



A Tale of Two Datasets: Representativeness and Generalisability of Inference for Samples of Networks

Journal:	<i>Journal of the American Statistical Association</i>
Manuscript ID	JASA-A&CS-2022-0039
Manuscript Type:	Article – Applications & Case Studies
Keywords:	Regression Diagnostics, Exponential-Family Random Graph Model, Missing Data, Network Size, Model-Based Inference
Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.	
data_and_code.zip	

SCHOLARONE™
Manuscripts

A Tale of Two Datasets: Representativeness and Generalisability of Inference for Samples of Networks

Abstract

The last two decades have witnessed considerable progress on foundational aspects of statistical network analysis, but less attention has been paid to the complex statistical issues arising in real-world applications. Here, we consider two samples of within-household contact networks in Belgium generated by different but complementary sampling designs: one smaller but with all contacts in each household observed, the other larger and more representative but recording contacts of only one person per household. We wish to combine their strengths to learn the social forces that shape household contact formation and facilitate simulation for prediction of disease spread, while generalising to the population of households in the region.

To accomplish this, we introduce a flexible framework for specifying multi-network models in the exponential family class and identify the requirements for inference and prediction under this framework to be consistent, identifiable, and generalisable, even when data are incomplete; explore how these requirements may be violated in practice; and develop a suite of quantitative and graphical diagnostics for detecting violations and suggesting improvements to a candidate model. We report on the effects of network size, geography, and household roles on household contact patterns (activity, heterogeneity in activity, and triadic closure).

Keywords: exponential-family random graph model, ERGM, missing data, network size, model-based inference, regression diagnostics

1 Introduction

2 Networks of human interaction provide invaluable insights into epidemiology of directly
3 transmitted infectious disease, and there is a great deal of interest in translating network
4 data into epidemic models (Keeling and Eames 2005, for a review). It is common to
5 focus on epidemiologically important settings such as households (Goeyvaerts et al. 2018;
6 Grijalva et al. 2015) and schools (Mastrandrea, Fournet, and Barrat 2015, for example),
7 and such data are often used as they are for simulating disease spread and evaluating the
8 impact of intervention strategies (Cencetti et al. 2021, for example). However, observation
9 of larger and broader epidemiologically relevant networks is limited by time, resources, and
10 considerations such as privacy, so it is often indirect or incomplete in a variety of ways,
11 therefore requiring statistical models to learn network structure from the available data
12 and reconstruct (simulate) networks consistent with it (Krivitsky and Morris 2017).

13 Exponential-Family Random Graph Models (ERGMs), also called p^* models (Wasser-
14 man and Pattison 1996; Lusher, Koskinen, and Robins 2012; Schweinberger et al. 2020,
15 among many), are a popular framework for specifying probability models for networks, pos-
16 tulating an exponential family on the sample space of graphs. Most applications of ERG
17 modelling concern a single, completely observed population network, but methods for in-
18 complete or indirect observation exist (Handcock and Gile 2010; Krivitsky and Morris
19 2017, among others). At the same time, the questions of asymptotics and inference—
20 particularly under varying network sizes—have been debated in the literature (Schwein-
21 berger et al. 2020).

22 Increasingly, networks are collected in samples, however. Examples include social net-
23 works, such as multiple classrooms (Lubbers 2003; Stewart et al. 2019), multiple households
24

(Goeyvaerts et al. 2018; Grijalva et al. 2015), and multiple persons' social support networks (Ersig, Hadley, and Koehly 2011); but also other types of networks such as connections among brain regions for multiple subjects (Schweinberger et al. 2020, Sec. 8).

Though simpler mathematically, inference from samples of independent, non-overlapping networks is no less substantively challenging. In practice, the straightforward "i.i.d." inference scenario (e.g., brain networks) is relatively rare, and a far more prevalent one—particularly for social and contact networks—is that of multiple non-overlapping settings, of similar nature and using the same notion of a relationship, but having varying sizes, compositions, and exogenous influences. This variation is important because size and composition of networks can have profound effects on their structure (Krivitsky, Handcock, and Morris 2011). Furthermore, the selection of networks to be observed may itself be a complex process, and the selected networks themselves may be incompletely observed, requiring network model inference to be integrated with survey sampling inference.

Samples of networks also provide opportunities. With only one network, ways to diagnose how well the model fits—and how well the inference generalises—are limited to working within that network: Hunter, Goodreau, and Handcock (2008) proposed a form of lack-of-fit testing that compares the observed network features not explicitly in the model to the distribution of those features simulated from the fitted model, and Koskinen et al. (2018) leveraged missing data techniques to compute an analogue of Cook's distance for each actor (the effect of observing each actor's observed relations on parameter estimates). On the other hand, models for independent (if heterogeneous) samples of networks can be diagnosed using familiar techniques developed for regression—provided those techniques can be adapted to networks, including partially observed networks.

Thus, ERGM analysis of samples of networks is increasingly common. One popular

1
2
3 approach is meta-analysis, pooling individual networks' estimates (Lubbers 2003). This
4 approach is impractical for large samples of small networks, because the model may be
5 nonidentifiable on each network individually (Vega Yon, Slaughter, and de la Haye 2021).
6 More recent is multilevel (hierarchical) modelling: Zijlstra, Van Duijn, and Snijders (2006)
7 developed it for the related p^2 model, and Slaughter and Koehly (2016) for a Bayesian
8 ERGM with random effects. Vega Yon, Slaughter, and de la Haye (2021) described exact
9 maximum likelihood inference for samples of very small networks. Also, when modelling
10 a time series of networks, *transitions* between successive networks are typically treated as
11 conditionally independent (Leifeld, Cranmer, and Desmarais 2018, for example).

12
13 However, assessing an ERGM's goodness-of-fit for a sample of networks has tended to
14 be limited to comparing distributions of observed network statistics to expected (Slaughter
15 and Koehly 2016; Stewart et al. 2019, for example) and replicating diagnostics of Hunter et
16 al. for each network (Vega Yon, Slaughter, and de la Haye 2021). Little attention has been
17 paid to methods appropriate for partially observed networks, large samples of networks,
18 and to identifying precisely how the model is misspecified.

19 Here, we consider two samples of within-household contact networks: one more complete
20 but restricted to households with a young child, the other larger and more representative
21 but with only one member's relations observed in each household, and both heterogeneous
22 in household sizes and compositions. We wish to fit a probability model to these samples
23 to pool their information and combine their strengths, which will allow us to learn about
24 the social forces affecting the formation of their contacts (i.e., inference) and predict their
25 unobserved relations or other households in the population (i.e., prediction and simulation).
26
27 More generally, we seek to answer three questions:

- 28
29
30 1. What are we estimating when we jointly fit a model to multiple networks?

- 1
- 2
- 3 2. What do we need to assume to combine information from multiple networks?
- 4
- 5 3. How do we test these assumptions?
- 6
- 7

8 In Section 2, we begin to address Question 1 by describing the datasets and applying the
9 principles of model-based survey sampling inference to make explicit assumptions associated
10 with inference from samples of networks that were previously left implicit. In Section 3, we
11 review ERGM inference for missing data, introduce a parametrisation for jointly modelling
12 an ensemble of networks, and discuss its inferential properties—and the requirements for
13 valid inference, addressing Question 2. We then consider in Section 4 the different ways in
14 which these requirements may be violated and combine missing data theory with classic
15 generalised linear model (GLM) diagnostics to produce tools for diagnosing lack of fit
16 in the proposed framework, addressing Question 3; and in addition propose fast model
17 selection techniques for ERGMs for ensembles of networks. Finally, in Section 5, we apply
18 these techniques to our original problem. Additional information can be found in the
19 appendices, referenced throughout. References to appendix figures and tables are prefixed:
20 e.g., Figure B5 is in Appendix B.

32 2 Data and Inferential Questions

33 2.1 Study Designs

34 Two paper-based surveys (Hoang et al. 2021; Goeyvaerts et al. 2018) were conducted in
35 Belgium in 2010–2011, using similar survey instruments but differing in sampling design. In
36 both surveys, recruited by random-digit dialling, participants (or their guardians) reported
37 their and their household members' demographic information and recorded their contacts
38 over the course of one day, including the contacts' ages and genders. Approximate duration
39

	1	2	3	4
Female, 40	1		1	1
Male, 41	2	1		1
Male, 13	3	1	1	
Female, 11	4	1	1	0

(a) H dataset: Contacts among household members for household #11.

	1	2	3	4
Male, 26	1		0	0
Female, 54	2	0		?
Male, 57	3	0	?	
Female, 23	4	1	?	?

(b) E dataset: Report of within-household contacts by ego #8.

Figure 1: Example observation units from the two datasets. Household composition is observed for both, but whereas the H households' relations are completely observed, the alter-alter relations in the E households are missing by design.

and frequency of each contact was also recorded, but for the remainder of this work, we focus our attention on presence or absence of contacts involving skin-to-skin touching.

The first major difference between the surveys is that whereas in the *egocentric* (E) survey (Hoang et al. 2021), only one participant per household was sampled; in the other (Goeyvaerts et al. 2018), the whole *household* (H) was enrolled. This within-household sampling design impacts profoundly the information available about the households: while all contacts are known for the networks in the H dataset (Figure 1a), only contacts incident on the one respondent in the household are known for the E dataset (Figure 1b), though, importantly (Krivitsky and Morris 2017), enough information was collected to identify these contacts uniquely within the household.

The second major difference is that the H survey was restricted to households with a child aged 12 or under, whereas the E survey was not. More incidentally, the surveys differed slightly in their geographical localisation: both surveys included households in the Flemish (Dutch-speaking) areas of Belgium, but only the H survey included participants

1
2
3 from the (majority-French-speaking) Brussels-Capital region of Belgium. Also, both sur-
4 veys called for fine-grained stratification by age, but in practice, the surveyor was not able
5 to adhere to it exactly, and in the *H* survey, households for which any members' contacts
6 were not successfully recorded were dropped altogether.
7
8
9

10 11 2.2 Descriptive Statistics 12

13 Dataset *H* comprises 318 households of size 2–7 for a total of 1266 members/respondents.
14 Requiring less effort per household to collect, *E* comprises 1258 respondents whose house-
15 holds (ranging in size 2–8) have a total of 4000 members with 53% of the households'
16 relationship states observed.
17
18

19 In the *H* dataset, individuals in their mid 20s are underrepresented; *E* is more repre-
20 sentative in this respect, except for omitting individuals living alone or in shared housing.
21 Both datasets' households are on average gender-balanced, but *E*'s respondents have a dif-
22 ferent age and gender distribution from its household members in general, with women aged
23 25–55 overrepresented and adolescents of both genders underrepresented. Most households
24 (*H*: 86%, *E*: 75%) were observed on a weekday; 11% of the *H* households are in Brussels.
25 In the *E* dataset, 34% had a child 12 or under and so could have been in the *H* dataset.
26
27

28 With respect to structural properties of these networks, networks are on average dense
29 (*H*: 93% *E*: 73%), with the *H* dataset's networks being more dense on average (*P*-val. <
30 0.0001). The *H* dataset's networks exhibit high triadic closure (global clustering coefficient
31 $\frac{3 \times \# \text{ triangles}}{\# \text{2-stars}}$ averaging 92%); it cannot be estimated on the partially observed *E* dataset.
32
33

34 More information can be found in Appendix A.
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

2.3 Implications for Inference

E dataset is representative but comprises egocentric, incomplete networks; *H* dataset is very selective but of complete networks. *E* generalises better to the population of Flanders. *H* allows higher-order (e.g., triadic) effects to be estimated, and includes Brussels. Combining information from multiple surveys with different strengths is not uncommon, and a variety of approaches can be used (Elliott, Raghunathan, and Schenker 2018).

These data and the substantive problem are particularly amenable to a model-based approach: the ERGM framework seamlessly integrates exogenous (e.g., age) and endogenous (e.g., friend-of-a-friend) effects likely to be relevant. The model-based approach is also feasible: unlike some egocentric data (Krivitsky and Morris 2017) each respondent's contacts in *E* could be identified uniquely within the household, so model-based inference of Handcock and Gile (2010) is possible.

For the purposes of prediction (e.g., given the distribution of household composition, how would an infection brought home from school spread?) we require the analysis to generalise to the population of households in Flanders and Brussels. The missing information principle (Orchard and Woodbury 1972; Breckling et al. 1994) suggests that if the model is accurate enough, it can be generalised to the population despite the heterogeneous and biased sample. More precisely, we require that the sampling process be *ignorable* or, if viewed as a missing data process, *missing at random*: the unobserved relationship states must be conditionally independent of the selection process given the model and what is observed (Rubin 1976; Handcock and Gile 2010). In our case, this also means that the model must render the dataset from which the network had come ignorable. To accomplish this, it must also account for network size and composition effects.

For the purposes of inference (e.g., do mothers have more contact with their children

than fathers?), we in addition require consistency and a sampling distribution for the parameter estimates. Fortunately, we can treat these networks as an independent sample: the probability that any member of any of the households in either sample has interacted with a member of one of the other households in either sample is low, and such an interaction is unlikely to affect within-household information in a systematic way in the first place. However, there are further nuances, discussed in Section 4.1.

3 Model Specification and Inference

Here, we review the ERGM framework and inference for it, then describe a framework for specifying ERGMs for multiple networks and its inference in turn. We refer the reader to the text book by Lusher, Koskinen, and Robins (2012) and a review by Schweinberger et al. (2020) for detailed discussions of ERGMs' formulation, interpretation, and inference.

3.1 Exponential-Family Random Graph Models for Completely and Partially Observed Networks

Let $N = \{1, 2, \dots, n\}$, for $n \geq 2$, be the set of actors whose relations are of interest. Since physical contacts are inherently two-way, we will focus on undirected graphs: the set of potential relations of interest $\mathbb{Y} \subseteq \{\{i, j\} \in N \times N : i \neq j\}$ is a subset of the set of *dyads*—distinct unordered pairs of actors. Then, the set of possible graphs of interest $\mathcal{Y} \subseteq 2^{\mathbb{Y}}$ (the set of all possible subsets of \mathbb{Y}). We use $\mathbf{y} \in \mathcal{Y}$ for the graph data structure, and $y_{i,j} \in \{0, 1\}$ as indicator of i and j being connected in \mathbf{y} (with $y_{i,j} \equiv y_{j,i}$).

An ERGM is specified by its sample space \mathcal{Y} , a collection $\mathbf{x} \in \mathcal{X}$ of quantitative and categorical exogenous attributes of actors (e.g., age and gender) or dyads (e.g., distance) used as predictors, and a (sufficient by construction) statistic $\mathbf{g} : \mathcal{Y} \times \mathcal{X} \mapsto \mathbb{R}^p$. This statis-

tic operationalises network features believed or hypothesised to be affecting the network process. With free model parameters $\boldsymbol{\theta} \in \mathbb{R}^p$, a random graph $\mathbf{Y} \sim \text{ERGM}_{\mathcal{Y}, \mathbf{x}, \mathbf{g}}(\boldsymbol{\theta})$ if

$$\Pr_{\mathcal{Y}, \mathbf{x}, \mathbf{g}}(\mathbf{Y} = \mathbf{y}; \boldsymbol{\theta}) = \exp\{\boldsymbol{\theta} \cdot \mathbf{g}(\mathbf{y}, \mathbf{x})\}/\kappa_{\mathcal{Y}, \mathbf{x}, \mathbf{g}}(\boldsymbol{\theta}), \quad \mathbf{y} \in \mathcal{Y},$$

where $\kappa_{\mathcal{Y}, \mathbf{x}, \mathbf{g}}(\boldsymbol{\theta}) = \sum_{\mathbf{y}' \in \mathcal{Y}} \exp\{\boldsymbol{\theta} \cdot \mathbf{g}(\mathbf{y}', \mathbf{x})\}$ is the normalising constant. For the sake of brevity, we will omit specification elements “ \mathcal{Y} ”, “ \mathbf{x} ”, “ \mathbf{g} ”, and “ $\boldsymbol{\theta}$ ” where unambiguous.

Network statistics that we will use in this work include the edge count $|\mathbf{y}|$ used to model propensity to have relations; edge counts within or between exogenous groups of actors to model homophily and other types of mixing; and endogenous effects: count of 2-stars $g_{2\text{-star}}(\mathbf{y}) = \sum_{i=1}^n \binom{|\mathbf{y}_i|}{2}$ (where $|\mathbf{y}_i|$ is the degree—number of ties incident on actor i) to model degree heterogeneity and count of triangles $g_{\text{triangles}}(\mathbf{y}) = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n y_{i,j} y_{j,k} y_{k,i}/6$ to model triadic closure. Ordinarily, we would be wary of using the latter two, because of their well-known tendency to induce badly behaved “degenerate” models in large networks and instead use less degeneracy-prone effects (Schweinberger et al. 2020, Sec. 3.1 for context and history). However, this application’s networks are very small and thus largely unaffected, so we use them for their simplicity.

With respect to this ERGM, we may take expectations $\mathbb{E}(\cdot)$ and variances $\text{Var}(\cdot)$, including those of the sufficient statistic: let $\boldsymbol{\mu}(\boldsymbol{\theta}) \stackrel{\text{def}}{=} \mathbb{E}\{\mathbf{g}(\mathbf{Y})\}$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta}) \stackrel{\text{def}}{=} \text{Var}\{\mathbf{g}(\mathbf{Y})\}$.

Given a network \mathbf{y} , an ERGM is typically estimated by maximum likelihood, with $l(\boldsymbol{\theta}) \stackrel{\text{def}}{=} \log \Pr(\mathbf{Y} = \mathbf{y}; \boldsymbol{\theta})$ and Fisher information $\mathcal{I}(\boldsymbol{\theta}) = -l''(\boldsymbol{\theta}) = \boldsymbol{\Sigma}(\boldsymbol{\theta})$. For most interesting models, the normalising constant $\kappa(\boldsymbol{\theta})$ is intractable, and estimation requires MCMC-based techniques (see Schweinberger et al. 2020, Sec. 1.2.1 references).

If the network is incompletely observed, likelihood estimation proceeds as follows (Handcock and Gile 2010): to the unobserved true population network \mathbf{y} , an observation process $\text{obs}(\cdot)$ (deterministic or conditioned-on) is applied, producing $\mathbf{y}^{\text{obs}} \stackrel{\text{def}}{=} \text{obs}(\mathbf{y})$, an observed

1
2
3 data structure with $y_{i,j}^{\text{obs}} \in \{0, 1, \text{NA}\}$ representing observed-absent, observed-present, and
4 unobserved potential relations, respectively. For the E dataset, $\text{obs}(\mathbf{y})$ is such that
5
6

$$7 \quad y_{i,j}^{\text{obs}} \equiv \begin{cases} y_{i,j} & \text{if } i = 1 \vee j = 1, \\ 8 \quad \text{NA} & \text{otherwise.} \end{cases}$$

9
10
11 Let $\mathcal{Y}(\mathbf{y}^{\text{obs}}) \stackrel{\text{def}}{=} \{\mathbf{y}' \in \mathcal{Y} : \text{obs}(\mathbf{y}') = \mathbf{y}^{\text{obs}}\}$: all complete networks that could have
12 produced \mathbf{y}^{obs} or, equivalently, all possible imputations of unobserved relations of \mathbf{y}^{obs} ; and
13 define conditional expectation
14
15

$$16 \quad \boldsymbol{\mu}(\boldsymbol{\theta} | \mathbf{y}^{\text{obs}}) \stackrel{\text{def}}{=} \mathbb{E}\{\mathbf{g}(\mathbf{Y}) | \mathbf{Y} \in \mathcal{Y}(\mathbf{y}^{\text{obs}})\}$$

17
18
19 and, analogously, conditional covariance $\boldsymbol{\Sigma}(\boldsymbol{\theta} | \mathbf{y}^{\text{obs}})$. Then, under noninformative sampling
20 and/or missingness at random, the face-value log-likelihood is $l(\boldsymbol{\theta}) = \log \sum_{\mathbf{Y} \in \mathcal{Y}(\mathbf{y}^{\text{obs}})} \Pr_{\mathbf{g}}(\mathbf{Y} =$
21 $\mathbf{y}; \boldsymbol{\theta})$ and *observed* information is (Orchard and Woodbury 1972; Sundberg 1974; Handcock
22 and Gile 2010)

$$23 \quad \mathcal{I}^{\text{obs}}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}(\boldsymbol{\theta} | \mathbf{y}^{\text{obs}}). \quad (1)$$

24
25 Unlike the completely observed case, (1) is not the Fisher information, because it depends
26 on the data (\mathbf{y}^{obs}). The Fisher information, then, also takes the expectation over the
27 possible values of \mathbf{y}^{obs} under the model:

$$28 \quad \mathcal{I}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}(\boldsymbol{\theta}) - \mathbb{E}_{\mathbf{Y}}[\boldsymbol{\Sigma}\{\boldsymbol{\theta} | \text{obs}(\mathbf{Y})\}] \quad (2a)$$

$$29 \quad = \text{Var}_{\mathbf{Y}}[\boldsymbol{\mu}\{\boldsymbol{\theta} | \text{obs}(\mathbf{Y})\}]. \quad (2b)$$

41 3.2 Multivariate Linear Model for ERGM Parameters

42 Consider a sample of networks indexed $s = 1, \dots, S$, and for each, let \mathbf{z}_s be row q -vector of
43 network covariates of interest, which may include its size, various compositional properties,
44
45

and stratum membership, and let β a $q \times p$ parameter matrix. Set network-level parameters $\theta_s \stackrel{\text{def}}{=} (\mathbf{z}_s \beta)^\top$. Then, we write that jointly

$$(\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_S) \sim \text{ERGM}_{\mathbf{z}, \vec{\mathbf{y}}, \vec{\mathbf{x}}, \vec{\mathbf{g}}}(\beta)$$

if $\mathbf{Y}_s \stackrel{\text{ind}}{\sim} \text{ERGM}_{\mathbf{y}_s, \mathbf{x}_s, \mathbf{g}_s}(\theta_s)$. Thus, the components of model specification (sample space, sufficient statistic, and any covariates) are allowed to vary arbitrarily between networks, but the ERGM parameters are parametrised by a matrix, analogously to multivariate linear regression models (e.g., Johnson and Wichern 2007, Ch. 7). Here, we wrote q and p to be the same for all networks, but no generality is lost if the implementation allows fixing selected elements of β at 0. Strictly speaking, this can be viewed as a submodel of the model proposed by Slaughter and Koehly (2016) in a Bayesian framework, which can be reexpressed, *a priori*, as $\theta_s \stackrel{\text{i.i.d.}}{\sim} \text{MVN}\{(\mathbf{z}_s \beta)^\top, \Sigma\}$, with fixed effects possible if Σ has 0 elements on the diagonal.

The elements of β control how network covariates affect ERGM parameters for both exogenous and endogenous network properties, and they do so in a familiar fashion, analogous to the linear predictor of a GLM—and as we show in later sections, we can borrow familiar GLM diagnostics as well.

Example: Network size effects For a given type of social setting (e.g., classroom, household), bigger networks will typically have lower density ($|\mathbf{y}|/\{n(n - 1)/2\}$ for undirected networks), with mean degree ($|\mathbf{y}|/\{n/2\}$) being close to invariant to size; but the “default” ERGM behaviour is to preserve network density (Krivitsky, Handcock, and Morris 2011) so that mean degree grows in proportion to n . Krivitsky et al. proposed to adjust this behaviour by an offset term of the form $-\log(n)|\mathbf{y}|$: other things being equal, the odds of a relation in a network of size n would be scaled by n^{-1} , stabilising the mean degree.

1
2
3 But, their result is asymptotic, reliant on sparsity, and only adjusts lower-order properties
4 (density, mixing, and, fortuitously, degree distribution).
5
6

7 Butts and Almquist (2015) proposed that, given a sample of networks, the effect of
8 network size on density could be estimated from data by multiplying the $\log(n)$ above
9 by a free parameter, say, γ , rather than by -1 , making the mean degree approximately
10 proportional to $n^{\gamma+1}$. Here, we can do this by setting $z_{s,k} = \log(n_s)$ and $g_{s,l}(\mathbf{y}_s) = |\mathbf{y}_s|$ for
11 some indices k and l ; then $\gamma \equiv \beta_{k,l}$. Considering that our networks are small and dense, we
12 can model a nonlinear network size effect by adding a quadratic covariate $z_{s,k+1} = \log^2(n_s)$
13 with $\beta_{k+1,l}$ then becoming its coefficient. Alternatively, orthogonal polynomial contrasts, a
14 spline, or dummy variables could be used.
15
16
17
18
19
20

21 22 23 3.3 Inference

24 We now describe this framework's inferential properties. Let \mathbf{I}_d be an identity matrix of
25 dimension d ; let \otimes be the Kronecker product; and let $Z_s \stackrel{\text{def}}{=} \mathbf{I}_p \otimes \mathbf{z}_s$. Then, we can reexpress
26 $\boldsymbol{\theta}_s = \mathbf{z}_s^\top \boldsymbol{\beta}^\top \equiv Z_s \text{vec}(\boldsymbol{\beta})$, for an exponential family with a complete-data likelihood
27
28

$$29 \quad 30 \quad 31 \quad 32 \quad \mathcal{L}(\boldsymbol{\beta}) = \exp \left\{ \text{vec}(\boldsymbol{\beta}) \cdot \sum_{s=1}^S Z_s^\top \mathbf{g}_s(\mathbf{y}_s) \right\} / \kappa(\boldsymbol{\beta}, \mathbf{z}).$$

33 Let $\boldsymbol{\mu}_s(\boldsymbol{\beta} \mid \mathbf{y}_s^{\text{obs}}) \stackrel{\text{def}}{=} \mathbb{E}\{\mathbf{g}_s(\mathbf{Y}_s) \mid \mathbf{Y}_s \in \mathcal{Y}(\mathbf{y}_s^{\text{obs}}); (\mathbf{z}_s \boldsymbol{\beta})^\top\}$ and analogously for $\boldsymbol{\mu}_s(\boldsymbol{\beta})$, $\boldsymbol{\Sigma}_s(\boldsymbol{\beta} \mid$
34 $\mathbf{y}_s^{\text{obs}})$, and $\boldsymbol{\Sigma}_s(\boldsymbol{\beta})$. Then, its partially observed Fisher information is
35
36

$$37 \quad 38 \quad 39 \quad 40 \quad \mathcal{I}(\text{vec } \boldsymbol{\beta}) = \sum_{s=1}^S Z_s^\top \text{Var}_{\mathbf{Y}_s}[\boldsymbol{\mu}\{\boldsymbol{\beta} \mid \text{obs}(\mathbf{Y}_s)\}] Z_s.$$

41 For those networks in the sample that are completely observed, $\boldsymbol{\mu}_s(\boldsymbol{\beta} \mid \mathbf{y}_s^{\text{obs}}) \equiv \mathbf{g}_s(\mathbf{y}_s)$
42 and $\text{Var}_{\mathbf{Y}_s}[\boldsymbol{\mu}\{\boldsymbol{\beta} \mid \text{obs}(\mathbf{Y}_s)\}] \equiv \boldsymbol{\Sigma}_s(\boldsymbol{\beta})$. Since this is an independent sample of networks,
43 consistency and asymptotic normality of $\hat{\boldsymbol{\beta}}$ in S can be shown (Sundberg 1974), provided
44
45
46

1
2
3 the sampling process is noninformative and $\mathcal{I}(\text{vec } \boldsymbol{\beta})$ is nonsingular asymptotically, which
4 requires the model to be identifiable.
5
6

7 4 Diagnosing Multivariate Linear ERGMs 8

9
10 Whether or not the estimation can be consistent and the inference be generalised to a
11 broader population of households depends on the model being identifiable given available
12 data and on its goodness-of-fit—both of which must take into account that at least some
13 of the networks in the sample are partially observed. Here, we discuss likely causes and
14 diagnostics for nonidentifiability, develop a generalisation to residual diagnostics for par-
15 tially observed networks, and consider a variety of ways in which the model may fit poorly
16 to the sample and how to diagnose them.
17
18

19 4.1 Causes and Diagnostics for Nonidentifiability 20

21 The key condition for consistency by Sundberg is that it must be nonsingular. Substan-
22 tively, there is a number of reasons this condition might not be satisfied.
23
24

25 **Nonidentifiable model specification** A model may erroneously contain a relationship
26 type or other network feature that is not possible in any potentially sampled network. For
27 a trivial example, counting the number of connections between adults and children is not
28 meaningful in a survey of households without children, and similarly counting 2-stars in
29 households of size 2. Similarly, given the large number of potential network features, and a
30 large number of potential network-level covariates, it is not difficult to inadvertently specify
31 a model that is not full rank. An example of this is network size as a covariate in a sampling
32 process that observes networks of only one distinct size; or a quadratic network size effect
33 if only two distinct sizes are observed. Then the minuend of (2a) (i.e., $\Sigma(\boldsymbol{\beta})$), respectively,
34
35
36
37
38
39
40
41
42
43
44
45

1
2
3 has zeros on the diagonal or linear dependence, and the model is not identified even under
4 complete observation.
5

6 This form of nonidentifiability can generally be detected during the course of estimation
7 by examining the variance–covariance matrices of simulated sufficient statistics.
8

9
10 **Network observation process not informative of the model** If the sampling process
11 entails partially observed networks, some combinations of observation process and model
12 specification may make an otherwise identifiable model nonidentifiable.
13
14

15 **Example 1** Consider an undirected network with actors partitioned into groups A and
16 B . A 3-parameter model whose statistic comprises the counts of all edges, of edges within
17 group A , and of edges between members of A and members of B is identifiable, and its $\Sigma(\boldsymbol{\theta})$
18 is full-rank. But, if only relationships incident on members of group A (A – A and A – B) are
19 observed, while B – B relations are missing by design, then the elements of $\mu(\boldsymbol{\theta} | \mathbf{y}^{\text{obs}})$ are
20 affinely dependent, and so $\mathcal{I}(\boldsymbol{\theta})$ is singular. (See Appendix B.1.)
21
22

23 **Example 2** For an i.i.d. sample of S 3-node undirected networks, it is possible to
24 estimate a 3-parameter model with edges, 2-stars, and triangles sufficient statistic; but not
25 if any one of the 3 possible relations is unobserved in each network: a direct enumeration
26 of the sample space in Appendix B.2 shows that (2b) is singular.
27
28

29 This type of nonidentifiability is more insidious. Its main symptom is that intermediate
30 estimates of the difference in (2a) are not positive definite; but the algorithm of Handcock
31 and Gile (2010) obtains this difference by subtracting the two simulated variance–covariance
32 matrices, and for data with high missingness fraction and models with many parameters
33 in particular, this symptom could be triggered by Monte Carlo error.
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

4.2 Residual Diagnostics for Partially Observed Networks

Traditional model diagnostics—whether for linear regression or for ERGMs (Hunter, Goodreau, and Handcock 2008)—work by comparing the observed data points to those predicted by the fitted model. The approach of Hunter, Goodreau, and Handcock in particular is to simulate networks from the fitted model, and compare the statistics of the simulated networks—particularly those statistics *not* in the original model—to their observed values. If the observed value falls outside of the range of the simulated, lack of fit is indicated. However, most of the networks in E dataset are partially observed, and this means that there is no “true” observed value for a network feature. We therefore develop the theory for diagnostics for partially observed networks.

For notational convenience, let $\vec{y} = [\mathbf{y}_s]_{s=1}^S$ refer to a vector of completely observed networks. Consider a real-valued function $t(\vec{y})$ that evaluates a particular network feature of interest, either cumulatively over all of the networks or for a specific network. Analogously to $\mu(\cdot)$ and $\Sigma(\cdot)$ in Section 3.1, let $\tau(\beta) \stackrel{\text{def}}{=} \mathbb{E}\{t(\vec{Y}); \beta\}$ and $\Psi(\beta) \stackrel{\text{def}}{=} \text{Var}\{t(\vec{Y}); \beta\}$, and likewise for the conditional expectations.

We can form a standardised (Pearson) residual for $t(\vec{y})$ by evaluating

$$R_t = \{t(\vec{y}) - \tau(\hat{\beta})\} / \sqrt{\Psi(\hat{\beta})}, \quad (3a)$$

with the expectation and the variance estimated by simulating from the fitted model. Under the true model, this residual would, by construction, have mean 0 and variance close to 1; this also facilitates outlier detection.

If the networks are not completely observed, $t(\vec{y})$ cannot be evaluated directly, it is natural to replace it with its empirical best predictor (Hunter et al. 2008; Stewart et

1
2
3 al. 2019; Krivitsky et al. 2021),
4
5

$$\tau(\hat{\beta} \mid \vec{y}^{\text{obs}}) \stackrel{\text{def}}{=} \mathbb{E}\{t(\vec{Y}) \mid \vec{Y} \in \mathcal{Y}(\vec{y}^{\text{obs}})\},$$

6
7 where \vec{y}^{obs} is defined analogously to \vec{y} . Then,
8
9

$$R_t = \{\tau(\hat{\beta} \mid \vec{y}^{\text{obs}}) - \tau(\hat{\beta})\} / \sqrt{\text{Var}_{\vec{Y}}[\tau(\hat{\beta} \mid \text{obs}(\vec{Y}))]}, \quad (3b)$$

10
11 Estimating the variance in the divisor in (3b) is not trivial. We discuss it in Appendix C.1.
12
13

14 4.3 Causes and Diagnostics for Lack-of-Fit

15
16 **Within-network** It may be the case that the within-network model fits poorly. Network
17 statistics used for diagnostics by Hunter, Goodreau, and Handcock (2008) include the full
18 degree distribution, the counts of shared partners (i.e., for a given pair of connected actors,
19 how many common connections do they have?), and the distribution of geodesic distances.
20 All of these can be used as $t(\cdot)$, but it may be impractical for two reasons. Firstly, family
21 networks are relatively small and very dense. This makes the statistics typically used
22 less than informative. Secondly, the sheer number of networks in the dataset means that
23 diagnosing each network individually is impractical, but, at the same time, pooling their
24 within-network diagnostics is likely to wash out any effects because of their heterogeneity.
25
26

27 Nonetheless, even if a statistic is suboptimal and difficult to interpret, for a model that
28 fits well, $R_{t,s}$ will still have mean 0 and variance close to 1.
29
30

31 **Between-network** It may be the case that the model for the network-level parameters
32 (θ_s) as a function of global parameters (β) fits poorly: in particular, it may fail to account
33 for network size and composition effects. At network level, the model has a form similar to
34 that of a GLM. We can thus use the developments of Section 4.2 directly to make familiar
35 diagnostic plots: for some statistic $t_s(\vec{y}) \stackrel{\text{def}}{=} t(y_s)$ (e.g., density), one can plot residuals R_{t_s}
36
37

for $s = 1, \dots, S$ against their respective $\mathbb{E}\{t_s(\vec{Y})\}$ (the fitted values) or against a candidate predictor. Or, we can plot $\sqrt{|R_{t_s}|}$ instead for a scale–location plot, analogously to the standard diagnostic plots in R (R Core Team 2021).

One can also use these residuals to bypass the computational cost of refitting an ERGM for each candidate specification: regressing R_{t_s} s on their respective $z_{s,\text{new}}$ can be used to assess the explanatory power of a candidate predictor z_{new} . Since the individual networks are independent, the residuals should be nearly independent as well.

Between-dataset If we wish for the fitted model to generalise and render the sampling designs ignorable, the model must account for differences in datasets without incorporating dataset effects directly. This can be done via a hypothesis test, such as a simulation Score test, along the lines of that described by Krivitsky (2012) in the context of valued ERGMs, by testing the significance of an explicit dataset effect without refitting the model. Details are given in Appendix C.2.

Non-systematic heterogeneity Lastly, even if there is no systematic bias in the model, there may be between-network heterogeneity due to unobserved factors. The above-described Pearson residuals incidentally provide us with a way to tell whether there is any heterogeneity left to explain: if there is none, R_{t_s} in (3) will, by construction, have mean 0 and variance around 1.

5 Application

We now return to the datasets we have introduced in Section 2, discuss model specification, and report resulting model diagnostics and results.

We have implemented the methodology described in an extension to the `ergm` package (Hunter et al. 2008; Krivitsky et al. 2021) for the R (R Core Team 2021) statistical envi-

1
2
3 ronment. To make this methodology accessible to a broad audience, we have published our
4 implementation in an R package, `ergm.multi`. Materials needed to reproduce the analysis
5 are included in the supplementary file, and the most recent versions of the packages can
6 be found on the Comprehensive R Archive Network (R Core Team 2021) or the Statnet
7 Project software repository (<https://github.com/statnet/>).
8
9
10
11

12 5.1 Model 13 14

15 A model used to join these two datasets must be one that is substantively meaningful and
16 interpretable and accounts for within-network conditional dependence among the relations;
17 while making the network size, composition, and dataset effects ignorable to enable gener-
18 alisable inference. We therefore dedicate a significant amount of attention to formulating
19 and justifying each of its elements.
20
21
22
23

24 **Design effects** If we wish to generalise our inference to the population of households,
25 our model must make any design effects ignorable. The substantively motivated effects
26 described below already control for some of those. For example, recall that the *H* dataset's
27 households were omitted if there was nonresponse from even a single member. To the extent
28 that the nonresponse rate is a function of household size (i.e., the bigger the household, the
29 more likely there is at least one nonresponse), a model that accurately controls for network
30 size will reduce the noninformativeness of this nonresponse.
31
32
33
34
35
36

37 To account for the fact that only *H* dataset incorporated households in Brussels, we
38 must add an indicator of the household being in Brussels as a covariate.
39
40

41 Both surveys' selection was strongly affected in different ways by participants' and
42 household members' ages, particularly children. We can adjust for the latter by including
43 effects of the presence of a child in a household as a covariate, or as a part of the model for
44
45
46
47

interaction of different household roles. More generally, when modelling age effects, there is a tension between interpretability and accuracy: family roles are most conveniently modelled with discrete age categories; but outside of a few critical ages defined exogenously (e.g., school attendance, legal adulthood, and retirement), age effects are likely to be continuous, best modelled semiparametrically (e.g., with splines). Our compromise is to use relatively fine-grained edge categories, discussed in the next section.

Household roles In order to study interactions pattern among household members, each member has been assigned to a category, constructed to resemble household roles, unfortunately not available in the original survey. A first classification was done according to age: young child (below 6 years of age), preadolescent (between 6 and 12 years of age), adolescent (between 13 years of age and 18 years of age), younger adult (between 19 years of age and 24 years of age), older adult (between 25 years of age and 60 years of age), and senior (older than 60 years of age). We further divided older adults according to gender, to investigate the possibility of gender-specific interactions (Goeyvaerts et al. 2018), ending with a total of 7 categories (young child, preadolescent, adolescent, younger adult, older female adult, older male adult, senior). The age cut of 12 was chosen to specifically account for the design effect.

We then modelled mixing counting the number of contacts between pairs of household roles. Our data about some age categories are limited, both for the design reasons already discussed and because some age groups, particularly young adults and seniors, have a higher propensity to live alone or in a dormitory or a nursing home and thus be excluded from both datasets. For this reason, and for reasons of substantive parsimony, not every combination was modelled individually. In particular, interaction counts of adolescents with young children and preadolescents were combined, as were those of young adults

1
2
3 with preadolescents and adolescents. (Young adults with young children were retained as a
4 separate count, because their chances of being parents of a young child were relatively high.)
5 Similarly, interactions of young adults with both older adult categories were combined, and
6 so were all interactions of seniors with non-seniors.
7
8

9
10 To further incorporate the design effect of only households with a child 12 or under
11 being eligible for inclusion in H , we added an effect for mixing between older male adults
12 and older female adults in those households that did *not* have such a child.
13
14

15
16 **Network size effects** As discussed in Section 3.2, the effect of network size on our
17 networks is not trivial; we use the polynomial effects of $\log n_s$ on density described there
18 to model it. In the analysis of the H dataset by Goeyvaerts et al. (2018), three different
19 density parameters were used depending on household size.
20
21

22
23 **Other network-level effects** Some of the surveys were conducted on a weekend and
24 others on a weekday (Table A1). Past literature (Goeyvaerts et al. 2018, for example)
25 suggests that contact patterns may differ depending on the day, and so we incorporate this
26 effect as well.
27
28

29 Contact patterns may differ systematically between families that live in detached hous-
30 ing and families that live in apartments. This potential effect has received limited attention
31 in the literature to date. Our data do not include housing type but do include postal codes.
32 The population densities in those postal codes can then be used as a proxy for housing
33 type. We use this potential predictor to illustrate the technique proposed in Section 4.3 of
34 regressing the network-level residuals on potential network-level predictors. Alternatively,
35 we may ask whether or not the post code is in any of Belgium's larger cities.
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Endogenous effects To model actor heterogeneity and triadic closure, we use the 2-stars count and the triangles count, given in Section 3.1. We reiterate the cautions about ERGM degeneracy, and further guard against it by allowing their coefficients to vary with network size as suggested by Goeyvaerts et al. (2018), using the same approach as for density.

An additional caveat here is that the E dataset, by virtue of only containing relations incident on one individual per household, does not contain information about triadic closure. (See Section 4.1 Example 2 and Appendix B.2.) We thus assume that net of all other effects, the effect of triadic closure on a household of a given size that does not have a child is the same as the effect of triadic closure on a household of that size that does have a child. It is not possible to test this assumption with the available data.

Other household properties—location, population density, and weekday—may also be predictive not only of density but also of higher-order structural features and can be incorporated, but not without cost. We thus begin with a model containing only density effects of location in Brussels and of whether the survey had been done on a weekend, and use its diagnostics to suggest further effects.

We call this starting model *Model 1*.

5.2 Diagnostics

We now apply the techniques from Section 4 to the proposed model. To validate our diagnostic techniques, we also fit a number of reduced models: in Appendix D, we demonstrate how we can identify their deficiencies.

Additional substantive models As proposed in Section 5.1, we regressed Pearson residuals edges, 2-stars, and triangles on a number of candidate covariates, with full results given in Table E16. We found the linear effect of log-population-density to be the most promising

covariate. Interestingly, its apparent effect (higher χ^2 or lower P -value) is on the 2-star and the triangle statistic, rather than the edge count statistic. To test whether this distinction has predictive power, we fit two more models: *Model 2*, adding the effect of population density on edges, 2-stars, and triangles; and *Model 3*, adding its effect on edges only. For completeness, we also fit *Model 4*, adding effects of indicators of Brussels postal code, survey being on a weekend, and population density on edges, 2-stars, and triangles, all.

Specifications for all models are summarised in Table E1, and their complete results and diagnostics are provided in Appendix E.

Unaccounted-for between-dataset differences For visual diagnostics, it is helpful to distinguish two “subdatasets” of the E dataset: those households with a child 12 years old or under—that *could have been* in the H dataset (E_Y), and those without, that could not (E_N). Selected residual plots are provided in Figure 2. We also provide smoothing lines for each subdataset individually: these lines diverging would indicate that the model had failed to account for some systematic difference between the datasets.

Overall, the distributions of the residuals are skewed downward. This is to be expected: underlying statistics are highly skewed. We see that the residuals for the H dataset and the E_Y subdataset coincide, and that the residual for the E_N subdataset diverges slightly. The apples-to-apples comparison of H and E_Y suggests that the model fuses the two datasets well, though it does not represent E_N households perfectly.

The simulated P -values for edges and 2-stars are, respectively, 0.015 and 0.071, with the omnibus P -value 0.036. Thus, at conventional significance level, it appears that there are differences in network density that are not accounted for by the model.

We examine the practical significance of this via Figure 3. Focusing on the two comparable groups (H dataset and subset of E with a child present), we see that there may be

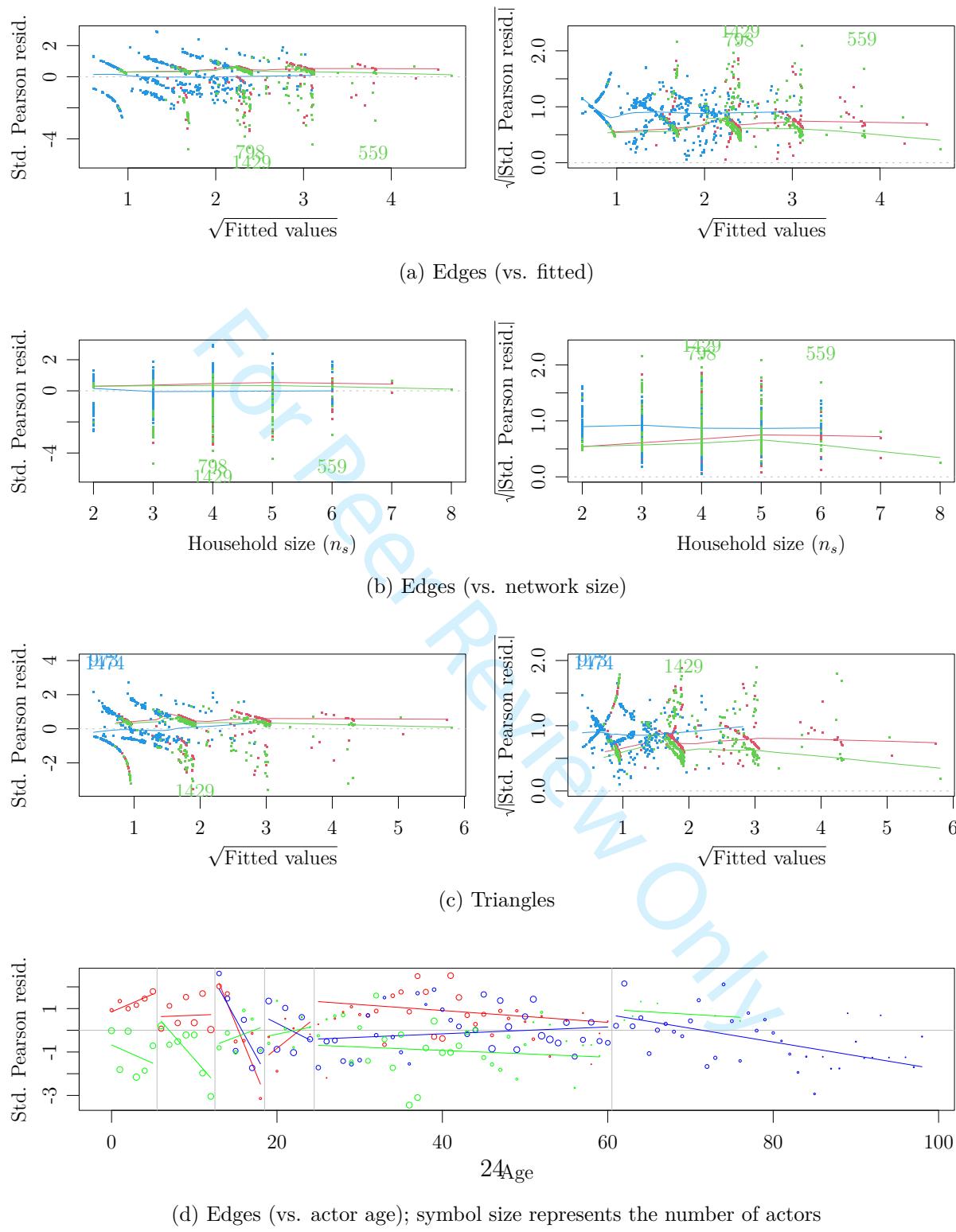


Figure 2: Selected Pearson residual plots of network statistics for *Model 1*. (● all H
 ● E w/ child ● E w/o child)

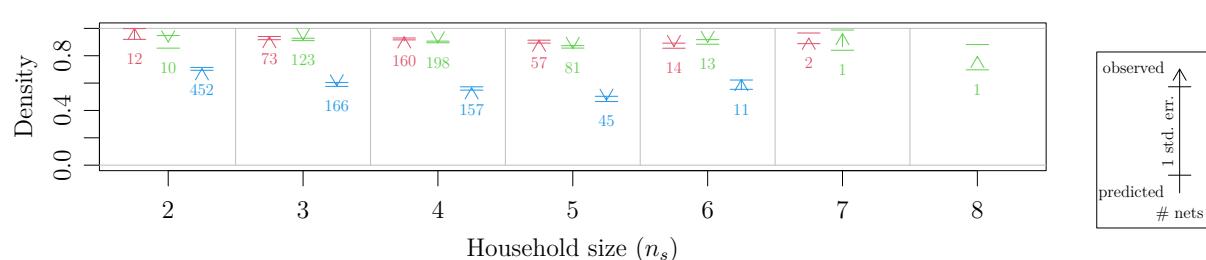


Figure 3: Average prediction errors of density in *Model 1*. Values are averaged over the networks grouped by size and subdataset. (● all H ● E w/ child ● E w/o child)

some systematic underestimation of network density for H households and underestimation for E households (taking into account the number of households of each network size); but also that the differences are small in magnitude and are likely to be of no practical significance.

Outliers Our residual plots reveal some outlier networks. For example, observation 1429 (E household #1111) highlighted in Figure 2a and other panels stands out because it is a family of four with two young children but no contacts reported by the father on the day of the survey. Similarly, the respondent in observation 798 (E #480) is an infant with no reported physical contacts—which could indicate a recording error.

Network size effects The residuals against the network size are also shown in Figure 2b. A model that fails to account for network size would display a linear or curved pattern in the residuals. We see no evidence of such a pattern.

We further test this by regressing the residual for the edges, 2-stars, and triangles statistics for each network on that network's size, weighting them by the inverse of their estimated variance (from divisors in (3)). If model controls for network size effects adequately, these residuals will have mean zero for each network size, which we test using an

1
2
3 omnibus ANOVA test, and we find that none are statistically significant (P -vals. 0.80, 0.69,
4 and 0.69, respectively): network size appears to be accounted for.
5
6

7 **Non-systematic heterogeneity** We report the standard deviations of Pearson residuals
8 for edges, 2-stars, and triangles, both for the dataset as a whole and for each of the sub-
9 datasets, in Table E18. We find no evidence of overdispersion, though standard deviations
10 being generally less than 1 suggest some degree of overfitting.
11
12

13 **Continuous age effects** We plot Pearson residuals of the total number of within-
14 household contacts of individuals of age a against a in Figure 2d. We see, as with other
15 diagnostics, that H dataset density tends to be slightly underpredicted, whereas compa-
16 rable households in the E dataset are overpredicted, and there are slight trends in the
17 residuals within age categories—a consequence of discretisation. We also observe that a
18 senior’s propensity to interact appears to drop off as they become older, which the cutoff at
19 60 does not capture. On the other hand, the number of actors affected is relatively small.
20
21

22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 5.3 Results

Model comparison Table 1 gives the parameter estimates for *Model 1*, our initial model, and *Model 2*, suggested by our residual diagnostics. Parameter estimates for *Model 3* and *Model 4* and measures of fit are given in Tables E23, E27, and E2, respectively. AIC is indifferent between *Model 1* and *Model 2* within margin of error and prefers them over *Model 3*, and *Model 4*. The evidence for the effect of population density on the edge count is stronger in *Model 2* ($\hat{\beta} = 0.18$, SE = 0.071, P -val. = 0.01), in the presence of its effect on 2-stars and triangles, than it is in *Model 3* ($\hat{\beta} = 0.05$, SE = 0.033, P -val. = 0.17). This is as predicted by the residual regression discussed in Section 5.2: there does not appear to be a net effect of population density on the density of the contact networks in general, but

1
2
3 there is an effect on higher-order properties, which through being correlated with network
4 density make its effect apparently significant.
5

6 Notably, an omnibus test of all population density effects in *Model 2* is not significant at
7 conventional level ($\chi^2 = 7.4$, df = 3, $P\text{-val.} = 0.06$), so whether population density in fact
8 has an effect—or whether we had found the best possible way to model it—is questionable;
9 we leave these questions for future work, except to suggest that type of housing should be
10 considered for future data collection.
11
12

13
14
15 **Substantive interpretation** We discuss results primarily from *Model 1*, though *Model 2*
16 yields the same conclusions. Only a few of the effects are interpretable in isolation. In
17 particular, we can conclude with some confidence ($-\hat{\beta} = 0.15$, SE = 0.062, $P\text{-val.} = 0.015$)
18 that weekends have a positive effect on the number of contacts that are observed in the
19 household, in line with prior literature (Grijalva et al. 2015). Presence of a child in a
20 household is associated with a higher propensity of older adult females and older adult
21 males (i.e., likely the parents) to interact with each other (if no child: $\hat{\beta} = -0.8$, SE =
22 0.25, $P\text{-val.} = 0.0014$). We do not find an effect of the household being in Brussels on
23 network density ($\hat{\beta} = 0.1$, SE = 0.20, $P\text{-val.} = 0.54$).
24
25

26 The estimated polynomial log-network-size effects are shown in Figure 4. There is
27 strong evidence that network size effects are present ($\chi^2 = 1403.2$, df = 6, $P\text{-val.} < 0.0001$),
28 including quadratic effects of network size ($\chi^2 = 9.7$, df = 3, $P\text{-val.} = 0.022$), though it
29 is not as clear whether quadratic effects of network size on 2-stars or triangles are ($\chi^2 =$
30 5.2, df = 2, $P\text{-val.} = 0.074$). In the presence of edge count and triangle count, 2-star effects
31 are not statistically significant ($\chi^2 = 5.3$, df = 3, $P\text{-val.} = 0.15$) but triangle effects (in the
32 presence of others) are ($\chi^2 = 347.3$, df = 3, $P\text{-val.} < 0.0001$).
33
34

35 Interactions between age categories reveal tendencies in interaction between household
36

Table 1: Parameter estimates (and standard errors) for *Model 1* and *Model 2*. Standard errors for AIC, BIC, and log-likelihood are due to MCMC error.

Effect ($\times z_s$)	<i>Model 1</i>	<i>Model 2</i>
edges $\times \log(n_s)$	-12.03 (3.80)**	-11.97 (0.88)***
$\times \log^2(n_s)$	4.57 (1.72)**	4.51 (0.44)***
$\times \log(\text{pop. dens. in post code})$		0.18 (0.07)*
if Brussels post code	0.12 (0.20)	0.01 (0.22)
if on weekend	0.15 (0.06)*	0.15 (0.06)*
2-stars	0.74 (1.19)	4.89 (0.72)***
$\times \log(n_s)$	-0.77 (0.76)	-6.77 (0.59)***
$\times \log^2(n_s)$	0.04 (0.21)	1.92 (0.22)***
$\times \log(\text{pop. dens. in post code})$		-0.08 (0.07)
triangles	7.56 (0.60)***	-4.56 (1.03)***
$\times \log(n_s)$	-5.74 (1.13)***	10.65 (0.75)***
$\times \log^2(n_s)$	1.42 (0.85)	-3.68 (0.44)***
$\times \log(\text{pop. dens. in post code})$		0.12 (0.17)
Young Child with Young Child	7.01 (1.90)***	8.39 (0.37)***
Young Child with Preadolescent	7.84 (1.91)***	9.21 (0.40)***
Preadolescent with Preadolescent	6.89 (1.88)***	8.25 (0.26)***
Adolescent with Adolescent	6.15 (1.86)***	7.49 (0.31)***
Young Child with Young Adult	7.79 (2.13)***	9.44 (1.12)***
Young Adult with Young Adult	6.82 (1.82)***	8.25 (0.34)***
Young Child with Female Older Adult	9.02 (1.86)***	10.39 (0.37)***
Preadolescent with Female Older Adult	8.45 (1.86)***	9.83 (0.25)***
Adolescent with Female Older Adult	7.89 (1.87)***	9.29 (0.24)***
Female Older Adult with Female Older Adul	6.12 (1.88)**	7.50 (0.31)***
Young Child with Male Older Adult	8.07 (1.87)***	9.50 (0.29)***
Preadolescent with Male Older Adult	7.54 (1.86)***	8.93 (0.22)***
Adolescent with Male Older Adult	7.12 (1.85)***	8.54 (0.23)***
Female Older Adult with Male Older Adult	8.29 (1.87)***	9.70 (0.26)***
if child absent	-0.80 (0.25)**	-0.84 (0.25)***
Male Older Adult with Male Older Adult	5.62 (1.88)**	7.04 (0.33)***
Senior with Senior	6.55 (1.82)***	7.88 (0.22)***
Adolescent with Young Child or Preadolescent	6.83 (1.87)***	8.19 (0.23)***
Young Adult with Preadolescent or Adolescent	6.29 (1.85)***	7.69 (0.25)***
Young Adult with Older Adult	6.79 (1.87)***	8.15 (0.17)***
Senior with Non-Senior	6.97 (1.86)***	8.34 (0.19)***
AIC	3390 (0.5)	3390 (0.5)
BIC	3590 (0.5)	3610 (0.5)
log-likelihood	-1664 (0.2)	-1661 (0.2)

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

members. We report the parameter estimates with a more intuitive layout in Figure 5. Since we do not use a baseline category (“intercept”), it is not meaningful to interpret them in isolation but only in contrast with each other. We can, for example, conclude that older female adults (i.e. mothers) tend to interact more than older male adults (i.e. fathers) with young children ($\hat{\Delta} = 0.9$, SE = 0.46, P -val. = 0.04), preadolescents ($\hat{\Delta} = 0.9$, SE = 0.29, P -val. = 0.0018), and adolescents ($\hat{\Delta} = 0.8$, SE = 0.28, P -val. = 0.0055). We thus confirm similar findings by Goeyvaerts et al. (2018).

6 Conclusion

Motivated by two collections of networks representing the same phenomena but collected using very different sampling designs, we developed a framework for combining the strengths of the two, facilitating population-wide simulation of household networks. In the process, we identified the requirements of this procedure and developed generally applicable techniques for specifying and diagnosing models for large samples of networks, techniques that, through their relationship to GLMs, can be used by researchers from a wide variety of disciplines.

In addition, the techniques we have developed do not rely on the networks being completely observed. To make this methodology accessible to a broad audience, we produced a user-friendly R package `ergm.multi`.

Our two surveys were conducted in Flanders and Brussels in 2010–2011. It is important to design and analyse household surveys in different settings with different inclusion criteria—but, ideally, compatible measurement instruments—to gain further insights on the contact patterns and the effects of endogenous factors such as triadic closure, exogenous individual attributes such as age, and exogenous household attributes such as size and type of residence. This work provides a foundation for identifying and testing these

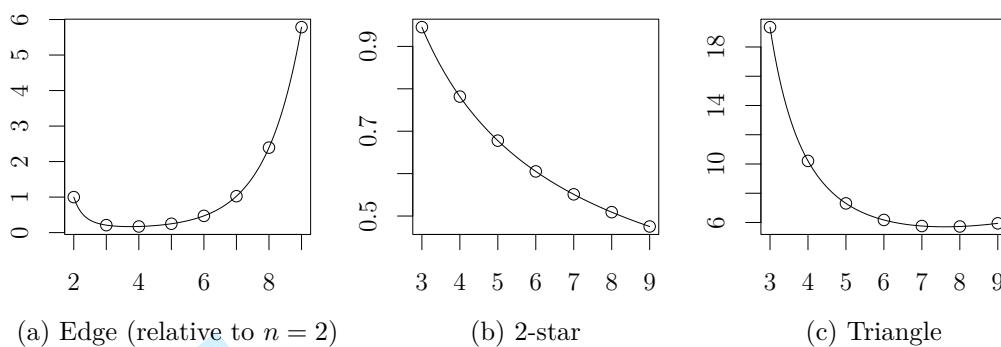


Figure 4: Estimated effects of network size on conditional odds of an instance of a graph feature. Two-stars and triangles are only possible for $n \geq 3$.

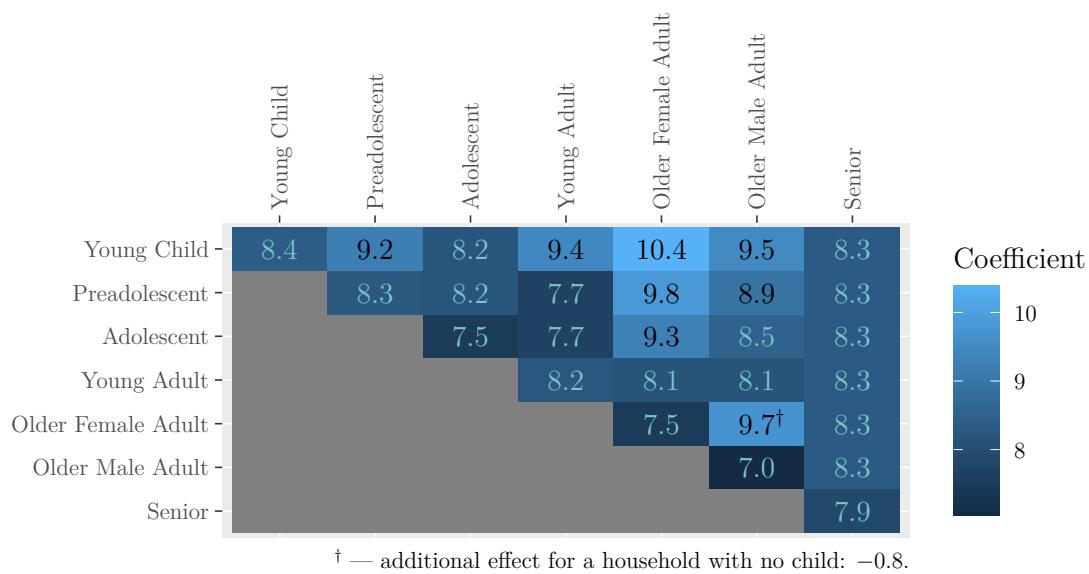


Figure 5: Parameter estimates for mixing by family role. Because there is no “intercept” effect in the model, testing them against 0 is not meaningful.

1
2
3 effects and for confirming the validity of the analysis—and opens the door to design of
4 future cost-effective yet highly informative hybrid network studies.
5
6

7 A number of methodological research directions remain. In our work, we used Pearson
8 residuals. Other types of residuals, such as deviance, tend to be better behaved and could,
9 perhaps, be derived for this family of models. Similarly, Cook's distance may be possible
10 to compute inexpensively for each network using the approach of Koskinen et al. (2018).
11
12

13 We did not find evidence of non-systematic heterogeneity of networks. Where such is
14 present, it can be accounted for in a mixed effects framework (Slaughter and Koehly 2016)
15 at an additional computational cost, or perhaps by constructing ERGM sufficient statistics
16 to absorb the variation (Butts 2017; Krivitsky 2012). Alternatively, quasi-likelihood and
17 generalised estimating equation approaches may be extended to samples of networks.
18
19

20 ERGM computational and diagnostic techniques are agnostic to the structure of the
21 sample space, so these approaches directly generalise to directed, temporal, valued, and
22 multilayer network scenarios. For our two surveys in particular, physical contact was not
23 the only relational measurement: the respondents were also asked about the approximate
24 duration of interaction (close proximity) on an ordinal scale (time ranges). Along similar
25 lines, the techniques for calculating standardised residuals under partially observed data
26 may be applicable to other domains that involve modelling independent samples of units
27 which are themselves partially observed.
28
29

30 SUPPLEMENTARY MATERIAL 31 32

33 Appendices A–E: Further details, discussion, and results. (PDF file)
34
35

36 Code and data: Materials (including R packages) to reproduce the analysis. (ZIP file)
37
38

References

- Breckling, J. U., R. L. Chambers, A. H. Dorfman, S.-M. Tam, and A. H. Welsh. 1994. “Maximum Likelihood Inference from Sample Survey Data.” *International Statistical Review* 62 (3): 349–363.
- Butts, C. T. 2017. “Baseline Mixture Models for Social Networks,” arXiv: 1710.02773v1.
- Butts, C. T., and Z. W. Almquist. 2015. “A Flexible Parameterization for Baseline Mean Degree in Multiple-network ERGMs.” *The Journal of Mathematical Sociology* 39 (3): 163–167.
- Cencetti, G., G. Santin, A. Longa, E. Pigani, A. Barrat, C. Cattuto, S. Lehmann, M. Salath'e, and B. Lepri. 2021. “Digital Proximity Tracing on Empirical Contact Networks for Pandemic Control.” *Nature Communications* 12 (1): 1655.
- Elliott, M. R., T. E. Raghunathan, and N. Schenker. 2018. “Combining Estimates from Multiple Surveys.” In *Wiley StatsRef: Statistics Reference Online*.
- Ersig, A. L., D. W. Hadley, and L. M. Koehly. 2011. “Understanding Patterns of Health Communication in Families at Risk for Hereditary Nonpolyposis Colorectal Cancer: Examining the Effect of Conclusive Versus Indeterminate Genetic Test Results.” *Health Communication* 26 (7): 587–594.
- Goeyvaerts, N., E. Santermans, G. Potter, A. Torneri, K. Van Kerckhove, L. Willem, M. Aerts, P. Beutels, and N. Hens. 2018. “Household Members Do Not Contact Each Other at Random: Implications for Infectious Disease Modelling.” *Proceedings of the Royal Society of B: Biological Sciences* 285:20182201.

- 1
2
3 Grijalva, C. G., N. Goeyvaerts, H. Verastegui, K. M. Edwards, A. I. Gil, C. F. Lanata,
4 and N. Hens. 2015. "A Household-based Study of Contact Networks Relevant for the
5 Spread of Infectious Diseases in the Highlands of Peru." *PLoS ONE* 10 (3): 1–14.
6
7 Handcock, M. S., and K. J. Gile. 2010. "Modeling Social Networks from Sampled Data."
8
9 *Annals of Applied Statistics* 4 (1): 5–25.
10
11 Hoang, T. V., P. Coletti, Y. W. Kifle, K. V. Kerckhove, S. Vercruyse, L. Willem, P. Beutels,
12 and N. Hens. 2021. "Close Contact Infection Dynamics Over Time: Insights from a
13 Second Large-scale Social Contact Survey in Flanders, Belgium, in 2010–2011." *BMC*
14
15 *Infectious Diseases* 21 (1): 274.
16
17 Hunter, D. R., S. M. Goodreau, and M. S. Handcock. 2008. "Goodness of Fit for Social
18 Network Models." *Journal of the American Statistical Association* 103 (481): 248–258.
19
20 Hunter, D. R., M. S. Handcock, C. T. Butts, S. M. Goodreau, and M. Morris. 2008. "ergm:
21 A Package to Fit, Simulate and Diagnose Exponential-family Models for Networks."
22
23 *Journal of Statistical Software* 24 (3): 1–29.
24
25 Johnson, R. A., and D. W. Wichern. 2007. *Applied Multivariate Statistical Analysis*. 6th.
26
27 Upper Saddle River: Prentice Hall.
28
29 Keeling, M. J., and K. T. D. Eames. 2005. "Networks and Epidemic Models." *Journal of*
30
31 *The Royal Society Interface* 2 (4): 295–307.
32
33 Koskinen, J., P. Wang, G. Robins, and P. Pattison. 2018. "Outliers and Influential Obser-
34
35 vations in Exponential Random Graph Models." 83 (4): 809–830.
36
37 Krivitsky, P. N. 2012. "Exponential-family Random Graph Models for Valued Networks."
38
39 *Electronic Journal of Statistics* 6:1100–1128.
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 Krivitsky, P. N., M. S. Handcock, and M. Morris. 2011. "Adjusting for Network Size
4 and Composition Effects in Exponential-family Random Graph Models." *Statistical
5 Methodology* 8 (4): 319–339.
6
7
8 Krivitsky, P. N., D. R. Hunter, M. Morris, and C. Klumb. 2021. "ergm 4.0: New Features
9 and Improvements," arXiv: 2106.04997v1.
10
11
12 Krivitsky, P. N., and M. Morris. 2017. "Inference for Social Network Models from Egocentrically-
13 sampled Data, with Application to Understanding Persistent Racial Disparities in HIV
14 Prevalence in the US." *Annals of Applied Statistics* 11 (1): 427–455.
15
16
17
18 Leifeld, P., S. J. Cranmer, and B. A. Desmarais. 2018. "Temporal Exponential Random
19 Graph Models with btergm: Estimation and Bootstrap Confidence Intervals." *Journal
20 of Statistical Software* 83 (6): 1–36.
21
22
23 Lubbers, M. J. 2003. "Group Composition and Network Structure in School Classes: A
24 Multilevel Application of the p* Model." *Social Networks* 25 (4): 309–332.
25
26
27 Lusher, D., J. Koskinen, and G. Robins. 2012. *Exponential Random Graph Models for Social
28 Networks: Theory, Methods, and Applications*. Cambridge University Press.
29
30
31 Mastrandrea, R., J. Fournet, and A. Barrat. 2015. "Contact Patterns in a High School:
32 A Comparison between Data Collected Using Wearable Sensors, Contact Diaries and
33 Friendship Surveys." *PLoS ONE* 10 (9): e0136497.
34
35
36
37
38
39
40 Orchard, T., and M. A. Woodbury. 1972. "A Missing Information Principle: Theory and
41 Applications." In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics
42 and Probability, Volume 1: Theory of Statistics*, 697–715. Berkeley: University of
43 California Press.
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rubin, D. B. 1976. "Inference and Missing Data." *Biometrika* 63 (3): 581–592.
- Schweinberger, M., P. N. Krivitsky, C. T. Butts, and J. R. Stewart. 2020. "Exponential-family Models of Random Graphs: Inference in Finite, Super and Infinite Population Scenarios." *Statistical Science* 35 (4): 627–662.
- Slaughter, A. J., and L. M. Koehly. 2016. "Multilevel Models for Social Networks: Hierarchical Bayesian Approaches to Exponential Random Graph Modeling." *Social Networks* 44:334–345.
- Stewart, J., M. Schweinberger, M. Bojanowski, and M. Morris. 2019. "Multilevel Network Data Facilitate Statistical Inference for Curved ERGMs with Geometrically Weighted Terms." *Social Networks* 59:98–119.
- Sundberg, R. 1974. "Maximum Likelihood Theory for Incomplete Data from an Exponential Family." *Scandinavian Journal of Statistics* 1 (2): 49–58.
- Vega Yon, G. G., A. Slaughter, and K. de la Haye. 2021. "Exponential Random Graph Models for Little Networks." *Social Networks* 64:225–238.
- Wasserman, S. S., and P. Pattison. 1996. "Logit Models and Logistic Regressions for Social Networks: I. An Introduction to Markov Graphs and p^* ." *Psychometrika* 61 (3): 401–425.
- Zijlstra, B. J. H., M. A. J. Van Duijn, and T. A. B. Snijders. 2006. "The Multilevel p_2 Model: A Random Effects Model for the Analysis of Multiple Social Networks." *Methodology* 2 (1): 42.

1
2
3
4
5
6
7
8 **Supplement to**
9
10 “**A Tale of Two Datasets:**
11 **Representativeness and**
12 **Generalisability of Inference for**
13 **Samples of Networks”:**
14 **Appendices A–E**

15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

FOR PEER REVIEW ONLY

Contents

A Additional Data Summaries	3
B Examples of Nonidentifiability due to Network Sampling	6
C Details of Diagnostic Methods	10
D Capacity to Detect Misspecification	14
E Additional Results and Diagnostics	18

A Additional Data Summaries

Data summaries critical to the analysis were included in Section 2.2 in the body of the article. This appendix provides additional summaries and visualisations.

Figure A1 shows the distributions of household sizes and ages and genders of household members in the two datasets. Differences between the two are due to design, with individuals aged around 25 years are certainly underrepresented in the *H* dataset (Panel (a)), as they are unlikely to be parents nor children of parents of young children. The members of the households in the *E* dataset (Panel (b)) are more representative of the population (though not completely so, as individuals living alone ($n_s = 1$) are excluded); but survey respondents (Panel (b), bottom) have a different age and gender distribution from *E* household members in general, with females aged 25–55 years overrepresented and adolescents of both genders underrepresented.

More household and structural summaries are provided in Table A1.

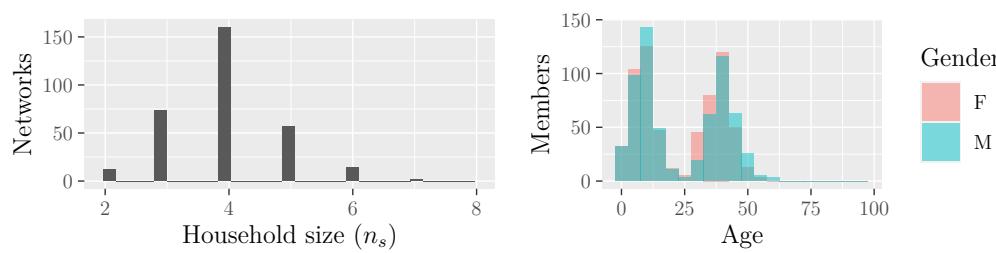
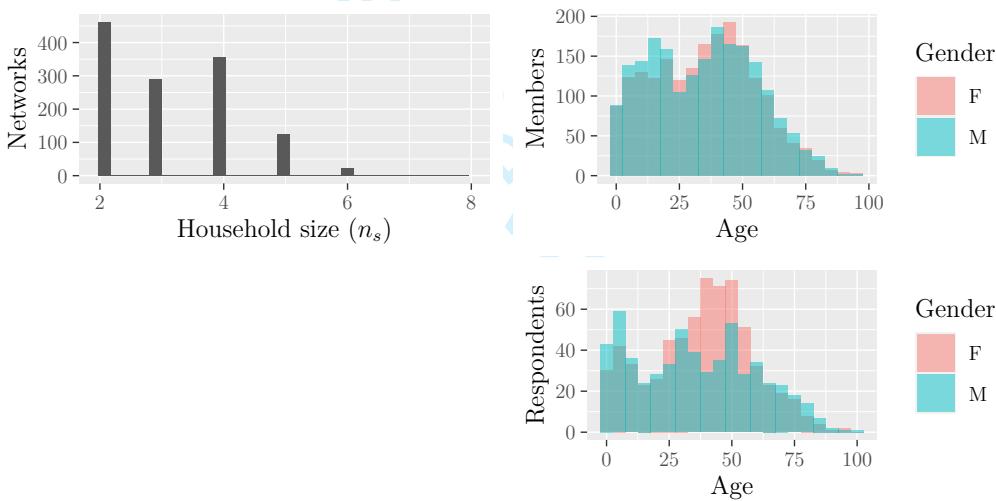
(a) H dataset: 318 households with a total of 1266 members/respondents(b) E dataset: 1258 households/respondents with a total of 4000 members

Figure A1: Household size distribution (left), household members' age and gender distribution (right), and respondents' age and gender distribution (bottom, E only).

Table A1: Miscellaneous dataset composition information. In the H dataset, every household member is a “respondent”. Clustering coefficient is only meaningful for $n_s \geq 3$ and cannot be computed for the E dataset households.

	H	E
Household members	1266 (100%)	4000 (100%)
Female	636 (50%)	1959 (49%)
Male	630 (50%)	2041 (51%)
Respondents		1258 (100%)
Female		677 (54%)
Male		581 (46%)
Child ≤ 12 present in household	318 (100%)	427 (34%)
Household observed on weekend	44 (14%)	316 (25%)
Household located in Brussels	36 (11%)	0 (0%)
Mean (std. dev.) net. density	0.93 (0.15)	0.73 (0.40)
Mean (std. dev.) net. clustering coef.	0.92 (0.21)	

B Examples of Nonidentifiability due to Network Sampling

B.1 Example 1: Exogenous Subgroups

Recall, an undirected network whose actor set N is partitioned into disjoint sets A and B . For actor sets $X \subseteq N$ and $Y \subseteq N$, let $|\mathbf{y}_{X,Y}|$ be the number of edges between actors in sets X and Y , and let $|\mathbf{y}|$ be the total edge count, and consider an ERGM with 3 parameters, whose sufficient statistic is

$$\mathbf{g}(\mathbf{y}) = [|\mathbf{y}|, |\mathbf{y}_{A,A}|, |\mathbf{y}_{A,B}|]^\top.$$

This statistic does not induce dependence among the dyads, and so under this model,

$$\begin{aligned} |\mathbf{Y}_{A,A}| &\sim \text{Binomial}\left(\binom{|A|}{2}, \text{logit}^{-1}(\theta_1 + \theta_2)\right) \\ |\mathbf{Y}_{A,B}| &\sim \text{Binomial}\left(|A||B|, \text{logit}^{-1}(\theta_1 + \theta_3)\right) \\ |\mathbf{Y}_{B,B}| &\sim \text{Binomial}\left(\binom{|B|}{2}, \text{logit}^{-1}(\theta_1)\right), \end{aligned}$$

all mutually independent. Call their variances $\sigma_{A,A}$, $\sigma_{A,B}$, and $\sigma_{B,B}$, respectively. Since $|\mathbf{y}| \equiv |\mathbf{y}_{A,A}| + |\mathbf{y}_{A,B}| + |\mathbf{y}_{B,B}|$,

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{bmatrix} \sigma_{A,A} + \sigma_{A,B} + \sigma_{B,B} & \sigma_{A,A} & \sigma_{A,B} \\ \sigma_{A,A} & \sigma_{A,A} & 0 \\ \sigma_{A,B} & 0 & \sigma_{A,B} \end{bmatrix},$$

and $|\boldsymbol{\Sigma}(\boldsymbol{\theta})| = \sigma_{A,A}\sigma_{A,B}\sigma_{B,B}$, nonsingular as long as all elements of $\boldsymbol{\theta}$ are finite.

Now, consider an observation process in which only relationships incident on members of group A (A - A and A - B) are observed and B - B relations are missing by design. B - B relationships are not explicitly part of $\mathbf{g}(\mathbf{y})$, and, in fact, \mathbf{y}^{obs} contains all of the information needed to compute $g_2(\cdot)$ and $g_3(\cdot)$. Then,

$$\begin{aligned} \boldsymbol{\mu}(\boldsymbol{\theta} | \mathbf{y}^{\text{obs}}) &= [|\mathbf{y}_{A,A}| + |\mathbf{y}_{A,B}| + \mathbb{E}\{|\mathbf{Y}_{B,B}| \mid \mathbf{Y} \in \mathcal{Y}(\mathbf{y}^{\text{obs}})\}, |\mathbf{y}_{A,A}|, |\mathbf{y}_{A,B}|]^\top \\ &= [|\mathbf{y}_{A,A}| + |\mathbf{y}_{A,B}| + \binom{|B|}{2}p_{B,B}, |\mathbf{y}_{A,A}|, |\mathbf{y}_{A,B}|]^\top, \end{aligned}$$

because $|\mathbf{Y}_{B,B}|$ does not depend on \mathbf{y}^{obs} . Thus, $\mu_1(\boldsymbol{\theta} \mid \mathbf{y}^{\text{obs}})$ is an affine combination of $\mu_2(\boldsymbol{\theta} \mid \mathbf{y}^{\text{obs}})$ and $\mu_3(\boldsymbol{\theta} \mid \mathbf{y}^{\text{obs}})$ for all \mathbf{y}^{obs} , making $\mathcal{I}(\boldsymbol{\theta})$ singular. More explicitly, from (2b),

$$\begin{aligned}\mathcal{I}(\boldsymbol{\theta}) &= \text{Var}_{\mathbf{Y}}[\boldsymbol{\mu}\{\boldsymbol{\theta} \mid \text{obs}(\mathbf{Y})\}] \\ &= \text{Var}\left[|\mathbf{Y}_{A,A}| + |\mathbf{Y}_{A,B}| + \binom{|B|}{2} p_{B,B}, |\mathbf{Y}_{A,A}|, |\mathbf{Y}_{A,B}|\right]^{\top} \\ &= \begin{bmatrix} \sigma_{A,A} + \sigma_{A,B} + 0 & \sigma_{A,A} & \sigma_{A,B} \\ \sigma_{A,A} & \sigma_{A,A} & 0 \\ \sigma_{A,B} & 0 & \sigma_{A,B} \end{bmatrix}.\end{aligned}$$

Here, the first column of $\mathcal{I}(\boldsymbol{\theta})$ is the sum of its second and third.

B.2 Example 2: Triadic Effects

Now, consider an independent, identically distributed sample of S undirected networks of size $n = 3$ and an ERGM with the following sufficient statistic, counting edges, 2-stars, and triangles in network \mathbf{y} :

$$\mathbf{g}(\mathbf{y}) = \left[|\mathbf{y}|, \sum_{i=1}^n \binom{|\mathbf{y}_i|}{2}, \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n y_{i,j} y_{j,k} y_{k,i} / 6\right],$$

where $|\mathbf{y}_i|$ is the number of ties incident on actor i . Each network in the sample has a sample space of size $|\mathcal{Y}| = 2^{\binom{3}{2}} = 8$, enumerated in Table B1.

For simplicity, we will set $\boldsymbol{\theta} = \mathbf{0}$, making all networks equiprobable. Then, we can evaluate the population covariance matrix and obtain

$$\mathcal{I}(\mathbf{0}) = \frac{S}{64} \begin{bmatrix} 48 & 48 & 12 \\ 48 & 60 & 18 \\ 12 & 18 & 7 \end{bmatrix}$$

and $|\mathcal{I}(\mathbf{0})| = (9/4096)S^3$. Notably, for $S = 1$, the MLE for θ_3 will be infinite due to the sufficient statistic being at its highest or lowest possible value, but this is unrelated to nonidentifiability.

Now, consider an egocentric missing data regime, where relations between Actor 1 and the others are observed but the relationship between Actor 2 and Actor 3 ($Y_{2,3}$) is missing. The conditional expectations $\boldsymbol{\mu}(\boldsymbol{\theta} \mid \mathbf{y}^{\text{obs}})$ and probabilities $\Pr(\mathbf{Y}^{\text{obs}} = \mathbf{y}^{\text{obs}})$ for every possible \mathbf{y}^{obs} are given in Table B2. Given those, the covariance matrix $\text{Var}_{\mathbf{Y}}[\boldsymbol{\mu}\{\boldsymbol{\theta} \mid \text{obs}(\mathbf{Y})\}]$ gives

$$\mathcal{I}(\mathbf{0}) = \frac{S}{64} \begin{bmatrix} 32 & 32 & 8 \\ 32 & 36 & 10 \\ 8 & 10 & 3 \end{bmatrix}$$

1
2
3
4
5
6
7
8 and $|\mathcal{I}(\mathbf{0})| = 0$.
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13 Table B1: Possible undirected networks of size 3 and their edge, 2-star, and
14 triangle statistics.

$y_{1,2}$	$y_{1,3}$	$y_{2,3}$	edges	2-stars	triangles
0	0	0	0	0	0
0	0	1	1	0	0
0	1	0	1	0	0
1	0	0	1	0	0
0	1	1	2	1	0
1	0	1	2	1	0
1	1	0	2	1	0
1	1	1	3	3	1

27
28
29
30
31
32
33
34
35
36
37 Table B2: Conditional expectations of edge, 2-star, and triangle statistics
38 for undirected networks of size 3 with $Y_{2,3}$ unobserved and their probabilities
39 under the null model.

$\Pr(Y_{1,2} = y_{1,2}, Y_{1,3} = y_{1,3})$	$y_{1,2}$	$y_{1,3}$	edges	2-stars	triangles
0.25	0	0	0.5	0.0	0.0
0.25	0	1	1.5	0.5	0.0
0.25	1	0	1.5	0.5	0.0
0.25	1	1	2.5	2.0	0.5

C Mathematical and Computational Details of Diagnostic Methods

C.1 Estimating the Variance of the Residuals for Partially Observed Networks

Here, we discuss estimation of $\text{Var}_{\vec{\mathbf{Y}}}[\tau\{\hat{\beta} \mid \text{obs}(\vec{\mathbf{Y}})\}]$ in (3b) in Section 4.2. This is a nested expectation, which calls for nested simulation of network statistics: first, we simulate a sample of complete networks $\vec{\mathbf{Y}}^{(1)}, \dots, \vec{\mathbf{Y}}^{(R_1)} \sim \text{ERGM}_{z, \vec{\mathcal{Y}}, \vec{x}, \vec{g}}(\beta)$, and then, for each $r_1 \in \{1, \dots, R_1\}$, simulate a sample $\vec{\mathbf{Y}}^{(r_1,1)}, \dots, \vec{\mathbf{Y}}^{(r_1,R_2)} \sim \text{ERGM}_{z, \vec{\mathcal{Y}}\{\text{obs}(\vec{\mathbf{Y}}^{(r_1)})\}, \vec{x}, \vec{g}}(\beta)$. We consider two ways to use these realisations to obtain the estimator.

Direct method Substituting simulated values into the expectation and the variance yields a consistent (as $R_1 \rightarrow \infty$ and $R_2 \rightarrow \infty$) estimator. This direct estimator for the variance of the conditional expectation in the divisor in (3b) has the following form:

$$\frac{1}{R_1 - 1} \sum_{r_1=1}^{R_1} \left\{ \overline{t(\vec{\mathbf{Y}}^{(r_1,:)})} - \overline{t(\vec{\mathbf{Y}}^{(:,,:)})} \right\}^2,$$

where $\overline{t(\vec{\mathbf{Y}}^{(r_1,:)})} \stackrel{\text{def}}{=} R_2^{-1} \sum_{r_2=1}^{R_2} t(\vec{\mathbf{Y}}^{(r_1,r_2)})$, an estimator of the inner expectation for a given $\text{obs}(\vec{\mathbf{Y}}^{(r_1)})$, and $\overline{t(\vec{\mathbf{Y}}^{(:,,:)})} \stackrel{\text{def}}{=} R_1^{-1} \sum_{r_1=1}^{R_1} \overline{t(\vec{\mathbf{Y}}^{(r_1,:)})}$, the grand mean. Here, we derive the approximate expression for its bias due to the variance capturing the additional variability (decreasing in R_2) in estimating the inner expectation by simulation.

Assuming that R_1 is sufficiently large that the variance of $\overline{t(\vec{\mathbf{Y}}^{(:,,:)})}$ is negligible, and since $\mathbb{E}\{t(\vec{\mathbf{Y}}^{(r_1,:)}) - t(\vec{\mathbf{Y}}^{(:,,:)})\} = 0$,

$$\begin{aligned} \mathbb{E}\{\overline{t(\vec{\mathbf{Y}}^{(r_1,:)})} - \overline{t(\vec{\mathbf{Y}}^{(:,,:)})}\}^2 &= \text{Var}\{\overline{t(\vec{\mathbf{Y}}^{(r_1,:)})} - \overline{t(\vec{\mathbf{Y}}^{(:,,:)})}\} \approx \text{Var}\{\overline{t(\vec{\mathbf{Y}}^{(r_1,:)})}\} \\ &\approx \text{Var}[\mathbb{E}\{\overline{\mathbf{g}(\vec{\mathbf{Y}}^{(r_1,:)})} \mid \vec{\mathbf{Y}}^{(r_1,1)}, \dots, \vec{\mathbf{Y}}^{(r_1,R_2)} \in \mathcal{Y}((\vec{\mathbf{Y}}^{(r_1)})^{\text{obs}})\}] \\ &\quad + \mathbb{E}[\text{Var}\{\overline{\mathbf{g}(\vec{\mathbf{Y}}^{(r_1,:)})} \mid \vec{\mathbf{Y}}^{(r_1,1)}, \dots, \vec{\mathbf{Y}}^{(r_1,R_2)} \in \mathcal{Y}((\vec{\mathbf{Y}}^{(r_1)})^{\text{obs}})\}] \\ &\approx \text{Var}[\mathbb{E}\{\mathbf{g}(\vec{\mathbf{Y}}') \mid \vec{\mathbf{Y}}' \in \mathcal{Y}((\vec{\mathbf{Y}})^{\text{obs}})\}] \\ &\quad + \frac{1}{R_2} \mathbb{E}[\text{Var}\{\mathbf{g}(\vec{\mathbf{Y}}') \mid \vec{\mathbf{Y}}' \in \mathcal{Y}((\vec{\mathbf{Y}})^{\text{obs}})\}] \end{aligned}$$

Therefore, this estimator is biased, with bias decreasing as a function of R_2 .
An adjusted estimator could then be

$$\frac{1}{R_1 - 1} \sum_{r_1=1}^{R_1} \{\overline{t(\vec{Y}^{(r_1,:)})} - \overline{t(\vec{Y}^{(:,,:)})}\}^2 - \frac{1}{R_2} \frac{1}{R_1} \sum_{r_1=1}^{R_1} \frac{1}{R_2 - 1} \sum_{r_2=1}^{R_2} \{t(\vec{Y}^{(r_1,r_2)}) - \overline{t(\vec{Y}^{(r_1,:)})}\}^2.$$

Law of Total Variance method This approach takes advantage of the Law of Total Variance to write:

$$\begin{aligned} \text{Var}_{\vec{Y}}[\tau\{\hat{\beta} \mid \text{obs}(\vec{Y})\}] &= \Psi(\hat{\beta}) - \mathbb{E}_{\vec{Y}}[\Psi\{\hat{\beta} \mid \text{obs}(\vec{Y})\}], \\ &\approx \frac{1}{R_1 R_2 - 1} \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \{t(\vec{Y}^{(i,j)}) - \overline{t(\vec{Y}^{(:,,:)})}\}^2 \\ &\quad - \frac{1}{R_1} \sum_{r_1=1}^{R_1} \frac{1}{R_2 - 1} \sum_{r_2=1}^{R_2} \{t(\vec{Y}^{(i,j)}) - \overline{t(\vec{Y}^{(i,:)})}\}^2, \end{aligned}$$

an unbiased (except perhaps from MCMC autocorrelation) estimator with an added benefit that some of the simulation error in the minuend and the subtrahend cancels. We derive its properties as follows.

Suppose that each $\vec{Y}^{(r_1)}$ induces a distribution of $t(\vec{Y}^{(r_1,r_2)})$ with conditional variance $\sigma_{r_1}^2$ (i.e., $\text{Var}\{t(\vec{Y}^{(r_1,r_2)}) \mid \vec{Y}^{(r_1,r_2)} \in \mathcal{Y}((\vec{Y}^{(r_1)})^{\text{obs}})\} = \sigma_{r_1}^2$). The unbiased estimator is

$$\tilde{\sigma}_{r_1}^2 = \frac{1}{R_2 - 1} \sum_{r_2=1}^{R_2} \{t(\vec{Y}^{(r_1,r_2)}) - \overline{t(\vec{Y}^{(r_1,:)})}\}^2$$

We wish to estimate the expected variance and, under the unconstrained process, $\mathbb{E}(\sigma_{r_1}^2) = \sigma^2$. Suppose that $\text{Var}(\sigma_{r_1}^2) = \alpha$. Then, the estimator

$$\tilde{\sigma}^2 = \frac{1}{R_1} \sum_{r_1=1}^{R_1} \tilde{\sigma}_{r_1}^2$$

has expectation

$$\begin{aligned} \mathbb{E}(\tilde{\sigma}^2) &= \frac{1}{R_1} \sum_{r_1=1}^{R_1} \mathbb{E}(\tilde{\sigma}_{r_1}^2) = \frac{1}{R_1} \sum_{r_1=1}^{R_1} \mathbb{E}_{\sigma_{r_1}^2} \{\mathbb{E}_{\tilde{\sigma}_{r_1}^2}(\tilde{\sigma}_{r_1}^2 \mid \sigma_{r_1}^2)\} \\ &= \frac{1}{R_1} \sum_{r_1=1}^{R_1} \mathbb{E}_{\sigma_{r_1}^2} \{\sigma_{r_1}^2\} = \sigma^2, \end{aligned}$$

so it is unbiased regardless of R_2 , R_1 , and the distribution of $t(\vec{Y}^{(r_1, r_2)})$.

To approximate its variance, we must assume a conditional distribution $t(\vec{Y}^{(r_1, r_2)}) | \vec{Y}^{(r_1)}$. If it is approximately normal,

$$\tilde{\sigma}_{r_1}^2 | \sigma_{r_1}^2 \sim \frac{\sigma_{r_1}^2}{R_2 - 1} \chi_{R_2 - 1}^2.$$

Its variance is

$$\begin{aligned} \mathbb{V}\text{ar}(\tilde{\sigma}^2) &= \frac{1}{R_1^2} \sum_{r_1=1}^{R_1} \mathbb{V}\text{ar}(\tilde{\sigma}_{r_1}^2) \\ &= \frac{1}{R_1^2} \sum_{r_1=1}^{R_1} [\mathbb{E}_{\sigma_{r_1}^2} \{ \mathbb{V}\text{ar}_{\tilde{\sigma}_{r_1}^2} (\tilde{\sigma}_{r_1}^2 | \sigma_{r_1}^2) \} + \mathbb{V}\text{ar}_{\sigma_{r_1}^2} \{ \mathbb{E}_{\tilde{\sigma}_{r_1}^2} (\tilde{\sigma}_{r_1}^2 | \sigma_{r_1}^2) \}] \\ &= \frac{1}{R_1^2} \sum_{r_1=1}^{R_1} [\mathbb{E}_{\sigma_{r_1}^2} \left\{ \frac{2\sigma_{r_1}^4(R_2 - 1)}{(R_2 - 1)^2} \right\} + \mathbb{V}\text{ar}_{\sigma_{r_1}^2} \{ \sigma_{r_1}^2 \}] \\ &= \frac{1}{R_1^2} \sum_{r_1=1}^{R_1} \left[\frac{2}{R_2 - 1} \mathbb{E}_{\sigma_{r_1}^2} \{ \sigma_{r_1}^4 \} + \alpha \right] \\ &= \frac{1}{R_1^2} \sum_{r_1=1}^{R_1} \left[\frac{2}{R_2 - 1} \{ \mathbb{E}_{\sigma_{r_1}^2}^2 (\sigma_{r_1}^2) + \mathbb{V}\text{ar}(\sigma_{r_1}^2) \} + \alpha \right] \\ &= \frac{1}{R_1^2} \sum_{r_1=1}^{R_1} \left[\frac{2}{R_2 - 1} \{ \sigma^4 + \alpha \} + \alpha \right] = \frac{1}{R_1} \left[\frac{2}{R_2 - 1} \sigma^4 + \left(\frac{2}{R_2 - 1} + 1 \right) \alpha \right]. \end{aligned}$$

A corollary of this is that increasing R_2 can only increase the precision up to a point, whereas as long as $R_2 \geq 2$, R_1 increases precision without limit. Nonetheless, if $\sigma^4 \gg \alpha$, it may be worthwhile to increase R_2 , since conditional simulation is computationally cheaper.

C.2 Simulation-based Score Test for Dataset Effects

Here, we describe two variants of the simulation Score test for dataset effects. Here, the null hypothesis is the candidate model, which does not include explicit dataset effects, and the alternative is the same model with an explicit dataset effect. Rejection of this null hypothesis therefore indicates that the dataset effect is not ignorable.

Due to the exponential family nature of the model, the test works by specifying a statistic $t_{\text{dataset}}(\vec{y})$ that is not a part of the candidate model and that contains an explicit dataset effect, simulate complete network datasets

1
2
3
4
5
6
7
8 $\vec{\mathbf{Y}}^{(1)}, \dots, \vec{\mathbf{Y}}^{(R)} \stackrel{\text{i.i.d.}}{\sim} \text{ERGM}(\hat{\beta})$ from the candidate model, and evaluate the
9 sample quantile $q \stackrel{\text{def}}{=} \sum_{r=1}^R \mathbb{I}\{t_{\text{dataset}}(\vec{\mathbf{y}}) \leq t_{\text{dataset}}(\vec{\mathbf{Y}}^{(r)})\}$. Then, $2 \min(q, 1-q)$
10 is a Score test P -value for the null hypothesis $\theta_{\text{dataset}} = 0$.

11 In our application, a straightforward $t_{\text{dataset}}(\vec{\mathbf{y}}) = \sum_{s \in H} |\mathbf{y}_s|$: separate
12 density for networks in the H dataset, which only depends on the completely
13 observed networks. (Applying it to partially observed networks would require
14 further adjustment along the lines of Section 4.2.)

15 A joint test for several network features is also possible: if $\mathbf{t}_{\text{dataset}}(\vec{\mathbf{y}})$
16 is vector-valued, then, with \mathbf{m} the estimate of $\mathbb{E}\{\mathbf{t}_{\text{dataset}}(\vec{\mathbf{Y}})\}$ and V the
17 estimate of $\text{Var}\{\mathbf{t}_{\text{dataset}}(\vec{\mathbf{Y}})\}$, both obtained by simulation, $\chi^2 = \{\mathbf{m} -$
18 $\mathbf{t}_{\text{dataset}}(\vec{\mathbf{y}})\}^\top V^{-1} \{\mathbf{m} - \mathbf{t}_{\text{dataset}}(\vec{\mathbf{y}})\}$ provides an omnibus score test. Unlike the
19 quantile-based test, this test relies on approximate normality of $\mathbf{t}_{\text{dataset}}(\vec{\mathbf{Y}})$.
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

D Capacity to Detect Misspecification

To illustrate misspecification detection, we fit a number of reduced models based on *Model 1*:

Model 0 model with only edge count, 2-star, and triangle statistics, with their linear and quadratic network size effects;

Model 1i dyad-independent submodel of *Model 1*, excluding 2-star and triangle effects;

Model 1n *Model 1* without linear or quadratic network size effects.

A comparison of their specification to that of the other models is given in Table E1.

Hypothesis testing for network size misspecification In *Model 1n*, which lacks network size effects, tests on residuals reject the hypothesis of fit for edges and triangles (*P*-vals. 0.022 and 0.035, respectively), but not 2-stars (*P*-val. = 0.069). In *Model 0*, which does incorporate them, none do (*P*-vals. 0.88, 0.82, and 0.84, respectively). *Model 1i*, which has network size effects for density (edge count) but not other features rejects hypothesis of fit for triangles (*P*-val. = 0.0005) but not for edges or 2-stars (*P*-vals. 0.40 and 0.079, respectively). Thus, the method appears to be capable of detecting and identifying network size effect misspecification with some detail.

Residual plots for dataset effects Turning to residual plots, Figure D1 shows the edge residual plots for *Model 1* (Panel (a)), *Model 1n* (Panel (b)), and *Model 0* (Panel (c)). Observe that unlike the *Model 1*'s, *Model 0*'s residuals for households without a child are systematically different from those with a child (from either dataset). Thus, the dataset effect is not ignorable. On the other hand, the lack of network size effects in *Model 1n* does not appear to be noticeable.

Density error plots for network size and dataset effects Figure D2 provides density effects plots for *Model 1*, *Model 1n*, and *Model 0*. We see in Panel (b) (*Model 1n*) that removing the network size effects does not significantly affect density predictions, as the other 25 parameters adjust to compensate, but we can observe a fairly consistent trend in density errors as a function of network size: for very small and large networks, the densities are underpredicted, but for mid-sized networks, they are overpredicted.

Table D1: Standard deviations of Pearson residuals for the specified network statistic under each model.

Model	p	edges	2-stars	triangles
<i>Model 0</i>	9	0.95	0.94	0.94
<i>Model 1i</i>	25	1.09	1.01	0.91
<i>Model 1n</i>	25	0.95	0.94	0.91
<i>Model 1</i>	31	0.96	0.95	0.93

(This pattern is not found in *Model 1*'s Panel (a).) *Model 0* (Panel (c)), as expected, shows large density prediction errors, since it does not attempt to account for household composition.

Pearson residual standard deviation for unaccounted-for heterogeneity Lastly, we consider the proposed diagnostic for unexplained heterogeneity: that Pearson residual variances exceed 1. Residual standard deviations are given in Table D1 for *Model 1* and the three reduced models. Interestingly, all the dyad-dependent models—even *Model 0*—have residual standard deviations less than 1. This is likely because both the 2-star and the triangle statistic create positive dependence among relations in the network, absorbing the variation along the lines of Butts (2017). For small networks in particular, it might not be possible to distinguish anything but the most severe network-level heterogeneity from more local heterogeneity. On the other hand, *Model 1i*, despite having as many parameters as *Model 1n* and almost thrice as many as *Model 0*, lacks dyad-dependent effects and cannot absorb this variation. Accordingly, its Pearson residual standard deviations exceed 1 for edges and 2-stars.

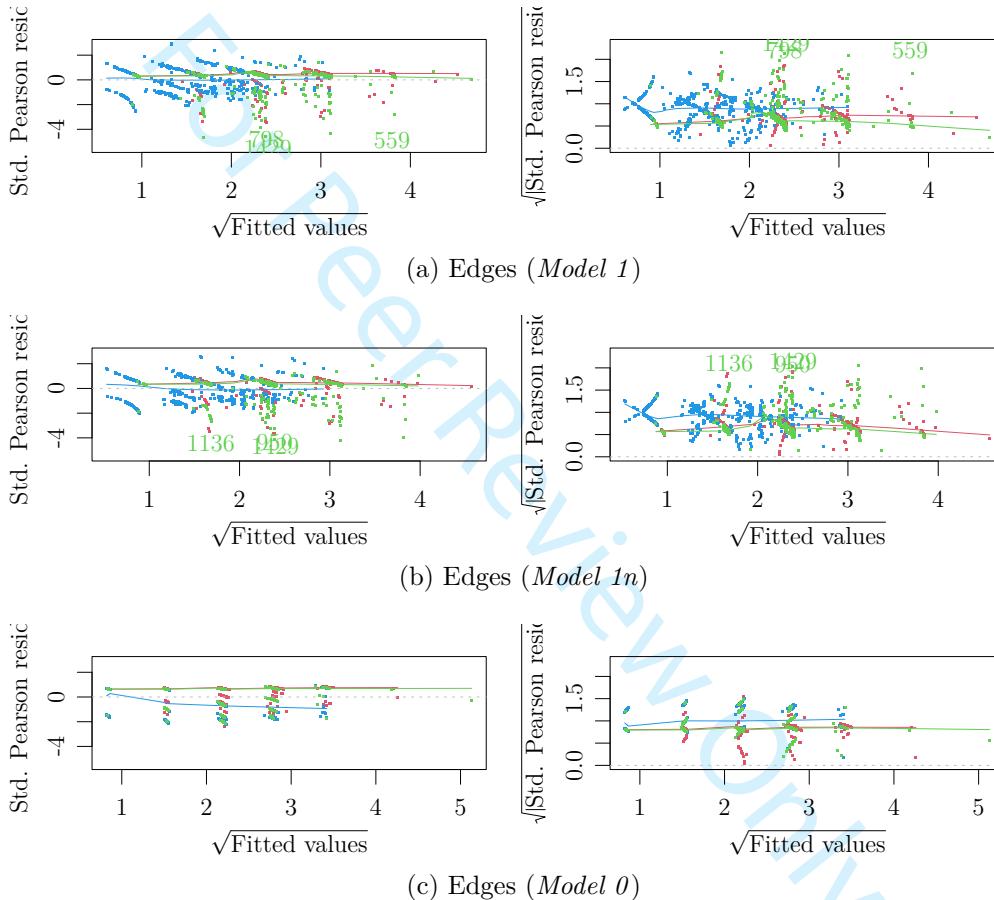


Figure D1: Pearson residual plots of network statistics for *Model 1* and its reduced models. (● all H ● E w/ child ● E w/o child)

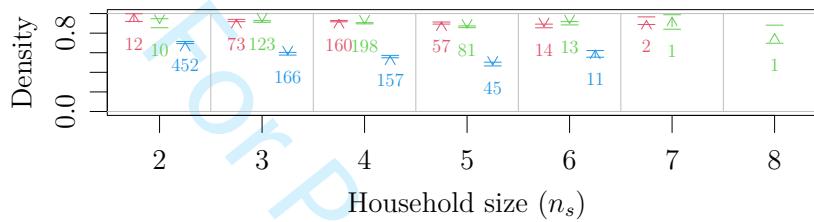
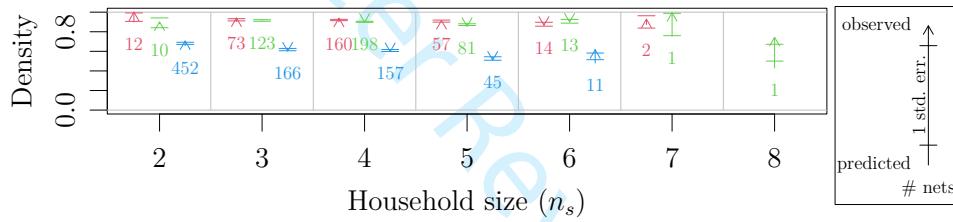
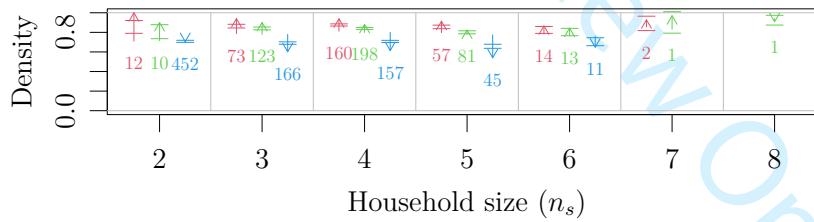
(a) *Model 1: Full model*(b) *Model 1n: Model 1 without network size effects*(c) *Model 0: Minimal model with network size effects only*

Figure D2: Average prediction errors of density. Values are averaged over the networks grouped by size and subdataset. (● all H ● E w/ child ● E w/o child)

E Additional Results and Diagnostics

Seven models, four (1 , 2 , 3 , and 4) substantive and three (0 , $1i$, and $1n$) to study misspecification detection were fit. Their motivation is described in Section 5.2 in the body of the article and in Appendix D, respectively.

Table E1 shows a side-by-side comparison of which effects were included in each model, and Table E2 gives their fit summaries (number of parameters, log-likelihood, AIC, and BIC).

In the following subsections, we provide the details for each of these models:

- full coefficient table;
- plots of residuals vs. fitted values and network size for edge, 2-star, and triangle counts;
- plot of residuals for edges incident on actors of each age against age;
- plot of prediction errors in density by network size and subdataset;
- Score tests for dataset effect net of the model;
- results from regressing residuals for edges, 2-stars, and triangles on candidate network-level predictors; and
- sample standard deviations of Pearson residuals for some network statistics under the model.

Table E1: Effects included in each model.

Effect ($\times z_s$)	0	1 <i>i</i>	1 <i>n</i>	1	2	3	4
edges	✓						
$\times \log(n_s)$		✓	✓	✓	✓	✓	✓
$\times \log^2(n_s)$		✓	✓	✓	✓	✓	✓
if Brussels post code		✓	✓	✓	✓	✓	✓
if on weekend		✓	✓	✓	✓	✓	✓
$\times \log(\text{pop. dens. in post code})$				✓	✓	✓	✓
2-stars	✓		✓	✓	✓	✓	✓
$\times \log(n_s)$		✓		✓	✓	✓	✓
$\times \log^2(n_s)$		✓		✓	✓	✓	✓
$\times \log(\text{pop. dens. in post code})$				✓			
if Brussels post code							✓
if on weekend							✓
triangles		✓		✓	✓	✓	✓
$\times \log(n_s)$		✓		✓	✓	✓	✓
$\times \log^2(n_s)$		✓		✓	✓	✓	✓
$\times \log(\text{pop. dens. in post code})$				✓			
if Brussels post code							✓
if on weekend							✓
Young Child with Young Child		✓	✓	✓	✓	✓	✓
Young Child with Preadolescent		✓	✓	✓	✓	✓	✓
Preadolescent with Preadolescent		✓	✓	✓	✓	✓	✓
Adolescent with Adolescent		✓	✓	✓	✓	✓	✓
Young Child with Young Adult		✓	✓	✓	✓	✓	✓
Young Adult with Young Adult		✓	✓	✓	✓	✓	✓
Young Child with Female Older Adult		✓	✓	✓	✓	✓	✓
Preadolescent with Female Older Adult		✓	✓	✓	✓	✓	✓
Adolescent with Female Older Adult		✓	✓	✓	✓	✓	✓
Female Older Adult with Female Older Adult		✓	✓	✓	✓	✓	✓
Young Child with Male Older Adult		✓	✓	✓	✓	✓	✓
Preadolescent with Male Older Adult		✓	✓	✓	✓	✓	✓
Adolescent with Male Older Adult		✓	✓	✓	✓	✓	✓
Female Older Adult with Male Older Adult		✓	✓	✓	✓	✓	✓
if child absent		✓	✓	✓	✓	✓	✓
Male Older Adult with Male Older Adult		✓	✓	✓	✓	✓	✓
Senior with Senior		✓	✓	✓	✓	✓	✓
Adolescent with Young Child or Preadolescent		✓	✓	✓	✓	✓	✓
Young Adult with Preadolescent or Adolescent		✓	✓	✓	✓	✓	✓
Young Adult with Older Adult		✓	✓	✓	✓	✓	✓
Senior with non-Senior		✓	✓	✓	✓	✓	✓

Table E2: Model fit summaries. MCMC standard errors are given in parentheses.

Model	p	log-likelihood	AIC	BIC
<i>Model 0</i>	9	-1920 (0.2)	3859 (0.5)	3917 (0.5)
<i>Model 1i</i>	25	-1851 (0.0)	3753 (0.0)	3915 (0.0)
<i>Model 1n</i>	25	-1696 (0.2)	3442 (0.5)	3604 (0.5)
<i>Model 1</i>	31	-1664 (0.2)	3390 (0.5)	3590 (0.5)
<i>Model 2</i>	34	-1661 (0.2)	3390 (0.5)	3610 (0.5)
<i>Model 3</i>	32	-1663 (0.2)	3391 (0.5)	3598 (0.5)
<i>Model 4</i>	38	-1659 (0.2)	3393 (0.4)	3639 (0.4)

1
2
3
4
5
6
7
8 **E.1 Full results and diagnostics for *Model 0* (the min-**
9 **imal model)**

12 Table E3: Parameter estimates (and standard errors) for *Model 0*.
13
14

Effect ($\times z_s$)	Coefficient (SE)
edges	3.93 (2.05)
$\times \log(n_s)$	-5.53 (4.23)
$\times \log^2(n_s)$	1.69 (1.88)
2-stars	0.48 (4.15)
$\times \log(n_s)$	-1.13 (4.56)
$\times \log^2(n_s)$	0.38 (1.24)
triangles	7.67 (9.00)
$\times \log(n_s)$	-4.44 (10.52)
$\times \log^2(n_s)$	0.61 (3.06)

26 Significance: *** ≤ 0.001 <** ≤ 0.01 <* ≤ 0.05
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54

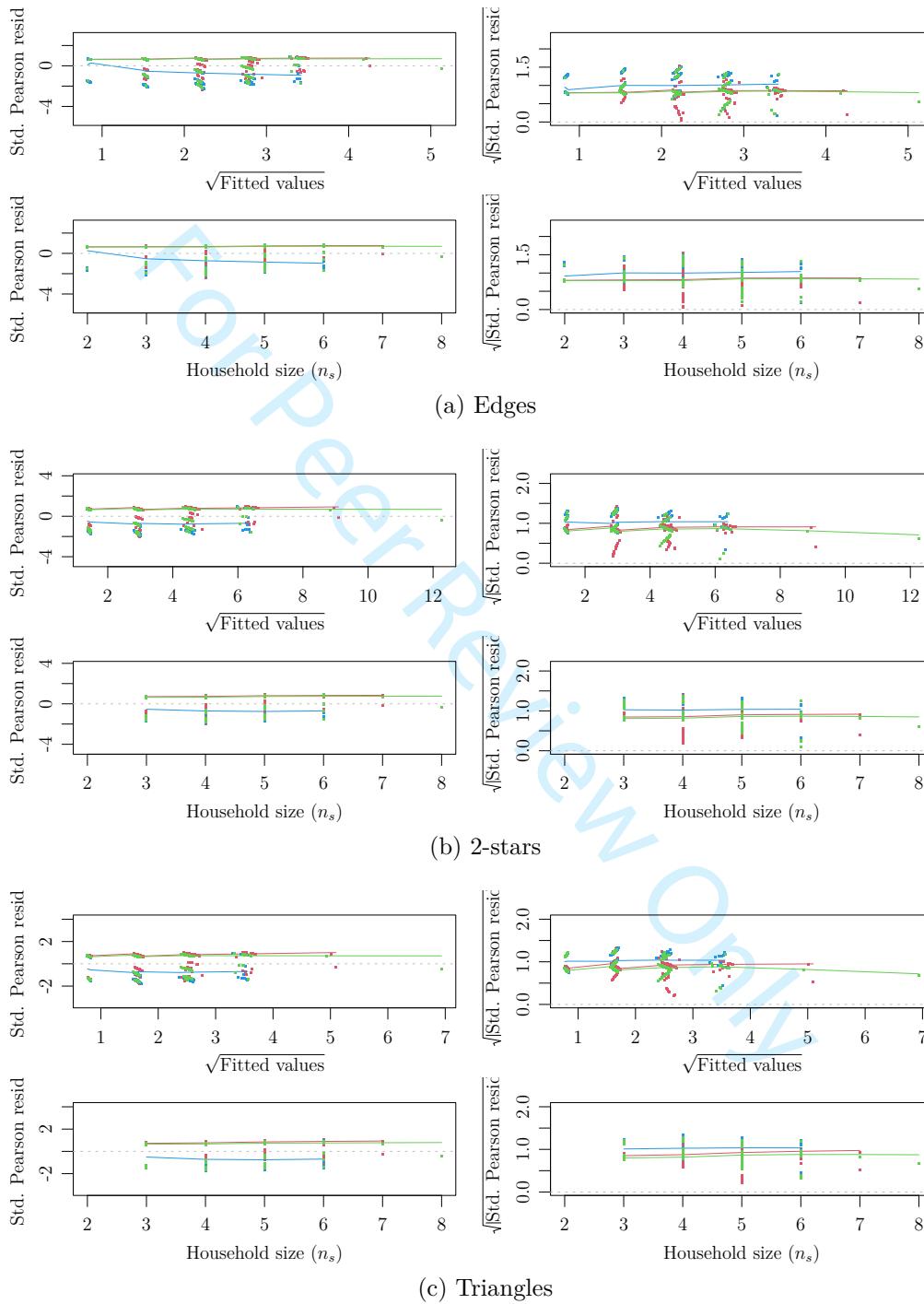


Figure E1: Residual plots of network statistics against fitted values and network size for *Model 0*. (● all H ● E w/ child ● E w/o child)

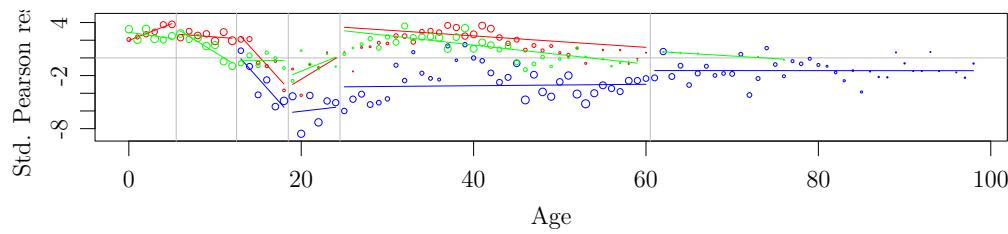


Figure E2: Residual plot for *Model 0* of edges incident on actors of a given age against age. (● all H ● E w/ child ● E w/o child)

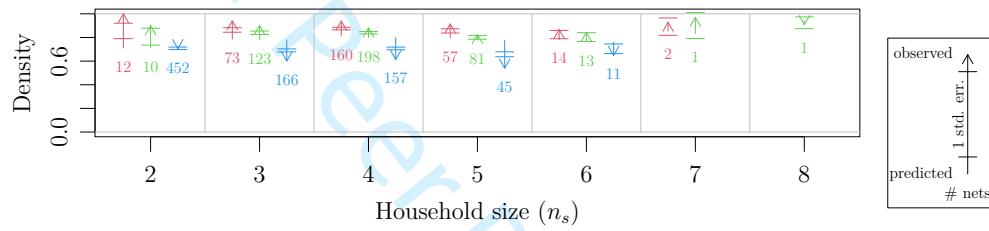


Figure E3: Average prediction errors of density in *Model 0*. Values are averaged over the networks grouped by size and subdataset. (● all H ● E w/ child ● E w/o child)

Table E4: Regression of residuals of *Model 0* for each network structural property on each candidate predictor.

Effect ($\times z_s$)	Wald χ^2 (df)	P-val.
edges		
$\times \log(\text{pop. dens. in post code})$ (linear)	4.8 (1)	0.03
$\times \log(\text{pop. dens. in post code})$ (quadratic)	5.2 (2)	0.07
if city post code	4.3 (1)	0.04
if Brussels post code	7.8 (1)	0.01
if weekend	0.8 (1)	0.38
2-stars		
$\times \log(\text{pop. dens. in post code})$ (linear)	17.1 (1)	< 0.01
$\times \log(\text{pop. dens. in post code})$ (quadratic)	17.7 (2)	< 0.01
if city post code	16.1 (1)	< 0.01
if Brussels post code	6.6 (1)	0.01
if weekend	2.1 (1)	0.15
triangles		
$\times \log(\text{pop. dens. in post code})$ (linear)	17.7 (1)	< 0.01
$\times \log(\text{pop. dens. in post code})$ (quadratic)	18.3 (2)	< 0.01
if city post code	15.3 (1)	< 0.01
if Brussels post code	6.8 (1)	0.01
if weekend	2.3 (1)	0.13

Table E5: Score tests of the null hypothesis of no effect of dataset H on the specified network statistic over and above *Model 0*.

Statistic	Score χ^2 (df)	P-val.
Omnibus	60.8 (2)	< 0.0001
edges	nonparam.	< 0.0001
2-stars	nonparam.	< 0.0001

Table E6: Sample standard deviations of Pearson residuals for edge, two-star, and triangle counts in *Model 0* for the dataset and its subsets. Values substantially greater than 1 indicate overdispersion relative to the model.

Effect ($\times z_s$)	Overall	H	E (w/ child)	E (w/o child)
edges	0.95	0.53	0.71	1.07
2-stars	0.94	0.62	0.75	1.00
triangles	0.94	0.70	0.76	0.97

1
2
3
4
5
6
7
8 **E.2 Full results and diagnostics for *Model 1i* (dyad-
9 independent model)**
10
11

12 Table E7: Parameter estimates (and standard errors) for *Model 1i*.
13
14

Effect ($\times z_s$)	Coefficient (SE)
edges $\times \log(n_s)$	-5.01 (0.98)***
$\times \log^2(n_s)$	1.88 (0.39)***
if Brussels post code	0.43 (0.30)
if on weekend	0.22 (0.11)*
Young Child with Young Child	5.57 (0.71)***
Young Child with Preadolescent	6.27 (0.73)***
Preadolescent with Preadolescent	5.09 (0.66)***
Adolescent with Adolescent	3.16 (0.68)***
Young Child with Young Adult	5.27 (1.21)***
Young Adult with Young Adult	3.12 (0.70)***
Young Child with Female Older Adult	6.89 (0.71)***
Preadolescent with Female Older Adult	6.02 (0.65)***
Adolescent with Female Older Adult	4.65 (0.64)***
Female Older Adult with Female Older Adult	2.81 (0.67)***
Young Child with Male Older Adult	6.16 (0.67)***
Preadolescent with Male Older Adult	5.43 (0.64)***
Adolescent with Male Older Adult	4.17 (0.64)***
Female Older Adult with Male Older Adult	6.05 (0.66)***
if child absent	-2.18 (0.24)***
Male Older Adult with Male Older Adult	2.25 (0.68)***
Senior with Senior	2.94 (0.53)***
Adolescent with Young Child or Preadolescent	4.49 (0.64)***
Young Adult with Preadolescent or Adolescent	2.52 (0.65)***
Young Adult with Older Adult	3.23 (0.62)***
Senior with non-Senior	3.36 (0.59)***

43 Significance: *** ≤ 0.001 ** ≤ 0.01 * ≤ 0.05
44
45
46
47
48
49
50
51
52
53
54

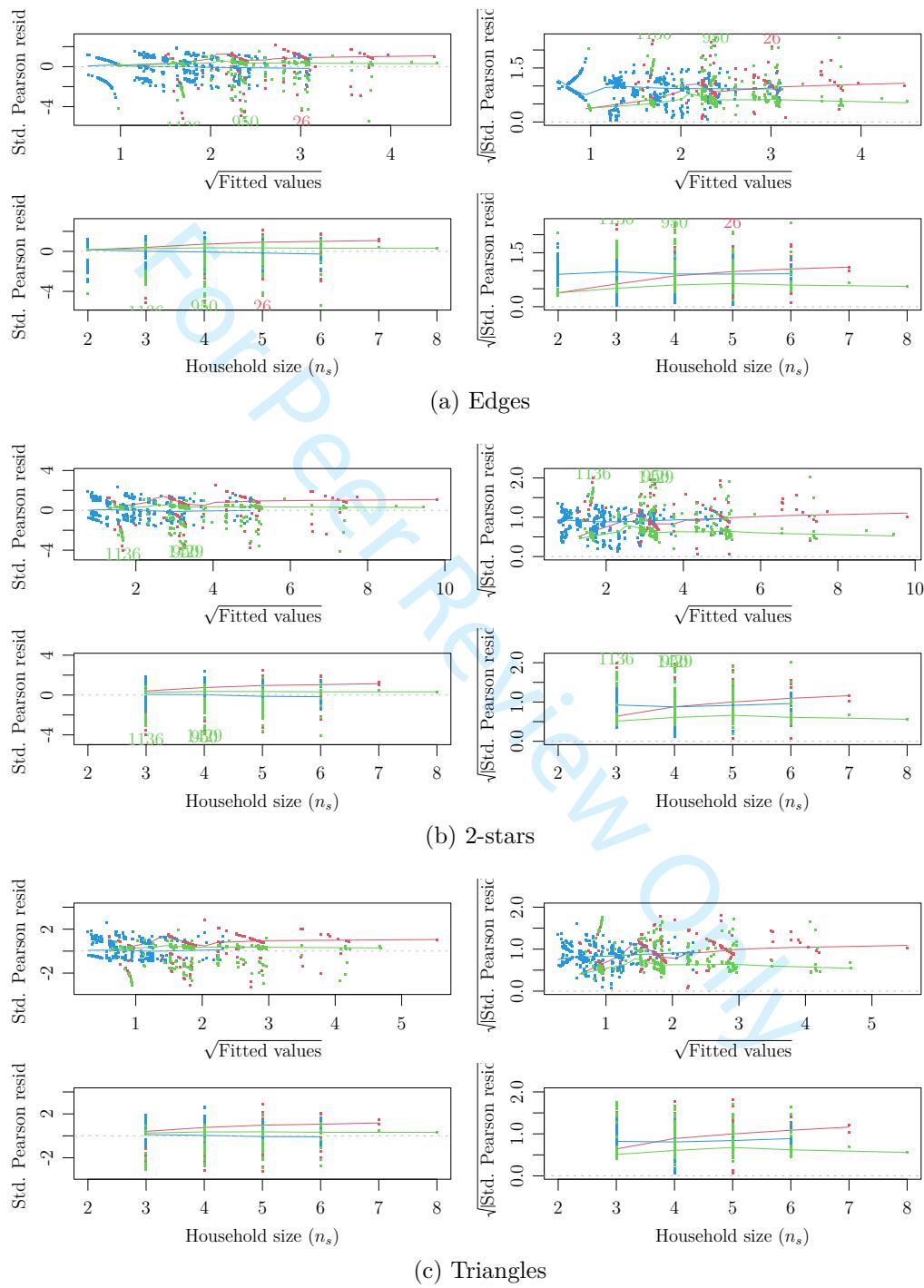


Figure E4: Residual plots of network statistics against fitted values and network size for *Model 1i*. (● all H ● E w/ child ● E w/o child)

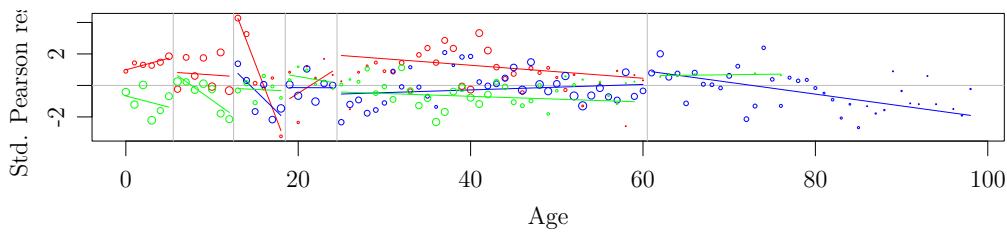


Figure E5: Residual plot for *Model 1i* of edges incident on actors of a given age against age. (● all H ● E w/ child ● E w/o child)

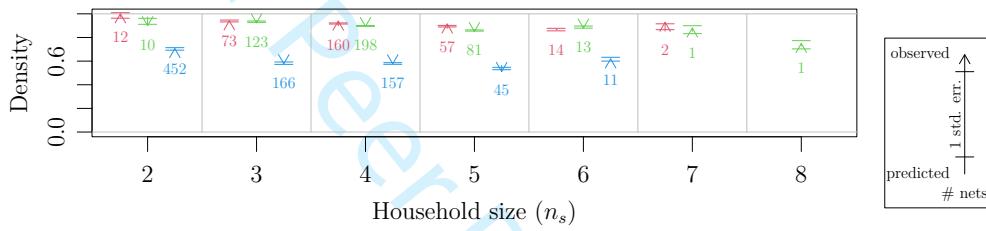


Figure E6: Average prediction errors of density in *Model 1i*. Values are averaged over the networks grouped by size and subdataset. (● all H ● E w/ child ● E w/o child)

Table E8: Regression of residuals of *Model 1i* for each network structural property on each candidate predictor.

Effect ($\times z_s$)	Wald χ^2 (df)	P-val.
edges		
$\times \log(\text{pop. dens. in post code})$ (linear)	4.2 (1)	0.04
$\times \log(\text{pop. dens. in post code})$ (quadratic)	5.8 (2)	0.06
if city post code	2.6 (1)	0.11
if Brussels post code	0.3 (1)	0.61
if weekend	1.2 (1)	0.27
2-stars		
$\times \log(\text{pop. dens. in post code})$ (linear)	3.0 (1)	0.08
$\times \log(\text{pop. dens. in post code})$ (quadratic)	3.8 (2)	0.15
if city post code	2.5 (1)	0.11
if Brussels post code	0.1 (1)	0.78
if weekend	1.6 (1)	0.21
triangles		
$\times \log(\text{pop. dens. in post code})$ (linear)	2.7 (1)	0.10
$\times \log(\text{pop. dens. in post code})$ (quadratic)	3.5 (2)	0.18
if city post code	1.8 (1)	0.18
if Brussels post code	0.1 (1)	0.77
if weekend	0.9 (1)	0.33

Table E9: Score tests of the null hypothesis of no effect of dataset H on the specified network statistic over and above *Model 1i*.

Statistic	Score χ^2 (df)	P-val.
Omnibus	49.3 (2)	< 0.0001
edges	nonparam.	< 0.0001
2-stars	nonparam.	< 0.0001

Table E10: Sample standard deviations of Pearson residuals for edge, two-star, and triangle counts in *Model 1i* for the dataset and its subsets. Values substantially greater than 1 indicate overdispersion relative to the model.

Effect ($\times z_s$)	Overall	H	E (w/ child)	E (w/o child)
edges	1.09	1.16	1.14	1.01
2-stars	1.01	1.05	1.00	0.93
triangles	0.91	0.96	0.89	0.82

1
2
3
4
5
6
7
8 **E.3 Full results and diagnostics for *Model 1n* (*Model 1*
9 without network size effects)**

10
11
12 Table E11: Parameter estimates (and standard errors) for *Model 1n*.
13
14

Effect ($\times z_s$)	Coefficient (SE)
edges if Brussels post code	0.12 (0.20)
if on weekend	0.19 (0.07)**
2-stars	-0.97 (0.07)***
triangles	2.90 (0.18)***
Young Child with Young Child	0.58 (0.35)
Young Child with Preadolescent	1.33 (0.40)***
Preadolescent with Preadolescent	0.46 (0.24)
Adolescent with Adolescent	-0.13 (0.29)
Young Child with Young Adult	1.27 (0.90)
Young Adult with Young Adult	0.49 (0.31)
Young Child with Female Older Adult	2.71 (0.34)***
Preadolescent with Female Older Adult	2.04 (0.21)***
Adolescent with Female Older Adult	1.26 (0.18)***
Female Older Adult with Female Older Adult	-0.39 (0.30)
Young Child with Male Older Adult	1.83 (0.26)***
Preadolescent with Male Older Adult	1.15 (0.18)***
Adolescent with Male Older Adult	0.67 (0.18)***
Female Older Adult with Male Older Adult	1.95 (0.22)***
if child absent	-0.92 (0.24)***
Male Older Adult with Male Older Adult	-0.89 (0.32)**
Senior with Senior	0.35 (0.16)*
Adolescent with Young Child or Preadolescent	0.46 (0.21)*
Young Adult with Preadolescent or Adolescent	0.18 (0.23)
Young Adult with Older Adult	0.22 (0.11)*
Senior with non-Senior	0.55 (0.14)***

43 Significance: *** ≤ 0.001 <** ≤ 0.01 <* ≤ 0.05
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

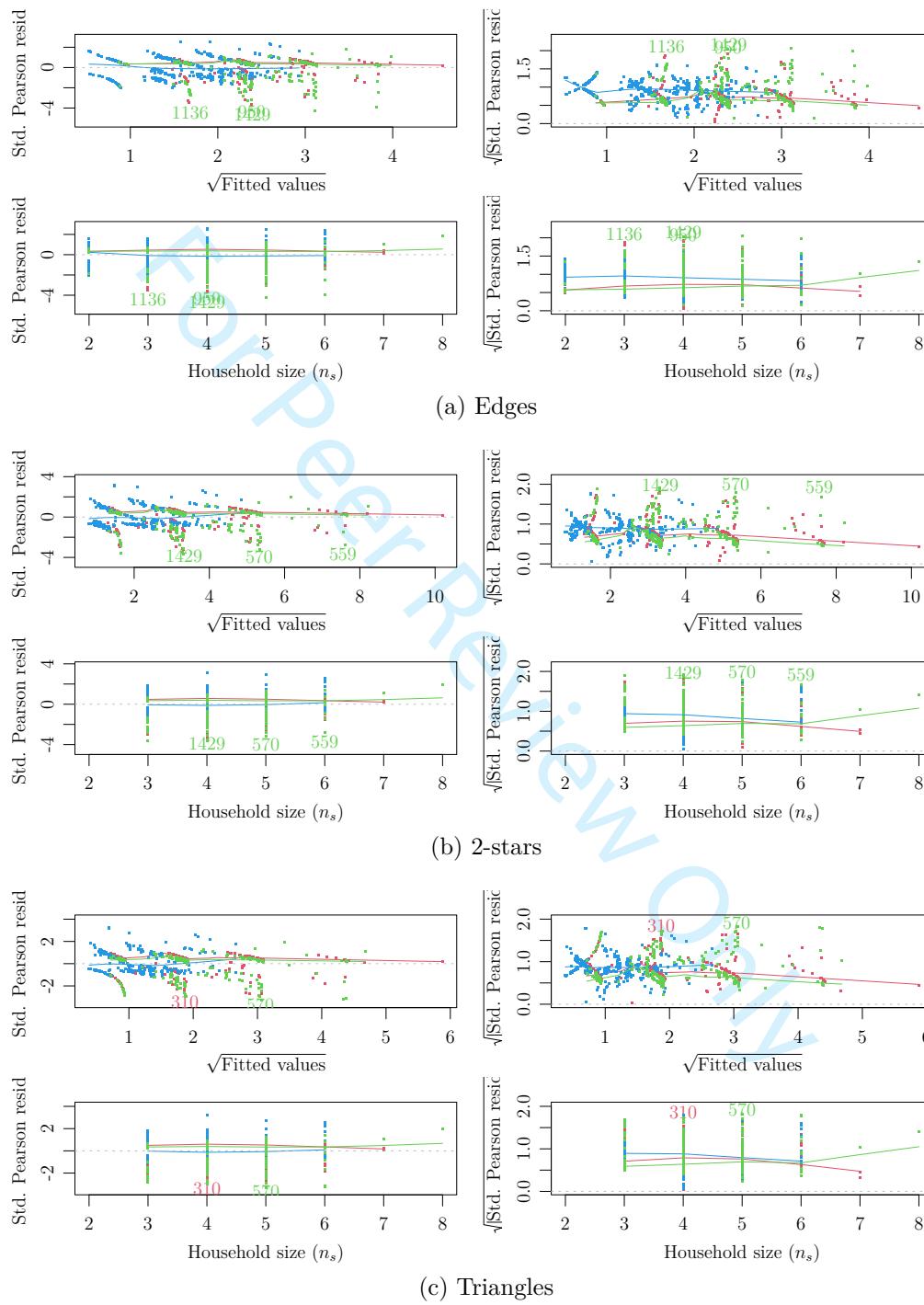


Figure E7: Residual plots of network statistics against fitted values and network size for *Model 1n*. (● all H ● E w/ child ● E w/o child)

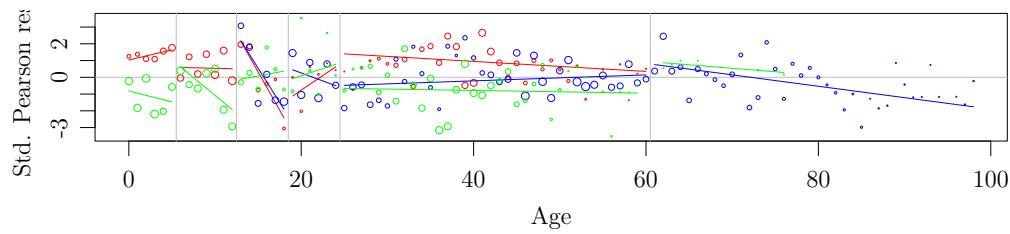


Figure E8: Residual plot for *Model 1n* of edges incident on actors of a given age against age. (● all H ● E w/ child ● E w/o child)

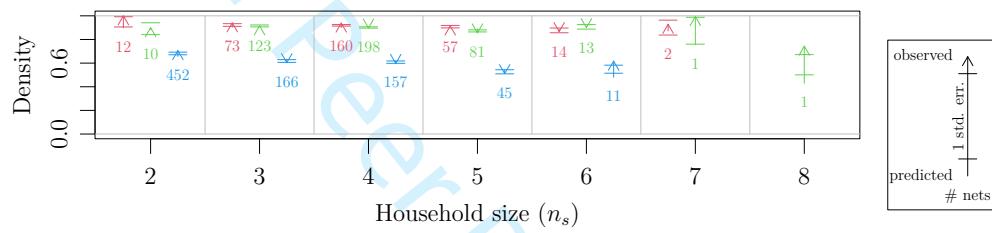


Figure E9: Average prediction errors of density in *Model 1n*. Values are averaged over the networks grouped by size and subdataset. (● all H ● E w/ child ● E w/o child)

Table E12: Regression of residuals of *Model 1n* for each network structural property on each candidate predictor.

Effect ($\times z_s$)	Wald χ^2 (df)	P-val.
edges		
$\times \log(\text{pop. dens. in post code})$ (linear)	4.5 (1)	0.03
$\times \log(\text{pop. dens. in post code})$ (quadratic)	5.2 (2)	0.08
if city post code	3.5 (1)	0.06
if Brussels post code	0.6 (1)	0.43
if weekend	1.5 (1)	0.23
2-stars		
$\times \log(\text{pop. dens. in post code})$ (linear)	5.7 (1)	0.02
$\times \log(\text{pop. dens. in post code})$ (quadratic)	6.1 (2)	0.05
if city post code	4.6 (1)	0.03
if Brussels post code	0.7 (1)	0.41
if weekend	1.1 (1)	0.28
triangles		
$\times \log(\text{pop. dens. in post code})$ (linear)	5.8 (1)	0.02
$\times \log(\text{pop. dens. in post code})$ (quadratic)	6.1 (2)	0.05
if city post code	4.2 (1)	0.04
if Brussels post code	0.9 (1)	0.34
if weekend	0.6 (1)	0.45

Table E13: Score tests of the null hypothesis of no effect of dataset H on the specified network statistic over and above *Model 1n*.

Statistic	Score χ^2 (df)	P-val.
Omnibus	10.2 (2)	0.0061
edges	nonparam.	0.0199
2-stars	nonparam.	0.1503

Table E14: Sample standard deviations of Pearson residuals for edge, two-star, and triangle counts in *Model 1n* for the dataset and its subsets. Values substantially greater than 1 indicate overdispersion relative to the model.

Effect ($\times z_s$)	Overall	H	E (w/ child)	E (w/o child)
edges	0.95	0.78	1.02	0.97
2-stars	0.94	0.79	1.00	0.95
triangles	0.91	0.81	0.97	0.91

E.4 Full results and diagnostics for *Model 1*

Table E15: Parameter estimates (and standard errors) for *Model 1*.

Effect ($\times z_s$)	Coefficient (SE)
edges $\times \log(n_s)$	-12.03 (3.80)**
$\times \log^2(n_s)$	4.57 (1.72)**
if Brussels post code	0.12 (0.20)
if on weekend	0.15 (0.06)*
2-stars	0.74 (1.19)
$\times \log(n_s)$	-0.77 (0.76)
$\times \log^2(n_s)$	0.04 (0.21)
triangles	7.56 (0.60)***
$\times \log(n_s)$	-5.74 (1.13)***
$\times \log^2(n_s)$	1.42 (0.85)
Young Child with Young Child	7.01 (1.90)***
Young Child with Preadolescent	7.84 (1.91)***
Preadolescent with Preadolescent	6.89 (1.88)***
Adolescent with Adolescent	6.15 (1.86)***
Young Child with Young Adult	7.79 (2.13)***
Young Adult with Young Adult	6.82 (1.82)***
Young Child with Female Older Adult	9.02 (1.86)***
Preadolescent with Female Older Adult	8.45 (1.86)***
Adolescent with Female Older Adult	7.89 (1.87)***
Female Older Adult with Female Older Adult	6.12 (1.88)**
Young Child with Male Older Adult	8.07 (1.87)***
Preadolescent with Male Older Adult	7.54 (1.86)***
Adolescent with Male Older Adult	7.12 (1.85)***
Female Older Adult with Male Older Adult	8.29 (1.87)***
if child absent	-0.80 (0.25)**
Male Older Adult with Male Older Adult	5.62 (1.88)**
Senior with Senior	6.55 (1.82)***
Adolescent with Young Child or Preadolescent	6.83 (1.87)***
Young Adult with Preadolescent or Adolescent	6.29 (1.85)***
Young Adult with Older Adult	6.79 (1.87)***
Senior with non-Senior	6.97 (1.86)***

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

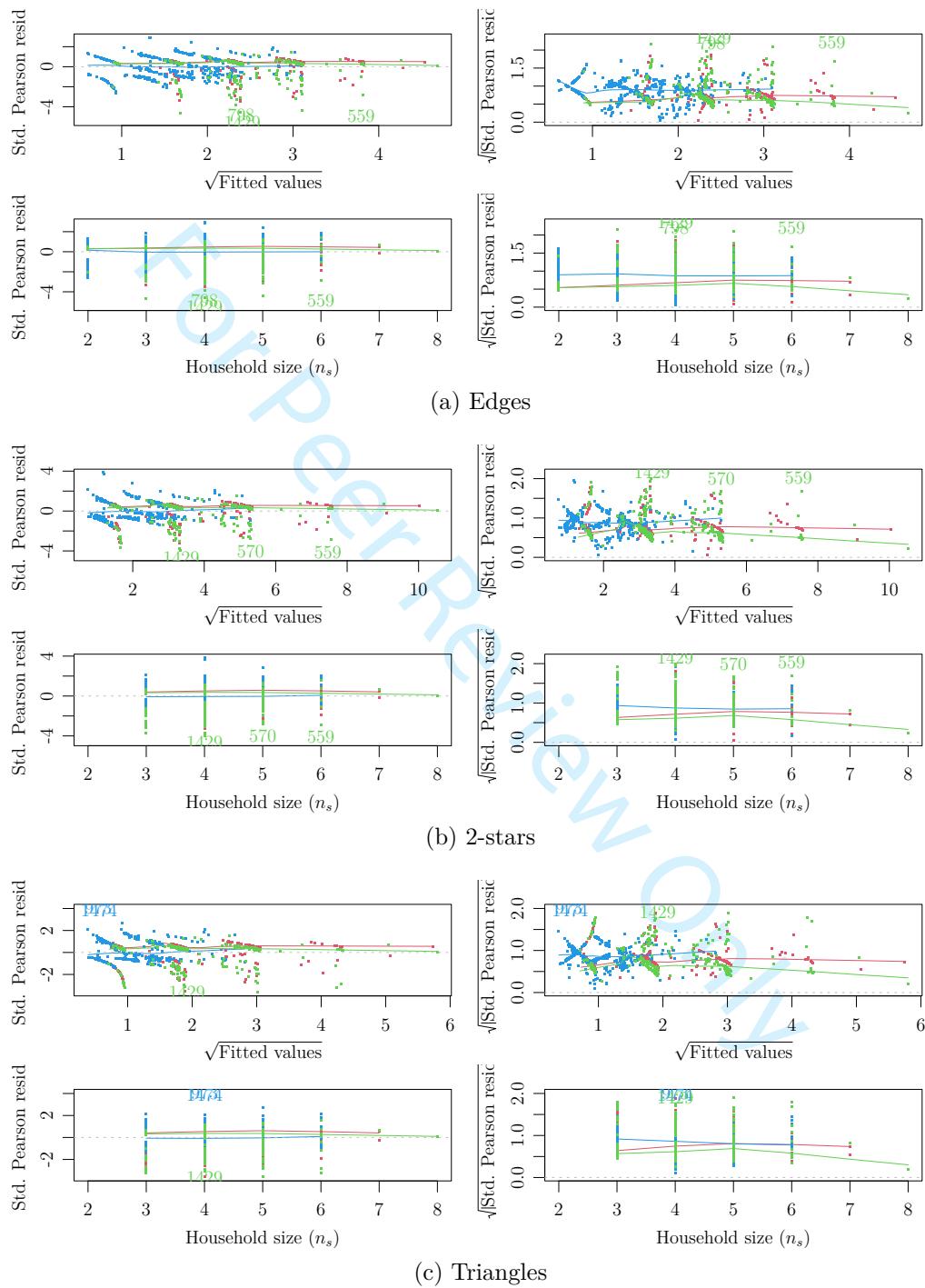


Figure E10: Residual plots of network statistics against fitted values and network size for *Model 1*. (● all H ● E w/ child ● E w/o child)

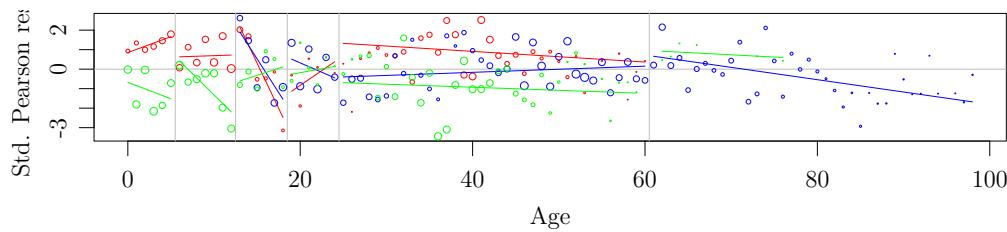


Figure E11: Residual plot for *Model 1* of edges incident on actors of a given age against age. (● all H ● E w/ child ● E w/o child)

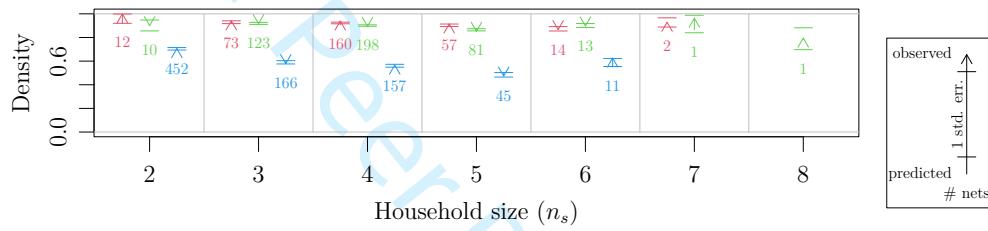


Figure E12: Average prediction errors of density in *Model 1*. Values are averaged over the networks grouped by size and subdataset. (● all H ● E w/ child ● E w/o child)

Table E16: Regression of residuals of *Model 1* for each network structural property on each candidate predictor.

Effect ($\times z_s$)	Wald χ^2 (df)	P-val.
edges		
$\times \log(\text{pop. dens. in post code})$ (linear)	3.2 (1)	0.08
$\times \log(\text{pop. dens. in post code})$ (quadratic)	3.6 (2)	0.16
if city post code	2.7 (1)	0.10
if Brussels post code	0.8 (1)	0.37
if weekend	0.8 (1)	0.39
2-stars		
$\times \log(\text{pop. dens. in post code})$ (linear)	5.0 (1)	0.02
$\times \log(\text{pop. dens. in post code})$ (quadratic)	5.5 (2)	0.06
if city post code	4.2 (1)	0.04
if Brussels post code	0.4 (1)	0.51
if weekend	0.8 (1)	0.37
triangles		
$\times \log(\text{pop. dens. in post code})$ (linear)	5.4 (1)	0.02
$\times \log(\text{pop. dens. in post code})$ (quadratic)	5.8 (2)	0.05
if city post code	4.1 (1)	0.04
if Brussels post code	0.6 (1)	0.44
if weekend	0.4 (1)	0.51

Table E17: Score tests of the null hypothesis of no effect of dataset H on the specified network statistic over and above *Model 1*.

Statistic	Score χ^2 (df)	P-val.
Omnibus	6.7 (2)	0.0357
edges	nonparam.	0.0149
2-stars	nonparam.	0.0709

Table E18: Sample standard deviations of Pearson residuals for edge, two-star, and triangle counts in *Model 1* for the dataset and its subsets. Values substantially greater than 1 indicate overdispersion relative to the model.

Effect ($\times z_s$)	Overall	H	E (w/ child)	E (w/o child)
edges	0.96	0.74	1.05	0.98
2-stars	0.95	0.79	1.03	0.96
triangles	0.93	0.83	0.98	0.94

E.5 Full results and diagnostics for *Model 2*

Table E19: Parameter estimates (and standard errors) for *Model 2*.

Effect ($\times z_s$)	Coefficient (SE)
edges $\times \log(n_s)$	-11.97 (0.88)***
$\times \log^2(n_s)$	4.51 (0.44)***
$\times \log(\text{pop. dens. in post code})$	0.18 (0.07)*
if Brussels post code	0.01 (0.22)
if on weekend	0.15 (0.06)*
2-stars	4.89 (0.72)***
$\times \log(n_s)$	-6.77 (0.59)***
$\times \log^2(n_s)$	1.92 (0.22)***
$\times \log(\text{pop. dens. in post code})$	-0.08 (0.07)
triangles	-4.56 (1.03)***
$\times \log(n_s)$	10.65 (0.75)***
$\times \log^2(n_s)$	-3.68 (0.44)***
$\times \log(\text{pop. dens. in post code})$	0.12 (0.17)
Young Child with Young Child	8.39 (0.37)***
Young Child with Preadolescent	9.21 (0.40)***
Preadolescent with Preadolescent	8.25 (0.26)***
Adolescent with Adolescent	7.49 (0.31)***
Young Child with Young Adult	9.44 (1.12)***
Young Adult with Young Adult	8.25 (0.34)***
Young Child with Female Older Adult	10.39 (0.37)***
Preadolescent with Female Older Adult	9.83 (0.25)***
Adolescent with Female Older Adult	9.29 (0.24)***
Female Older Adult with Female Older Adult	7.50 (0.31)***
Young Child with Male Older Adult	9.50 (0.29)***
Preadolescent with Male Older Adult	8.93 (0.22)***
Adolescent with Male Older Adult	8.54 (0.23)***
Female Older Adult with Male Older Adult	9.70 (0.26)***
if child absent	-0.84 (0.25)***
Male Older Adult with Male Older Adult	7.04 (0.33)***
Senior with Senior	7.88 (0.22)***
Adolescent with Young Child or Preadolescent	8.19 (0.23)***
Young Adult with Preadolescent or Adolescent	7.69 (0.25)***
Young Adult with Older Adult	8.15 (0.17)***
Senior with non-Senior	8.34 (0.19)***

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

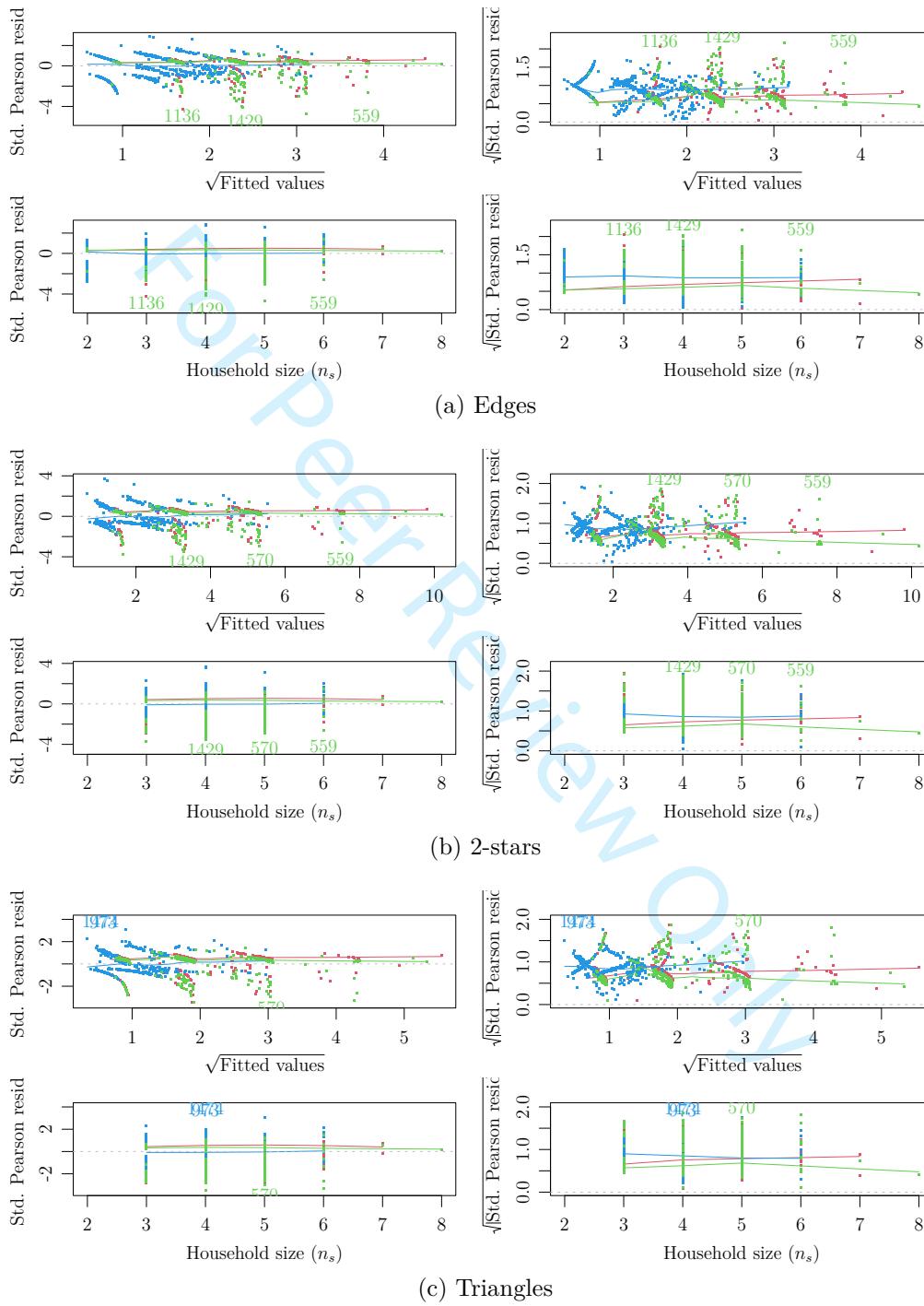


Figure E13: Residual plots of network statistics against fitted values and network size for *Model 2*. (● all H ● E w/ child ● E w/o child)

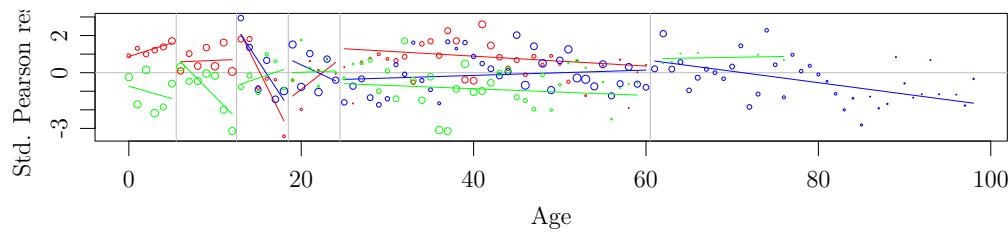


Figure E14: Residual plot for *Model 2* of edges incident on actors of a given age against age. (● all H ● E w/ child ● E w/o child)

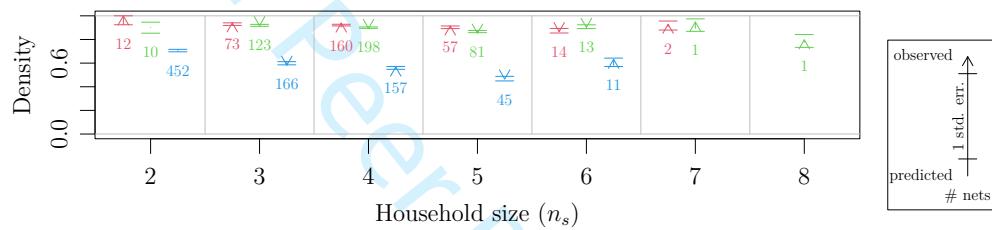


Figure E15: Average prediction errors of density in *Model 2*. Values are averaged over the networks grouped by size and subdataset. (● all H
● E w/ child ● E w/o child)

Table E20: Regression of residuals of *Model 2* for each network structural property on each candidate predictor.

Effect ($\times z_s$)	Wald χ^2 (df)	P-val.
edges		
$\times \log(\text{pop. dens. in post code})$ (linear)	0.0 (1)	0.96
$\times \log(\text{pop. dens. in post code})$ (quadratic)	0.0 (2)	1.00
if city post code	0.6 (1)	0.43
if Brussels post code	0.3 (1)	0.58
if weekend	0.9 (1)	0.36
2-stars		
$\times \log(\text{pop. dens. in post code})$ (linear)	0.8 (1)	0.36
$\times \log(\text{pop. dens. in post code})$ (quadratic)	0.9 (2)	0.65
if city post code	2.1 (1)	0.15
if Brussels post code	0.1 (1)	0.79
if weekend	0.7 (1)	0.41
triangles		
$\times \log(\text{pop. dens. in post code})$ (linear)	0.9 (1)	0.33
$\times \log(\text{pop. dens. in post code})$ (quadratic)	0.9 (2)	0.63
if city post code	2.0 (1)	0.15
if Brussels post code	0.2 (1)	0.68
if weekend	0.3 (1)	0.62

Table E21: Score tests of the null hypothesis of no effect of dataset H on the specified network statistic over and above *Model 2*.

Statistic	Score χ^2 (df)	P-val.
Omnibus	6.0 (2)	0.0489
edges	nonparam.	0.0232
2-stars	nonparam.	0.0876

Table E22: Sample standard deviations of Pearson residuals for edge, two-star, and triangle counts in *Model 2* for the dataset and its subsets. Values substantially greater than 1 indicate overdispersion relative to the model.

Effect ($\times z_s$)	Overall	H	E (w/ child)	E (w/o child)
edges	0.96	0.76	1.02	0.99
2-stars	0.94	0.80	1.00	0.95
triangles	0.92	0.82	0.96	0.93

E.6 Full results and diagnostics for *Model 3*

Table E23: Parameter estimates (and standard errors) for *Model 3*.

Effect ($\times z_s$)	Coefficient (SE)
edges $\times \log(n_s)$	-7.53 (2.90)**
$\times \log^2(n_s)$	2.59 (1.29)*
$\times \log(\text{pop. dens. in post code})$	0.05 (0.03)
if Brussels post code	0.02 (0.22)
if on weekend	0.15 (0.06)*
2-stars	0.68 (0.55)
$\times \log(n_s)$	-1.59 (0.63)*
$\times \log^2(n_s)$	0.58 (0.33)
triangles	4.67 (1.44)**
$\times \log(n_s)$	-0.71 (2.50)
$\times \log^2(n_s)$	-0.59 (1.07)
Young Child with Young Child	5.09 (1.49)***
Young Child with Preadolescent	5.90 (1.50)***
Preadolescent with Preadolescent	4.97 (1.47)***
Adolescent with Adolescent	4.25 (1.46)**
Young Child with Young Adult	5.75 (1.74)***
Young Adult with Young Adult	4.96 (1.46)***
Young Child with Female Older Adult	7.12 (1.48)***
Preadolescent with Female Older Adult	6.55 (1.46)***
Adolescent with Female Older Adult	5.98 (1.47)***
Female Older Adult with Female Older Adult	4.19 (1.48)**
Young Child with Male Older Adult	6.18 (1.46)***
Preadolescent with Male Older Adult	5.64 (1.45)***
Adolescent with Male Older Adult	5.23 (1.45)***
Female Older Adult with Male Older Adult	6.40 (1.45)***
if child absent	-0.79 (0.25)**
Male Older Adult with Male Older Adult	3.73 (1.49)*
Senior with Senior	4.69 (1.43)**
Adolescent with Young Child or Preadolescent	4.91 (1.46)***
Young Adult with Preadolescent or Adolescent	4.42 (1.44)**
Young Adult with Older Adult	4.86 (1.46)***
Senior with non-Senior	5.07 (1.45)***

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

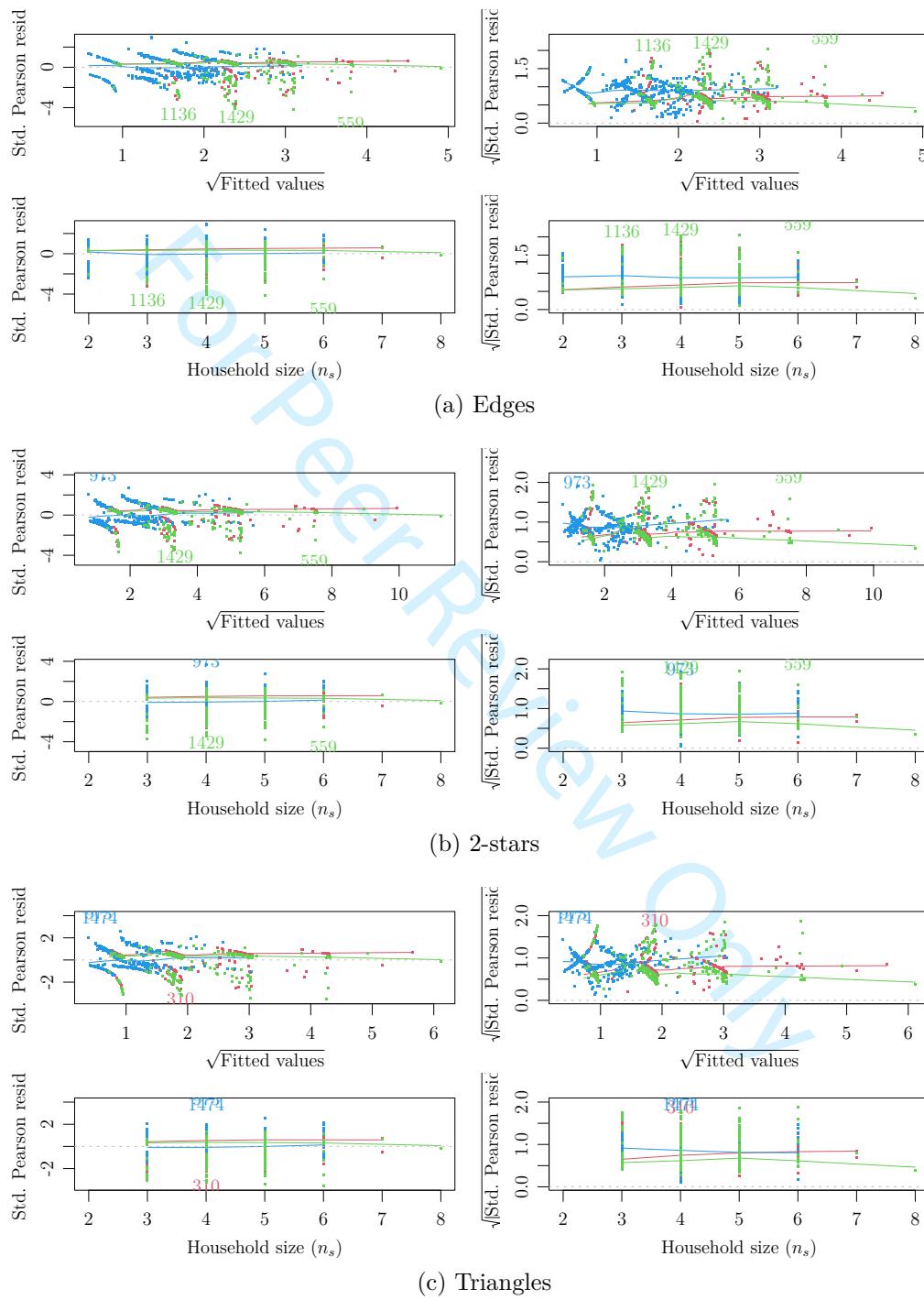


Figure E16: Residual plots of network statistics against fitted values and network size for *Model 3*. (● all H ● E w/ child ● E w/o child)

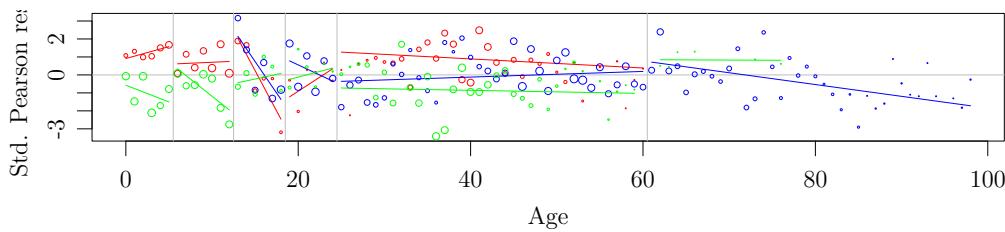


Figure E17: Residual plot for *Model 3* of edges incident on actors of a given age against age. (● all H ● E w/ child ● E w/o child)

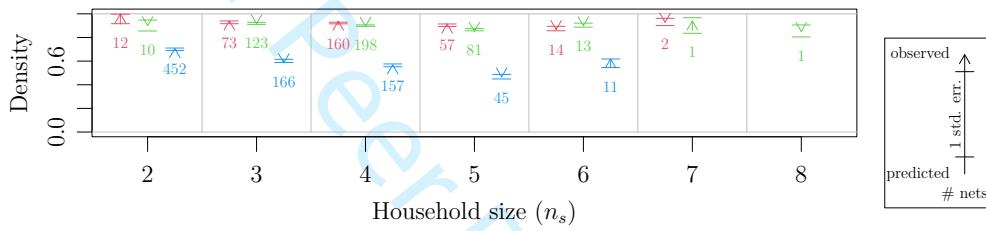


Figure E18: Average prediction errors of density in *Model 3*. Values are averaged over the networks grouped by size and subdataset. (● all H ● E w/ child ● E w/o child)

Table E24: Regression of residuals of *Model 3* for each network structural property on each candidate predictor.

Effect ($\times z_s$)	Wald χ^2 (df)	P-val.
edges		
$\times \log(\text{pop. dens. in post code})$ (linear)	2.0 (1)	0.15
$\times \log(\text{pop. dens. in post code})$ (quadratic)	2.3 (2)	0.32
if city post code	2.1 (1)	0.14
if Brussels post code	0.9 (1)	0.35
if weekend	0.7 (1)	0.39
2-stars		
$\times \log(\text{pop. dens. in post code})$ (linear)	3.8 (1)	0.05
$\times \log(\text{pop. dens. in post code})$ (quadratic)	4.0 (2)	0.14
if city post code	3.8 (1)	0.05
if Brussels post code	0.5 (1)	0.48
if weekend	0.9 (1)	0.34
triangles		
$\times \log(\text{pop. dens. in post code})$ (linear)	3.8 (1)	0.05
$\times \log(\text{pop. dens. in post code})$ (quadratic)	4.0 (2)	0.14
if city post code	3.5 (1)	0.06
if Brussels post code	0.6 (1)	0.43
if weekend	0.5 (1)	0.49

Table E25: Score tests of the null hypothesis of no effect of dataset H on the specified network statistic over and above *Model 3*.

Statistic	Score χ^2 (df)	P-val.
Omnibus	6.4 (2)	0.0400
edges	nonparam.	0.0188
2-stars	nonparam.	0.0927

Table E26: Sample standard deviations of Pearson residuals for edge, two-star, and triangle counts in *Model 3* for the dataset and its subsets. Values substantially greater than 1 indicate overdispersion relative to the model.

Effect ($\times z_s$)	Overall	H	E (w/ child)	E (w/o child)
edges	0.95	0.73	1.03	0.98
2-stars	0.94	0.78	1.00	0.96
triangles	0.92	0.82	0.97	0.94

For Peer Review Only

E.7 Full results and diagnostics for *Model 4*

Table E27: Parameter estimates (and standard errors) for *Model 4*.

Effect ($\times z_s$)	Coefficient (SE)
edges $\times \log(n_s)$	-13.42 (0.88)***
$\times \log^2(n_s)$	5.18 (0.45)***
$\times \log(\text{pop. dens. in post code})$	0.13 (0.07)
if Brussels post code	0.02 (0.53)
if on weekend	0.08 (0.13)
2-stars	3.23 (0.71)***
$\times \log(n_s)$	-4.05 (0.56)***
$\times \log^2(n_s)$	0.95 (0.22)***
$\times \log(\text{pop. dens. in post code})$	-0.06 (0.08)
if Brussels post code	0.05 (0.37)
if on weekend	-0.25 (0.18)
triangles	1.74 (1.13)
$\times \log(n_s)$	1.68 (0.82)*
$\times \log^2(n_s)$	-0.69 (0.46)
$\times \log(\text{pop. dens. in post code})$	0.09 (0.20)
if Brussels post code	-0.11 (0.83)
if on weekend	0.77 (0.48)
Young Child with Young Child	8.64 (0.36)***
Young Child with Preadolescent	9.46 (0.40)***
Preadolescent with Preadolescent	8.51 (0.25)***
Adolescent with Adolescent	7.74 (0.30)***
Young Child with Young Adult	9.47 (0.96)***
Young Adult with Young Adult	8.45 (0.35)***
Young Child with Female Older Adult	10.62 (0.36)***
Preadolescent with Female Older Adult	10.06 (0.24)***
Adolescent with Female Older Adult	9.53 (0.23)***
Female Older Adult with Female Older Adult	7.74 (0.31)***
Young Child with Male Older Adult	9.70 (0.29)***
Preadolescent with Male Older Adult	9.18 (0.22)***
Adolescent with Male Older Adult	8.77 (0.22)***
Female Older Adult with Male Older Adult	9.91 (0.26)***
if child absent	-0.79 (0.25)**
Male Older Adult with Male Older Adult	7.25 (0.33)***
Senior with Senior	8.16 (0.22)***
Adolescent with Young Child or Preadolescent	8.46 (0.22)***
Young Adult with Preadolescent or Adolescent	7.92 (0.25)***
Young Adult with Older Adult	8.41 (0.17)***
Senior with non-Senior	8.59 (0.18)***

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

For Peer Review Only

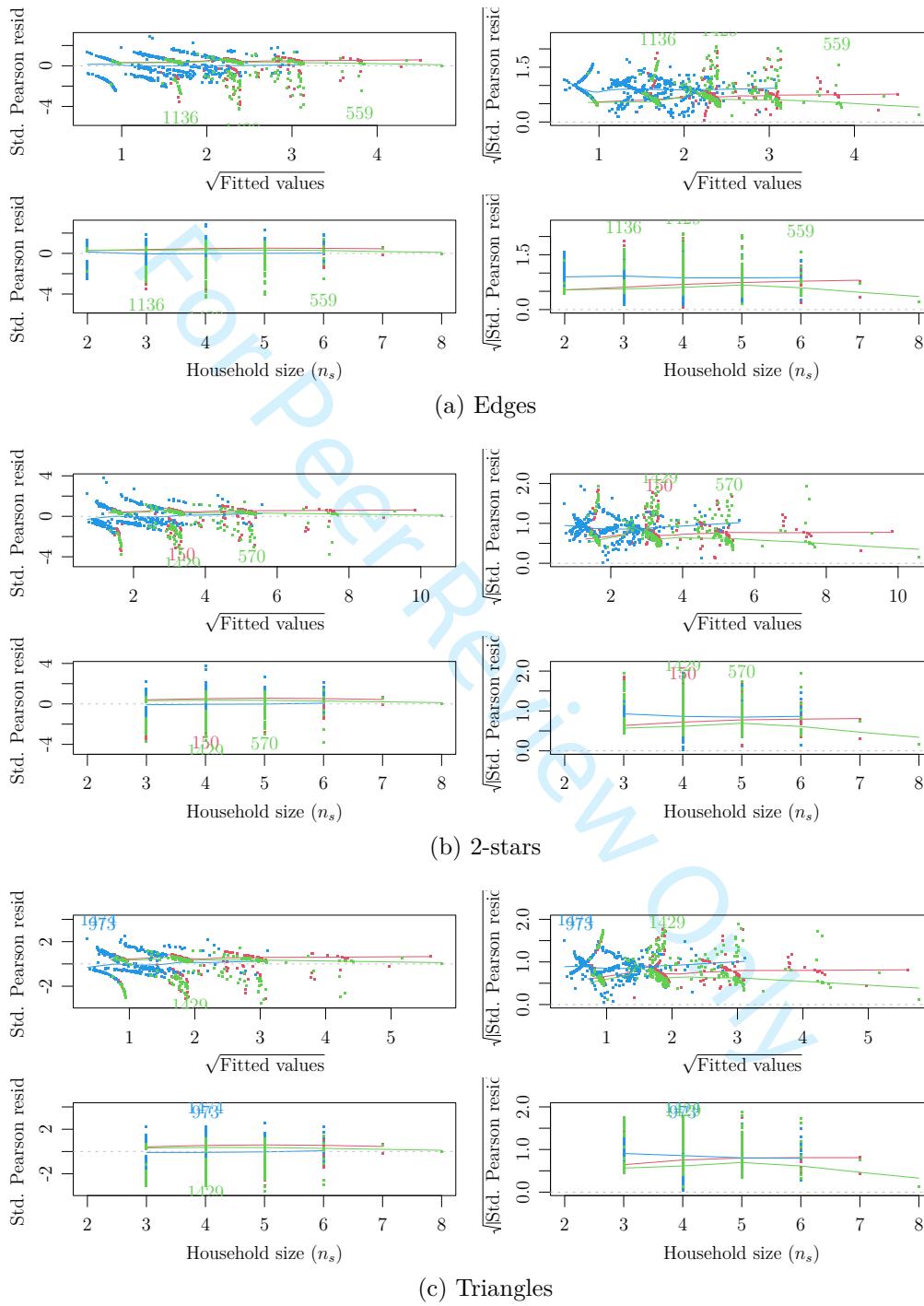


Figure E19: Residual plots of network statistics against fitted values and network size for *Model 4*. (● all H ● E w/ child ● E w/o child)

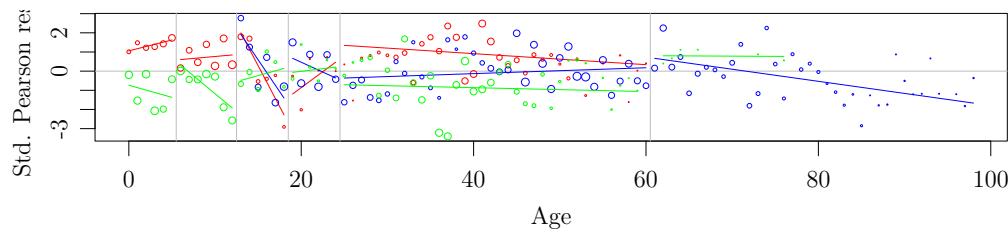


Figure E20: Residual plot for *Model 4* of edges incident on actors of a given age against age. (● all H ● E w/ child ● E w/o child)

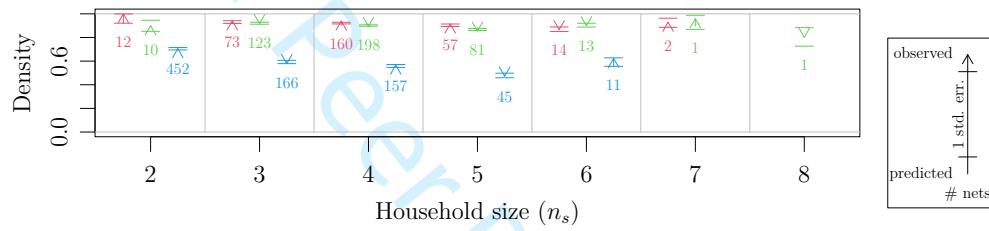


Figure E21: Average prediction errors of density in *Model 4*. Values are averaged over the networks grouped by size and subdataset. (● all H ● E w/ child ● E w/o child)

Table E28: Regression of residuals of *Model 4* for each network structural property on each candidate predictor.

Effect ($\times z_s$)	Wald χ^2 (df)	P-val.
edges		
$\times \log(\text{pop. dens. in post code})$ (linear)	0.4 (1)	0.55
$\times \log(\text{pop. dens. in post code})$ (quadratic)	0.4 (2)	0.83
if city post code	1.1 (1)	0.30
if Brussels post code	0.6 (1)	0.45
if weekend	1.3 (1)	0.26
2-stars		
$\times \log(\text{pop. dens. in post code})$ (linear)	1.7 (1)	0.19
$\times \log(\text{pop. dens. in post code})$ (quadratic)	1.7 (2)	0.42
if city post code	2.5 (1)	0.11
if Brussels post code	0.2 (1)	0.62
if weekend	1.0 (1)	0.33
triangles		
$\times \log(\text{pop. dens. in post code})$ (linear)	1.9 (1)	0.17
$\times \log(\text{pop. dens. in post code})$ (quadratic)	1.9 (2)	0.39
if city post code	2.5 (1)	0.11
if Brussels post code	0.4 (1)	0.53
if weekend	0.5 (1)	0.46

Table E29: Score tests of the null hypothesis of no effect of dataset H on the specified network statistic over and above *Model 4*.

Statistic	Score χ^2 (df)	P-val.
Omnibus	6.3 (2)	0.0424
edges	nonparam.	0.0135
2-stars	nonparam.	0.0522

Table E30: Sample standard deviations of Pearson residuals for edge, two-star, and triangle counts in *Model 4* for the dataset and its subsets. Values substantially greater than 1 indicate overdispersion relative to the model.

Effect ($\times z_s$)	Overall	H	E (w/ child)	E (w/o child)
edges	0.96	0.75	1.04	0.98
2-stars	0.94	0.80	1.00	0.95
triangles	0.92	0.83	0.96	0.93

References

- Butts, C. T. 2017. "Baseline Mixture Models for Social Networks," arXiv: 1710.02773v1.