $\{r, s, t\}$, we see that

$$\begin{aligned} n\mathrm{cov}\left(\widetilde{\phi}_3, \widetilde{\phi}_4\right) &\to \mathrm{cov}\left(\xi_{jk}\xi_{jl}, \xi_{sj}\xi_{tj}\right) + 2\mathrm{cov}\left(\xi_{jk}\xi_{jl}, \xi_{jr}\xi_{tr}\right) \\ &\quad + 2\mathrm{cov}\left(\xi_{jk}\xi_{jl}, \xi_{sk}\xi_{tk}\right) + 4\mathrm{cov}\left(\xi_{jk}\xi_{jl}, \xi_{kr}\xi_{tr}\right), \\ &= C(3, 4)_1 + 2C(3, 4)_2 + 2C(3, 4)_3 + 4C(3, 4)_4, \end{aligned} \tag{S45}$$

where '$\to$' denotes convergence in the limit as n goes to infinity. Under appropriate definition of $C(v, w)_i$ for $i \in \{1, 2, 3, 4\}$, the same argument applies when both v and w are one of $\{3, 4\}$. When $w = 5$ (relation products of the form $\{\xi_{jk}\xi_{kl}\}$ and $\{\xi_{jk}\xi_{lj}\}$) and $v = 3$, however, we then must consider covariances $\mathrm{cov}\left(\xi_{jk}\xi_{jl}, \xi_{rs}\xi_{tr}\right)$ and $\mathrm{cov}\left(\xi_{jk}\xi_{jl}, \xi_{sr}\xi_{rt}\right)$ from $w = 5$, which doubles the coefficients in (S45). This accounts for $\beta_w = 2$ when $w = 5$ and $\beta_w = 1$ otherwise in (S42). The same argument applies when $v = 5$.

We now analyze both v and w in $\{1, 2\}$, corresponding to variance and the reciprocal covariance (a) in Figure 1. In this case, both $|\Theta_v| = |\Theta_w| = n(n-1)$. Taking $v = 1$ and $w = 1$ as an example, in the limit, the covariance in (S43) is

$$n\mathrm{cov}\left(\widetilde{\phi}_1, \widetilde{\phi}_2\right) =_a n^{-3} \sum_{jk} \sum_{rs} \mathrm{cov}\left(\xi_{jk}^2, \xi_{rs}^2\right). \tag{S46}$$

Again, we only consider covariances corresponding to pairs of relation products that share a single actor as only these covariances survive in the limit. There are four possible ways that $\{j, k\}$ shares exactly one actor with $\{r, s\}$. We define the three unique covariances and their multiplicities corresponding to the four ways that $\{j, k\}$ shares exactly one actor with $\{r, s\}$ as follows:

- When $r = j$, we define the covariance $F(1,1)_1 = \text{cov}\left(\xi_{jk}^2, \xi_{js}^2\right)$, of which there is one out of the four possibilities;
- When $s = k$, we define the covariance $F(1,1)_2 = \text{cov}\left(\xi_{jk}^2, \xi_{rk}^2\right)$, of which there is one;
- When $s = j$, we define the covariance $F(1,2)_3 = \text{cov}\left(\xi_{jk}^2, \xi_{rj}^2\right)$, which is the same as when $r = k$, accounting for the remaining two possibilities.

Now, the fact that there are $n^3 + \Theta(n^2)$ covariances in the sum (S46) corresponding to each of the four possible ways that $\{j,k\}$ shares exactly one actor with $\{r,s\}$ gives that

$$\begin{aligned} n\text{cov}\left(\widetilde{\phi}_1, \widetilde{\phi}_1\right) &\to \text{cov}\left(\xi_{jk}^2, \xi_{js}^2\right) + \text{cov}\left(\xi_{jk}^2, \xi_{rk}^2\right) + 2\text{cov}\left(\xi_{jk}^2, \xi_{rj}^2\right) \\ &= F(1,1)_1 + F(1,1)_2 + 2F(1,1)_3. \end{aligned}$$

Of course, the same argument applies to any v and w both in $\{1,2\}$. In the case where $v = 1$ and $w = 2$, by symmetry, $F(1,2)_1 = F(1,2)_2$. Similarly, for $v = w = 2$, all $F(2,2)_1 = F(2,2)_2 = F(2,2)_3$.

Similar counting arguments to those in the previous paragraphs apply when one of v, w is in $\{3,4,5\}$ and the other is in $\{1,2\}$. As an example, consider $v = 1$ and $w = 3$. Once again, only pairs of relations that share a single actor will remain in the limit. Then, in the limit, the covariance in (S43) becomes

$$n\text{cov}\left(\widetilde{\phi}_2, \widetilde{\phi}_3\right) =_a n^{-4} \sum_{jk} \sum_{rs} \sum_{t \notin \{r,s\}} \text{cov}\left(\xi_{jk}^2, \xi_{rs}\xi_{rt}\right). \tag{S47}$$

Now, there are six ways in which the first pair of relations share an actor with the second pair, i.e. all sets with exactly one actor from $\{j,k\}$ equal to exactly one other from $\{r,s,t\}$. We define the covariances corresponding to the six possibilities below:

- When $r = j$, we define the covariance $D(1,3)_1 = \text{cov}\left(\xi_{jk}^2, \xi_{js}\xi_{jt}\right)$, of which there is one out of the six possibilities;
- When $r = k$, we define the covariance $D(1,3)_2 = \text{cov}\left(\xi_{jk}^2, \xi_{ks}\xi_{kt}\right)$, of which there is one;
- The overlaps where $s = j$ and $t = j$ result in the same covariance (multiplicity two), which we define $D(1,3)_3 = \text{cov}\left(\xi_{jk}^2, \xi_{rj}\xi_{rt}\right)$;
- The overlaps where $s = k$ and $t = k$ result in the same covariance (multiplicity two), which we define $D(1,3)_4 = \text{cov}\left(\xi_{jk}^2, \xi_{rk}\xi_{rt}\right)$.

Then, noting that there are $n^4 + \Theta(n^3)$ covariances in the sum (S47) corresponding to each of the six possible ways that exactly one actor from $\{j,k\}$ is equal to exactly one other from $\{r,s,t\}$, we have that $n\text{cov}\left(\widetilde{\phi}_2, \widetilde{\phi}_3\right)$ converges to

$$\begin{aligned} &\text{cov}\left(\xi_{jk}^2, \xi_{js}\xi_{jt}\right) + \text{cov}\left(\xi_{jk}^2, \xi_{ks}\xi_{kt}\right) + 2\text{cov}\left(\xi_{jk}^2, \xi_{rj}\xi_{rt}\right) + 2\text{cov}\left(\xi_{jk}^2, \xi_{rk}\xi_{rt}\right), \\ &= D(1,3)_1 + D(1,3)_2 + 2D(1,3)_3 + 2D(1,3)_4. \end{aligned} \tag{S48}$$

When $D(v,w)_k$ for $k \in \{1,2,3,4\}$ is appropriately defined, the same argument applies for all settings where one of v, w is in $\{3,4\}$ and the other is in $\{1,2\}$. When $w = 5$ (relation products of the form $\{\xi_{jk}\xi_{kl}\}$ and $\{\xi_{jk}\xi_{lj}\}$), however, we then must consider covariances in (S47) $\text{cov}(\xi_{jk}^2, \xi_{rs}\xi_{tr})$ and $\text{cov}(\xi_{jk}^2, \xi_{sr}\xi_{rs})$, which doubles the coefficients in (S48). This accounts for $\beta_w = 2$ when $w = 5$ and $\beta_w = 1$ otherwise in (S42). When $v = 2$, for example, we have the simplification that $D(2,3)_1 = D(2,3)_2$ and $D(2,3)_3 = D(2,3)_4$. □

The expressions for the estimators based on the errors are simpler to analyze than those based on the residuals. For example, when comparing the mean-square errors of the exchangeable and dyadic clustering estimators, it is desirable to analyze $MSE_\xi(\widetilde{V}_E \mid X)$ instead of $MSE_\xi(\widehat{V}_E \mid X)$. The following lemma allows us to do just this. This lemma states that mean-square errors of the estimators based on the errors are asymptotically equivalent to the mean-square error of those based on the residuals. The proof consists of first evaluating the mean-square error conditional on X. We then show that $n^3 MSE_\xi(\widetilde{V}_E \mid X)$ converges in X-probability to a nonzero constant in general and that $n^3 MSE_\xi(\widetilde{V}_E \mid X) - n^3 MSE_\xi(\widehat{V}_E \mid X)$ converges in X-probability to zero, implying that the difference between $MSE_\xi(\widetilde{V}_E \mid X)$ and $MSE_\xi(\widehat{V}_E \mid X)$ is asymptotically negligible. We repeat the procedure for $MSE_\xi(\widetilde{V}_{DC} \mid X)$ and $MSE_\xi(\widehat{V}_{DC} \mid X)$.

Lemma S17. *Assuming $E\left(|x_{jk}^{(l)}|^8\right) < L^8 < \infty$ for all $(l = 1, \ldots, p)$ and under the assumptions of Theorem S9, the mean-square error for both the exchangeable and dyadic clustering estimators based on the residuals is asymptotically equivalent to the mean-square error of each respective estimator based on the errors. That is,*

$$n^3 MSE_\xi\left(\widehat{V}_E \mid X\right) = n^3 MSE_\xi\left(\widetilde{V}_E \mid X\right) + O_p(n^{-1/2}) = O_p(1), \tag{S49}$$

and analogously for dyadic clustering.

Proof. We will focus on the exchangeable estimator first, and then the dyadic clustering estimator. Throughout, we drop the conditioning on X in the mean-square error as it is understood, for example $MSE_\xi(\widehat{V}_E) = MSE_\xi(\widehat{V}_E \mid X)$.

We begin with the exchangeable estimator. By definition, the mean-square error of the exchangeable estimator is

$$\begin{aligned} MSE_\xi\left(\widehat{V}_E\right) &= E\left\{\left(\widehat{V}_E - V^*\right)^2 \mid X\right\}, \\ &= MSE_\xi\left(\widetilde{V}_E\right) + E\left\{\left(\widehat{V}_E - \widetilde{V}_E\right)^2 \mid X\right\} + 2E\left\{\left(\widehat{V}_E - \widetilde{V}_E\right)\left(\widetilde{V}_E - V^*\right) \mid X\right\}, \end{aligned} \tag{S50}$$

where $V^* = \text{var}(\widehat{\beta})$, the true variance of $\widehat{\beta}$. By the Cauchy-Schwarz inequality,

$$E\left\{\left(\widehat{V}_E - \widetilde{V}_E\right)\left(\widetilde{V}_E - V^*\right) \mid X\right\} \leq \left[MSE_\xi\left(\widetilde{V}_E\right) E\left\{\left(\widehat{V}_E - \widetilde{V}_E\right)^2 \mid X\right\}\right]^{1/2}. \tag{S51}$$

If we show that $n^3 MSE_\xi\left(\widetilde{V}_E\right)$ converges in X-probability to a constant, i.e. $MSE_\xi\left(\widetilde{V}_E\right) = O_p(n^{-3})$, and that $E\left\{\left(\widehat{V}_E - \widetilde{V}_E\right)^2 \mid X\right\} = O_p(n^{-4})$, then (S51) implies that the third ad-

ditive term of (S50) is $O_p(n^{-7/2})$. This is sufficient to establish (S49). We begin with showing $n^3 MSE_\xi\left(\widetilde{V}_E\right) = O_p(1)$. By definition, the scaled mean-square error is $n^3 MSE_\xi\left(\widetilde{V}_E\right) = n^3 E\left\{\text{tr}\left(\widetilde{V}_E^2\right) \mid X\right\}$, which is equal to

$$\sum_{v=1}^{5}\sum_{w=1}^{5} n\text{cov}\left(\widetilde{\phi}_v, \widetilde{\phi}_w\right) \text{tr}\left\{\left(\frac{X^TX}{n^2}\right)^{-1}\left(\frac{X^T\mathcal{S}_vX}{n^3}\right)\left(\frac{X^TX}{n^2}\right)^{-2}\left(\frac{X^T\mathcal{S}_wX}{n^3}\right)\left(\frac{X^TX}{n^2}\right)^{-1}\right\}. \tag{S52}$$

By Lemma S16, $n\text{cov}(\widetilde{\phi}_v, \widetilde{\phi}_w)$ converges to a finite constant for every $(v, w) \in \{1, \ldots, 5\} \times \{1, \ldots, 5\}$. The convergence in probability of each multiplicative term in (S52) containing X is defined by assumption (B1); only those with both v and w in $\{3, 4, 5\}$ survive in the limit as these have $|\Theta_v| = \Theta(n^3)$ whereas $|\Theta_v| = \Theta(n^2)$ for $v \in \{1, 2\}$. Thus, we have that

$$n^3 MSE_\xi\left(\widetilde{V}_E\right) \to_{pr_X} \sum_{v=3}^{5}\sum_{w=3}^{5}\sum_{i=1}^{4} \alpha_i\beta_v\beta_w C(v, w)_i \text{tr}\left(M_1^{-1}M_vM_1^{-2}M_wM_1^{-1}\right), \tag{S53}$$

which is finite., and where '$\to_{pr_X}$' denotes convergence in X-probability.

It remains to show that $E\left\{\left(\widehat{V}_E - \widetilde{V}_E\right)^2 \mid X\right\} = O_p(n^{-4})$. To establish this fact, it is sufficient to show that $\widehat{V}_E - \widetilde{V}_E = O_p(n^{-2})$, and then, by the continuous mapping theorem, $\left(\widehat{V}_E - \widetilde{V}_E\right)^2 = O_p(n^{-4})$, which implies the desired result. Writing directly,

$$\widehat{V}_E - \widetilde{V}_E = \frac{1}{n}\sum_{v=1}^{5}\left(\widehat{\phi}_v - \widetilde{\phi}_v\right)\left(\frac{X^TX}{n^2}\right)^{-1}\left(\frac{X^T\mathcal{S}_vX}{n^3}\right)\left(\frac{X^TX}{n^2}\right)^{-1}. \tag{S54}$$

By assumption (B1), the multiplicative terms involving X converge in probability to constants. To establish $\widehat{V}_E - \widetilde{V}_E = O_p(n^{-2})$, it is sufficient show that $\widehat{\phi}_v - \widetilde{\phi}_v = O_p(n^{-1})$ for all $v \in \{1, \ldots, 5\}$. Writing this expression directly, the difference $\widehat{\phi}_v - \widetilde{\phi}_v$ is

$$-(\widehat{\beta} - \beta)^T\left(\sum_{(jk,lm)\in\Theta_v}\frac{x_{jk}\xi_{lm} + x_{lm}\xi_{jk}}{|\Theta_v|}\right) + (\widehat{\beta} - \beta)^T\left(\sum_{(jk,lm)\in\Theta_v}\frac{x_{jk}x_{lm}^T}{|\Theta_v|}\right)(\widehat{\beta} - \beta). \tag{S55}$$

By Theorem S9, $\widehat{\beta} - \beta = O_p(n^{-1/2})$. Also, by assumption (B1), the sum involving X in the second term converges in probability to a constant; thus, the second additive term in (S55) is $O_p(n^{-1})$. Turning to the first additive term, we notice its expectation is zero, that is $E(x_{jk}\xi_{lm}) = 0$ for all relations jk and lm. The variance is

$$\text{var}\left(\sum_{(jk,lm)\in\Theta_v}\frac{x_{jk}\xi_{lm}}{|\Theta_v|}\right) = \frac{1}{|\Theta_v|^2}\sum_{(jk,lm)\in\Theta_v}\sum_{(rs,tu)\in\Theta_v} E\left(x_{jk}x_{rs}^T\xi_{lm}\xi_{tu}\right) = O(n^{-1}),$$

where we use the fact that $E\left(x_{jk}x_{rs}^T\xi_{lm}\xi_{tu}\right)$ is only nonzero when relation lm shares an actor with relation tu since $E\left(\xi_{lm}\xi_{tu}\right) = 0$ whenever $\{j, k\} \cap \{l, m\} = \varnothing$ and X is independent ξ by assumption (B3). This fact removes at least a factor of n from the sum. Thus, we have that $\sum_{(jk,lm)\in\Theta_v} x_{jk}\xi_{lm} + x_{lm}\xi_{jk}/|\Theta_v| = O_p(n^{-1/2})$, which gives that $\widehat{\phi}_v - \widetilde{\phi}_v = O_p(n^{-1})$ and

$$E\left\{\left(\widehat{V}_E - \widetilde{V}_E\right)^2 \mid X\right\} = O_p(n^{-4}),$$

which establishes (S49) for the exchangeable estimator.

The same argument following (S50) applies to the dyadic clustering estimator. To establish (S49) for the dyadic clustering estimator, it is thus sufficient to show that $n^3 MSE_\xi\left(\widetilde{V}_{DC}\right)$ converges in X-probability to a constant and that $E\left\{\left(\widehat{V}_{DC}-\widetilde{V}_{DC}\right)^2 \mid X\right\} = O_p(n^{-4})$. We begin with the former. By definition, $n^3 MSE_\xi\left(\widetilde{V}_{DC}\right)$ is

$$\frac{1}{n^5}\sum_{(jk,lm)\in\Theta_0}\sum_{(rs,tu)\in\Theta_0}\operatorname{cov}\left(\xi_{jk}\xi_{lm},\xi_{rs}\xi_{tu}\right)\times\ldots \\ \ldots\times\operatorname{tr}\left\{\left(\frac{X^TX}{n^2}\right)^{-1}x_{jk}x_{lm}^T\left(\frac{X^TX}{n^2}\right)^{-2}x_{rs}x_{tu}^T\left(\frac{X^TX}{n^2}\right)^{-1}\right\}, \tag{S56}$$

where Θ_0 is the set of relation pairs that share an actor in any manner. Then, substituting the asymptotic values for $\operatorname{cov}\left(\xi_{jk}\xi_{lm},\xi_{rs}\xi_{tu}\right)$ from Lemma S16 and separating the sum by the five ways that two relations may share an actor,

$$n^3 MSE_\xi\left(\widetilde{V}_{DC}\right) =_a \frac{1}{n^5}\sum_{v=3}^{5}\sum_{w=3}^{5}\sum_{i=1}^{4}\sum_{T(v,w)_i}C(v,w)_i\; x_{lm}^T M_1^{-2}x_{rs}x_{tu}^T M_1^{-2}x_{jk}, \tag{S57}$$

where '$=_a$' denotes equality in the limit and $T(v,w)_i$ is the set of relations (jk,lm,rs,tu) such that $(jk,lm)\in\Theta_v$ and $(rs,tu)\in\Theta_w$ and such that the pairs of relations (jk,lm) and (rs,tu) share a single actor as appropriate for $C(v,w)_i$ in Lemma S16. In (S57), we substitute the limit of $\left(X^TX/n^2\right)^{-2}$ from assumption (B1). Also in (S57), only terms with v and w both in $\{3,4,5\}$ survive in the limit as the set $|T(v,w)_i| = \Theta(n^5)$ (as detailed in Lemma S16), while the order is less for either v or w in $\{1,2\}$, so these terms vanish in the limit. Evaluating the vector products, the expression on the right hand side of (S57) equal to

$$\sum_{v=3}^{5}\sum_{w=3}^{5}\sum_{i=1}^{4}\sum_{a=1}^{p}\sum_{b=1}^{p}\sum_{c=1}^{p}\sum_{d=1}^{p}C(v,w)_i\;\left(m_1^{-2}\right)_{ab}\left(m_1^{-2}\right)_{cd}\left(\frac{1}{n^5}\sum_{T(v,w)_i}x_{lm}^{(a)}x_{rs}^{(b)}x_{tu}^{(c)}x_{jk}^{(d)}\right), \tag{S58}$$

where $\left(m_1^{-2}\right)_{ab}$ is the (a,b) entry in M_1^{-2}, e.g., and $x_{jk}^{(a)}$ is the entry in X pertaining to column a and relation jk. Further, the variance $\operatorname{var}\left(n^{-5}\sum_{T(v,w)_i}x_{lm}^{(a)}x_{rs}^{(b)}x_{tu}^{(c)}x_{jk}^{(d)}\right)$ is

$$\frac{1}{n^{10}}\sum_{T(v,w)_i}\sum_{U(v,w)_i}\operatorname{cov}\left(x_{lm}^{(a)}x_{rs}^{(b)}x_{tu}^{(c)}x_{jk}^{(d)},x_{ef}^{(a)}x_{gh}^{(b)}x_{np}^{(c)}x_{yz}^{(d)}\right), \tag{S59}$$
$$=\frac{\Theta(n^9)}{n^{10}}\to 0,$$

where $U(v,w)_i = T(v,w)_i$ and (lm,rs,tu,jk) indexes the first sum and (ef,gh,np,yz) indexes the second sum. The convergence is the result of the independence portion of assumption in (B1) and the bounded moment assumption on X. The variance in (S59) converges to zero for every set of covariates $\{a,b,c,d\}$, every relation type v and w both in $\{3,4,5\}$, and every covariance type

$i \in \{1,2,3,4\}$. Thus, provided that the expectation of $n^{-5}\sum_{T(v,w)_i} x_{lm}^{(a)} x_{rs}^{(b)} x_{tu}^{(c)} x_{jk}^{(d)}$ converges to a constant, this expression converges in probability to that same constant. This expectation is

$$E\left(x_{lm}^{(a)} x_{rs}^{(b)} x_{tu}^{(c)} x_{jk}^{(d)}\right) = \operatorname{cov}\left(x_{jk}^{(d)} x_{lm}^{(a)}, x_{rs}^{(b)} x_{tu}^{(c)}\right) + (m_v)_{ad}\,(m_w)_{bc}\,, \tag{S60}$$

where $(m_v)_{ij}$ is the (i,j) entry in M_v and we use the symmetry of M_v for all $v \in \{3,4,5\}$. Unlike ξ, for a given $i \in \{1,2,3,4\}$, the covariances in (S60) may not be the same for two relation sets in $T(v,w)_i$. However, by assumption (B1) and taking $i = 1$, $v = 3$, and $w = 4$ for example, we still have that

$$\operatorname{cov}\left(x_{jl}^{(d)} x_{sj}^{(a)}, x_{tj}^{(b)} x_{jk}^{(c)}\right) = \operatorname{cov}\left(x_{ef}^{(d)} x_{ge}^{(a)}, x_{he}^{(b)} x_{ep}^{(c)}\right), \tag{S61}$$

for $|\{j,k,l,s,t,e,f,g,h,p\}| = 10$. That is, covariances that share an actor in the same way are still equal. So, for fixed $i \in \{1,2,3,4\}$ and pair of v and w both in $\{3,4,5\}$, we may collect the α_i possible covariances and average them to attain the convergent value. We thus define the limit

$$\frac{1}{n^5}\sum_{T(v,w)_i} \operatorname{cov}\left(x_{jk}^{(d)} x_{lm}^{(a)}, x_{rs}^{(b)} x_{tu}^{(c)}\right) \to \alpha_i\beta_v\beta_w \frac{1}{\alpha_i}\sum_{W(v,w)_i} \operatorname{cov}\left(x_{jk}^{(d)} x_{lm}^{(a)}, x_{rs}^{(b)} x_{tu}^{(c)}\right), \tag{S62}$$

$$:= \alpha_i\beta_v\beta_w C_X^{(d,a,b,c)}(v,w)_i, \tag{S63}$$

where $W(v,w)_i$ is the set of α_i ways that (jk, lm, rs, tu) correspond to $T(v,w)_i$. For example, when $i = 4$, $v = 3$, and $w = 4$, $W(v,w)_i$ contains four index sets corresponding to the four multiplicities of $C(v,w)_4$ as defined in Lemma (3). The convergence of (S62) results from (S61). As the average over $W(v,w)_i$ is over a finite number of terms since each α_i is bounded, there is no possibility of divergence. The covariances in (S62) are finite by assumption (B2) and $\beta_v\beta_w$ arises from the asymptotic limit of $n^{-5}|T(v,w)_i|$. Taking (S62) together with (S60), we have the convergence of the expectation of $n^{-5}\sum_{T(v,w)_i} x_{lm}^{(a)} x_{rs}^{(b)} x_{tu}^{(c)} x_{jk}^{(d)}$. Along with (S59), convergence of the expectation of $n^{-5}\sum_{T(v,w)_i} x_{lm}^{(a)} x_{rs}^{(b)} x_{tu}^{(c)} x_{jk}^{(d)}$ establishes the convergence of $n^{-5}\sum_{T(v,w)_i} x_{lm}^{(a)} x_{rs}^{(b)} x_{tu}^{(c)} x_{jk}^{(d)}$ to the same limit.

Now, for a particular v and w both in $\{3,4,5\}$ and $i \in \{1,2,3,4\}$, we collect the set of $\{C_X^{(d,a,b,c)}(v,w)_i\}$ for every a, b, c, and d in $\{1,2,\ldots,p\}$ from (S63) into a $p^2 \times p^2$ matrix defined $D_X(v,w)_i$. Substituting this definition into (S58) while noting each $\{M_v\}_{v=1}^5$ is symmetric, the convergent value for the dyadic clustering estimator is

$$n^3 MSE_\xi\left(\widetilde{V}_{DC}\right) \to_{pr_X} \sum_{v=3}^{5}\sum_{w=3}^{5}\sum_{i=1}^{4} \alpha_i\beta_v\beta_w C(v,w)_i \times \ldots \tag{S64}$$

$$\ldots \times \left\{\operatorname{vec}\left(M_1^{-2}\right)^T D_X(v,w)_i \operatorname{vec}\left(M_1^{-2}\right) + \operatorname{tr}\left(M_1^{-2} M_v M_1^{-2} M_w\right)\right\}.$$

Since the convergent value in (S64) is a finite constant, it remains to show that

$$E\left\{\left(\widehat{V}_{DC} - \widetilde{V}_{DC}\right)^2 \;\middle|\; X\right\} = O_p(n^{-4}).$$

As with the exchangeable estimator, it is sufficient to show that $\widehat{V}_{DC} - \widetilde{V}_{DC} = O_p(n^{-2})$. Using the residual definition $e_{jk} = \xi_{jk} + x_{jk}^T(\beta - \widehat{\beta})$, the expression for $\widehat{V}_{DC} - \widetilde{V}_{DC}$ is

$$\begin{aligned}
&\frac{1}{n}\sum_{v=1}^{5}\sum_{(jk,lm)\in\Theta_v}\left(\frac{X^TX}{n^2}\right)^{-1}\left\{\frac{x_{jk}x_{lm}^T}{n^3}\left(e_{jk}e_{lm} - \xi_{jk}\xi_{lm}\right)\right\}\left(\frac{X^TX}{n^2}\right)^{-1} \qquad \text{(S65)}\\
&\quad =_a \frac{1}{n}\sum_{v=1}^{5}\sum_{(jk,lm)\in\Theta_v} M_1^{-1}\left[\frac{x_{jk}x_{lm}^T}{n^3}\left\{x_{jk}^T\left(\widehat{\beta}-\beta\right)\left(\widehat{\beta}-\beta\right)^T x_{lm}\right\} - 2x_{jk}^T\left(\widehat{\beta}-\beta\right)\xi_{lm}\right]M_1^{-1},
\end{aligned}$$

where we substitute the convergence of $\left(X^TX/n^2\right)^{-1}$ to M_1^{-1} and have used the exchangeability property to get the factor of two on the second additive term in the center of (S65). Analyzing the first additive term in the center of (S65),

$$\begin{aligned}
&\sum_{v=1}^{5}\sum_{(jk,lm)\in\Theta_v}\frac{x_{jk}x_{lm}^T}{n^3}\left\{x_{jk}^T\left(\widehat{\beta}-\beta\right)\left(\widehat{\beta}-\beta\right)^T x_{lm}\right\}\\
&\quad = \sum_{jk}\sum_{lm\in\Theta_{jk}}\left(\frac{x_{jk}x_{jk}^T}{n^2}\right)\left(\widehat{\beta}-\beta\right)\left(\widehat{\beta}-\beta\right)^T\left(\frac{x_{lm}x_{lm}^T}{n}\right), \qquad \text{(S66)}\\
&\quad = \Theta_p(1)O_p(n^{-1/2})O_p(n^{-1/2})\Theta_p(1) = O_p(n^{-1}),
\end{aligned}$$

recalling the notation that Θ_{jk} is the set of all relations that share an actor with relation jk and that $|\Theta_{jk}| = \Theta(n)$. We attain the convergence rate by noting that the X-terms in (S66) converge in probability to constants by assumption (B1) and $\widehat{\beta} - \beta = O_p(n^{-1/2})$ by Theorem S9. The convergences in (S66) are for $p \times p$ matrices; these convergences are element-wise.

We now analyze the convergence rate of the second additive term in the center of (S65),

$$\begin{aligned}
\sum_{v=1}^{5}\sum_{(jk,lm)\in\Theta_v}\frac{x_{jk}x_{lm}^T}{n^3}\left\{x_{jk}^T\left(\widehat{\beta}-\beta\right)\xi_{lm}\right\} &= \sum_{lm}\sum_{jk\in\Theta_{lm}}\left(\frac{x_{jk}x_{jk}^T}{n}\right)\left(\frac{\xi_{lm}x_{lm}^T}{n^2}\right)\left(\widehat{\beta}-\beta\right),\\
&= \Theta_p(1)O_p(n^{-1/2})O_p(n^{-1/2}) = O_p(n^{-1}).
\end{aligned}$$

Again, the convergence of the first multiplicative term is a result of assumption (B1) and $\widehat{\beta} - \beta = O_p(n^{-1/2})$ by Theorem S9. The mean $n^{-2}\sum_{lm} x_{lm}\xi_{lm}$ is expectation zero and $O_p(n^{-1/2})$ by previous arguments, for example in (S34). Thus, we have $\widehat{V}_{DC} - \widetilde{V}_{DC} = O_p(n^{-2})$, and the dyadic clustering estimator satisfies the relation in (S49). □

S10.2 Proof of Theorem S11

We now establish that the mean-square error of the exchangeable estimator is less than that of the dyadic clustering estimator with high probability. To do so, we show that the value to which the difference in mean-square errors converges is nonnegative. Throughout the proof, we drop the conditioning on X in the mean-square error as it is understood, for example $MSE_\xi(\widehat{V}_E) = MSE_\xi(\widehat{V}_E \mid X)$.

The asymptotic difference in mean-square errorss is as follows, where we substitute the expressions for the estimators based on the errors in (S53) and (S64), as justified by Lemma S17:

$$n^3\left\{MSE_\xi\left(\widehat{V}_{DC}\right)-MSE_\xi\left(\widehat{V}_E\right)\right\}$$
$$\to_{pr_X}\sum_{v=3}^{5}\sum_{w=3}^{5}\sum_{i=1}^{4}\alpha_i\beta_v\beta_w C(v,w)_i\text{vec}\left(M_1^{-2}\right)^T D_X(v,w)_i\text{vec}\left(M_1^{-2}\right). \tag{S67}$$

It remains to show that this is a nonnegative constant. To do so, we show that the matrix in the quadratic form in (S67) is the limit of a variance matrix, and thus positive semi-definite. We will show that the scaled variance

$$\frac{1}{n^5}\text{var}\left\{\sum_{v=1}^{5}\sum_{jk,lm\in\Theta_v}\xi_{jk}\xi_{lm}\text{vec}\left(x_{jk}x_{lm}^T\right)\right\}$$
$$=\frac{1}{n^5}\sum_{v=1}^{5}\sum_{w=1}^{5}\sum_{jk,lm\in\Theta_v}\sum_{rs,tu\in\Theta_w}\text{cov}\left\{\xi_{jk}\xi_{lm}\text{vec}\left(x_{jk}x_{lm}^T\right),\ \xi_{rs}\xi_{tu}\text{vec}\left(x_{rs}x_{tu}^T\right)\right\}, \tag{S68}$$

converges to the desired matrix. First, the sum in (S68) is $\Theta(n^5)$ as the relations jk and lm must share at least one actor with the relations rs and tu for the covariance to be nonzero. Then, by the arguments in Lemma S16, only pairs of relations that share a single actor survive in the limit. Finally, by assumption (B3), X is independent ξ and the variance in (S68) is asymptotically equivalent to

$$\frac{1}{n^5}\sum_{v=3}^{5}\sum_{w=3}^{5}\sum_{jk,lm\in\Theta_v}\sum_{rs,tu\in\Theta_w}E\left\{\left(\xi_{jk}\xi_{lm}-\phi_v\right)\left(\xi_{rs}\xi_{tu}-\phi_w\right)\right\}\times\ldots$$
$$\ldots\times E\left[\left\{\text{vec}\left(x_{jk}x_{lm}^T\right)-\text{vec}\left(M_v\right)\right\}\left\{\text{vec}\left(x_{rs}x_{tu}^T\right)-\text{vec}\left(M_w\right)\right\}^T\right],$$

where only terms with both v and w in $\{3,4,5\}$ survive in the limit. Then, by assumptions (A1) and (B1) and applying Lemma S16, the variance converges to

$$\frac{1}{n^5}\text{var}\left\{\sum_{v=1}^{5}\sum_{jk,lm\in\Theta_v}\xi_{jk}\xi_{lm}\text{vec}\left(x_{jk}x_{lm}^T\right)\right\}\to\sum_{v=3}^{5}\sum_{w=3}^{5}\sum_{i=1}^{4}\alpha_i\beta_v\beta_w C(v,w)_i D_X(v,w)_i, \tag{S69}$$

where we substitute the definition of $D_X(v,w)_i$ following (S63). Thus, the matrix in (S69) is positive semi-definite. Now, (S67) becomes

$$n^3\left\{MSE_\xi\left(\widehat{V}_{DC}\right)-MSE_\xi\left(\widehat{V}_E\right)\right\}$$
$$\to_{pr_X}\text{vec}\left(M_1^{-2}\right)^T\left\{\sum_{v=3}^{5}\sum_{w=3}^{5}\sum_{i=1}^{4}\alpha_i\beta_v\beta_w C(v,w)_i D_X(v,w)_i\right\}\text{vec}\left(M_1^{-2}\right)\geq 0.$$

□

S11 Simulation study

S11.1 Details of simulation study

We simulated data from a linear regression model with an exchangeable error model, independent and identically distributed errors, and a non-exchangeable error model. For each error setting, we employed the following three-covariate regression model:

$$y_{ij} = \beta_1 + \beta_2 1_{(x_{2i} \in C)} 1_{(x_{2j} \in C)} + \beta_3 |x_{3i} - x_{3j}| + \beta_4 x_{4ij} + \xi_{ij}, \tag{S70}$$

where $1_{(.)}$ denotes the indicator function. In this model, β_1 is an intercept; β_2 is a coefficient on a binary indicator of whether individuals i and j both belong to a pre-specified class $C \subset (1, \ldots, n)$; β_3 is a coefficient on the absolute difference of a continuous, actor-specific covariate x_{3i}; and β_4 is that for a pair-specific continuous covariate x_{4ij}. For all simulations, we fixed true coefficients $\beta = (1, 1, 1, 1)^T$. We drew each x_{2i} from a Bernoulli$(1/2)$ distribution independently. In the rare event that $x_{2i} = x_{2j}$ for all (i, j) pairs, one realization x_{2k} was randomly flipped to a 1 or 0. All x_{3i} and x_{4ij} were drawn independently from a standard normal distribution.

Each error setting was specified to have the same total variance:

$$\sum_{ij} \text{var}(\xi_{ij}) = 3n(n-1).$$

This variance was chosen so that the variance of the error would be similar to that of the regression mean model $\beta^T x_{ij}$. In the independent and identifically distributed errors setting, $\xi_{ij} \sim_{iid} \text{N}(0, 3)$ for all (i, j). To generate the non-exchangeable errors, a mean-zero random effect was added to the upper left quadrant of var(ξ) under independent and identically distributed errors. The errors for the non-exchangeable error setting may be written

$$\xi_{ij} = \tau 1_{(i \leq \lfloor n/2 \rfloor)} 1_{(j \leq \lfloor n/2 \rfloor)} + \epsilon_{ij}, \quad \tau \sim \text{N}\left(0, \frac{9n}{4\lfloor n/2 \rfloor}\right), \quad \epsilon_{ij} \sim_{iid} \text{N}(0, 3/4),$$

where '$1_{(j \leq \lfloor n/2 \rfloor)}$' is an indicator of index i less than or equal to the floor of $n/2$, for example. Since noise was added to actor relations in the same position in Ξ in each simulation run, the distribution of the relations was not exchangeable: the distribution of the errors would be different for a reordering of the rows and columns of the error array.

Finally, the distribution of the exchangeable (bilinear mixed effects model) error setting is defined in Hoff (2005) and detailed in full below. The bilinear mixed effects model is a generalization of the "additive common shocks" error structure used in simulation studies to justify the dyadic clustering estimator (Cameron and Miller (2014); Aronow et al. (2015); Tabord-Meehan (2018)).

$$y_{ij} = \beta^T x_{ij} + \xi_{ij}, \qquad \xi_{ij} = a_i + b_j + z_i^T z_j + \gamma_{(ij)} + \epsilon_{ij}, \tag{S71}$$

$$(a_i, b_i) \sim \text{N}(0_2, \Sigma_{ab}), \qquad \Sigma_{ab} = \begin{pmatrix} \sigma_a^2 & \rho_{ab}\sigma_a\sigma_b \\ \rho_{ab}\sigma_a\sigma_b & \sigma_b^2 \end{pmatrix},$$

$$z_i, z_j \sim \text{N}_d(0, \sigma_z^2 I_d), \qquad \gamma_{(ij)} = \gamma_{(ji)} \sim \text{N}(0, \sigma_\gamma^2), \qquad \epsilon_{ij} \sim \text{N}(0, \sigma_\epsilon^2),$$

where a_i, b_j, z_i, z_j, and ϵ_{ij} are independent. We selected the dimension of the latent space to be $d = 2$, the correlation between sender and receiver effects as $\rho_{ab} = 1/2$, and the sender variance to be

Table S2: Approximate standard deviations for exchangeable error setting.

σ_ϵ	σ_a	σ_b	σ_γ	σ_z
0.866	0.957	0.677	0.677	0.677

twice that of receiver variance: $\sigma_a^2 = 2\sigma_b^2$. We further specified $\sigma_z = \sigma_\gamma = \sigma_b$. Finally, we selected $\sigma_\epsilon^2 = 3/4$. With the aforementioned choices, the restriction $\sum_{ij} \text{var}(\xi_{ij}) = 3n(n-1)$ generated a quadratic equation in σ_b. The standard deviations that resulted from solving this quadratic equation are shown in Table S2.

We generated 500 random realizations of covariates for relational data sets with numbers of actors $n = 20, 40, 80, 160, 320$. For each covariate realization, 1,000 random error realizations were generated for each of the three error settings: independent and identically distributed, exchangeable, and non-exchangeable. Using the model (S70), a simulated data set was created from each covariate realization and error realization pair. The regression model was fit to each data set using ordinary least squares, and standard errors were estimated using the exchangeable and dyadic clustering sandwich variance estimators. Confidence interval coverage was estimated for each covariate realization by counting the fraction of confidence intervals that contained the true coefficient. Bias and variance of the standard error estimators was estimated for each covariate realization relative to the known true standard errors of the ordinary least squares estimator, conditional on the covariate realization.

S11.2 Bias in simulation study

In addition to the estimated coverages in Figure 2, we plot the estimated bias of the standard error estimators in Figure S2. We note that the bias of the dyadic clustering estimator is generally substantially larger in absolute value, especially in the independent and exchangeable error cases, and most standard error estimators are biased downwards.

To verify Theorem 2, we compared the biases of the dyadic clustering and exchangeable estimators of the variances of the regression coefficients in the simulation studies with exchangeable and independent errors. We computed the mean bias for each variance estimator across all simulations for each error structure, data set size, coefficient, and estimator. Then, we took the ratio of mean absolute biases of the dyadic clustering variance estimator to that of the exchangeable estimator. The ratios, averaged for all n, are given in Table S3. We note that, in all cases, the mean bias of the dyadic clustering estimator was negative, and larger in absolute value than that of the exchangeable estimator. The fact that every ratio is greater than 2 confirms Theorem 2.

Table S3: Mean (across values of n) ratio of average (across covariate and error realizations) absolute bias in dyadic clustering variance estimator to average absolute bias of exchangeable variance estimator for independent and exchangeable error models.

	binary	positive	real-valued
independent	96.48	5.03	123.45
exchangeable	3.04	5.43	5.02

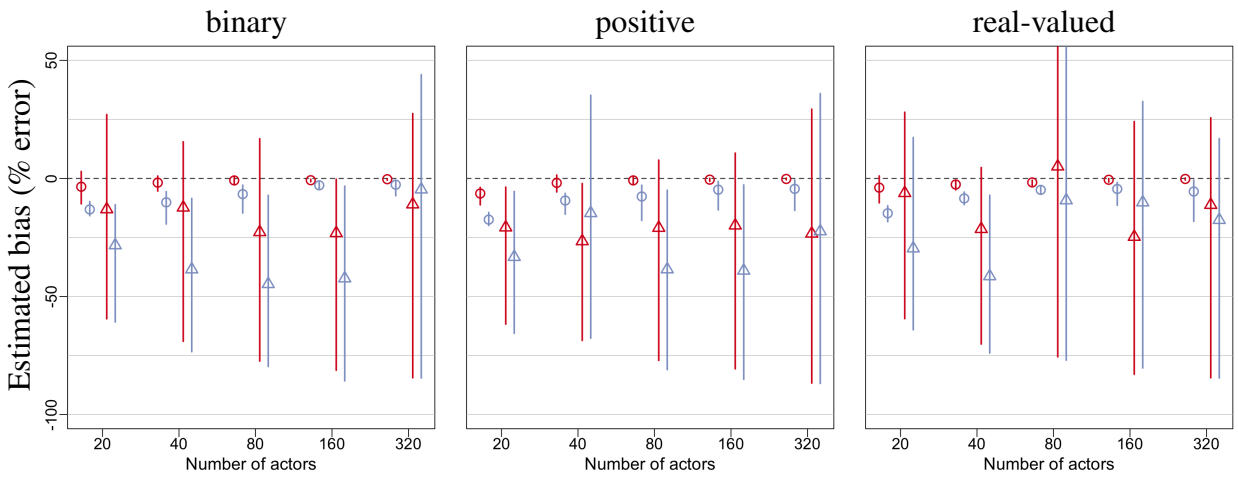

Figure S2: Estimated bias of standard error estimators for each of three covariates (binary, positive, and real-valued) when the errors are generated from exchangeable (circles) and non-exchangeable (triangle) error models. Points represent mean estimated percent error (of standard error estimators) and lines represent 95% quantiles of percent error for exchangeable (red) and dyadic clustering (blue) estimators.

S12 Dyadic clustering covariance matrix invertibility

Ideally, for a covariance matrix estimate $\widehat{\Omega}$ to be of utmost utility, it must be invertible. For example, if we wish to employ feasible generalized least squares, the estimate of the the covariance matrix must be nonsingular. However, in many cases the dyadic clustering estimator is singular and hence cannot be inverted. In cases when the dyadic clustering estimator $\widehat{\Omega}_{DC}$ is singular, it can still be used to produce the dyadic clustering estimator $\widehat{V}_{DC}$ in (5).

Theorem S18. *The dyadic clustering estimate of the error variance, $\widehat{\Omega}_{DC}$, is singular for directed data with $R = 1$.*

Proof. The dyadic clustering estimator can be written as the Hadamard product between the outer product of the residuals and a matrix of indicators of whether the dyad indices share a member,

$$\widehat{\Omega}_{DC} = ee^T \circ 1\{(i,j) \cap (k,l) \neq \varnothing\},$$

where $1\{(i,j) \cap (k,l) \neq \varnothing\}$ is an $n(n-1) \times n(n-1)$ matrix of indicators of indices (i,j) sharing an index with (k,l). The rank of the outer product of the residuals is one: $\text{rank}(ee^T) = 1$. The rank of the indicator matrix is at most $n(n-1)/2$, since the indices (i,j) share a member with an arbitrary pair (k,l) if and only if the indices (j,i) do as well. Thus, the column of $1\{(i,j) \cap (k,l) \neq \varnothing\}$ corresponding to (i,j) is the same as that corresponding to (j,i).

For any two square matrices of equal size A and B, $\text{rank}(A \circ B) \leq \text{rank}(A)\text{rank}(B)$. Thus,

$$\begin{aligned}
\text{rank}(\widehat{\Omega}_{DC}) &= \text{rank}[ee^T \circ 1\{(i,j) \cap (k,l) \neq \varnothing\}], \\
&\leq \text{rank}(ee^T)\text{rank}[1\{(i,j) \cap (k,l) \neq \varnothing\}], \\
&\leq \frac{n(n-1)}{2}.
\end{aligned}$$

$\widehat{\Omega}_{DC}$ is therefore not full rank. □

Remark S19. *Theorem S18 does not hold for undirected data when $R = 1$. If the data are undirected, then the bound does not guarantee singularity of $\widehat{\Omega}_{DC}$ since the dimension of $\widehat{\Omega}_{DC}$ is exactly $n(n-1)/2$. In practice, we find that $\widehat{\Omega}_{DC}$ may be full rank in this case.*

Remark S20. *The result of Theorem S18 holds for both directed and undirected data when $R > 1$. In this case, the column in the indicator matrix $1\{(i,j) \cap (k,l) \neq \varnothing\}$ corresponding to the indices (i,j,s) is the same as that column corresponding to (i,j,r) for all values of $r = 1, \ldots, R$. Thus, again $\widehat{\Omega}_{DC}$ is not full rank.*

S13 Efficient inversion of the exchangeable covariance matrix

To perform the generalized least squares estimation procedure as described in Section 5, we must invert the exchangeable covariance matrix Ω_E as defined in Figure S1. For now, we work in the case where $R = 1$. Since Ω_E is a real symmetric matrix, its inverse is real and symmetric as well. However, we can say more about the patterns in the inverse Ω_E^{-1}. Recall that Ω_E has at most six unique terms; call these parameters ϕ. We find that the inverse Ω_E^{-1} has at most six unique terms as well. If we define the parameters in Ω_E^{-1} as p, we can write

$$\Omega_E(\phi)\Omega_E^{-1}(p) = I, \quad \phi, p \in \mathbb{R}^6,$$

where I is the $n(n-1) \times n(n-1)$ identity. Lastly, we make the conjecture that the parameter pattern in Ω_E^{-1} is exactly the same as that in Ω_E; we find this conjecture to be true in practice. One caveat is that the locations in which we assume zeros in Ω_E are not zero in Ω_E^{-1} in general.

We can find the inverse parameters p from ϕ without inverting the entire matrix Ω_E by instead solving the following linear system

$$C(\phi, n)p = (1,0,0,0,0,0)^T, \quad C(\phi, n) \in \mathbb{R}^{6\times 6},$$

where $C(\phi, n)$ is a set of six linear equations based on the parameters ϕ and the number of actors n and is depicted in Figure S3. Thus, we replace the need to invert the $n(n-1) \times n(n-1)$ matrix Ω_E by the inversion of the 6×6 matrix $C(\phi, n)$. Using this procedure, the computational cost associated with finding the inverse of Ω_E is independent of the number of actors n.

Now consider the case of array data with $R > 1$. Inversion of the exchangeable covariance Ω_E in Figure 1 requires consideration of the patterns in the block matrices. Recalling that Ω_E is parametrized by twelve terms when $R > 1$, we denote the first six parameters as $\phi^{(1)}$ and the second six $\phi^{(2)}$, corresponding to Ω_1 and Ω_2 respectively. Again the inverse Ω^{-1} has the exact same block matrix pattern as Ω. Thus, the inverse may be parametrized by $p^{(1)}$ and $p^{(2)}$, each with length six, defined by the following linear equations.

$$\begin{aligned} C(\phi^{(1)}, n)p^{(1)} + (R-1)C(\phi^{(2)}, n)p^{(2)} &= (1,0,0,0,0,0)^T, \\ C(\phi^{(2)}, n)p^{(1)} + C(\phi^{(1)}, n)p^{(2)} + (R-2)C(\phi^{(2)}, n)p^{(2)} &= 0. \end{aligned}$$

This is twelve linear equations in $p^{(1)}$ and $p^{(2)}$. In this formulation we reduce a $Rn(n-1) \times Rn(n-1)$ inversion to a 12×12 inversion for calculation of Ω_E^{-1}. Again, there is no dependence of the complexity of the inversion on the array dimensions n and R.

$$\begin{bmatrix}
\phi_1 & \phi_2 & (n-2)\phi_3 & (n-2)\phi_4 & 2(n-2)\phi_5 & (n-2)(n-3)\phi_6 \\
\phi_2 & \phi_1 & (n-2)\phi_5 & (n-2)\phi_5 & (n-2)(\phi_3+\phi_4) & (n-2)(n-3)\phi_6 \\
\phi_3 & \phi_5 & \phi_1+(n-3)\phi_3 & \phi_5+(n-3)\phi_6 & \phi_2+\phi_4+(n-3)(\phi_5+\phi_6) & (n-3)\left(\phi_4+\phi_5+(n-4)\phi_6\right) \\
\phi_4 & \phi_5 & \phi_5+(n-3)\phi_6 & \phi_1+(n-3)\phi_4 & \phi_2+\phi_3+(n-3)(\phi_5+\phi_6) & (n-3)\left(\phi_3+\phi_5+(n-4)\phi_6\right) \\
\phi_5 & \phi_4 & \phi_2+(n-3)\phi_5 & \phi_3+(n-3)\phi_6 & \phi_1+\phi_5+(n-3)(\phi_4+\phi_6) & (n-3)\left(\phi_3+\phi_5+(n-4)\phi_6\right) \\
\phi_6 & \phi_6 & \phi_4+\phi_5+(n-4)\phi_6 & \phi_3+\phi_5+(n-4)\phi_6 & \phi_3+\phi_4+2\phi_5+2(n-4)\phi_6 & \phi_1+\phi_2+(n-4)\left(\phi_3+\phi_4+2\phi_5+(n-5)\phi_6\right)
\end{bmatrix}$$

Figure S3: Matrix $C(\boldsymbol{\phi}, n)$.

S14 Trade data analysis

S14.1 Trade model details

We analyzed international trade among 58 countries over 20 years. The data were analyzed and made available (DOI:10.1214/10-AOAS403) by Westveld and Hoff (2011), which used a modified gravity mean model to represent the logarithm of yearly trade between each pair of countries as linear function of seven covariates in years 1981-2000. The gravity model, proposed by Tinbergen (1962), posits that the total trade between countries is proportional to overall economic activity of the countries weighted by the inverse of the distance between them (raised to a power). Following Ward and Hoff (2007) we added an indicator for whether the nations' militaries cooperated in the given year and a measure of democracy, i.e. polity, which ranges from 0 (highly authoritarian) to 20 (highly democratic). The complete model has the form:

$$\begin{aligned}\text{logTrade}_{ijt} =& \beta_{0t} + \beta_{1t}\text{logGDP}_{it} + \beta_{2t}\text{logGDP}_{jt} + \beta_{3t}\text{logD}_{ijt} \\ &+ \beta_{4t}\,\text{Pol}_{it} + \beta_{5t}\,\text{Pol}_{jt} + \beta_{6t}\,\text{CC}_{ijt} + \beta_{7t}\,(\text{Pol}_{it} \times \text{Pol}_{jt}) + \epsilon_{ijt},\end{aligned} \tag{S72}$$

where logTrade_{ijt} is the (log) volume of trade sent from country i to country j in year t; logGDP_{it} and logGDP_{jt} are the (log) Gross Domestic Product of nations i and j, respectively; logD_{ijt} is the (log) geographic distance between nations; CC_{ijt} is the measure of cooperation in conflict (coded as +1 if nations were on the same side of a dispute and -1 if they were on opposing sides); and Pol_{it} and Pol_{jt} are the polity measures for i and j, respectively, where polity ranges from 0 (highly authoritarian) to 20 (highly democratic).

S14.2 Estimation

Let β denote the length $8T$ vector of regression coefficients defined in (S72), where $t = 1, \ldots, T$ and $T = 20$ is the final time period. We estimated β using feasible generalized least squares as in (2), where y is a vector consisting of (logTrade_{ijt}) where $(i = 1, \ldots, n; j = 1, \ldots, n; i \neq j; t = 1, \ldots, T)$, X is a matrix of corresponding covariates defined by the model in (S72), and we assume Ω is jointly exchangeable. We initialized β using ordinary least squares. Then, we estimated Ω using the residuals and the joint exchangeability assumption, as in (6) and (7). We then used (2) with $\Omega = \widehat{\Omega}$ to re-estimate β. We iterated estimates of Ω and β until convergence. We defined convergence as an absolute change in weighted residual inner product being less than a particular tolerance $\epsilon > 0$. Explicitly, convergence is defined to occur when

$$\begin{aligned}&|Q^{\gamma} - Q^{\gamma-1}| < \epsilon, \\ &Q^{\gamma} = (y - X\widehat{\beta}^{\gamma})^T(\widehat{\Omega}^{\gamma})^{-1}(y - X\widehat{\beta}^{\gamma}),\ (\gamma = 1, 2, \ldots),\end{aligned}$$

where $\widehat{\beta}^{\gamma}$ and $\widehat{\Omega}^{\gamma}$ are the estimators at iteration γ of the regression coefficients and error covariance matrix, respectively. For the trade data analysis, we take $\epsilon = 10^{-6}$. After convergence, we set $\Omega = \widetilde{\Omega}$ in (4) to obtain standard errors for the estimated regression coefficients, which is the standard approach in feasible generalized least squares.

S14.3 In-sample evaluation

We compared the coefficients and confidence intervals resulting from the proposed feasible generalized least squares approach to the estimated coefficients and credible intervals from the mixed effects model of Westveld and Hoff (2011), and to coefficients estimated by ordinary least squares and confidence intervals using the dyadic clustering estimator $\widehat{V}_{DC}$. The estimated coefficients and corresponding 95% confidence intervals and posterior credible intervals are shown in Figure S4.

S14.4 Prediction study

To compare the ability of the proposed exchangeable approach and the model from Westveld and Hoff (2011) – which we refer to as the hierarchical, longitudinal mixed effects model – to represent the trade data, we examined the out-of-sample predictive performance of the estimators. We detail the computation of predicted values here. First, we establish some notation for this section. Recall that the trade data set has covariate measures in X that vary by year. Thus, we rewrite the linear model in (S72) as

$$y_t = X_t\beta_t + \Xi_t, \quad (t = 1, \ldots, T),$$

where y_t represents the $n(n-1)$ vector of relations among the $n = 58$ countries in year t, X_t is a matrix eight covariates corresponding to year t, and Ξ_t is a vector of errors for year t.

We estimated both models on the first 4 through 19 years of data, and then used these estimates to predict the trade in the following year. To generate predictions from each model, we computed the conditional expectation $E(y_T \mid \{y_t\}_{t=1}^{T-1})$ based on the assumption that y_T and (y_r), for $r = 1, \ldots, T-1$, are jointly normal. The exchangeable approach and mixed effects model correspond to different models of the variance-covariance matrices of Ξ_t and the covariances matrices between vectors Ξ_t and Ξ_{t+h}. As a baseline, we included ordinary least squares, assuming independence of each year, i.e. Ξ_t independent of Ξ_{t+h}.

We first discuss the ordinary least squares estimation and prediction procedure used for this study, as this approach is the simplest. We estimated the coefficients in the prediction time period, β_T, with the coefficients from the previous time period and assumed independent and identically distributed entries in all Ξ_t. Thus, the ordinary least squares estimator of the trade in year T is

$$E(y_T \mid \{y_t\}_{t=1}^{T-1})_{OLS} = X_T\widehat{\beta}_{T-1}.$$

For the exchangeable procedure, we again set $\widehat{\beta}_T = \widehat{\beta}_{T-1}$. Based on the model underlying the exchangeable procedure (Figure 6(a)), the variance $\text{var}(y_t) = \Omega_1$ for all $t = 1, \ldots, T$ and the covariance $\text{cov}(y_t, y_{t+h}) = \Omega_2$ for all h. Further, it can be shown that the precision corresponding to the concatenated vector $z_{T-1}^T = (y_1^T, y_2^T, \ldots, y_{T-1}^T)$ is of the same pattern as the variance $\text{var}(z_{T-1})$, which has Ω_1 along the diagonal blocks and Ω_2 in the off-diagonal blocks (as in Figure 6(a)). We define the diagonal blocks of $\text{var}(z_{T-1})^{-1}$ as Ψ_1 and the off-diagonal blocks Ψ_2. Then, when the relations $\{y_t\}_{t=1}^T$ are jointly normally distributed, we define the prediction from the exchangeable procedure is

$$E(y_T \mid \{y_t\}_{t=1}^{T-1})_E = X_T\widehat{\beta}_{T-1} + \Omega_2\{\Psi_1 + (T-2)\Psi_2\}\sum_{t=1}^{T-1}\left(y_t - X_t\widehat{\beta}_t\right).$$

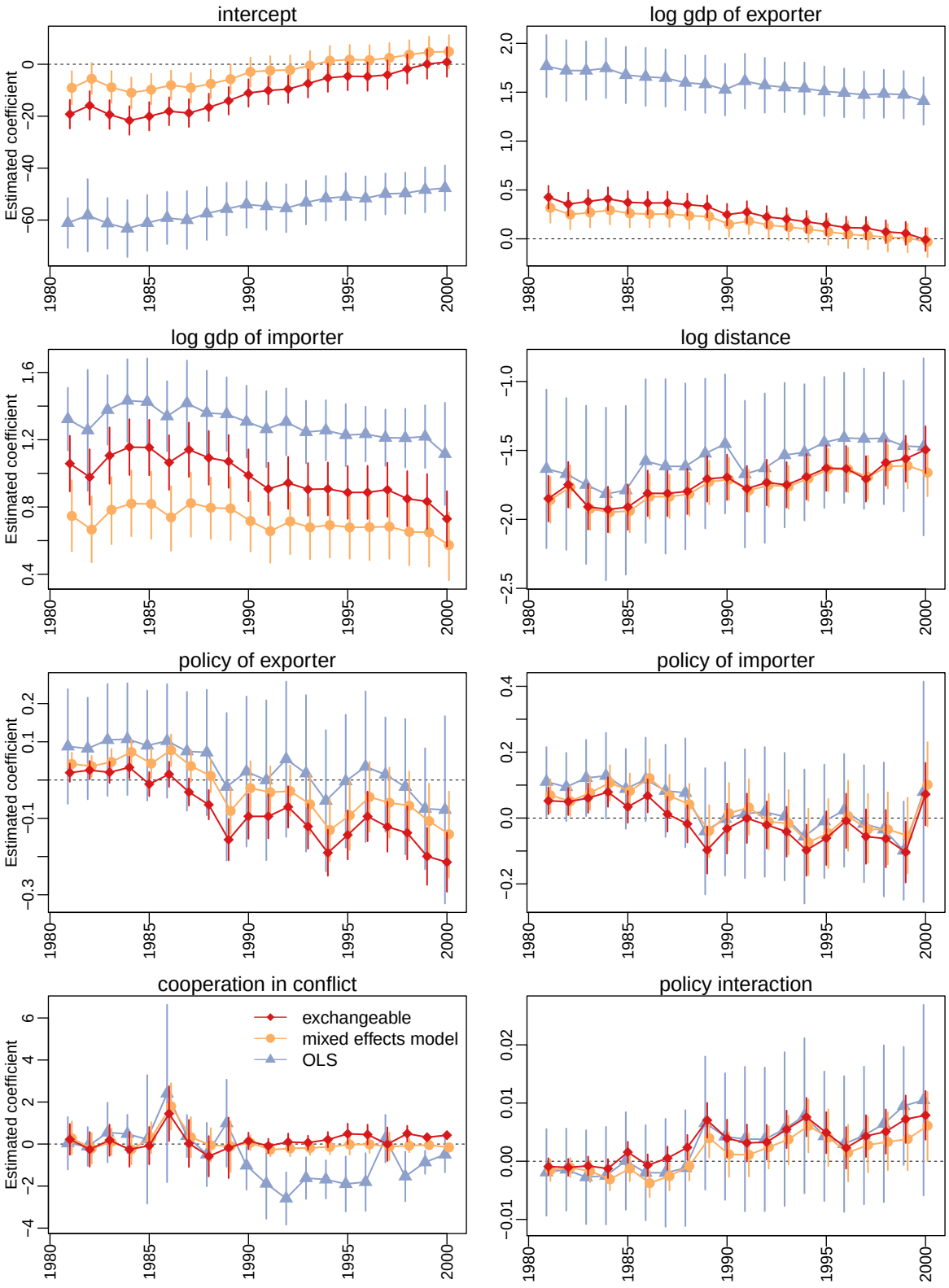

Figure S4: Estimated coefficients and 95% confidence/credible intervals using three different estimation techniques: exchangeable (red/diamonds), mixed effects model (orange/circles), and ordinary least squares with dyadic clustering estimator ("OLS," triangles/blue).

Finally, to detail predictions from the mixed effects model, we first review some terms used in Westveld and Hoff (2011). The proposed model is

$$y_{ij,t} = x_{ij,t}^T \beta_t + s_{i,t} + r_{j,t} + g_{ij,t},$$

where $s_{i,t}$ and $r_{j,t}$ are sender and receiver random effects, respectively, and $g_{ij,t}$ is a reciprocal random effect. These effects evolve according to autoregressive order one processes, such that

$$\begin{bmatrix} s_{i,t} \\ r_{i,t} \end{bmatrix} = \Phi_{sr} \begin{bmatrix} s_{i,t-1} \\ r_{i,t-1} \end{bmatrix} + \varepsilon_{i,t}, \qquad (i = 1, \ldots, T;\ t = 1, \ldots, T-1),$$
$$\begin{bmatrix} g_{ij,t} \\ g_{ji,t} \end{bmatrix} = \Phi_{g} \begin{bmatrix} g_{ij,t-1} \\ g_{ji,t-1} \end{bmatrix} + e_{ij,t}, \qquad (i, j = 1, \ldots, T;\ i \neq j;\ t = 1, \ldots, T-1),$$

where Φ_{sr} is a 2×2 autoregressive matrix and Φ_g is a symmetric autoregressive matrix. The error terms $\varepsilon_{i,t}$ and $e_{ij,t}$ are mean-zero independent bivariate Gaussian random variables. Please see Westveld and Hoff (2011) for more details of the model. With the notation and model defined, and again setting $\widehat{\beta}_T = \widehat{\beta}_{T-1}$, algebra reveals that the prediction from the mixed effects model is

$$E(y_{ij,T} \mid \{y_t\}_{t=1}^{T-1})_{MEM} = x_{ij,T}^T \widehat{\beta}_{T-1} + \begin{bmatrix} 1 & 0 \end{bmatrix}^T \widehat{\Phi}_{sr} \begin{bmatrix} \widehat{s}_{i,T-1} \\ \widehat{r}_{i,T-1} \end{bmatrix} \ldots$$
$$+ \begin{bmatrix} 0 & 1 \end{bmatrix}^T \widehat{\Phi}_{sr} \begin{bmatrix} \widehat{s}_{j,T-1} \\ \widehat{r}_{j,T-1} \end{bmatrix} + \begin{bmatrix} 1 & 0 \end{bmatrix}^T \widehat{\Phi}_{g} \begin{bmatrix} \widehat{g}_{ij,T-1} \\ \widehat{g}_{ji,T-1} \end{bmatrix}.$$

To estimate the mixed effects model, Westveld and Hoff (2011) generated a Markov chain of length 55,000, of which the first 10,000 were discarded and every 20th iteration saved, giving 2,250 samples to approximate the posterior distributions of β_{T-1}, Φ_{sr}, Φ_g, and the random effects at $t = T - 1$ in the above equations. The mean of these 2,250 samples from the joint posterior distribution, for each pair $i \neq j$, was used to construct $E(y_T \mid \{y_t\}_{t=1}^{T-1})_{MEM}$.

We evaluate performance in the predicted year using the mean square prediction error, defined

$$MSPE = \frac{1}{n(n-1)} \left|\left| E(y_T \mid \{y_t\}_{t=1}^{T-1}) - y_T \right|\right|_2^2,$$

where the expectation is replaced by the expectation of the appropriate prediction estimator (ordinary least squares, exchangeable, or mixed effects).

S15 A test for exchangeability

S15.1 Motivation and details

In this section we briefly motivate and describe a permutation test for assumption of exchangeability of the errors. For simplicity, we focus on the case where ordinary least squares is used to estimate regression coefficients β, and where $R = 1$.

If the errors are truly non-exchangeable in a way that impacts inference, then the dyadic clustering estimator should be extreme relative to the distribution of the dyadic clustering estimator under the assumption of exchangeability. The idea of the proposed test is to generate the null (exchangeable) distribution using permutations that mimic the exchangeability assumption.

If the exchangeable assumption is correct, then by definition, a simultaneous permutation of the rows and columns of the error matrix leaves the distribution unchanged. That is,

$$pr(\xi_{ij}) = pr(\xi_{\pi(i)\pi(j)}),$$

where $pr(\xi_{ij})$ denotes the joint distribution of the set (ξ_{ij}) over all relations such that $i \neq j$. Although the errors are unobservable, the residuals are observable and can be used to approximate the errors. Thus, the residuals should be approximately exchangeable, such that

$$pr(e_{ij}) \approx pr(e_{\pi(i)\pi(j)}),$$

where $e_{ij} = y_{ij} - \widehat{\beta}^T x_{ij}$ is the residual for relation (i, j). We rewrite the dyadic clustering estimator of the variance var$(\widehat{\beta} \mid X)$ as

$$\widehat{V}_{DC} = \sum_{(jk,lm)\in\Theta_0} e_{jk} e_{lm} \widetilde{x}_{jk} \widetilde{x}_{lm}^T, \tag{S73}$$

where Θ_0 represents the set of pairs of relations (jk, lm) that share an actor and where $\widetilde{x}_{jk} = (X^T X)^{-1} x_{jk}$. Applying the joint exchangeability assumption to the residuals, one may generate an approximate sample of the dyadic clustering estimator under the null assumption of exchangeability

$$\widetilde{V}_{DC}^{(\pi)} \approx \sum_{(jk,lm)\in\Theta_0} e_{\pi(j)\pi(k)} e_{\pi(l)\pi(m)} \widetilde{x}_{jk} \widetilde{x}_{lm}^T. \tag{S74}$$

We use an L_2 norm to evaluate the discrepancy between the observed dyadic clustering estimator based on the observed residual matrix in (S73) and the distribution of dyadic clustering estimators under the exchangeable null in (S74). The squared distance can be computed using

$$\bar{d}^{(\pi)} = ||\widehat{V}_{DC} - \bar{V}_{DC}^{(\pi)}||_2^2. \tag{S75}$$

where $\bar{V}_{DC}^{(\pi)}$ is the element-wise sample mean dyadic clustering estimator of the exchangeable permutations of $\widetilde{V}_{DC}^{(\pi)}$ in (S74). A distribution of these squared distances under the exchangeable null may be obtained by replacing the dyadic clustering estimator based on the observed residuals with the permuted samples $\widetilde{V}_{DC}^{(\pi)}$,

$$d^{(\pi)} = ||\widetilde{V}_{DC}^{(\pi)} - \bar{V}_{DC}^{(\pi)}||_2^2. \tag{S76}$$

If $\bar{d}^{(\pi)}$ in (S75) is extreme relative to the distribution of $d^{(\pi)}$ in (S76) under the exchangeable null, then this constitutes evidence that the errors are non-exchangeable. We obtain an empirical p-value by computing the fraction of the permuted, squared distances $d^{(\pi)}$ that are greater than $\bar{d}^{(\pi)}$.

S15.2 Evaluating the test

We simulated from exchangeable and non-exchangeable relational regression models to assess the ability of our proposed hypothesis testing procedure to detect non-exchangeability. We evaluated

the power and size of the test by simulating data from null and alternative models and evaluating the test on each simulated dataset.

We begin by specifying the null exchangeable model. In addition to an intercept, we include a single relational covariate, the coefficient of which is the target of inference. The model is specified as

$$
\begin{aligned}
y_{ij} &= \beta_0 + \beta_1 x_{ij} + \xi_{ij}, \\
v_i &\overset{iid}{\sim} N(0,1),\ i = 1, 2, \ldots, n, \\
x_{ij} &= \frac{1}{\sqrt{2}}(v_i - v_j), \\
w_i &\overset{iid}{\sim} N(0,1),\ i = 1, 2, \ldots, n, \\
\epsilon_{ij} &= \frac{1}{2}(w_i - w_j), \\
\xi_{ij} &= \epsilon_{ij} + \frac{1}{\sqrt{2}} z_{ij}, \\
z_{ij} &\overset{iid}{\sim} N(0,1).
\end{aligned}
\tag{S77}
$$

Such a model produces true variance parameters $\phi^T = [1, 1/2, 1/4, 1/4, -1/4, 0]$. Further, var$(x_{ij}) = 1$, such that the variance of the errors and variance of the covariate x_{ij} are equivalent. We take $\beta_0 = \beta_1 = 1$.

The non-exchangeable model we consider can be expressed as follows:

$$
\begin{aligned}
y_{ij} &= \beta_0 + \beta_1 x_{ij} + \xi_{ij}, \\
v_i &\sim N(0, (i/n)^2),\ i = 1, 2, \ldots, n, \\
x_{ij} &= \frac{1.7}{\sqrt{2}}(v_i - v_j), \\
w_i &\overset{\perp\!\!\!\perp}{\sim} N(0, v_i^2),\ i = 1, 2, \ldots, n, \\
\epsilon_{ij} &= 1.7(w_i - w_j), \\
\xi_{ij} &= (\epsilon_{ij} + \frac{1}{\sqrt{2}} z_{ij})/1.6, \\
z_{ij} &\overset{iid}{\sim} N(0,1).
\end{aligned}
\tag{S78}
$$

Under this model, both X and the errors are non-exchangeable and the variance of the errors errors depend on values of X. The numerical prefactors were chosen such that var$(x_{ij}) \approx 1$ and var$(\xi_{ij}) \approx 1$ as in the exchangeable model in (S77).

To evaluate the power and size of the test, we simulated 1,000 datasets from each of the null and alternative models in (S77) and (S78), respectively. For each simulated dataset, we evaluated the empirical two-sided p-value using 1,000 permutations of the form of (S74), and rejected the null hypothesis at the $\alpha = 0.05$ level. The test results in Table S4 show approximate size, i.e. the probability of rejecting the null when the null is true, of 0.025. Under the considered alternative distribution, Table S4 gives the estimate approximate power of about 0.839. Thus, for the given alternative model at moderate relational data size ($n = 50$), the proposed permutation test has a good chance of detecting the non-exchangeability in the alternative model, while failing to reject a true null with probability less than 0.05.

Table S4: Estimated probability of a Type I error (size) and one less the probability of a Type II error (power) for the proposed permutation test.

Size	Power
0.025	0.839