

Rejoinder to Discussion of “A Tale of Two Datasets:
Representativeness and Generalisability of Inference
for Samples of Networks”

Pavel N. Krivitsky^{*1}, Pietro Coletti², and Niel Hens^{2,3}

¹Department of Statistics and UNSW Data Science Hub
School of Mathematics and Statistics
University of New South Wales
Sydney, Australia

²I-BioStat
Data Science Institute
Hasselt University
Hasselt, Belgium

³Centre for Health Economics and Modelling Infectious Diseases
Vaccine and Infectious Disease Institute
University of Antwerp
Antwerp, Belgium

*p.krivitsky@unsw.edu.au

We thank Prof. Michael Stein for selecting our article for discussion and the four discussants for their praise and criticism alike. The discussants make a number of important points about both our particular analysis and the more general statistical and computational issues surrounding it, and we provide our thoughts on these in turn. It is our hope that this rejoinder will provide both additional guidance for practitioners and directions and suggestions for future research, both in methodology and in applications.

On some points, we can only concur. We see the link between simulation-based residuals and score tests drawn by Schweinberger and Fritz (2023, Sec. 3) as a promising way to formalize the approach proposed in Section 4.3 of the article to diagnose between-network lack-of-fit and suggest additional predictors by regressing residuals on candidates. Our argument is heuristic and empirical, whereas a rigorous characterization of precisely which augmented ERGM constitutes the score test alternative hypothesis for a particular residual regression test would certainly guide the model selection better or suggest better tests. Similarly, their suggestion (Sec. 4) to implement user-friendly score testing is well taken, though care must be taken to ensure that partially observed scenarios are handled correctly.

Niezink (2023, Sec. 2) makes a case for using family roles rather than age-gender categories as predictors of contacts and makes a point that some family compositions are categorically different from others, so the assumption of smooth network size effects implicit in our use of polynomials is questionable. Although we stand behind our model's adequacy based on the diagnostics we present, we agree in principle and had, in fact, considered the family role and categorical composition approach. Unfortunately, the uncertainty inherent in inferring roles and relations from demographics would create an errors-in-variables problem, greatly complicating the analysis. Thus, we reiterate the suggestion we made in the article: that future surveys collect this information directly.

Other points, we address in more detail below.

1 Sample Size, Network Size, and Power

Vega Yon (2023) points out that in determining the sample size requirements for inference for samples of networks, one must take into account both the number of networks to sample (S) and the sizes of the individual networks (n_s). He also notes that in our application, the size distribution of networks in the population is well-defined and exogenous.

We would add that even so, the study designer still has some discretion in oversampling or stratifying on network sizes and compositions. The relative amounts of (Fisher) information contained in networks of different sizes in turn depends on the scaling regime of the model and can range from constant in n_s to linear to quadratic, depending on what network feature (edge count,

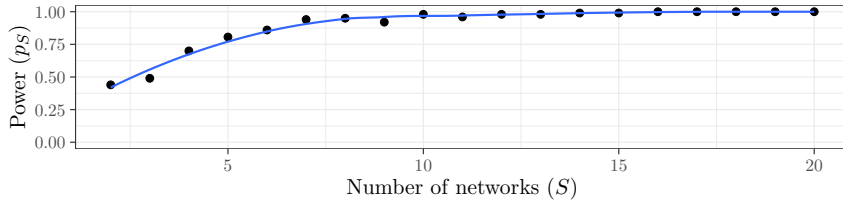


Figure 1: Empirical power p_S for detecting a gender homophily effect with an assumed $\theta_{\text{homophily}} = 1.1$ as a function of number of networks S , where each network has size 8 nodes and 20 edges in the scenario of Vega Yon (2023, Sec. 1).

mean degree, or density, respectively) is asymptotically invariant to n_s under the model (Krivitsky and Kolaczyk, 2015).

In the simplified situation of networks of the same size and with a given edge count, the approach suggested by Vega Yon (2023) can be used to infer the empirical power p_S (as defined by Vega Yon), which we illustrate in Figure 1. This is, of course, only a proof of concept, and should be adapted to the situation at hand.

2 Multicollinearity and Variance Inflation Factors

Vega Yon (2023) raises the question of multicollinearity and variance inflation in ERGMs for small networks. Outside of the extreme scenarios in which multicollinearity leads to outright nonidentifiability (i.e., Section 4.1 and Appendix C in the article), whether it and high VIFs are a serious concern depends on how the parameter estimates are intended to be used (O’Brien, 2007, for example), and we examine their impact on our analysis here.

In developing this discussion, we identified a calculation error that led to incorrect standard errors for the parameter estimates given in Table 1 in the article, with those related to 2-stars, triangles and their interactions with network size too small by an order of magnitude. None of the reported substantive conclusions about overall effects of 2-stars, triangles and network size change. We present the corrected standard errors in Table 1 and the corrected omnibus and contrast tests in Tables 2 and 3, respectively.

A hurdle in using ERGM VIF in our specific case is that it is only well-defined in models with an edge count (i.e., intercept) effect (Duxbury, 2018). Our models do not use an intercept: our mixing effects cover all of the possible age–gender group combinations, visualized in Figure 4 in the article. (In the language of ANOVA, we use the *means*, as opposed to the *effects*, parametrization.)

We can nonetheless examine the impact of multicollinearity on our anal-

Table 1: Parameter estimates and corrected standard errors for *Model 1* and *Model 2*.

Relationship Effect × Network-Level Effect	Coefficient (S.E.)	
	<i>Model 1</i>	<i>Model 2</i>
edges × log(n_s)	−14.28 (3.78) ^{***}	−13.78 (3.69) ^{***}
× log ² (n_s)	5.69 (1.71) ^{***}	5.47 (1.66) ^{**}
if Brussels post code	0.08 (0.19)	−0.02 (0.21)
× log(pop. dens. in post code)		0.04 (0.03)
if on weekend	0.14 (0.06) [*]	0.13 (0.06) [*]
2-stars	1.91 (4.95)	1.14 (4.55)
× log(n_s)	−2.15 (5.87)	−1.22 (5.38)
× log ² (n_s)	0.34 (1.72)	0.07 (1.58)
triangles	5.55 (11.55)	7.30 (10.72)
× log(n_s)	−3.46 (14.29)	−5.65 (13.28)
× log ² (n_s)	0.93 (4.38)	1.60 (4.09)
Young Child with Young Child	8.60 (1.88) ^{***}	8.66 (1.83) ^{***}
Young Child with Preadolescent	9.10 (1.88) ^{***}	9.15 (1.84) ^{***}
Preadolescent with Preadolescent	8.17 (1.85) ^{***}	8.24 (1.81) ^{***}
Adolescent with Adolescent	7.70 (1.84) ^{***}	7.75 (1.80) ^{***}
Young Child with Young Adult	9.64 (2.07) ^{***}	9.67 (2.05) ^{***}
Preadolescent with Young Adult	7.25 (1.85) ^{***}	7.28 (1.81) ^{***}
Adolescent with Young Adult	7.73 (1.80) ^{***}	7.82 (1.77) ^{***}
Young Adult with Young Adult	7.66 (1.85) ^{***}	7.70 (1.81) ^{***}
Young Child with Older Female Adult	10.26 (1.85) ^{***}	10.32 (1.81) ^{***}
Preadolescent with Older Female Adult	9.67 (1.85) ^{***}	9.73 (1.80) ^{***}
Adolescent with Older Female Adult	8.90 (1.84) ^{***}	8.96 (1.80) ^{***}
Older Female Adult with Older Female Adult	7.45 (1.87) ^{***}	7.50 (1.83) ^{***}
Young Child with Older Male Adult	9.09 (1.87) ^{***}	9.14 (1.82) ^{***}
Preadolescent with Older Male Adult	8.76 (1.83) ^{***}	8.83 (1.79) ^{***}
Adolescent with Older Male Adult	8.20 (1.85) ^{***}	8.26 (1.80) ^{***}
Older Female Adult with Older Male Adult	10.11 (1.84) ^{***}	10.17 (1.80) ^{***}
if child absent	−1.22 (0.30) ^{***}	−1.20 (0.30) ^{***}
Older Male Adult with Older Male Adult	6.59 (1.87) ^{***}	6.66 (1.82) ^{***}
Older Female Adult with Senior	8.12 (1.82) ^{***}	8.20 (1.78) ^{***}
Older Male Adult with Senior	7.51 (1.86) ^{***}	7.58 (1.81) ^{***}
Senior with Senior	7.82 (1.81) ^{***}	7.89 (1.77) ^{***}
Adolescent with Young Child or Preadolescent	8.07 (1.85) ^{***}	8.13 (1.80) ^{***}
Young Adult with Older Adult	8.02 (1.84) ^{***}	8.07 (1.80) ^{***}
Young Child or Preadolescent with Senior	8.29 (1.93) ^{***}	8.34 (1.89) ^{***}
Adolescent or Young Adult with Senior	9.93 (2.09) ^{***}	10.01 (2.08) ^{***}

Significance: *** ≤ 0.001 < ** ≤ 0.01 < * ≤ 0.05

Table 2: Selected omnibus tests for *Model 1*.

Effects	Wald χ^2 (df)	<i>P</i> -val.
any 2-star	20.9 (3)	< 0.001
any triangle	98.7 (3)	< 0.001
any $\log(n_s)$ or $\log^2(n_s)$	78.9 (6)	< 0.001
any $\log^2(n_s)$	17.7 (3)	0.001
2-star or triangle $\log^2(n_s)$	9.4 (2)	0.009

Table 3: Selected contrasts for *Model 1*.

Contrast	Estimate (S.E.)	<i>P</i> -val.
Older Female vs. Male Adults with Young Children	1.16 (0.47)	0.013
Older Female vs. Male Adults with Preadolescents	0.91 (0.29)	0.002
Older Female vs. Male Adults with Adolescents	0.70 (0.25)	0.006
Older Female > Male Adults with Seniors (one-tailed)	0.61 (0.31)	0.026

ysis directly. Figure 2 (left) shows $\widehat{\text{Cor}}(\text{vec } \hat{\beta})$ for a representative selection of parameters of *Model 1*. There are particularly strong correlations (positive and negative, following a “checkerboard” pattern across the polynomial network size effects) among the density (edges) and the endogenous (2-stars and triangles) effects and between those effects and the mixing effects. Among the mixing effects, all correlations are strong and positive. Referring back to Table 1, as expected we observe particularly high standard errors for some of these endogenous effects; and if the VIF were well-defined, it would certainly be very high for all but a few of the parameter estimates.

However, this turns out to have little effect on the substantive conclusions. The endogenous effects are not individually significant in Table 1, but when we perform Wald omnibus tests shown in Table 2 (and Sec. 5.3 of the article), we find that the alternating signs of the coefficients combine with the “checkerboard” correlations to produce highly significant χ^2 statistics. Then, when we visualize the network size effects, we take a linear combination of the estimates with positive coefficients—and the negative correlations result in the predicted effect for a given network size being far more precise than individual parameter estimates (Figure 3).

Lastly, but perhaps most importantly, when we compare different mixing cells by taking contrasts, a strong positive correlation between these estimates means that the contrasts in Table 3 are more precise than their constituents in Table 1.

To examine the VIFs themselves, we reparametrize *Model 1* to use an

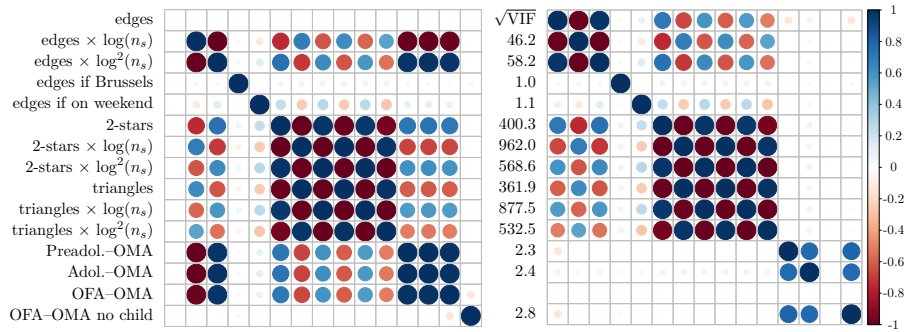


Figure 2: Correlation matrices of selected parameter estimates for *Model 1* (left) and its effects reparametrization with root-variance-inflation-factors (right). Most mixing effects have been omitted for space and follow the same pattern as those shown.

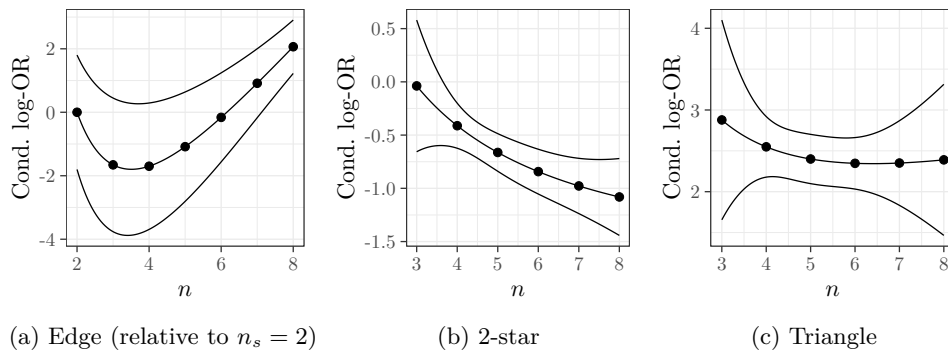


Figure 3: Estimated effects of network size on conditional log-odds of an instance of a graph feature with 1-standard-error bounds.

edge count statistic and omit one of the mixing cells (the baseline), specifically, the Older Female Adult – Older Male Adult cell, because it has the largest sample size. Figure 2 (right) shows the correlations and the root-VIFs that result. We see that the intercept has “absorbed” the correlation between the mixing effects and the network size and endogenous effects, as well as some of the correlation among the mixing effects. Thus, VIFs are extremely high for the endogenous and network-size effects but modest for the mixing effects. However, it is important to keep in mind that those latter VIFs are for very specific contrasts—those with the baseline cell—which are not of substantive interest.

3 Model Selection and Dyad-Dependent Effects

Niezink (2023, Sec. 1) highlights an important problem of efficient model selection for ERGMs. In our application in particular, most of the effects in the model were determined by substantive and inferential considerations, with only a few selected using diagnostics and AIC. More generally, however, the many possible relational effects representable by increasingly rich network data combined with the network-level effects on those effects can create a combinatorial explosion of possibilities.

Then, indeed, starting with dyad-independent effects and only adding dyad-dependent effects at the end as needed may turn out to be the only computationally feasible approach for many scenarios; but this approach is not without caveats. Model selection techniques such as stepwise regression can be sensitive to the order in which effects are considered for inclusion. In the case of ERGMs, theoretical and empirical studies are needed to confirm whether or not a model selection procedure that first selects dyad-independent effects and then selects dyad-dependent effects would produce systematically different results from one that selected dyad-dependent effects first or alongside. A corollary for our application would be that say, child effect choice may have been different if we had used residuals from a dyad-independent “*Model 0d*” to select it.

It may be possible to achieve the best of both worlds by substituting log-pseudolikelihood (Strauss and Ikeda, 1990) for the log-likelihood in the fit criterion (AIC or BIC); but whether this will also lead to systematically different results compared to selection using the true likelihood is an open question as well. (A further caveat is that for partially observed networks, the correct pseudolikelihood may not be that much easier to compute than the likelihood itself; see, for example, remarks by Krivitsky (2017, Sec. 5).)

4 Data for Between-household Relations

Between-household contacts, discussed by Schweinberger and Fritz (2023), present a distinct modeling challenge from data collection perspective. The demographic-level information collected in contact diaries (and egocentric surveys in general) typically does not enable unambiguous identification of contacted individuals. This necessitates an approach such as that of Krivitsky and Morris (2017), in which survey data are used to estimate the global network statistics of interest, which are then used to estimate ERGM parameters by sufficiency. However, while residing in the same household is the strongest form of propinquity, between-household contacts are likely to have a very strong geographic propinquity effect as well, which cannot be modeled unless the contacts' geographic information is collected.

Moreover, one mechanism for violation of the local dependence assumption is that the activities household members participate in affect the connections both between and within households. Future studies should therefore collect the necessary information about overlapping social circles or activities people engage in, and perhaps models of Wang et al. (2016) could be of use.

5 User Guidance

Niezink (2023) lists a number of frameworks and software tools that address a similar scenario to `ergm.multi` and asks which should be used when. Here, we would draw a distinction between the choice of a *statistical model* and the choice of a *software package* implementing the model. As we note in Section 3.2 of the article, our framework is a fixed-effects special case of that of Slaughter and Koehly (2016). Among the fixed-effects work listed, our reading of the work of Vega Yon et al. (2021) and Stewart et al. (2019) is that the classes of models they span are either equivalent to or are subclasses of ours, the latter perhaps of the curved case outlined in Appendix B of the article. Thus, as of this writing, the questions the practitioner should ask are 1) whether the networks in the sample can be safely assumed to be independent, and if not, whether between-network models such as those of Wang et al. (2016) need to be specified; and 2) whether the fixed-effects model is adequate or leaves unaccounted-for heterogeneity that calls for a mixed-effects model.

Statistical software packages, on the other hand, are ever-evolving and interdependent, and not all of the above-listed models come with user-friendly and flexible code (e.g., Slaughter and Koehly, 2016). Thus, the best advice we can offer to the practitioner here is to research the workshops, the tutorials, and the vignettes provided with the packages and on their web sites, as well as recent papers that apply them, because those

sources provide a continually updated view of a package’s capabilities.

On a more positive note, this interdependence can make capabilities “emerge” without additional developer effort: for example, package `Bergm` for Bayesian estimation of ERGMs (Caimo et al., 2022) does not have facilities for samples of networks, but it uses `ergm` (Krivitsky et al., 2023) as its back-end and so can use all ERGM specifications implemented for `ergm`—including those provided by `ergm.multi` for modeling samples of networks. Thus, their combination may already be able to fit at least some special cases of Slaughter and Koehly (2016).

Lastly, we concur that when demonstrating diagnostic techniques, it is important to demonstrate both fit and lack-of-fit look like. Appendix F of our article provides a sampling of those, but tutorials and similar can and should be developed as well.

References

- Alberto Caimo, Lampros Bouranis, Robert Krause, and Nial Friel. Statistical network analysis with `Bergm`. *Journal of Statistical Software*, 104(1), 2022. doi:10.18637/jss.v104.i01. [Cited on page 8.]
- Scott W. Duxbury. Diagnosing multicollinearity in exponential random graph models. *Sociological Methods & Research*, 50(2):491–530, jul 2018. doi:10.1177/0049124118782543. [Cited on page 2.]
- Pavel N. Krivitsky. Using contrastive divergence to seed Monte Carlo MLE for exponential-family random graph models. *Computational Statistics & Data Analysis*, 107:149–161, March 2017. doi:10.1016/j.csda.2016.10.015. [Cited on page 6.]
- Pavel N. Krivitsky and Eric D. Kolaczyk. On the question of effective sample size in network modeling: An asymptotic inquiry. *Statistical Science*, 30(2):184–198, 2015. doi:10.1214/14-STS502. [Cited on page 2.]
- Pavel N. Krivitsky and Martina Morris. Inference for social network models from egocentrically-sampled data, with application to understanding persistent racial disparities in HIV prevalence in the US. *Annals of Applied Statistics*, 11(1):427–455, 2017. doi:10.1214/16-AOAS1010. [Cited on page 7.]
- Pavel N. Krivitsky, David R. Hunter, Martina Morris, and Chad Klumb. `ergm` 4: New features for analyzing exponential-family random graph models. *Journal of Statistical Software*, 105(6):1–44, 2023. doi:10.18637/jss.v105.i06. [Cited on page 8.]
- Nynke M. D. Niezink. Discussion of “A Tale of Two Datasets: Representativeness and Generalisability of Inference for Samples of Networks” by

- Pavel N. Krivitsky, Pietro Coletti, and Niel Hens. *Journal of the American Statistical Association*, To appear, 2023. [Cited on pages 1, 6, and 7.]
- Robert M. O’Brien. A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, 41(5):673–690, mar 2007. doi:10.1007/s11135-006-9018-6. [Cited on page 2.]
- Michael Schweinberger and Cornelius Fritz. Discussion of “A Tale of Two Datasets: Representativeness and Generalisability of Inference for Samples of Networks” by Pavel N. Krivitsky, Pietro Coletti, and Niel Hens. *Journal of the American Statistical Association*, To appear, 2023. [Cited on pages 1 and 7.]
- Andrew J. Slaughter and Laura M. Koehly. Multilevel models for social networks: Hierarchical Bayesian approaches to exponential random graph modeling. *Social Networks*, 44:334–345, 2016. doi:10.1016/j.socnet.2015.11.002. [Cited on pages 7 and 8.]
- Jonathan Stewart, Michael Schweinberger, Michal Bojanowski, and Martina Morris. Multilevel network data facilitate statistical inference for curved ERGMs with geometrically weighted terms. *Social Networks*, 59:98–119, 2019. doi:https://doi.org/10.1016/j.socnet.2018.11.003. [Cited on page 7.]
- David Strauss and Michael Ikeda. Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association*, 85(409):204–212, 1990. [Cited on page 6.]
- George G. Vega Yon. Discussion of “A Tale of Two Datasets: Representativeness and Generalisability of Inference for Samples of Networks” by Pavel N. Krivitsky, Pietro Coletti, and Niel Hens. *Journal of the American Statistical Association*, To appear, 2023. [Cited on pages 1 and 2.]
- George G. Vega Yon, Andrew Slaughter, and Kayla de la Haye. Exponential random graph models for little networks. *Social Networks*, 64:225–238, 2021. doi:10.1016/j.socnet.2020.07.005. [Cited on page 7.]
- Peng Wang, Garry Robins, Philippa Pattison, and Emmanuel Lazega. Social selection models for multilevel networks. *Social Networks*, 44:346–362, 2016. ISSN 0378-8733. doi:10.1016/j.socnet.2014.12.003. [Cited on page 7.]