

On the analysis of a sample of exponential random graph model estimates

J.W.G. Simons^a,

Supervisor: prof. dr. M.A.J. van Duijn^b

^a*Department of Methodology and Statistics, University of Utrecht, P.O. Box 80140, 3508 TC Utrecht, The Netherlands*

^b*Department of Sociology, University of Groningen, Grote Rozenstraat 31, 9712 TG Groningen, The Netherlands*

ARTICLE INFO

Keywords:

Exponential random graph model (ERGM)

Meta-analysis

Goodness-of-fit (GOF)

Information criteria

Social network analysis

ABSTRACT

Sed ut perspiciatis unde omnis iste natus error sit voluptatem accusantium doloremque laudantium, totam rem aperiam, eaque ipsa quae ab illo inventore veritatis et quasi architecto beatae vitae dicta sunt explicabo. Nemo enim ipsam voluptatem quia voluptas sit aspernatur aut odit aut fugit, sed quia consequuntur magni dolores eos qui ratione voluptatem sequi nesciunt. Neque porro quisquam est, qui dolorem ipsum quia dolor sit amet, consectetur, adipisci velit, sed quia non numquam eius modi tempora incidunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim ad minima veniam, quis nostrum exercitationem ullam corporis suscipit laboriosam, nisi ut aliquid ex ea commodi consequatur? Quis autem vel eum iure reprehenderit qui in ea voluptate velit esse quam nihil molestiae consequatur, vel illum qui dolorem eum fugiat quo voluptas nulla pariatur?

1. Introduction

A primary objective of social network research is to establish knowledge on social network structures. Such knowledge is typically induced by deriving hypotheses from theory and evaluating whether these hold with respect to a social network sample. Hypotheses can be represented as parameters in a statistical social network model (see, e.g., Cox, 2006). As such, social network models can be used to examine a set of hypotheses by representing an observed social network structure within a theoretically delimited hypothetical space (Cox, 2006). The degree to which a network model is able to represent a network structure can be defined as its goodness-of-fit (GOF) (Hunter, Goodreau, & Handcock, 2008). External validity is defined as the degree to which the results of a network model over a sample can be applied to the population from which it was drawn (see, e.g., Crano, Brewer, & Lac, 2014). External validity prescribes that the results of a network model will not generalize to the population if the network sample to which it was applied does not sufficiently reflect the characteristics of that population (Crano et al., 2014). Study results are defined as being externally valid if the network model is able to capture the structure of the network sample and if that sample captures the characteristics of the network population (Crano et al., 2014). Externally valid results consequently express the efficacy of a theory for explaining a social network structure in the population.

In the extant literature an overall model over a social network sample is typically obtained by way of a meta-analysis (Lubbers & Snijders, 2007). A meta-analysis averages the respective parameters of the estimated network model for each of the observed networks in the sample. (Lubbers & Snijders, 2007). In the recent literature, goodness-of-fit (GOF) is subsequently used to quantify the degree to which the network model fits on the level of each individual network and that of the network sample. Model modifications

are consequently made when the fit of the network model is found to be unsatisfactory. After any such modifications a network model over the sample is again obtained by averaging over the - available - parameter estimates in the network sample.

A major disadvantage of the meta-analytic approach is that the procedure assumes parameters to be comparable across sample networks. If this assumption is violated, the parameter estimates of the general network model are likely to be inaccurate, jeopardizing external validity. The conditions under which this same-parameter assumption is violated are however not well-understood. Even if goodness-of-fit (GOF) is used to identify and address model misfit, it generally remains unclear whether consequent model averaging results in an accurate model. At present, no empirical research has been conducted on whether a meta-analysis can be used to obtain valid population level ERGM estimates. It is additionally unknown whether ERGM GOF indices can be used to reliably quantify the fit of such a meta-analysis.

As such, the objective of this study is to identify the conditions under which a meta-analysis produces a valid exponential random graph model (ERGM) with respect to a sample of completely observed social networks. It additionally seeks to investigate the conditions under which a subset of the currently available goodness-of-fit (GOF) measures are able to identify sub-par GOF on the level of the network sample. The ERGM is furthermore chosen as the social network model of interest because it is arguably the most dominant model class for modelling cross-sectional network data. Note however that the relevance of the topic is not limited to the cross-sectional modelling framework. On the contrary, the outlined problem is also relevant to longitudinal network models such as stochastic actor-oriented models (SAOMs) (Snijders, 2001; Rambaran, Dijkstra, & Veenstra, 2020).

The following research question is formulated: Under which conditions can a meta-analysis be used to obtain a

valid exponential random graph model (ERGM) representation of a network sample? And additionally: Under which conditions is goodness-of-fit (GOF) a reliable indicator for determining when and whether a meta-analysis of ERGM estimates can be used to represent a network sample? The remainder of the paper consists of five sections. First, an overview is provided of the ERGM and the meta-analysis. Then, two methods for evaluating GOF are presented and discussed. The overview section concludes by delineating two factors which are hypothesized to influence the validity of the meta-analysis. Second, the analytical strategy and the data that will be used to evaluate the effects of these two factors are discussed. In short, an empirically informed simulation study will be utilized. The results, conclusion, and discussion sections respectively present, interpret, and discuss the outcomes of the analysis.

2. Theory

2.1. The exponential random graph model

2.1.1. The exponential random graph model

Exponential random graph models (ERGMs) are statistical models which can be used to explicate the formation of cross-sectional social network structures (Lusher, Koskinen, & Robins, 2013). Unless indicated otherwise, the information presented in this section has been sourced from the book by Lusher et al. (2013). Assume a fixed and predetermined node set n , represented as $N = \{1, \dots, n\}$ with $i \in N$. Let J represent the set of possible vertices excluding self-loops for the node set n : $N, J = \{(i, j) : i, j \in N, i \neq j\}$. For any observed network x_{obs} , the set of vertices E realized in x_{obs} is a random subset of J . Thus, for any element $(i, j) \in J$, a random variable X_{ij} can be defined such that $X_{ij} = 1$ if $(i, j) \in E$ and $X_{ij} = 0$ if $(i, j) \notin E$. Denote this random variable X_{ij} as an edge-variable which can be represented by a stochastic adjacency matrix $X = [X_{ij}]$. The elements of the stochastic matrix X indicate whether pairs of vertices are adjacent or not. Finally denote the space of all possible stochastic adjacency matrices by \mathbf{X} , with a realization of X within \mathbf{X} being denoted by $x = [x_{ij}]$. Note that this definition of a network edge-variable applies to undirected networks, but will similarly apply to directed networks given minor notational adjustments.

Assume that an edge between any two nodes X_{ij} can be explained either by attributes of nodes or edge patterns between nodes. Refer to the first type of predictor as an exogenous covariate and the second type as an endogenous covariate. The effect of an exogenous covariate on X_{ij} can then be understood analogous to a logistic regression model. Start by defining a set of p exogenous covariates and respectively denote these as $w_{ij,1}, w_{ij,2}, \dots, w_{ij,p}$. Interpret the elements of p as a function of an attribute for two respective actors i and j . Since X_{ij} is a dichotomous response variable, a logistic regression model can be used to obtain estimates for the unknown parameter set $\theta_1, \theta_2, \dots, \theta_p$ such that these predict X_{ij} as a function of p . Writing the logistic regression function in terms of the logit:

$$\text{logit } \Pr(X_{ij} = 1 | \theta) = \log \frac{\Pr(X_{ij} = 1 | \theta)}{\Pr(X_{ij} = 0 | \theta)} = \theta_1 w_{ij,1} + \theta_2 w_{ij,2} + \dots + \theta_p w_{ij,p} \quad (1)$$

so that the probability of an edge X_{ij} is given by the set of predictors p weighted by a respective set of parameters θ . Positive parameter values in (1) indicate an increased probability of an edge, where negative parameter values indicate a decrease in that probability. The difference in the log-odds for two pairs of nodes (i, j) and (h, m) can consequently be defined as:

$$\frac{\text{logit } \Pr(X_{ij} = 1 | \theta)}{\text{logit } \Pr(X_{hm} = 1 | \theta)} = \theta_1 (w_{ij,1} - w_{hm,1}) + \theta_2 (w_{ij,2} - w_{hm,2}) + \dots + \theta_p (w_{ij,p} - w_{hm,p}). \quad (2)$$

Assume for illustration purposes that the pairs (i, j) and (h, m) differ only on the node attribute of the exogenous covariate $w_{ij,1}$. In that case equation (2) reduces to θ_1 or the odds-ratio. The larger the value of θ_1 , the greater the probability of the presence of an edge between the node pairs which are similar as opposed to non-similar on $w_{ij,1}$, ceteris paribus.

Define endogenous or structural covariates as the second type of predictor of the edge between any two nodes X_{ij} . Endogenous covariates represent counts of network configurations. A network configuration can be understood as a local sub-graph within the larger graph in which it is embedded. Examples of network configurations are counts of the triad census or the number of reciprocated vertices within a larger social network structure. The analogy with logistic regression breaks down when endogenous covariates are introduced because these variables induce dependencies amongst edge-variables. The issue of dependence is not addressed here, but the observation is made that ignoring dependence among observations negatively affects the accuracy of conventional statistical models. As such, in order to make inferences about network structures a model needs to be formulated which is able to incorporate dependencies among edge-variables. Exponential random graph models (ERGMs) incorporate possible dependencies amongst edge-variables by predicting the probability of an edge conditional on what is observed in the rest of the network. More formally, the ERGM models each edge-variable X_{ij} conditional on the other vertices in the network X_{-ij} . Write this conditional probability as $\Pr(X_{ij} = 1 | X_{-ij} = x_{-ij}, \theta)$. Concentrating on the exogenous predictors, the conditional logit becomes :

$$\log \frac{\Pr(X_{ij} = 1 | X_{-ij} = x_{-ij}, \theta)}{\Pr(X_{ij} = 0 | X_{-ij} = x_{-ij}, \theta)} = \theta_1 \delta_{ij,1}^+(x) + \theta_2 \delta_{ij,2}^+(x) + \dots + \theta_p \delta_{ij,p}^+(x) \quad (3)$$

The functions $\delta_{ij,k}^+(x)$ in (3) are defined as change statistics for the k th configuration. These change statistics represent the change in transitioning from a graph for which $X_{-ij} = x_{-ij}$ and $X_{ij} = 0$ to a graph for which $X_{-ij} = x_{-ij}$ and $X_{ij} = 1$. Adding any edge X_{ij} to a network graph results

in a possible change in any of the network configurations p . This change is captured by the relevant change statistic, which when weighted by a respective parameter results in a change in the probability for $X_{ij} = 1$. As was the case for the exogenous covariates, whether an increase or decrease in the configuration count results in a respective increase or decrease in the probability for $X_{ij} = 1$ depends on the sign of the parameter estimate.

Having outlined the key components of the exponential random graph model (ERGM), write an equivalent form of the exponential random graph model (ERGM) such that it gives a probability expression for all edge-variables simultaneously. Define this as the joint form of the ERGM :

$$\Pr(X = x \mid \theta) \equiv P_\theta(x) = \frac{1}{\kappa(\theta)} \exp\{\theta_1 z_1(x) + \theta_2 z_2(x) + \dots + \theta_p z_p(x)\} \quad (4)$$

The functions $z_k(x)$ are counts of configurations in the graph x , such that the corresponding change statistic for $z_k(x)$ is $\delta_{ij,k}^+(x) = z_k(\Delta_{ij}^+ x) - z_k(\Delta_{ij}^- x)$, where $\Delta_{ij}^+ x(\Delta_{ij}^- x)$ denotes a matrix x for which x_{ij} is constrained to be equal to one (zero). As outlined earlier, the parameters weight the relative importance of their respective configurations. The normalizing constant $\kappa(\theta) = \sum_{y \in X} \exp\{\theta_1 z_1(y) + \theta_2 z_2(y) + \dots + \theta_p z_p(y)\}$ ensures that the sum of $P_\theta(x)$ over all graphs is equal to one.

Note that equation (4) describes a probability distribution for all graphs with n nodes. Formally, the exponential random graph model (ERGM) obtains the space of all possible stochastic adjacency matrices \mathbf{X} for a set of nodes n , one of which is the empirically observed graph x_{obs} . This distribution of graphs over \mathbf{X} implies a parallel distribution of network statistics. In general, the inferential goal of the ERGM is to center this distribution of statistics over those of the observed network. Put differently, its goal is to maximize the probability of the observed graph. Note that as a corollary of this inferential goal, ERGM statistics are always well-represented by the model. Additionally note however that this corollary does not hold for - often more complex - auxiliary statistics. Define the distribution as centered on the observed values when the expected value of the networks statistics $E_\theta(z(X))$ is equal to the network statistics of the observed network x_{obs} : $E_\theta(z(X)) = z(x_{obs})$. Equivalently, write this as the moment equation $E_\theta(z(X)) - z(x_{obs}) = 0$, where solving the moment equation for θ provides the maximum likelihood estimates (MLEs), the parameter value set that provide maximal support to the observed network x_{obs} .

In most cases the moment equation does not have an analytical solution. As such, Markov chain Monte Carlo (MCMC) methods are generally required to obtain a solution to the moment equation. The general principle behind MCMC is to start with a provisional parameter vector θ and to use a Metropolis sampler to obtain a sample of graphs $x^{(1)}, x^{(2)}, \dots, x^{(M)}$. Please refer to section 12.2.2 in Lusher et al. (2013) for a working definition of the Metropolis sampler. Given a sample of graphs of size m which approximates

the graph space \mathbf{X} of $P_\theta(x)$, calculate the sample equivalent $\bar{z}_\theta = \frac{1}{M}(z(x^{(1)}) + z(x^{(2)}) + \dots + z(x^{(M)}))$ of $E_\theta(z(X))$, and evaluate the moment-equation $\bar{z}_\theta - z(x_{obs}) = 0$. If $\bar{z}_\theta - z(x_{obs}) \neq 0$, choose another value θ and repeat the process until a θ is found for which $\bar{z}_\theta - z(x_{obs}) = 0$, the MLE.

A number of MCMC procedures have been formulated for executing the algorithm described in the previous paragraph. The procedure used here was originally proposed by Geyer and Thompson (1992). The Geyer and Thompson (1992) method solves the moment equation by representing the graph space \mathbf{X} through a fixed sample of graphs for the provisional value of the parameter vector θ . Given a fixed sample from the model $P_\theta(x)$, the method applies Fisher scoring to find θ such that $\bar{z}_\theta - z(x_{obs}) = 0$. Since the sample of graphs is an approximation of \mathbf{X} , a weighted average of the networks statistics is used to determine \bar{z}_θ . The sample average $\bar{f}_\theta = w^{(1)} f(x^{(1)}) + w^{(2)} f(x^{(2)}) + \dots + w^{(M)} f(x^{(M)})$, of a function $f(x)$, with weights

$$w^m = \frac{e^{(\theta_1 - \tilde{\theta}_1)z_1(x^{(m)}) + \dots + (\theta_p - \tilde{\theta}_p)z_p(x^{(m)})}}{\sum_{k=1}^M e^{(\theta_1 - \tilde{\theta}_1)z_1(x^{(k)}) + \dots + (\theta_p - \tilde{\theta}_p)z_p(x^{(k)})}} \quad (5)$$

is an asymptotic approximation to the true expected value $E_\theta(f(X))$ when $\theta \approx \tilde{\theta}$. To solve the moment equation, a sequence of parameters $\theta^{(0)} = \tilde{\theta}, \theta^{(1)}, \theta^{(2)}, \dots, \theta^{(g)}$ is generated using Fisher scoring. The updating step of $\theta^{(g)}$ in the Fisher scoring algorithm has the form:

$$\theta^{(g)} = \theta^{(g-1)} - \frac{1}{D(\theta^{(g-1)})} \left\{ \sum_{m=1}^M w^m z(x^{(m)}) - z(x_{obs}) \right\} \quad (6)$$

and iterates until that $\theta^{(g-1)}$ for which $\sum_{m=1}^M w^m z(x^{(m)}) = E_\theta(z(X))$ such that $E_\theta(z(X)) - z(x_{obs}) = 0$, which is the MLE. The term $D(\theta)$ in the Fisher scoring algorithm is a scaling matrix for the difference between observed values and the expected values of the network statistics. To account for the fact that the network statistics differ in their sensitivity to changes in the parameter vector θ , $D(\theta)$ is set to the weighted sample co-variance: $\sum_m w^{(m)} z(x^{(m)}) z(x^{(m)})^T - [\sum_m w^{(m)} z(x^{(m)})] [\sum_m w^{(m)} z(x^{(m)})]^T$.

2.1.2. Meta-analysis of exponential random graph models

As noted earlier, a meta-analysis is typically used to obtain an exponential random graph model (ERGM) over a network sample. The basic idea is to use a weighted average to summarize over the results of multiple statistical models, here ERGMs (see, e.g., Veroniki et al., 2016). Assume a scenario where the goal is to obtain an ERGM over a network sample of size m , where an ERGM with parameters p is fitted to each network in the sample. A random-effects meta-analytic model subsequently expresses each $\hat{\theta}_{p_m}$ as the sum of three components: an underlying sample-specific true effect μ , a within-sample variation term ϵ_{p_m} which represents sampling error, and a between-sample variation term ζ_{p_m} which represents the variance of the distribution of the true

sample effects (see, e.g., Veroniki et al., 2016). Write the random-effects meta-level regression equation as:

$$\hat{\theta}_{p_m} = \mu_p + \epsilon_{p_m} + \zeta_{p_m}. \quad (7)$$

Now let μ_{θ_p} express the weighted average over $\hat{\theta}_{p_m}$ as a realization of one of the p ERGM parameters at the sample level. Under the random effects model the weight assigned to each sample network for μ_{θ_p} is

$$w_{p_m} = \frac{1}{v_{p_m} + \tau_p^2} \quad (8)$$

where w_{p_m} is the sum of the within-study variance v_{p_m} for sample m and the between-studies variance τ_p^2 (Veroniki et al., 2016). The weighted mean μ_{θ_p} is then computed for each p as

$$\mu_{\theta_p} = \frac{\sum_{i=1}^m w_{m_p} \theta_{p_m}}{\sum_{i=1}^m w_{m_p}} \quad (9)$$

where the default method for estimating τ_p^2 is restricted maximum likelihood (REML) (Veroniki et al., 2016). The reader is referred to Veroniki et al. (2016) for a discussion of the REML method. The variance of the combined effect is defined as the reciprocal of the sum of the weights:

$$v_{\theta_p} = \frac{1}{\sum_{i=1}^m w_{m_p}} \quad (10)$$

with associated standard error $\sigma_{\mu_{\theta_p}} = \sqrt{v_{\theta_p}}$ and 95 % confidence intervals $\mu_{\theta_p} \pm \sigma_{\mu_{\theta_p}}$ (Veroniki et al., 2016). The set $\{\mu_{\theta_p}, \sigma_{\mu_{\theta_p}}, \mu_{\theta_p} \pm \sigma_{\mu_{\theta_p}}\}$ can then be interpreted as representing a general ERGM over the network sample.

Note that the presented meta-analytic method assumes uni-variate probability distributions for each parameter μ_{θ_p} . In practice however, a realization for any μ_{θ_p} affects the probability distribution of each $\mu_{\theta_{-p}}$. As such, a multivariate meta-analytic methodology is likely more appropriate for obtaining valid estimates of the set μ_{θ_p} . For the sake of limiting the scope of the paper, such a method is here neither discussed nor applied.

2.2. Goodness-of-fit of the exponential random graph model

The goodness-of-fit (GOF) of an exponential random graph model (ERGM) can principally be understood as an indicator of extremity between the observed values in the (sample of) network(s) and the values expected under the ERGM (Hunter, Goodreau, & Handcock, 2008). For both the ERGM and the meta-analysis of multiple ERGMs, GOF is generally quantified by way of an empirical p-value. The GOF of an ERGM can additionally be defined in a relative sense by way of information criteria. These criteria enable comparison of the fit of differently parameterized ERGMs relative to a particular observed network.

2.2.1. Auxiliary statistics as a goodness-of-fit index

Goodness-of-fit (GOF) indices can be used to diagnose and improve the fit of an exponential random graph model (ERGM) to an observed network (Hunter, Goodreau, & Handcock, 2008). Multiple indices have been formulated for assessing the GOF of an ERGM with respect to a single observed social network structure. Examples of such indices are auxiliary statistics (Hunter, Goodreau, & Handcock, 2008), approximate Bayesian GOF (Lusher et al., 2013) and outlier analysis (Koskinen, Wang, Robins, & Pattison, 2018). Here, auxiliary statistics indices are used to quantify the GOF of an ERGM to an observed social network.

The auxiliary statistics goodness-of-fit (GOF) index was introduced by Hunter, Goodreau, and Handcock (2008). It quantifies the GOF of an exponential random graph model (ERGM) to an observed social network by simulating from the ERGM to investigate graph configurations which were not modeled explicitly, i.e., which were not included as statistics in the ERGM specification. Put differently, the index summarizes the discrepancy between what is observed in the social network graph and what is expected under the ERGM for a set of graph configurations or network statistics not included in the model.

The auxiliary statistics goodness-of-fit (GOF) index is calculated by simulating an empirical distribution of graphs from an exponential random graph (ERGM) for the auxiliary graph statistic(s) of interest and determining the p-value of the observed auxiliary statistic(s) in that distribution. Note that auxiliary statistics are subsets of general graph configurations. For example, the sixteen possible triadic configurations in a directed network are all classified under the triad census. In short, the triad census is a count of how many of the 16 possible types of triads are present in a directed graph. Additionally note that the choice of auxiliary graph statistic(s) determines which structural aspects of the networks are important for assessing the GOF of the ERGM. Hunter, Goodreau, and Handcock (2008) propose using the degree, the triad census, and the geodesic distance in most situations. If the fitted ERGM is sufficient for explaining a particular graph configuration in the observed social network, then the position of the associated observed auxiliary statistic $S_k(x_{obs})$ should not be located in the tails of the empirical distribution of simulated graphs. One can thus quantify GOF with regards to an observed auxiliary statistic by contextualizing its position in the distribution of empirically simulated graphs by way of a t-statistic: $t = [S_k(x_{obs}) - \bar{S}_k] / SD(S_k(x))$. Here, \bar{S}_k and $SD(S_k(x))$ are the mean and standard deviation of the distribution of the statistic over the simulated sample of graphs. An empirical p-value can consequently be obtained as the observed value of the statistic with respect to the distribution of simulated networks as: $p_{S_k(x)} = \Pr(S_k(x) \leq S_k(x_{obs}))$. If the resulting p-value is smaller than a particular α then the observed statistic is far from what is expected under the ERGM, where a value for α of 0.05 is typically taken as being extreme. Such a value would be indicative of a poor fit of the ERGM to the

observed social network structure for that particular graph configuration.

A Mahalanobis distance can secondly be used to obtain multivariate auxiliary statistic goodness-of-fit (GOF) indices for an exponential random graph model (ERGM). More specifically, the elements of each auxiliary statistic are now expressed as the set. Recall that the auxiliary statistics GOF index for an ERGM was determined by generating sampling distributions from that ERGM for an a-priori chosen set of auxiliary statistics. Then, for each network in the sample the t-value associated with the position of that sample network's auxiliary statistic in the general ERGM sampling distribution is calculated as: $t = [S_k(x_{obs}) - \bar{S}_k] / SD(S_k(x))$, with associated p-value: $p_{S_k(x)} = \Pr(S_k(x) \leq S_k(x_{obs}))$. Given these components, one can consequently calculate a Mahalanobis distance over the elements of each auxiliary statistic over the network sample:

$$D = \sqrt{(S_k(x_{obs}) - \bar{S}_k)' \Sigma^{-1} (S_k(x_{obs}) - \bar{S}_k)}. \quad (11)$$

These Mahalanobis distances can then be translated into p-values as

$$p_D = 1 - \frac{1}{2^{n/2} \Gamma(n/2)} D^{n/2-1} e^{-D^2/2}.$$

The Mahalanobis distances and their respective p-values can consequently be used to diagnose the fit of the ERGM on each auxiliary statistic.

2.2.2. Information criteria as a relative index of goodness-of-fit

In general, the objective of an information criterion is to select the statistical model which best approximates the underlying, unobserved process which generated the observed data (Wit, Heuvel, & Romeijn, 2012). Formally, this is achieved by minimizing the Kullback–Leibler divergence with respect to the observed data (Wit et al., 2012). In more straightforward terms, information criteria seek to maximize the fit of the model to the data, while simultaneously minimizing its complexity, where lower values indicate a model which is better able to strike this balance. (Wit et al., 2012). Note that information criteria are a relative quantity, and should be interpreted as expressing the relative efficacy of two or more differently parameterized models for striking a balance between complexity and fit.

The information criteria formulated by Akaike (1974) - Akaike's information criterion (AIC) - and (Schwarz, 1978) - Schwarz's information criterion (SIC) - will here be used to quantify relative GOF. The AIC and SIC can respectively be defined as

$$AIC(m) = -2\log P_{\hat{\theta}_m} + 2p, \quad (12)$$

$$SIC(m) = -2\log P_{\hat{\theta}_m} + \log S. \quad (13)$$

Here, m refers to a particular ERGM, $P_{\hat{\theta}_m}$ refers to the maximized likelihood under m , p refers to the number of parameters in the ERGM, and S refers to the size of the sample

(Wit et al., 2012). For a derivation of these two formulae with respect to the Kullback–Leibler divergence the reader is referred to Wit et al. (2012).

Note that there are two disadvantages associated with using information criteria over auxiliary statistics as goodness-of-fit (GOF) indices for exponential random graph models (ERGMs) (Wit et al., 2012). First, due to the fact the likelihood function cannot generally be evaluated analytically, the AIC and SIC are approximations of the true likelihood (Wit et al., 2012). The auxiliary statistics approach is additionally more informative because it provides a graphical overview of which graph configurations are fit well and which are not (Wit et al., 2012). Information criteria are nonetheless included because they enable comparison of the fit of differently parameterized ERGMs relative to a particular observed network. The auxiliary statistics GOF index does not have this feature, which justifies the inclusion of information criteria as a GOF index.

2.2.3. Goodness-of-fit of the exponential random graph model for a sample of networks

On the level of the network sample, goodness-of-fit (GOF) can be used to evaluate the degree to which the meta-analysis of the set of exponential random graph model (ERGM) estimates fits to the network sample. As such, it can be used as a measure of whether conclusions with regards to an overall ERGM can reasonably be extended to the network population.

The goodness-of-fit (GOF) of a meta-analysis for a set of exponential random graph model (ERGM) estimates can be defined in analogy to the GOF of the ERGM. In short, empirical distributions are simulated from the estimated meta-analysis ERGM parameters for a set of auxiliary statistics. As for the GOF of an ERGM, typically the in- and out-degree, the geodesic distance, and the triad-census are used (Hunter, Goodreau, & Handcock, 2008). The aggregate observed value for each auxiliary statistic over the network sample is subsequently positioned in the empirical distribution for that statistic. Uni- and multi-variate p-values can then be obtained per the method for obtaining GOF over a single ERGM. Finally note that relative GOF measures such as information criteria are redundant on the sample level.

2.3. On using a meta-analysis for summarizing exponential random graph model estimates

Having thus defined the form of the exponential random graph model (ERGM) and the goodness-of-fit (GOF) indices on the level of the individual network and the network sample, return to the research question: Under which conditions can a meta-analysis be used to obtain a valid exponential random graph model (ERGM) representation of a network sample? And additionally: Under which conditions is goodness-of-fit (GOF) a reliable indicator for determining when and whether a meta-analysis of ERGM estimates can be used to represent a network sample?

In order to formulate an approach to answering these two questions, assume a universe where the conditions un-

der which a meta-analysis of ERGM estimates is valid are known. Additionally assume that the goodness-of-fit (GOF) indices reliably detect optimal or sub-optimal fit of an ERGM on the level of the individual network and that of the sample. Assume that a researcher has obtained a network sample which sufficiently captures the characteristics of the population from which it was drawn. Given the sample, the researcher would proceed by fitting an exponential random graph model (ERGM) to each network in the sample to then obtain a meta-analysis of ERGM estimates over the ERGM set. The researcher would subsequently use GOF to diagnose the fit of the network-level ERGMs and the meta-analysis of ERGM estimates. If sub-par GOF is identified for the meta-analysis its parametrization or that of any of the ERGMs on the network-level can be altered to improve its fit. After an iterative process of any such modifications, the resulting meta-analysis would provide an optimal fit to the network sample. Since the sample was an adequate representation of the population, the researcher could then establish external validity. Put differently, the researcher would be able to evaluate theoretically derived hypotheses and determine whether the effect of a certain actor attribute or network feature is salient in the real world.

Since the conditions under which the parameter estimates of an exponential random graph model (ERGM) are decently accurate at the sample level are currently not well-understood, in practice it is generally not clear when an acceptable level of accuracy has been achieved. In order to shed tentative light on this issue, this study identifies two types of conditions which are hypothesized to affect the validity of the ERGM model parameters on the level of the network sample: the characteristics of the sample and the complexity of the ERGM.

2.3.1. Characteristics of the sample

The first factor which is hypothesized to affect the accuracy of a meta-analysis of exponential random graph model (ERGM) estimates are the characteristics of the sample. Two sample characteristics, its size and its heterogeneity, are considered. Sample size is a prominent condition in most if not all statistical undertakings, and here simply refers to the number of social networks in the sample. The heterogeneity of the sample refers to the degree to which the networks in the sample show different co-variate and structural properties. It can alternatively be defined as the number of outlying networks in the sample.

Traditional sampling theory prescribes that larger sample sizes result in increased model accuracy when estimating population parameters (Lohr, 2009). Since the empirical data concern wholly observed student classroom relationships (Vermeij, 2006), statistical independence can be assumed to hold between sample networks. On that basis parameter estimates for a meta-analysis of ERGM estimates are hypothesized to be more accurate for larger as opposed to smaller network samples.

As is the case for standard statistical models, the accuracy of a meta-analysis of exponential random graph model

(ERGM) estimates is reduced when it is applied to a sample which consists of many as opposed to a few or no outlying networks (Koskinen et al., 2018). This is due to the fact that the meta-analysis is unable to sufficiently capture all intricacies of all the networks in the sample (Koskinen et al., 2018). In more formal terms, ERGMs are not robust (Rehnberg, 2016). On that basis, it is hypothesized that the accuracy of the meta-analysis is reduced when it is applied to a sample with as opposed to without outlying networks, *ceteris paribus*.

A third highly salient characteristic of the sample are the sizes of the social networks in the sample itself. This sample characteristic can be seen as a form of heterogeneity, in that a higher variance in the respective network sizes could be hypothesized to reduce the accuracy of the meta-analysis of exponential random graph model (ERGM) estimates. Although network size is generally an important feature (Kolaczyk & Krivitsky, 2015), here this characteristic is kept constant in order to limit the scope of the simulation study.

2.3.2. Complexity of the exponential random graph model

The second factor relates to the complexity of the specification of the meta-analysis of the exponential random graph model (ERGM) estimates. This factor can be defined as the number and type of parameters that need to be included in the meta-analysis specification to provide an adequate representation of the network sample. The objective is to find that ERGM which provides a fit to the network sample which is as simple as possible while still being complex enough to capture the intricacies of each network in the sample. If the configuration of the meta-analysis is too simple or too complex, sample results will not correspond to the characteristics of the network population. Different aspects of the networks in the sample might furthermore also be more or less easy to model. Work by Van Duijn, Gile, and Handcock (2009) showed higher parameter accuracy for covariate as opposed to structural network effects. As such, covariate effects are more likely to be adequately modeled by a relatively simple parameterized ERGM, where a more complex ERGM is required to capture structural effects. Based on the previous line of reasoning, it is hypothesized that the accuracy of the meta-analysis decreases for an increased discrepancy between the complexity of the model and the complexity of the sample. The larger the discrepancy between the relative complexity of the model and the sample, the larger the inaccuracy of the meta-analysis ERGM estimates. It is furthermore hypothesized that ERGM accuracy is generally higher for covariate as opposed to structural network effects.

3. Analytical Strategy

A three-stage empirically informed simulation study will be used to provide an answer to the hypotheses and research questions. The first stage consists of generating empirically feasible social network samples to which the exponential random graph model (ERGM) and the respective goodness-

of-fit (GOF) indices can be applied. Put differently, in the first stage the population level process which generates observed social network structures needs to be emulated. It should be noted that simulating network variability is a delicate process where networks with nonsensical configurations are quickly induced for certain ERGM parameter combinations due to their high mutual dependence. As such, simulated network variability should be informed by empirical network variability. An empirical sample of 86 Dutch secondary class networks by Vermeij (2006) has been made available for the purposes of this study. Ethical approval has been obtained from the University Utrecht faculty ethical review board (FERB) for using these data. Two of the 86 networks - namely the 8th and the 16th - will be used to estimate two separate population level models. Each of the two classrooms consists of 25 pupils. The data on each pupil in each classroom are the relationship nominations with respect to the other pupils in the classroom and two covariate attributes. A variation on the five point summary as discussed in Luke (2015) is presented in table 1 for describing the relationship nomination variable across the two empirical classroom networks. As becomes apparent from density and component variables, the 8th network is slightly more dense than the 16th. The clustering coefficient indicates that students in the 16th network tend to cluster together more than in the 8th. Two students are finally isolated in the 16th network, where none are in the 8th. The reader is referred to section 2.1 in Vermeij (2006) for an exact description of how the relationship nomination variable was operationalized.

Table 1
Five-point summary of the 8th and 16th Vermeij classroom relationship nominations.

	Network 8	Network 16
Size	25	25
Density	0.17	0.15
Components	12	14
Diameter	4	5
Clustering coefficient	0.24	0.32
Isolates	0	2

The pupil attributes are the gender and classroom identification of each pupil, where classroom identification is defined as the sense of connection of the pupil to the classroom. On the gender attribute, a value of zero denotes an individual of the male biological sex, where a value of one denotes an individual of the female biological sex. In the 8th network, 14 (11) of the 25 students are male (female). In the 16th network, 13 (12) of the 25 students are male (female). Finally the classroom identification attribute is a z-score where negative (positive) values indicate less (more) identification with the classroom. In the 8th (16th) network the mean of the classroom identification variable is -0.70 (-0.61), with a standard deviation of 0.70 (0.84).

The “ergm” R-package will be used for estimating the parameters of the population-level exponential random graph models (ERGM) over the two empirical Vermij networks

(Hunter, Handcock, Butts, Goodreau, & Morris, 2008; Team, 2013). Each population-level ERGM is assigned a unique seed and has an MCMC burn-in of 168340, a proposal interval of 10240, and an MCMC sample size of 10240. Both ERGMs are parameterized on the basis of the following model terms: an edges term which models network density, a mutual term which models tie reciprocity, a geometrically-weighted edgewise shared partnerships (GWESP) term which models triad closure, a geometrically-weighted out-degree term (GWODEGREE) which models vertex tendencies for sending ties, a classroom identification sender term which models vertex tendencies for sending ties based on classroom identification, and a gender homophily term which models tie reciprocity based on gender similarity. This particular specification resulted from a trial-and-error process where the objective was to maximize the fit of the ERGM to the observed data with respect to the AIC and SIC while ensuring a uniform model parametrization for each population-level ERGM.

The next step in the first stage of the simulation study is to simulate network samples from the population-level ERGMs while varying the characteristics of the sample. To ensure reproducibility, each network in each sample is generated with a unique seed. With regards to the sample size condition, network samples of sizes 25 and 75 are generated respectively. The heterogeneity factor is secondly operationalised as the degree to which the networks are sampled from either a homogeneous or a mixed population-level ERGM. More specifically, in the first heterogeneity condition all networks in the sample are drawn from the population level ERGM estimated over the 16th classroom, where in the second condition 80% of the networks in the sample are drawn from the population ERGM estimated over the 16th population-level ERGM and 20% are drawn from the 8th population-level ERGM. The reader is referred to table 2 for an overview of the target statistics for each of the two population-level ERGMs. Target statistics are the mean-value ERGM parameter values on which the MCMC procedure tries to match the configuration of each sampled network.

Table 2
Target statistics of the population level ERGMs.

	Edges	Mutual	Gwesp	Gwodeg	Nodecov	Nodematch
ERGM 8	101	13	58.73	17.67	-115.25	34
ERGM 16	89	15	70.84	17.46	-101.44	36

The second stage of the simulation procedure consists of obtaining the exponential random graph model (ERGM) and the appropriate GOF measures on the level of the network and of the sample. Here the complexity factor is incorporated by varying the complexity of the exponential random graph model (ERGM) that is applied to set of simulated network samples. Three ERGMs are formulated: one with only the structural model terms - namely the edges, mutual, gwesp, and gwodegree parameters - one homogeneous model which is identical to the parametrization of the

Table 3

Mean (%) convergence failure for the structural, true and dense triadic models by sample size and heterogeneity.

		Structural model	True model	Dense triadic model
Sample size 25	Homogeneous sample	4.08 (16.32%)	2 (8.00%)	12.31 (49.26%)
	Heterogeneous sample	7.82 (31.29%)	4.10 (16.39%)	9.81 (39.23%)
Sample size 75	Heterogeneous sample	12.89 (17.18%)	7.52 (10.02%)	34.06 (45.42%)
	Homogeneous sample	22.18 (29.57%)	11.45 (15.27%)	29.45 (39.27%)

population-level ERGMs, and one where a dense triad parameter has been added to the population-level parametrization. Dense triads are defined as those triads in the triad census with either four or more ties (Block, Stadtfeld, & Snijders, 2019). To ensure acceptable convergence only the triad with exactly four ties - namely 120C - is added to the model. Furthermore, in order to establish a point of reference, the homogeneous model is obtained over each network in the sample - where a network is re-sampled if convergence is not achieved. Convergence of the structural and dense triad models is allowed to fail and tracked by way of a running counter. The degree to which the ERGM fails to converge in each of the respective cells is presented in the results section. The reader is referred to table 4 for an overview of the 12 simulation cells in the simulation study.

Table 4

Simulation matrix.

		Sample size	
Complexity	Heterogeneity	25	75
Structural	Homogeneous	Scenario 1	Scenario 7
	Mixed	Scenario 2	Scenario 8
True	Homogeneous	Scenario 3	Scenario 9
	Mixed	Scenario 4	Scenario 10
Dense triad	Homogeneous	Scenario 5	Scenario 11
	Mixed	Scenario 6	Scenario 12

The “*ergm*” R-package will subsequently be used to estimate exponential random graph (ERGM) parameters at the single network level (Hunter, Handcock, et al., 2008; Team, 2013). Each ERGM is assigned a unique seed and has an MCMC burn-in of 16834, a proposal interval of 1024, and an MCMC sample size of 1024. The “*metafor*” R-package will be used to obtain a uni-variate meta-analysis of ERGM estimates at the level of the network sample (Viechtbauer, 2010; Team, 2013). The step length of the Fisher scoring algorithm is set to 0.5 and the maximum number of iterations is set to 1000. The procedure for calculating auxiliary statistics goodness-of-fit (GOF) indices at the level of the individual network and the network sample will be hand-implemented in R. In line with (Hunter, Goodreau, & Handcock, 2008) the degree, the triad census, and the geodesic distance will be used to quantify auxiliary statistics GOF. The information criteria are provided by the “*ergm*” R-package at the level of the single network (Hunter, Handcock, et al., 2008; Team, 2013). The bias and RMSE for each fitted ERGM parameter relative to its respective population-level model is also cal-

culated. Note that the bias and RMSE are defined relative to the scale of the estimated parameter. To enable inter-cell comparison, the bias and RMSE are defined as the percentage deviation of the estimate relative to the true value. More specifically, the bias of each estimator is defined as the difference between the fitted and true value divided by the true value over the set of data replications

$$bias(\hat{\theta}) = \frac{1}{Z} \sum_z \left(\frac{\hat{\theta}_z - \theta}{\theta} \right), \quad (14)$$

where the RMSE is defined as the square root of the square of the difference divided by the true value over the set of data replications

$$RMSE(\hat{\theta}) = \sqrt{\frac{1}{Z} \sum_z \left(\frac{\hat{\theta}_z - \theta}{\theta} \right)^2}. \quad (15)$$

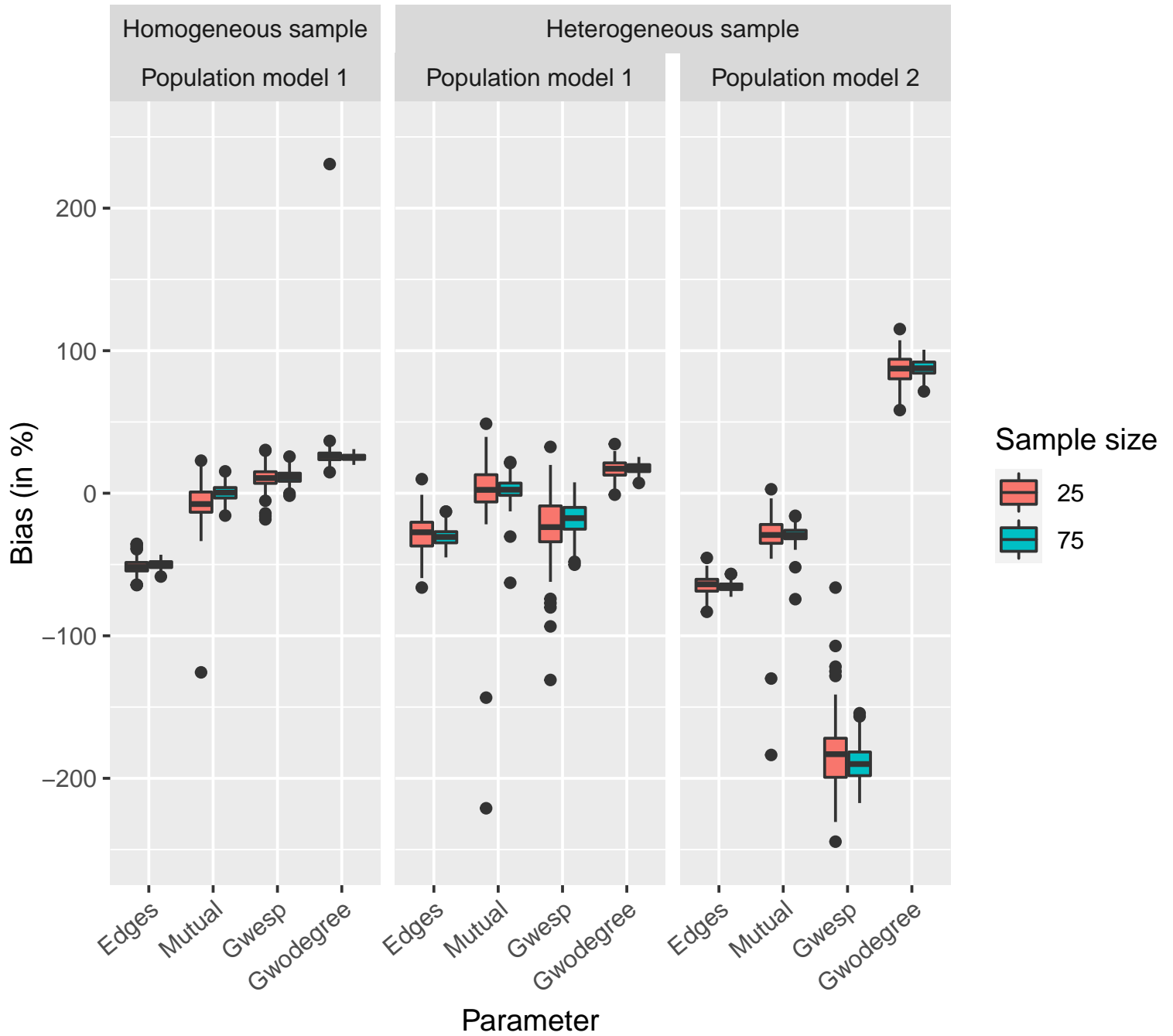
Finally, a meta-analysis of an ERGM and its respective diagnostics is ideally obtained a 1000 times for each cell in the simulation matrix. Due to constraints in time and computational power, a meta-analysis over each cell has been obtained 124 times.

The third and final stage of the simulation study uses box-plots to present the bias and over the set of meta-analyses. Box plots are additionally used to present the auxiliary statistics goodness-of-fit (GOF) for the meta-analysis over each cell for the first iteration. AIC and SIC difference scores for the structural and dense-triadic models relative to the true model are additionally inspected.

4. Results

Table 3 shows the mean tendency of the ERGM to fail to converge within each of the 12 simulation cells. Percentages are additionally presented in parentheses alongside the respective means to facilitate interpretation between the two sample size conditions. As becomes apparent from the table, the true model fails to converge the least, followed by the structural model and the dense triadic model. There are close to no percent-wise differences between the sample size conditions, where large percent-wise differences can be observed between the homogeneous and heterogeneous sample conditions. More specifically, for the structural and true model specifications, the MCMC procedure fails to converge about twice as often in the heterogeneous as in the homogeneous sample. For the dense triadic model no distinct percent-wise difference can be observed between the two heterogeneity conditions.

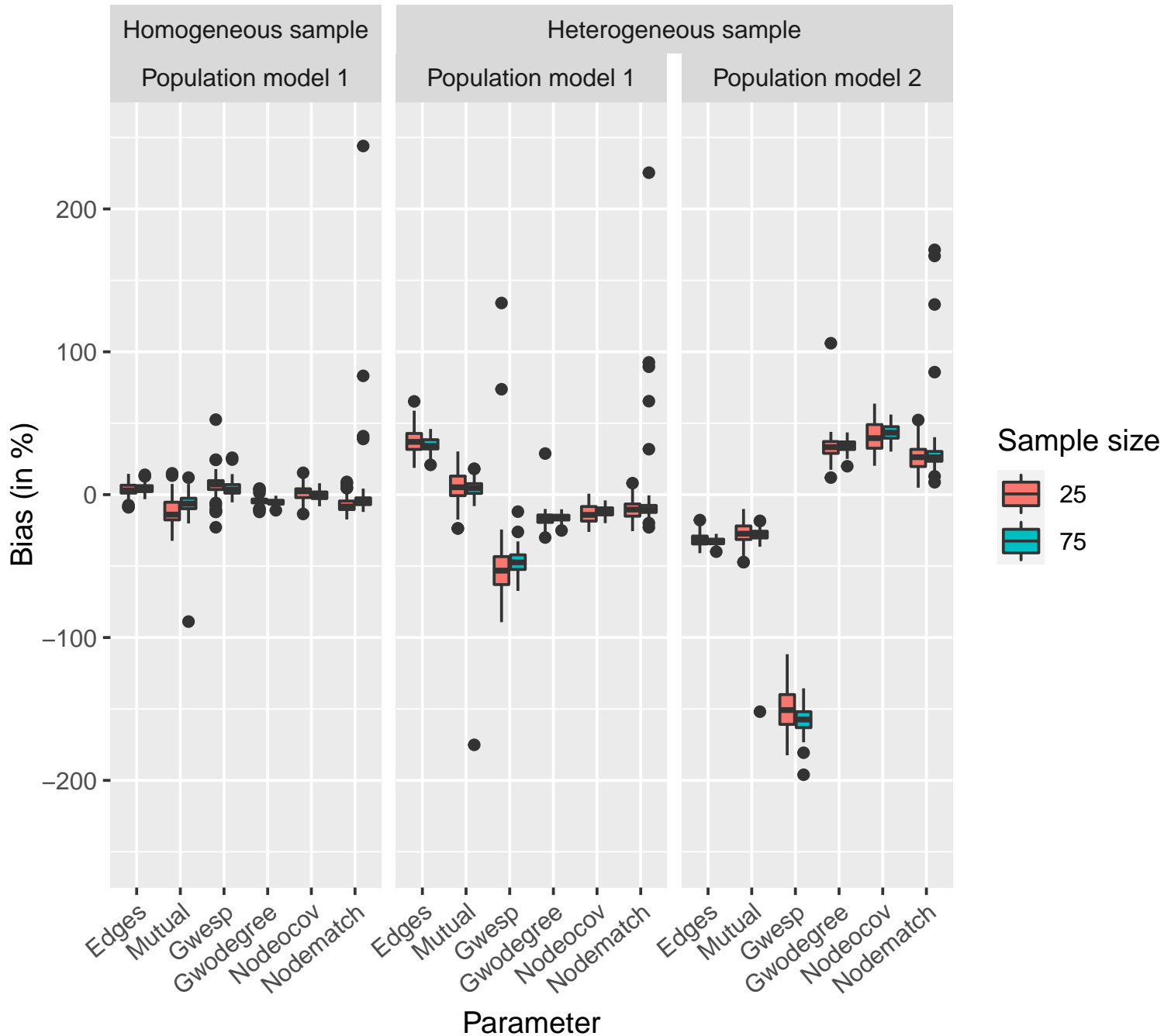
Figure 1: Relative bias of the parameters of the structural overall exponential random graph model by heterogeneity and sample size.



The reader is subsequently referred to figure 1 for an overview of the relative parameter bias of the structural meta-analysis model over the sample size and heterogeneity conditions for all 124 iterations. The first column of figure 1 shows that the structural specification underestimates the edges and mutual parameters whilst overestimating the geometrically weighted edgewise shared partners and (gwesp) and out-degree (gwodegree) parameters in the homogeneous sample. Put differently, the structural specification underestimates the number of edge realizations and reciprocated ties

while overestimating the degree of clustering and the number of edges sent. The bias is largest for the edges parameter. The different sets of parameter estimates are furthermore relatively precise. Put differently, the variance of the bias of the estimates between the respective meta-analyses is low. The first structural population model in the heterogeneous sample underestimates the edges and gwesp parameters whilst overestimating the gwodegree parameter. The mutual parameter is unbiased on average. None of the parameter estimates are furthermore markedly biased. Relative to the

Figure 2: Relative bias of the parameters of the true overall exponential random graph model by heterogeneity and sample size.

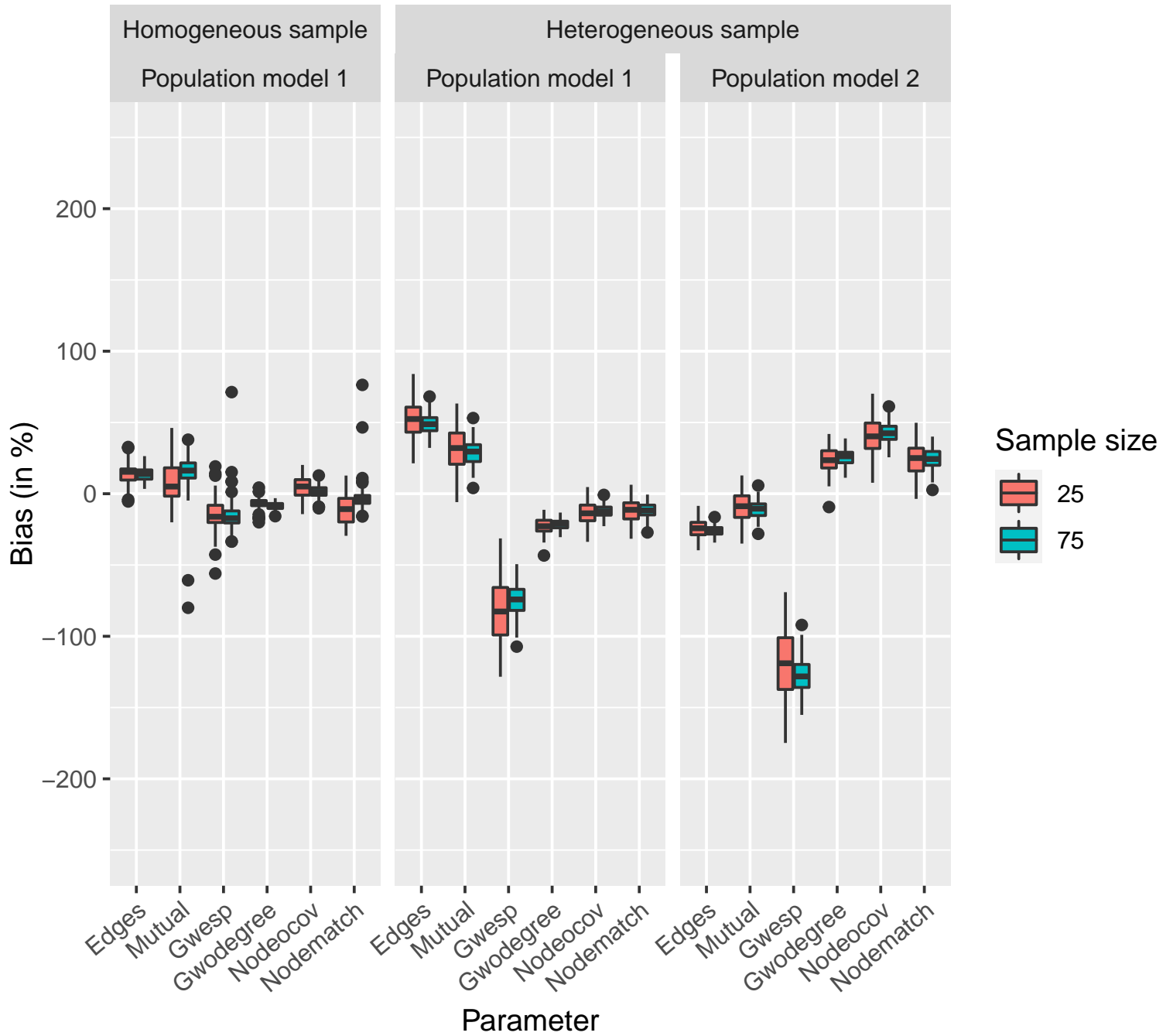


first population model in the homogeneous sample, the precision of all estimates has decreased. The second structural population model in the heterogeneous sample condition finally underestimates the edges, mutual and gwesp parameters whilst overestimating the gwodegree parameter. The gwesp parameter is markedly biased. The precision of the estimates is comparable to the precision of the parameters in the first population model in the heterogeneous sample. The bias for each of the parameters in the second population model is finally the largest for the three structural population models. The meta-analyses which have been estimated over

a sample of size 75 furthermore seem to be more precise than those which have been estimated over a sample of size 25, but not more unbiased.

Figure 2 shows the relative parameter bias for the true meta-analysis specification over the sample size and heterogeneity conditions for all 124 iterations. With regards to the homogeneous sample, the meta-analysis is close to unbiased for all parameters. Only the mutual term is underestimated slightly. As such, the true model seems to closely capture the density, tie reciprocation, vertex clustering, out-degree, out-degree by classroom identification, and gender homophily

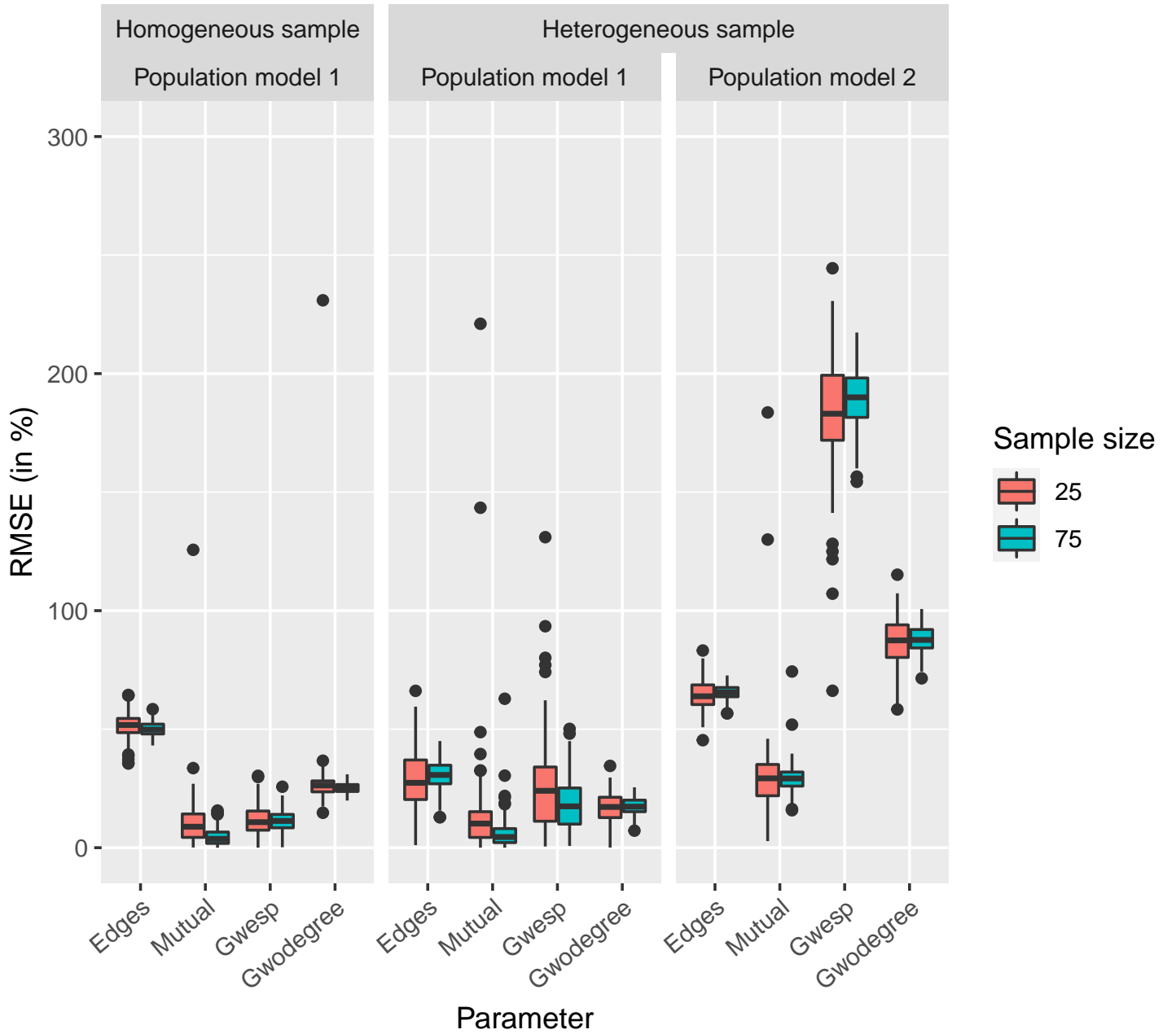
Figure 3: Relative bias of the parameters of the dense triadic overall exponential random graph model by heterogeneity and sample size.



effects in the population for a homogeneous sample. The set of estimates for each of the respective parameters is furthermore very precise. The first population model in the heterogeneous sample overestimates the edges parameter. The mutual parameter is close to unbiased. The model underestimates the gwesp, gwodegree and covariate parameters. Relative to the first population model in the homogeneous sample, the bias is slightly smaller for the mutual parameter and is approximately the same for the nodematch parameter. The bias for the edges, gwesp, gwodegree and nodeocov

terms is however larger. The precision is generally worse for the first population model in the heterogeneous as opposed to the homogeneous sample. The second population model in the heterogeneous sample underestimates the edges, mutual, and gwesp parameters, while overestimating the gwodegree and covariate parameters. The bias for the gwesp parameter is especially severe. The bias for the edges parameter is approximately equal to the bias of the edges parameter in the first population model in the heterogeneous sample, while the bias has increased for the remaining terms. The preci-

Figure 4: Relative RMSE of the parameters of the structural overall exponential random graph model by heterogeneity and sample size.

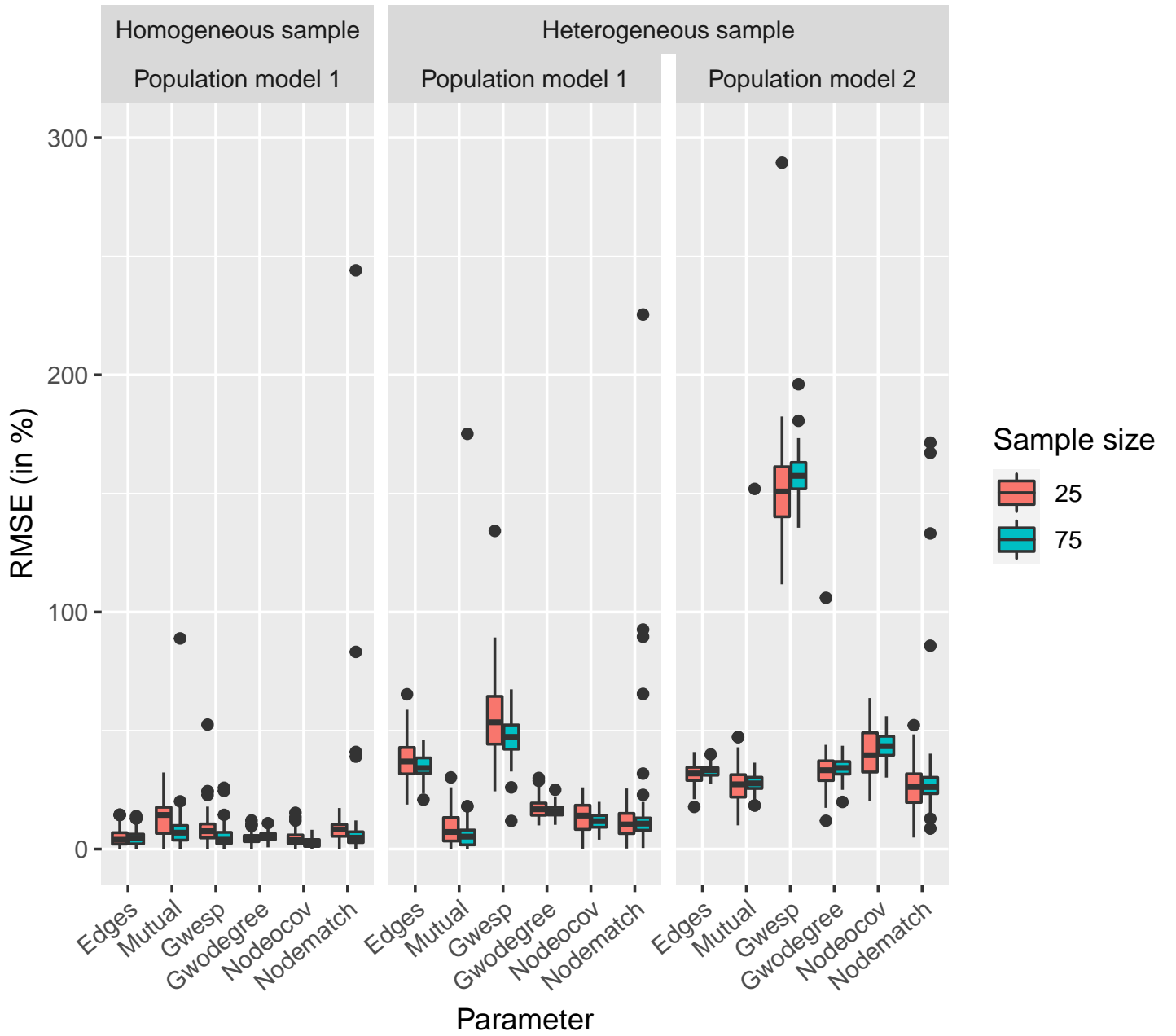


sion of the estimates of the gwesp parameters for both population models in the heterogeneous sample are markedly worse than the precision of gwesp estimates in the homogeneous sample. The precision of the first and second population models in the heterogeneous sample are approximately the same. With respect to the sample size, all three columns of the figure show that the meta-analyses are more precise for a sample of size 75 than a sample of size 25. Finally, in comparison to each of the structural meta-analysis model specifications in figure 1, the bias under each of the true meta-analysis specification is smaller for all parameter estimates except for the mutual term in the first and second population

models of the heterogeneous sample.

Figure 3 shows the relative parameter bias for the dense-triadic meta-analysis specification over the sample size and heterogeneity conditions for all 124 iterations. The first column of the figure shows that the dense triadic model slightly overestimates the edges parameter while slightly underestimating the gwesp, gwodegree and nodematch parameters in the homogeneous sample. The mutual and nodecov parameters are close to unbiased. In general, the dense triadic model shows little bias over the homogeneous sample. The estimates are furthermore relatively precise. The first population model in the heterogeneous sample overestimates

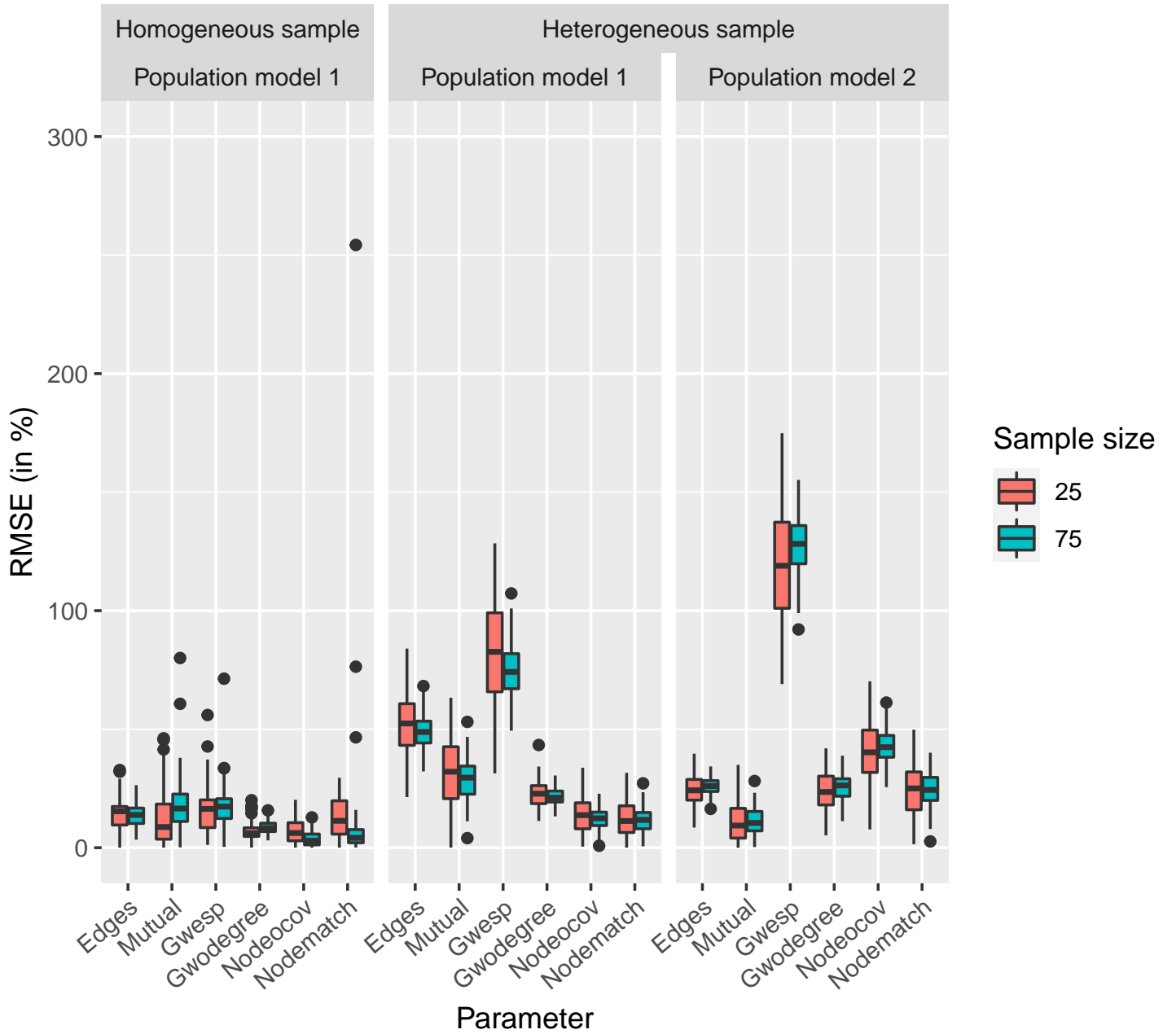
Figure 5: Relative RMSE of the parameters of the true overall exponential random graph model by heterogeneity and sample size.



the edges and mutual parameters while underestimating the gwesp, gwodegree, and covariate parameters. The relative bias is especially marked for the edges and gwesp parameters. Relative to the first population model in the homogeneous sample the relative bias has increased in the heterogeneous sample for all parameters except for the nodematch term. This is especially true for the edges, mutual, and gwesp terms. The estimates for the first population model in the heterogeneous sample are less precise for the edges and gwesp terms relative to the same model in the homogeneous sample, but approximately equal for the remaining terms. The second population model in the hetero-

neous sample underestimates the edges, mutual, and gwesp parameter while overestimating the gwodegree, nodecov, and nodematch parameters. The bias is especially marked for the gwesp term. Relative to the first population model in the heterogeneous sample, the bias is smaller for the edges and mutual parameters, but larger for the gwesp, gwodegree, and covariate parameters. The precision is slightly better for the edges and mutual terms, slightly worse for the gwodeg and covariate terms, and about the same for the gwesp term. In line with earlier findings, the meta-analysis estimates are more precise for a sample size of 75 than a sample size of 25 for all specifications of the dense-triadic meta-analysis. The

Figure 6: Relative RMSE of the parameters of the dense triadic overall exponential random graph model specification by heterogeneity and sample size.



dense triadic specification finally shows larger bias on all parameters in the homogeneous sample than the true specification. It is however less biased for the structural terms relative to the structural specification. The bias of the first population model under the dense triadic specification is slightly worse than that of the true specification on the structural terms. Little to no difference in bias can however be observed on the covariate terms. The relative bias of the first population model under the dense triadic specification in the heterogeneous sample is larger for all parameters relative to the structural specification. The bias of the second population model under the dense triadic specification is finally slightly

smaller for the mutual and gwesp terms relative to the true specification but equivalent for the remaining terms. The relative bias of the second population model under the dense triadic specification in the heterogeneous sample is smaller for all structural terms relative to the structural specification.

The reader is referred to figure 4 for an overview of the RMSE of the parameters for the structural specification over the sample size and heterogeneity conditions for all 124 iterations. The first column of figure 4 shows that the RMSE of the structural specification is lowest for the mutual term, followed by the gwesp, gwodegree, and edges terms respectively. Note that the variance of the bias is relatively low for

all structural specification parameters in the homogeneous sample. It is highest for the edges parameter. The precision of the variance of the bias for each of the parameters is generally very high. For the first population model in the heterogeneous sample, the RMSE is mostly equivalent to the RMSE of the first model in the homogeneous sample. The RMSE for the edges and gwodegree parameters is smaller, while it is larger for the gwesp parameter. The RMSE for the mutual term is approximately the same. The precision of the RMSE has however increased for all parameters. For the second population model in the heterogeneous sample, the RMSE is much larger than in either of the two first population models. The RMSE of the mutual term is the lowest, followed by the edges, gwodegree, and gwesp terms. The RMSE for the gwesp term is especially inflated, indicating that the variance of the bias for the gwesp estimate is very high. The precision of the variance of the bias is slightly higher than for the first population model in the heterogeneous sample, except for the gwesp term. The precision of the variance of the bias is finally consistently higher for meta-analyses applied to a sample size of 75 than a sample size of 25.

The reader is referred to figure 5 for an overview of the RMSE of the parameters for the true specification over the sample size and heterogeneity conditions for all 124 iterations. The first column shows that the RMSE is close to zero for all parameters in the true model specification. As such, there is little to no variance around the mean bias for any of these parameters. The precision of the RMSE for each parameter is furthermore very high. Relative to the first population model in the homogeneous sample, the RMSE for the first population model in the heterogeneous sample is equal for the mutual term but higher for the remaining terms. The RMSE for the edges and gwesp terms are markedly higher. The precision is also generally lower. The RMSE for the second population model in the heterogeneous sample is finally higher for all parameters relative to the first population model in the homogeneous sample. Although the RMSE of the edges parameter is somewhat lower, the RMSE of all remaining parameters is higher in the second as opposed to the first population model in the heterogeneous sample. The RMSE for the gwesp parameter is especially high. The precision of the RMSE in the second population model is slightly worse for the covariate terms, but otherwise comparable to the precision of the first population model in the heterogeneous sample. With regards to the sample size condition, the precision of the variance of the bias is consistently higher for meta-analyses applied to a sample size of 75 as opposed to one of 25. The RMSE of the true specification over the homogeneous samples is generally finally lower for the structural terms relative to the structural specification. The RMSE of the true specification for the first population model in the heterogeneous sample is close to equivalent to the RMSE for the structural specification, with the exception of the RMSE for the gwesp term, which is higher in the true specification. For the second population model in the heterogeneous sample, the RMSE of the true and structural specifications is equivalent for the mutual and gwodegree parameters, but

worse for the edges and gwesp parameters in the structural specification.

The reader is referred to figure 6 for an overview of the RMSE of the parameters for the dense triadic specification over the sample size and heterogeneity conditions for all 124 iterations. The first column of the figure shows slight RMSE values for all parameters in the true model specification. As such, there is some variance around the mean bias for these parameters under the dense triadic specification. The precision of the RMSE for each parameter is furthermore relatively high. For the first population model in the heterogeneous sample, the RMSE is relatively high for the edges, mutual, and gwesp parameters, but markedly lower for the gwodegree and covariate terms. The precision of the edges, mutual, and gwesp parameters is additionally quite low. For the second population model in the heterogeneous sample, the RMSE for the edges and gwesp terms are lower than for the first model in the heterogeneous sample, while the RMSE of the gwesp, gwodegree, and the covariate terms is higher. Analogously, the precision of the edges and mutual terms is improved relative to the first population model, while the precision of the gwodegree and the covariate terms is worse. The precision of the variance of the bias is finally consistently higher for meta-analyses applied to a sample size of 75 than a sample size of 25. Relative to the first true population model in the homogeneous sample, the RMSE of the dense triadic model is slightly higher. Relative to the structural model, the RMSE is lower for the edges and gwodegree parameters and about the same for the mutual and gwesp parameters. For the first population model in the heterogeneous sample, the dense triadic specification has a higher RMSE for the edges, mutual, and gwesp parameters, with there being close to no difference for the remaining parameters. For the structural parameters the dense triadic specification has a higher RMSE than the structural specification. For the second population model in the heterogeneous sample, the RMSE is close to equal between the dense triadic and true specifications, with the exception that the gwesp parameter has a much higher RMSE for the true specification. For the structural terms, the RMSE is higher for all parameters in the structural specification relative to the dense triadic specification.

With respect to the sample size condition, it can be concluded that the bias is generally not lower for a larger sample size, but that the RMSE is. It can additionally be concluded that the bias and RMSE are more precise for a higher as opposed to a larger sample size. With respect to the heterogeneity condition, it can be concluded that both the bias and RMSE tend to be larger in the heterogeneous sample as opposed to the homogeneous sample. The precision of the bias and RMSE is also lower in the heterogeneous sample. For the complexity condition with respect to the homogeneous sample, the bias and RMSE were lower for the true meta-analysis specification relative to the structural and dense triadic specifications. The bias and RMSE were furthermore lower for the dense triadic specification relative to the structural specification. For the first population model in

Figure 7: Auxiliary statistics goodness-of-fit of the overall exponential random graph model for the homogeneous sample.

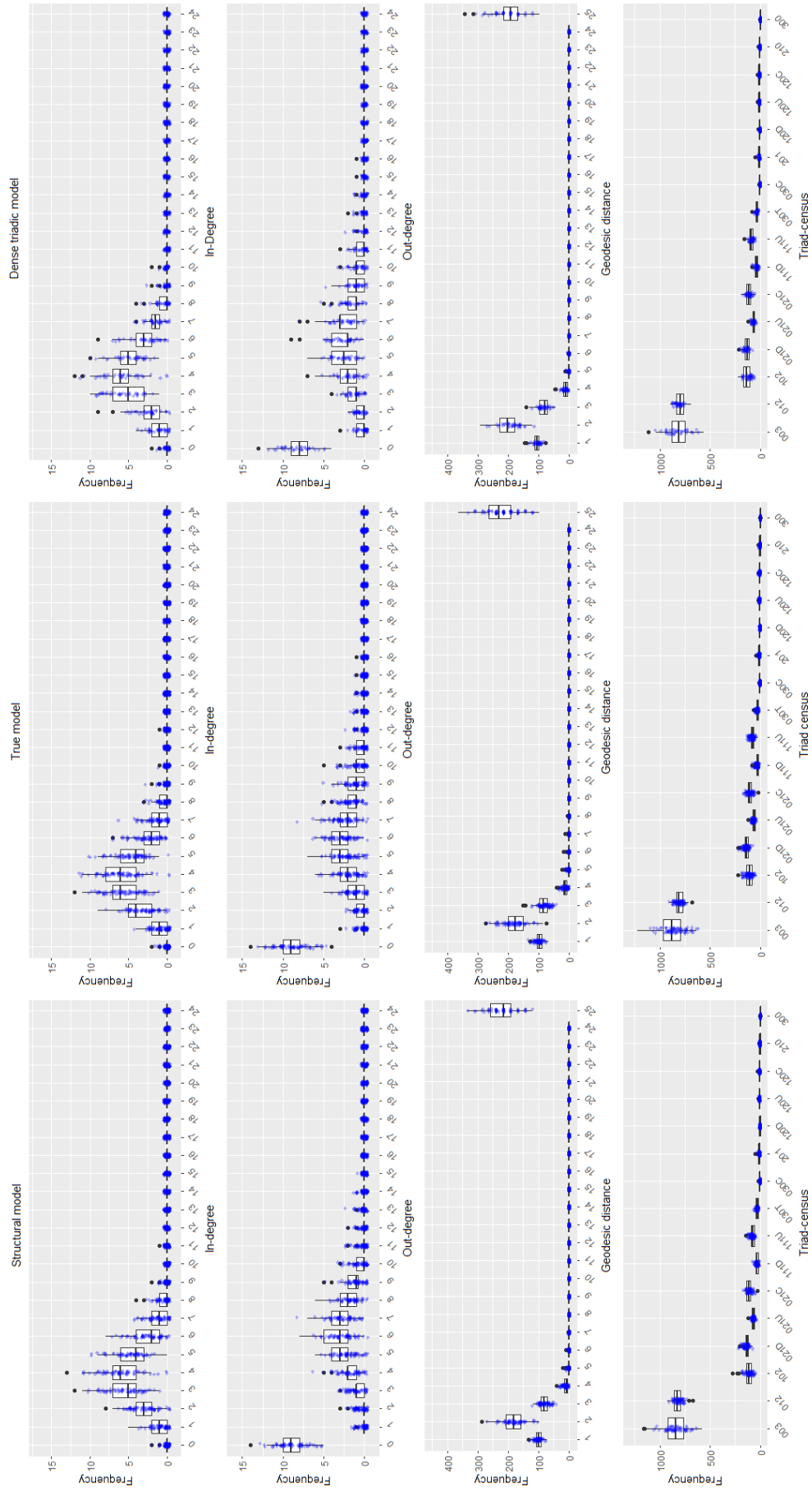


Figure 8: Auxiliary statistics goodness-of-fit of the overall exponential random graph model for the homogeneous sample.

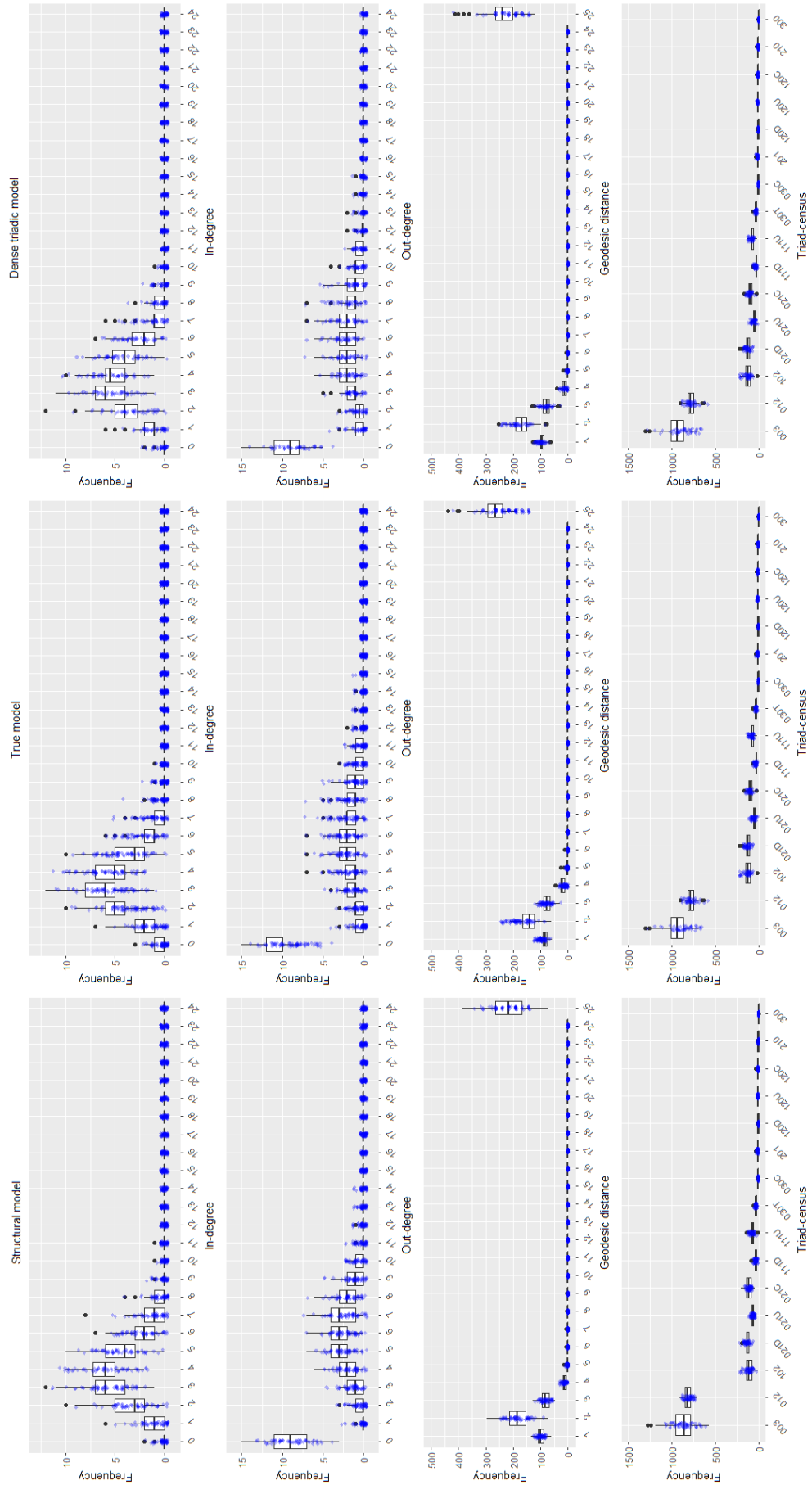
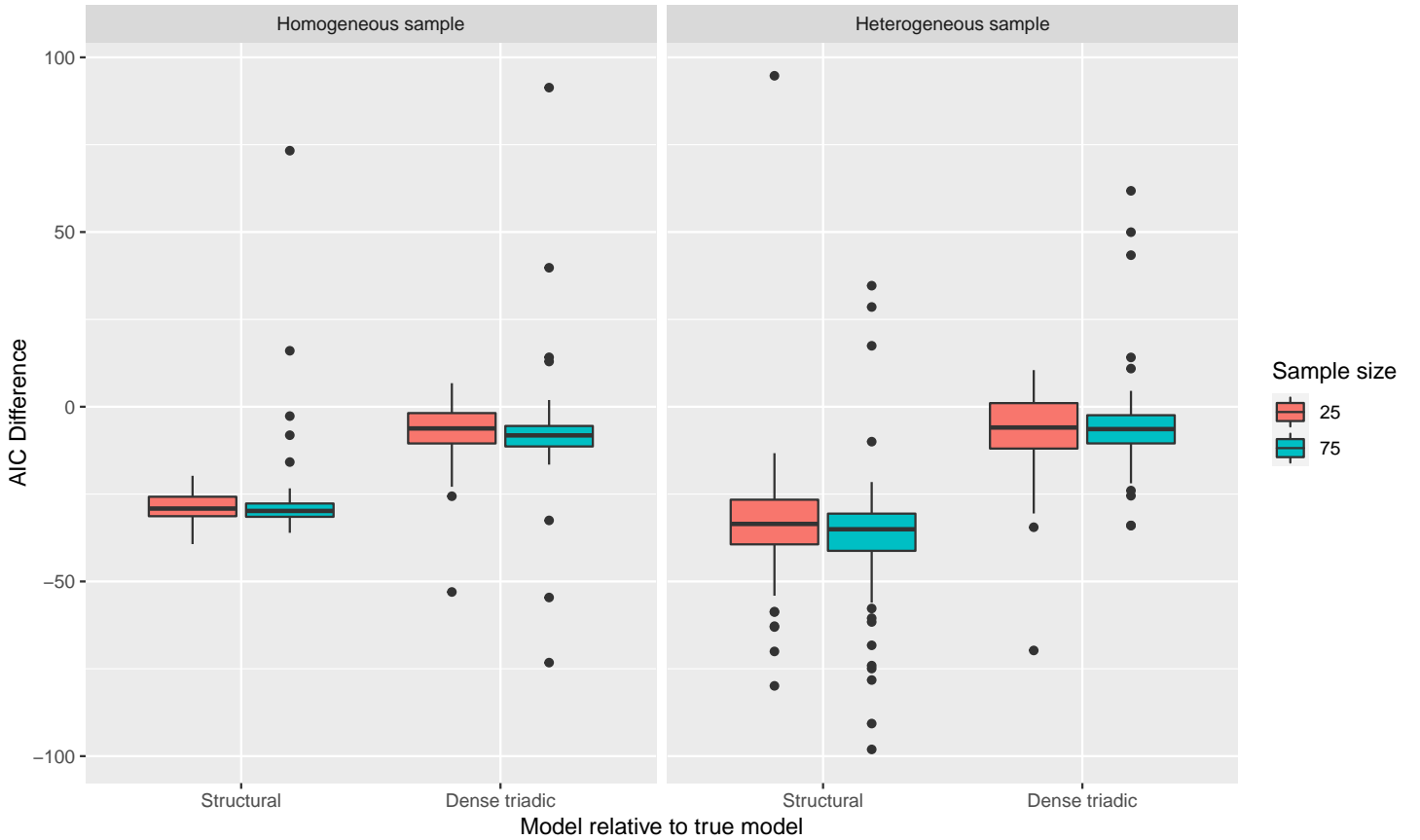


Figure 9: AIC difference score of the structural and dense triad model relative to the true model.



the heterogeneous sample, the bias and rmse of the structural specification is comparable to the true specification, while the bias and rmse of the true specification is generally better than that of the dense triadic specification. For the second population model in the heterogeneous sample, the bias and rmse of the structural specification is worse than in the true specification, while the bias and rmse of the true specification is approximately equivalent to that of the dense triadic specification. It is also generally the case that the bias and RMSE of the structural parameters are much worse than those of the covariate parameters. Finally, none of the null-hypotheses are rejected because no formal statistical testing was conducted.

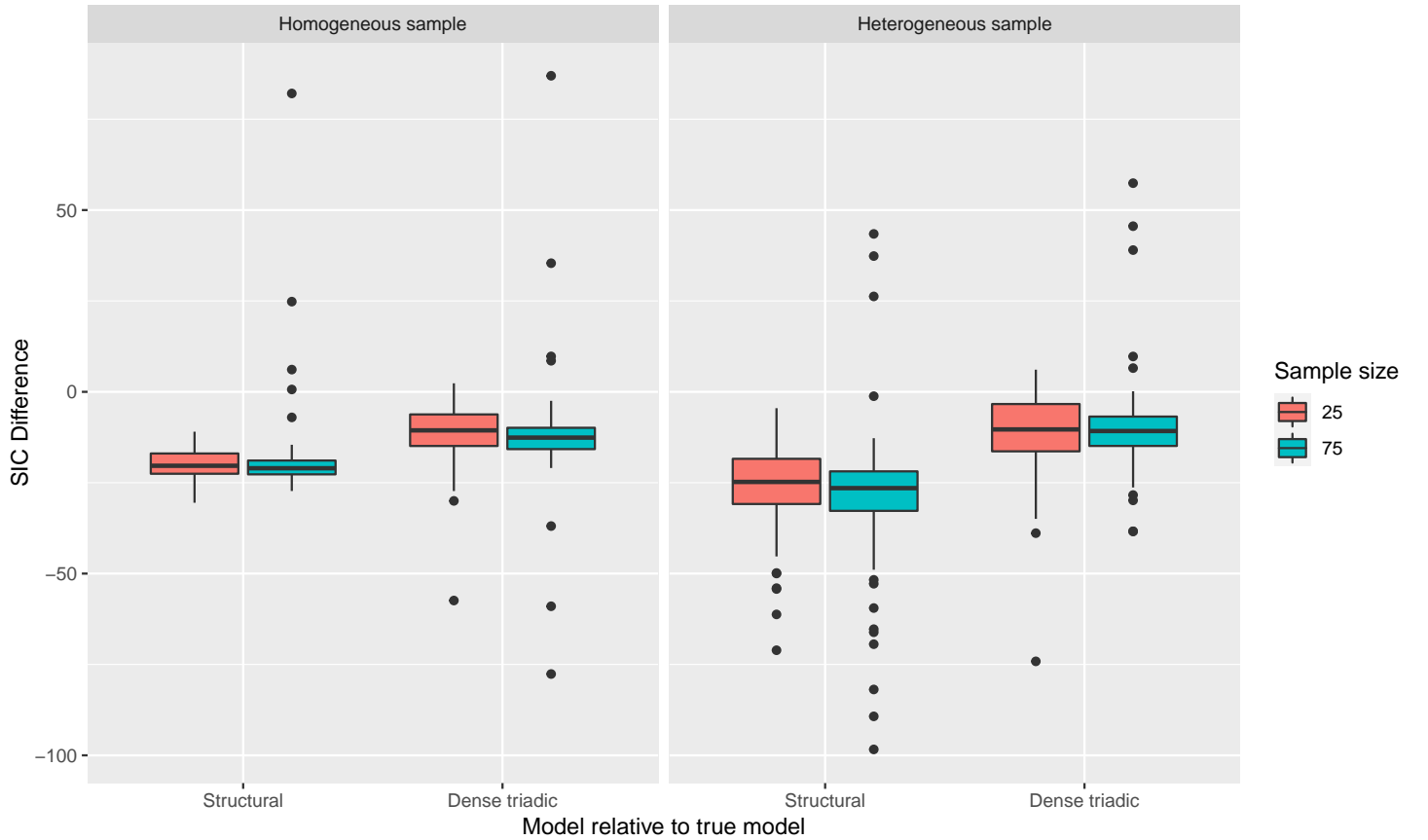
The reader is consequently referred to figure 7 for an overview of the auxiliary statistics goodness-of-fit of the overall exponential random graph model (ERGM) for the homogeneous sample. More specifically, the figure shows the respective auxiliary statistics goodness-of-fit (GOF) indices for the first of the 124 meta-analysis models for the homogeneous sample. Overlaid over the respective GOF indices for the respective complexity specifications are the observed values in the simulated network sample. As becomes apparent from the plot, the simulated and observed distributions tend to overlap. No egregious discrepancies can be discerned between what is simulated under the model and what is observed in the sample of the auxiliary statistics. This pat-

tern furthermore seems to hold for the structural, true, and dense triadic specifications.

A similar pattern can be discerned in figure 8, which shows the respective auxiliary statistics goodness-of-fit (GOF) indices for the first of the 124 meta-analysis models for the heterogeneous sample. It becomes clear from the plot that the simulated and observed distributions tend to overlap. No egregious discrepancies can be discerned between what is simulated under the model and what is observed in the sample of the auxiliary statistics. This pattern again seems to hold for the structural, true, and dense triadic meta-analysis specifications.

The reader is finally referred to figures 9 and 10 for an overview of AIC and SIC difference score of the structural and dense triad model relative to the true model. As becomes apparent from the first column of the two figures, the AIC and SIC are both higher for the structural and dense triadic model relative to the true model in the homogeneous sample. The difference is furthermore much more pronounced for the structural specification. The second column of figures 9 and 10 shows that the AIC and SIC are both higher for the structural and dense triadic model relative to the true model for the heterogeneous sample. The difference is again much more pronounced for the structural than the dense triadic model.

Figure 10: SIC difference score of the structural and dense triad model relative to the true model.



5. Conclusion

The objective of this study was to identify the conditions under which a meta-analysis produces a valid exponential random graph model (ERGM) with respect to a sample of completely observed social networks. Its second objective was to investigate the conditions under which a subset of the currently available goodness-of-fit (GOF) measures are able to identify sub-par GOF on the level of the network sample.

With regards to the first objective, two conditions were hypothesized to affect the validity of the ERGM model parameters on the level of the network sample: the characteristics of the sample and the complexity of the ERGM. With respect to the sample characteristics, it was hypothesized that a meta-analysis of ERGM estimates is more accurate for larger as opposed to smaller network samples and that the accuracy of the meta-analysis is reduced when it is applied to a sample with as opposed to no outlying networks. With respect to the complexity of the ERGM, the accuracy of a meta-analysis of ERGM estimates was hypothesized to decrease for an increased discrepancy between the complexity of the model and the complexity of the sample. The accuracy of covariate effects was additionally hypothesized to be higher than that of the structural network effects.

The results showed that a larger sample size did not decrease the bias of the meta-analysis estimates, but that it

did decrease its RMSE and increased the precision of both the bias and RMSE. As such, larger sample sizes are tentatively found to result in more precise meta-analysis exponential random graph model (ERGM) estimates which are clustered more closely around the mean bias. It was secondly shown that the bias and RMSE were larger and less precise for meta-analysis parameter estimates in heterogeneous samples than homogeneous samples. It can thus be tentatively concluded that a meta-analysis is more likely to be biased and less precise for samples which are drawn from a mixed network population than from a uniform network population. With respect to the complexity conditions, the bias and RMSE of the true meta-analysis specification was found to be lower than those of the structural and dense triadic specifications in the homogeneous sample. For the first population model in the heterogeneous sample, the bias and rmse of the true specification were found to be comparable to the structural specification, while being superior to the bias and RMSE of the dense triadic specification. For the second population model in the heterogeneous sample, the bias and RMSE of the true specification were superior to those of the structural specification, while being approximately equivalent to that of the dense triadic specification. As such, it can be concluded that for samples from a homogeneous network population, the specification which best approximates the underlying population model is generally

optimal. For a heterogeneous sample it is similarly concluded that ideally a specification should be used which most closely approximates the true underlying population model. Although the bias and RMSE of the structural specification was equivalent to the bias and RMSE of the true specification in the first population model, it was much worse in the second. Inversely, although the bias and RMSE of the dense triadic specification was equivalent to the bias and RMSE of the true specification in the second population model, it was much worse in the first. On average therefore, the true specification is the most likely to render optimal results in both sample heterogeneity types, although these will likely not be unbiased. It is finally concluded that the bias and RMSE of the structural parameters were generally much worse than those of the covariate parameters.

With regards to the auxiliary statistics goodness-of-fit (GOF) measure and the AIC and SIC criteria, the results indicated that both these GOF measures were reliable indicators in the context of the meta-analysis of ERGM estimates. More specifically, the auxiliary statistics GOF indices were found to closely resemble the observed distributions for the set of auxiliary statistics for the first meta-analysis in each cell. This finding results in the tentative conclusion that auxiliary statistics GOF functions as a reliable indicator for the fit of a meta-analysis of ERGM estimates. It should however be noted that discrepancies in the fit on the auxiliary statistics between the different model specification were not easily identifiable. Since it has been established that the structural and dense triadic specification are more biased than the true specification, the auxiliary statistics GOF should ideally have reflected this fact. This did not however not become clear from visual inspection of figures 7 and 8. The AIC and SIC were finally found to be reliable indicators of relative GOF for the meta-analysis of ERGM estimates.

6. Discussion

This study has a number of shortcomings. The first is that a uni-variate meta-analysis was used to obtain exponential random graph model (ERGM) estimates over the network sample instead of a multivariate meta-analysis. This is a key shortcoming because ERGM parameters are known to be inter-dependent. By not incorporating such dependencies in the weighted average of the meta-analysis, bias was likely introduced. As such, both applied researchers and statistical researchers would benefit from incorporating a multivariate meta-analytic method. A second major shortcoming is that no t- and p-values were presented for the auxiliary statistics goodness-of-fit (GOF) on both the level of the individual ERGM and the meta-analysis of the ERGM estimates. As a consequence no exact results were provided for the auxiliary statistics GOF indices on either of these levels. Future work could benefit from obtaining these measures and presenting them by way of statistical significance testing to strengthen potential conclusions. This point translates into the larger observation that no significance testing was employed to formally evaluate the set of presented hypotheses. This point also holds for the bias and RMSE of

the meta-analyses over the different conditions. Ideally, a multivariate analysis of variance (MANOVA) would have been used to separate within and between variance for the bias, RMSE, and the respective t-values, Mahalanobis distances, and p-values of the auxiliary statistics GOF at the level of the network and the network sample over the set of simulation cells. In that way formal hypotheses tests could have been executed by way of significance testing, which would likely have strengthened conclusions. As such any future work could strongly benefit from significance testing. A final point is that a meta-analysis would ideally have been obtained over each simulation cell a 1000 times to ensure proper statistical coverage, and not 124 times as was the case in this study. As such, a subset of meta-analysis units which belong to the target population of meta-analysis might have been omitted.

References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6), 716–723.
- Block, P., Stadtfeld, C., & Snijders, T. A. (2019). Forms of dependence: Comparing saoms and ergms from basic principles. *Sociological Methods & Research*, 48(1), 202–239.
- Cox, D. R. (2006). *Principles of statistical inference*. Cambridge university press.
- Crano, W. D., Brewer, M. B., & Lac, A. (2014). *Principles and methods of social research*. Routledge.
- Geyer, C. J., & Thompson, E. A. (1992). Constrained monte carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 54(3), 657–683.
- Hunter, D. R., Goodreau, S. M., & Handcock, M. S. (2008). Goodness of fit of social network models. *Journal of the American Statistical Association*, 103(481), 248–258.
- Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., & Morris, M. (2008). ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of statistical software*, 24(3), nihpa54860.
- Kolaczyk, E. D., & Krivitsky, P. N. (2015). On the question of effective sample size in network modeling: An asymptotic inquiry. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 30(2), 184.
- Koskinen, J., Wang, P., Robins, G., & Pattison, P. (2018). Outliers and influential observations in exponential random graph models. *Psychometrika*, 83(4), 809–830.
- Lohr, S. L. (2009). *Sampling: design and analysis*. Nelson Education.
- Lubbers, M. J., & Snijders, T. A. (2007). A comparison of various approaches to the exponential random graph model: A reanalysis of 102 student networks in school classes. *Social networks*, 29(4), 489–507.

- Luke, D. A. (2015). *A user's guide to network analysis in r*. Springer.
- Lusher, D., Koskinen, J., & Robins, G. (2013). *Exponential random graph models for social networks: Theory, methods, and applications*. Cambridge University Press.
- Rambaran, J. A., Dijkstra, J. K., & Veenstra, R. (2020). Bullying as a group process in childhood: A longitudinal social network analysis. *Child development*, 91(4), 1336–1352.
- Rehnberg, Z. (2016). *Exponential random graph models under measurement error*. Washington University in St. Louis.
- Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2), 461–464.
- Snijders, T. A. (2001). The statistical evaluation of social network dynamics. *Sociological methodology*, 31(1), 361–395.
- Team, R. C. (2013). *R: A language and environment for statistical computing*. Vienna, Austria.
- Van Duijn, M. A., Gile, K. J., & Handcock, M. S. (2009). A framework for the comparison of maximum pseudo-likelihood and maximum likelihood estimation of exponential family random graph models. *Social Networks*, 31(1), 52–62.
- Vermeij, L. (2006). *What's cooking: cultural boundaries among dutch teenagers of different ethnic origins in the context of school*. Utrecht University.
- Veroniki, A. A., Jackson, D., Viechtbauer, W., Bender, R., Bowden, J., Knapp, G., ... Salanti, G. (2016). Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Research synthesis methods*, 7(1), 55–79.
- Viechtbauer, W. (2010). Conducting meta-analyses in r with the metafor package. *Journal of statistical software*, 36(3), 1–48.
- Wit, E., Heuvel, E. v. d., & Romeijn, J.-W. (2012). 'all models are wrong...': an introduction to model uncertainty. *Statistica Neerlandica*, 66(3), 217–236.