

1 Exponential Random Graph models for Little Networks

2 Abstract

Statistical models for social networks have enabled researchers to study complex social phenomena that give rise to observed patterns of relationships among social actors and to gain a rich understanding of the interdependent nature of social ties and actors. Much of this research has focused on social networks within medium to large social groups. To date, these advances in statistical models for social networks, and in particular, of Exponential-Family Random Graph Models (ERGMS), have rarely been applied to the study of small networks, despite small network data in teams, families, and personal networks being common in many fields. In this paper, we revisit the estimation of ERGMs for small networks and propose using exhaustive enumeration when possible. We developed an R package that implements the estimation of pooled ERGMs for small networks using Maximum Likelihood Estimation (MLE), called “ergmito”. Based on the results of an extensive simulation study to assess the properties of the MLE estimator, we conclude that there are several benefits of direct MLE estimation compared to approximate methods and that this creates opportunities for valuable methodological innovations that can be applied to modeling social networks with ERGMs.

3 *Keywords:* exponential random graph models, small networks, exact
4 statistics, simulation study

5 1. Introduction

6 Statistical models for social networks have enabled researchers to study com-
7 plex social phenomena that give rise to observed patterns of relationships among
8 social actors, and to gain a rich understanding of the *interdependent* nature of
9 social ties and social actors [17, 25]. For example, this research has provided new

insights into the role that the attributes of social actors (e.g., their characteristics, beliefs, and decisions), and endogenous structural processes (e.g., social balance, and relationship reciprocity) play in shaping social networks across different populations and social settings, and how these social networks, in turn, influence individuals and groups.

Much of this research has focused on social networks within medium to large social groups: networks ranging from dozens or hundreds of members (e.g., classrooms and organizations) to millions (e.g., online social networks). However, modern advances in statistical models for social networks have rarely been applied to the study of small networks, despite small network data from teams, families, and personal (ego-centric) networks being common in many fields that study social phenomena [1, 3, 4, 10]. The study of small networks often uses descriptive statistics that summarize basic structural features of the network; for example, the density, degree distribution, or triad count. However, researchers in these fields are often interested in testing hypotheses about *why* localized social structures, such as reciprocity, balance, and homophily, emerge in these small groups. A key limitation to such work has been the availability of statistical models for networks that can flexibly test and control for the kind of dependencies inherent to network data. In this paper, we propose an approach for applying one of the most widely used statistical models for social networks—exponential random graph models, or ERGMs—to small graphs, to better enable new and rich research on “little networks”.

2. Exponential-Family Random Graph Models

Exponential-family random graph models (ERGMs) are one of the most popular tools used by social scientists to understand social networks and test hypotheses about these networks [6, 11, 21, 26, 33, and others]. In this family of models, an observed graph y , comprised of a set of nodes (vertices) and ties (edges), is characterized by a set of sufficient statistics defined on the graph, $s(y)$, and parameters θ . In a model that also includes node characteristics X , this leads to the following equation:






Representation	Description
	Mutual Ties (Reciprocity) $\sum_{i \neq j} y_{ij} y_{ji}$
	Transitive Triad (Balance) $\sum_{i \neq j \neq k} y_{ij} y_{jk} y_{ik}$
	Homophily $\sum_{i \neq j} y_{ij} \mathbf{1}(x_i = x_j)$
	Covariate Effect for Incoming Ties $\sum_{i \neq j} y_{ij} x_j$
	Four Cycle $\sum_{i \neq j \neq k \neq l} y_{ij} y_{jk} y_{kl} y_{li}$

Figure 1: Besides of the common edge count statistic (number of ties in a graph), ERGMs allow measuring other more complex structures that can be captured as sufficient statistics.

$$\Pr(Y = y \mid \theta, X) = \frac{\exp\{\theta^t s(y, X)\}}{\kappa(\theta, X)}, \quad \forall y \in \mathcal{Y} \quad (1)$$

Where $\kappa(\theta, X) = \sum_{y \in \mathcal{Y}} \exp\{\theta^t s(y, X)\}$ is the normalizing constant, and \mathcal{Y} is the support of the model which is usually assumed to include all graphs of the same type (e.g., directed or undirected) and size, and do not include self-ties. In the directed graph case, the size of \mathcal{Y} equals $2^{n(n-1)}$ possible graphs. This makes the exact calculation of $\kappa(\theta, X)$, and therefore of (1), computationally expensive. A sophisticated array of parameters can be specified for ERGMs that reflect social and structural process of interest to social scientists, such as social closure, connectivity, and other affiliation preferences. Figure 1 shows some examples of the structures (statistics) that can be estimated with ERGMs.

Although “small networks” is a topic mentioned several times in the literature on social network models [6, 25, 33], interest in larger social networks has dominated the field.¹ Thus, ERGM methods have been developed to accommodate larger networks (although they do not scale well to “very large”

¹This is perhaps because, as put by [25], small networks are considered to be “uninteresting special cases”

53 networks of several thousand nodes or more). One example of this is the calcu-
 54 lation of the likelihood function: rather than being calculated using exhaustive
 55 enumeration (which we will refer to as "exact likelihood"), the most popular
 56 software packages used for estimating these models apply simulation-based esti-
 57 mation methods. As a consequence, current methods used to estimate ERGMs
 58 for medium to large networks do not translate well to small network data (i.e., 6
 59 or fewer nodes if directed), and applications of these statistical network models
 60 to small networks are rare.

61 One major technical and theoretical issue in ERGM estimation generally,
 62 which is exacerbated with small networks, is the problem of *degeneracy*. Degen-
 63 erate models occur when the observed graph statistics lie in a region on or near
 64 the boundary of the support, and can be stressed when estimation depends on
 65 Monte Carlo Integration [8, 29]. Small networks, which are more likely to be
 66 nearly empty or nearly full, have a smaller region of support, and are more likely
 67 to be on or near the boundary of that support. For example, if we are trying to
 68 estimate an ERGM in a network with only three nodes, in the scenario where
 69 the graph is directed and does not allow for self-ties, the chances of obtaining a
 70 graph with either one or zero ties (i.e., empty or almost completely empty), or
 71 a graph with five or six ties (i.e., fully or almost fully connected) is about 20%
 72 using a uniform sampler.

73 Because researchers studying small networks often have observed *samples* of
 74 small networks (e.g., multiple team, family, or personal/egocentric networks), a
 75 common work-around to the issue of model degeneracy is to combine the inde-
 76 pendent small networks into a single larger block-diagonal graph. Estimation
 77 then proceeds by assuming that ties between blocks are impossible (i.e., treated
 78 as structural zeros in estimation). The major problems with this approach are
 79 that it can be complicated to fit, and difficult to extend. For example, the same
 80 set of constraints (the structural zeros) that allows for the model to be fit can
 81 also make the estimation procedure more difficult, and increase the possibility
 82 of sampling problems during MCMC estimation. More importantly, however,
 83 the block-diagonal approach can be difficult to extend.

84 A basic “complete pooling” model – that assumes a common data gener-
 85 ating process across all networks – is straightforward to define using existing
 86 tools for ERGM estimation. However, relaxing that assumption to allow for
 87 variability across graphs (i.e., unpooled or partially-pooled models) can become
 88 problematic; this would typically require the creation of block-wise node mem-
 89 bership attributes, and complex interaction terms involving subgraph statistics
 90 and node membership variables. Moreover, extending such a framework to not
 91 just allow for between-group variability, but to *predict* it (for example, as a
 92 function of additional group-level variables), is not straightforward in such an
 93 approach.

94 To overcome the challenges described above for fitting ERGMs to small net-
 95 works, we leverage the fact that in the case of small networks, the full likelihood
 96 function *is* tractable. This allows the direct estimation of model parameters
 97 without using Markov Chain Monte Carlo (MCMC) or other approximate meth-
 98 ods, avoiding some of the convergence issues associated with inference degen-
 99 eracy [8]. It also opens up the possibility of making it much easier to combine
 100 ERGMs with other statistical techniques, opening the door for richer methods
 101 for modeling and understanding small-group structure and dynamics. In this
 102 paper, we describe how modern computational power allows for the complete
 103 specification of the likelihood for small graphs, and how this specification al-
 104 lows us to use the standard tools of maximum likelihood estimation (MLE), as
 105 opposed to approximate methods, including confidence intervals and likelihood
 106 ratio tests. We present examples using these techniques, provide some initial
 107 Monte Carlo results on bias, type I error rates, and power, and discuss future
 108 extensions these techniques make feasible.

109 **3. ERGMitos: ERGMs for small networks**

110 With modern computers, calculating the exact likelihood function of an
 111 ERGM for a small network becomes computationally feasible. This has an im-
 112 portant implication: the process for estimating the parameters of an ERGM for

113 small networks can be done directly. Many innovative techniques have been de-
 114 veloped to handle models with intractable normalizing constants (e.g., MCMC-
 115 MLE, Bayesian techniques such as the exchange sampler, etc.), and often these
 116 techniques work quite well. Of course, no techniques are without tradeoffs;
 117 MCMC-MLE estimation can be very sensitive to starting values, and the qual-
 118 ity of standard errors can depend on the availability of analytic gradients [19].
 119 Bayesian techniques like the exchange sampler [18] may be comparatively slow,
 120 which may be an issue when many networks are to be analyzed.

121 Moreover, simulation-based methods may have particular susceptibilities to
 122 model degeneracy. As stated by [8, p. 7], "[i]f the model used to simulate
 123 the graphs is not close enough to produce realizations that cover the observed
 124 values of the statistics, the MC-MLE will not exist even in cases where the
 125 MLE does." For example, many common network models, such as triangle-based
 126 models, can lead to bimodal distributions of graph statistics that simulation-
 127 based methods have difficulty with, even when the MLE falls between the modes
 128 [13]. Therefore, even though the degeneracy issue is not completely avoided,
 129 a method based on exact (non-simulation) inference may not only provide a
 130 better solution (in general) by avoiding the additional uncertainty induced by
 131 simulations and approximations, but may help mitigate the degeneracy issue in
 132 cases where the MLE exists.

133 Of course, the statistical analysis of a single small network is going to be
 134 uninformative due to the small numbers of dyads, and a high restriction in the
 135 variability of possible subgraph statistics. Fortunately, small network data is
 136 typically collected from *samples* of small groups, which allows for the develop-
 137 ment of models to analyze structural variation both within and across small
 138 networks. If we assume that the sample of networks comes from a population
 139 of networks (groups) that are governed by the same data generating process, we
 140 end up with the following likelihood, defining a completely-pooled model:

$$\Pr(Y_1 = y_1, \dots, Y_P = y_P \mid \theta, X_1, \dots, X_P) = \prod_{p=1}^P \frac{\exp\{\theta^t s(y_p, X_p)\}}{\kappa_p(\theta, X_p)} \quad (2)$$

Where P denotes the number of networks used in the model, and $\kappa_p(\theta, X_p)$ is explicitly calculated, unlike existing approaches to ERGM estimation. We call this framework, which is a revisited version of ERGM in the case of small networks, *ERGMito*².

One complication with ERGMs involving multiple networks is that such models are usually not projective [23], and the magnitude of parameter estimates and their standard errors (at least with network statistics in standard use) are all implicitly conditioned on network size. This means caution is required when combining or comparing parameter estimates from different networks. Even the interpretation of the basic edge count parameter can become difficult; as pointed out in [15], assuming a common edge parameter in networks of different sizes is equivalent to assuming equal edge probabilities in an edge-only model. However, many real-world social networks are sparse, with a density that grows much more slowly than network size.

This is a significant concern in cases where samples of networks may span a wide range of sizes, and requires careful consideration of issues regarding choice of network statistics and the appropriateness of a completely-pooled approach (compared to, for example, a partially-pooled approach that allows for variability in estimates across networks). However, in the case outlined in this paper, we focus on a sample of networks that are of similar size, and we assume there is a common data-generating process. The requirements and assumptions of the completely-pooled model in (2) therefore can be assumed to hold.

The simulations and model fitting were conducted using the R package

²The *ito/ita* suffix is used in Spanish to denote small, or affection. We are especially grateful to George Barnett who proposed the name during the North American Social Networks Conference in 2018.

164 `ergmito`, which has been developed to implement the methods described in
 165 this paper.

166 4. Illustration with simulated data: `fivenets`

167 4.1. Data-generating-process and model fitting

168 In the following section we work with a simulated data set that was created
 169 using the data-generating-process of `ERGMitos`. This particular dataset, which
 170 we call “fivenets”, is included in the in the R package `ergmito`³. The data
 171 set contains five simulated networks with nodal attributes (we use gender in
 172 the following example), and the networks were generated using the following
 173 specification:

$$\Pr(Y = y \mid X, \theta) = \frac{\exp \left\{ \theta_{edges} \left(\sum_{i,j} y_{ij} \right) + \theta_{same} \left(\sum_{i,j} y_{ij} \mathbf{1}(X_i = X_j) \right) \right\}}{\kappa(\theta, X)}$$

174 where $\theta_{edges} = -2.0$ and $\theta_{same} = 2$. Using this equation we draw five networks
 175 of size five. The process of “homophily” is represented by a parameter that
 176 is defined as the number of ties in which ego and alter have the same gender,
 177 θ_{same} . Before drawing the networks we randomly generated the node attribute
 178 (gender) parameter to each vertex as a Bernoulli with parameter 0.5. [Figure 2](#)
 179 shows the generated networks, including their nodal attributes.

180 Using the `ergmito` R package, we fit three different models to the data⁴: (1)
 181 a Bernoulli graph, which is a model that only includes the “edges” parameter,
 182 (2) a model with “gender homophily” as its only parameter, and finally (3)
 183 a model including both “edges” and “gender homophily”, this is, the correct
 184 specification of the model. [Table 1](#) shows the estimation results of the three
 185 different specifications of the model and as expected, model (3) has the best

³The R package is available to be downloaded at BLIND REVIEW.

⁴Some details regarding the computational aspects of the model fitting process are provided in [Appendix A](#).

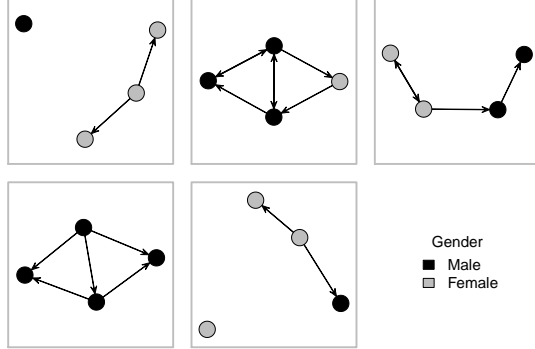


Figure 2: Fivenets data set. These graphs were randomly drawn from an ERGM distribution with two parameters: number of edges and gender homophily, with parameters equal to -2.0 and 2.0 respectively.

186 overall fit to the data. Furthermore, since all three models were fitted using
 187 MLE, we can compare the edgcount and homophily models with the full model
 188 using Likelihood Ratio tests [35]. The *goodness of fit* of this model is evaluated
 189 in the following section.

	Homophily	Edgcount	Full model
Edgcount		-0.69* (0.27)	-1.70** (0.54)
Homophily (on Gender)	-0.12 (0.34)		1.59* (0.64)
LR-test statistic (χ^2)	7.04**	13.72***	
AIC	85.06	78.38	73.34
BIC	87.15	80.48	77.53
Log Likelihood	-41.53	-38.19	-34.67
Num. networks	5	5	5
Convergence	0	0	0

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table 1: Fitted ERGMitos using the fivenets dataset. Looking at AICs and LR-test statistics, the full model (last column of the table) is the one with the best fit to the observed data. More over, the 95% level CI of each covers the true parameters: $\hat{\theta}_{edges} \in [-2.77, -0.64]$; $\hat{\theta}_{Homophily} \in [0.33, 2.85]$.

190 4.2. Goodness-of-fit in ERGMitos

191 Researchers that apply ERGMs should be familiar with the goodness-of-fit
 192 (GOF) diagnostics that are used to assess how well the estimated model can

193 reproduce graphs that are similar to the observed graph on a range of local and
 194 global graph statistics [12]. In the case of *ERGMitos* applied to small networks,
 195 local graph statistics will be more relevant than global statistics to assess GOF.
 196 For example, the graph geodesic distribution (i.e., the distribution of shortest-
 197 path lengths) is often used to assess GOF for larger networks, but this is clearly
 198 less relevant in the case of small networks (like in our case, containing at most 6
 199 nodes if directed, or 8 if undirected) because the shortest-path length between
 200 any two nodes typically lies between 1 and 2 steps. Therefore, we focus the
 201 GOF analysis on the parameters fit in the model as the bare-minimum of local
 202 graph statistics, as shown in Figure 3. An important difference in our approach
 203 compared to traditional GOF assessments for ERGMs is that we are able to
 204 enumerate the full support of the model, and so instead of showing a boxplot we
 205 present a 90% exact confidence interval per-statistic per network, comparing the
 206 fitted model’s distribution with the observed parameters. A detailed discussion
 207 of this aspect of the *ERGMitos* is presented at the end of this paper.

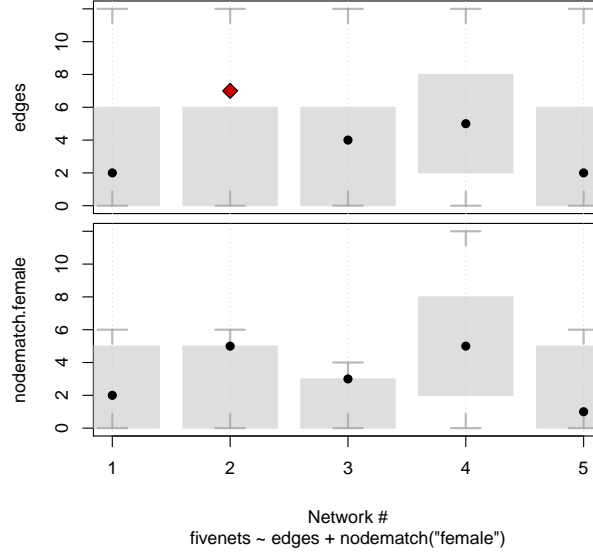
208 An important advantage of the *ERGMitos* over “regular” ERGMs is that
 209 we can observe the surface of the log-likelihood over different combinations
 210 of parameters in a rather straightforward way. This, together with the GOF
 211 analysis should be a routine step done after every *ERGMito* fit. Figure 4 shows
 212 the surface of the log-likelihood function around the solution parameters to the
 213 maximization problem.

214 The ability to calculate the surface of the exact likelihood function provides
 215 additional tools for assessing the quality of the estimated set of parameters. One
 216 good use of this diagnostic is to evaluate the roughness of the log-likelihood
 217 function, which in principle should give us an idea of the likelihood of the
 218 maximization process failing to reach a global maxima, or estimates being close
 219 to problematic (e.g., potentially degenerate) areas of the parameter space.

220 **5. Simulation study**

221 We conducted two sets of simulations studies in which we compare the per-
 222 formance of MLE with MC-MLE in terms of bias, power, and type I error rates.

Figure 3: Goodness-of-fit in ERGMito. This illustrates how the observed sufficient statistics of each one of the 5 networks (x-axis) locate in the overall estimated distribution based on the fitted ERGMito. The gray lines in each box show the minimum and maximum value that the sufficient statistics can take in each one of the 5 networks, whereas the dotted lines provide a 90% confidence interval. The dots are the observed statistics in each network.



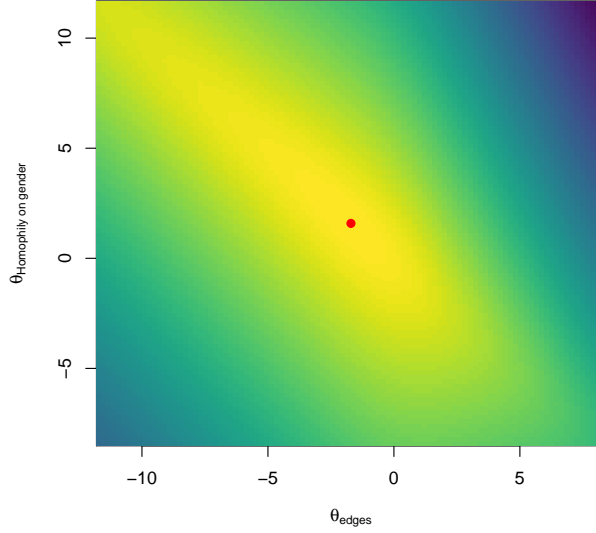
In the first set, we analyze empirical bias and empirical power of each estimator in an scenario in which the ERGM is defined by *edgcounts* and *transitive triads*. For the second set of simulations, we look at empirical type I error rates when ERGMs are misspecified by including a *transitive triad* term in the context of a data-generating-process that only includes an *edgcount* statistic.

The code used to reproduce this entire section can be found at BLIND REVIEW.

5.1. Empirical Bias and Power

Using the ERGMito R package, we generated 20,000 samples (datasets) consisting of several small networks defined by the *edges*, *edgcount*, and *ttriads*, number of transitive triads, parameters. Each sample was generated using different combinations of parameters. While all come from an ERGM model defined by *edgcounts* and number of transitive triads, for every sample we specified:

Figure 4: Surface of the log-likelihood function of the pooled ERGMito model. Lighter colors represent higher values while darker ones represent lower values. The red dot corresponds to the location of the MLE estimate of the model.



236 (1) population parameters for the ERGM, (2) size of the sample (number of
 237 networks in the sample), and (3) composition of the sample in terms of the
 238 combination of networks of size four and five. A detailed description of each
 239 one of these three components used to draw the samples follows:

- 240 1. **Population parameters** First we drew two numbers from a piece-wise
 241 Uniform distribution with values in $[-2, -.1] \cup [.1, 2]$, $(\theta_{edges}, \theta_{ttriads})$,
 242 which corresponded to the parameters associated to the statistics **edge-**
 243 **count** and **number of transitive triads**. This specifies the ERGM from
 244 which we will be drawing the networks from. This is akin to the approach
 245 taken by [22], although we took a more conservative approach than their
 246 ranges of $(-5, 0)$ and $(0, 5)$ for the parameters “edges” and “triangles” in
 247 order to increase the number of non-degenerate samples, this is, samples
 248 composed mostly of either empty or fully connected graphs.
- 249 2. **Number of networks per sample** Then, we specified the number of

networks to generate from the model defined in the previous step, using one of the following sample sizes $\{5, 10, 30, 50, 100, 150, 200, 300\}$. The 20,000 simulations were equally split across the various sample sizes (e.g., the overall simulation contained 2,500 samples with 5 networks, 2,500 samples with 10 networks, etc.)

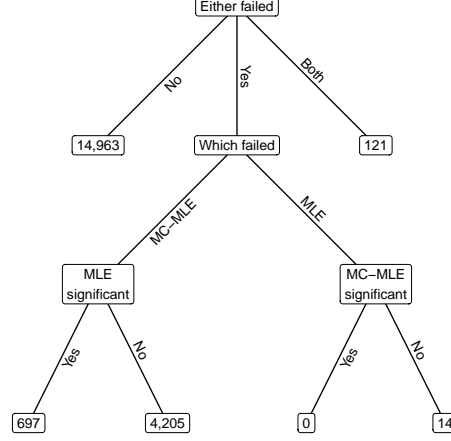
3. Number of nodes per network Finally, the composition of each sample in terms of the number of vertices that each network has was uniformly-random selected from the pairs $\{N, 0\}, \{N - 1, 1\}, \dots, \{1, N - 1\}, \{0, N\}$, where the first number of each pair is the number of networks of size 4, and the second is the number of networks of size 5 in the sample, for example, if the sample size selected in the previous step was 30, then the possible pairs to select from would be $\{30, 0\}, \{29, 1\}, \dots, \{1, 29\}, \{0, 30\}$, so samples in which all networks were of size 4 (meaning we draw the pair $\{30, 0\}$) or size 5 (again, selecting the pair $\{0, 30\}$) were equally likely.

For each one of the 20,000 simulated datasets, we then estimated the model using both MLE—as implemented in the *ERGMito* R package [30]—and MC-MLE—using *statnet*’s *ergm* R package [9, 14]. In the case of the latter, the pooled estimation was done by fitting what is known in the literature as a block-diagonal model in which (a) networks are stacked together in a single adjacency matrix, and (b) the sampling space for the MCMC process is constrained to sample from graphs where ties are only possible within blocks. We set the MCMC control parameters *interval* and *samplesize* to 2,048 (double the default values) to increase precision of the parameter estimates.

5.1.1. Empirical Bias and Power: Samples in which either of MC-MLE or MLE returned with an error

Of the models fit to the 20,000 datasets (each comprised of samples of small networks), in 14,963 cases (roughly a 75%) the models were successfully fit using both the MLE and MC-MLE estimators. In 121 cases, both implementations failed with an error; for the remaining cases (4916) the MC-MLE (ERGM) failed in most of them (4905 of 4916), while implementation of the MLE (ERGMito)

Figure 5: Distribution of the failed estimations. A failed estimation attempt refers to the case in which the implementation of the algorithm (ERGM for the MC-MLE, and ERGMito for the MLE method), returned with an error.

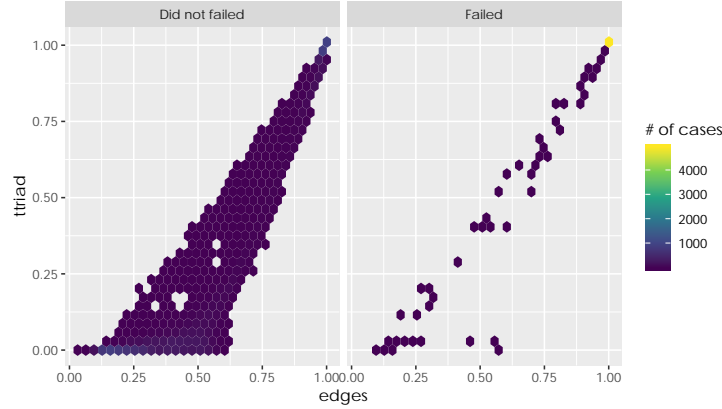


estimator failed in only 14 of these 4916 cases. Of the 4905 cases in which the MC-MLE failed, MLE was able to detect a significant effect in both parameters in 697 cases (i.e., 14 percent of these cases). Conversely, of the 14 cases in which the MLE failed, MC-MLE was not able to detect a significant effect in any of these cases. [Figure 5](#) summarizes the successful and failed models described above.

To gain insight into the problems causing the failed models, we considered the distribution of the sufficient statistics. These indicate that most of the problems arose in situations where the sufficient statistics were near to the boundary of the support. This is illustrated in [Figure 6](#), and shows that in most of the cases in which both methods failed, the sufficient statistics were mostly saturated, meaning they reached the maximum possible value because the cases sampled were largely composed of fully connected networks. Given errors arose when the set of sufficient statistics was close to saturation, both ERGM and ERGMito should consider those cases and test for possible failures during the estimation process.

We also considered if the rate of model failure was greater when the dataset contained a smaller sized sample of networks. [Table 2](#) shows the error rates by sample size for both implementation methods, and it is clearly higher for

Figure 6: Distribution of the average sufficient statistics per sample. Since samples can contain networks of sizes four and five, we have re-scaled the sufficient statistics counts by each network size’s corresponding maximum value so these range from zero to one.



298 smaller samples (particularly less than 30 networks), and this is especially true
 299 for MC-MLE.

300 Moreover, if the optimization algorithms are able to return anything, both
 301 ERGM and ERGMito try to compute variance-covariance matrices using the
 302 model’s log-likelihood, this being approximated in the case of MC-MLE. It is
 303 important to note that in the case that the Hessian matrix is not p.s.d (positive-
 304 semi-definite), both ERGM and ERGMito use the Moore-Penrose generalized
 305 inverse algorithm as implemented in the R package MASS [32]. For more on the
 306 interpretation of variance-covariance matrices when the Hessian is not p.s.d.,
 307 see [7]. Details on the evaluation of parameter estimates in ERGMito can be
 308 found in [Appendix B](#).

309 Finally, the reader must bear in mind that unexpected errors during the
 310 estimation process, despite being associated with properties of each estimation
 311 method, are issues that can be captured by the program, and thus, be managed
 312 in a more informative way so that the user has a better understanding regarding
 313 the possible causes of the problem.⁵

⁵We will be submitting bug reports to the ergm package’s issue tracker.

Sample size	P(Error)	
	MC-MLE	MLE
5	0.35	0.00
10	0.32	0.03
30	0.27	0.01
50	0.25	0.01
100	0.23	0.00
150	0.20	0.00
200	0.20	0.00
300	0.18	0.00

Table 2: Failure probability (software implementation) by sample size.

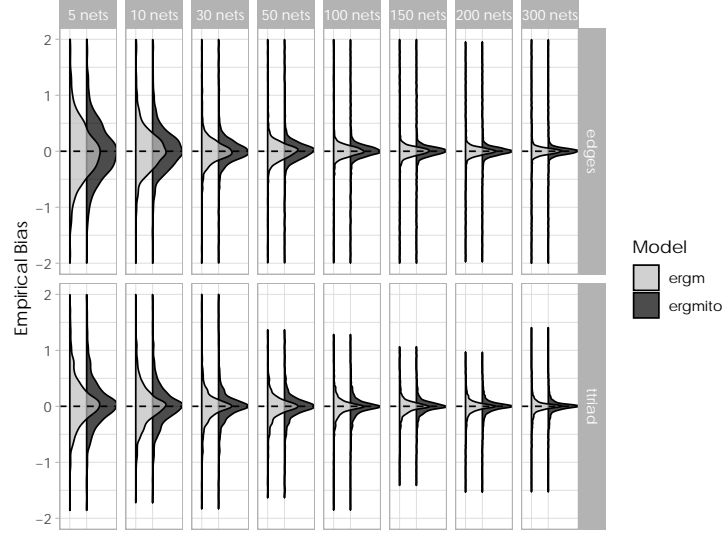
314 5.1.2. Empirical Bias and Power: Samples fitted successfully

315 Out of the 20,000 model fitting attempts on the simulated datasets, about
316 25% of those fit with either MLE or MC-MLE failed to return an estimate.
317 The results that follow only include those samples in which both MLE and MC-
318 MLE methods returned without an error. As shown in Figure 7, both estimation
319 methods behaved similarly in terms of empirical bias in the models studied here.
320 As the size of the sample of networks in the dataset increased (i.e., when there
321 were more networks within the sample), the empirical bias of both MC-MLE
322 and MLE decreased, as expected.

323 As shown in Figure 8, both MC-MLE and MLE behave very similarly with
324 empirical power levels increasing as the size of the sample of networks increases
325 in the dataset size (the x-axis of each sub-figure), and as the effect size increases
326 (the y-axis of each sub-figure). Moreover, we compared power levels using two-
327 sample proportion tests and found no significant differences between the power
328 levels of MC-MLE and MLE. It is interesting to note that, in the case of ef-
329 fect sizes of magnitude [0.5, 1.0), the discovery rate for the `ttriad` parameter
330 reaches nearby 0.75 for sample sizes between 30 to 50 networks, which is a rather
331 common sample size in the study of small networks such as teams, families, and
332 sometimes ego-networks.

333 Figure 9 shows the effect of the composition of the sample in each dataset, in
334 terms of the proportion of networks of size five (vs. size four), and the number

Figure 7: Empirical distribution of the bias per model parameter, for MC-MLE and MLE estimation methods. In general we see that the parameter estimates' bias is centered around zero and both MC-MLE (ERGM) and MLE (ERGMito) have about the same bias in our simulation study.

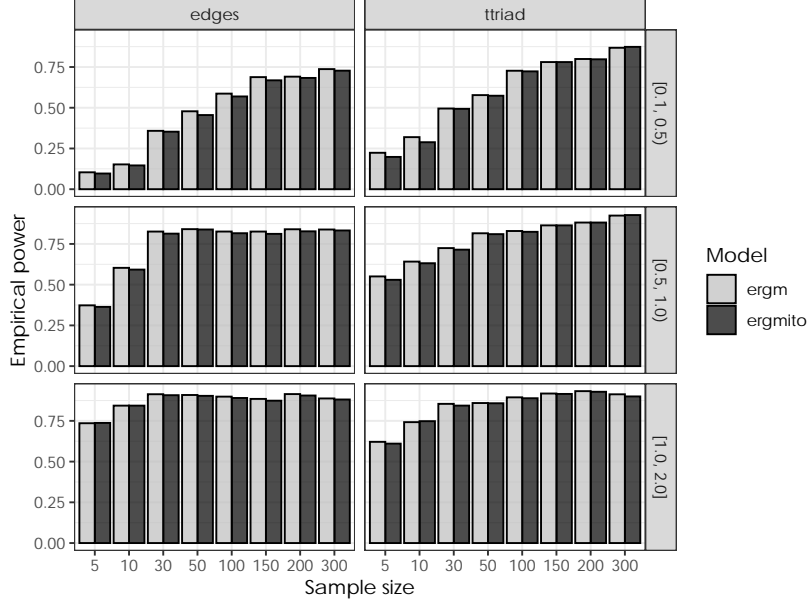


335 of networks per dataset. In this case, we observe no meaningful patterns that
 336 would indicate the dataset composition is related to power.

337 One notable difference between the two estimation methods was in compu-
 338 tational time: using MLE makes the estimation process significantly faster. As
 339 shown in Figure 10, MLE (ERGMito) is orders of magnitude faster than MC-
 340 MLE (ERGM). Therefore, while both estimators show very similar properties
 341 in terms of power and bias, practitioners will benefit by using MLE (ERGMito)
 342 when modeling small networks because it may substantially reduce the estima-
 343 tion time.⁶

⁶While this is mostly true, there are some scenarios in which the speed gains may not be as dramatic as it has been showed here. The biggest computational bottleneck that ERGMito faces is the calculation of the full support of the sufficient statistics. In the case of structure-only statistics, ERGMito, and actually *ergm*, computes the full distribution very quickly, but, as the model starts to become more complex, such calculation becomes more expensive.

Figure 8: Empirical power by dataset size and effect size (the later considering only magnitude), for ERGM and ERGMito estimation methods. Power increases for both MC-MLE (ERGM) and MLE (ERGMito) with increases in the size of the dataset and effect size. There are indistinguishable differences in power between the two estimation methods.

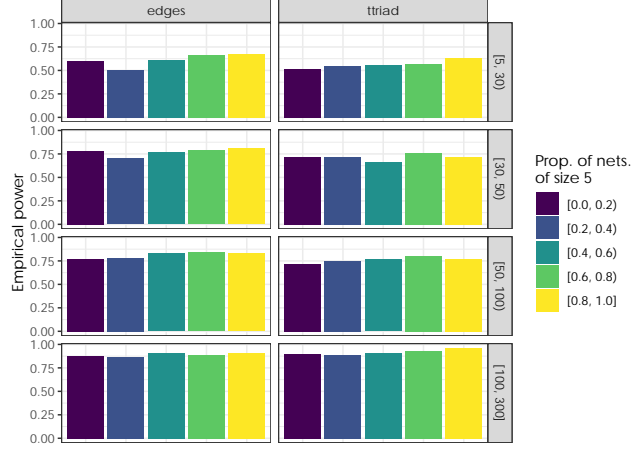


344 5.2. Type I error rates

345 Using the same procedure described in subsection 5.1, we simulated 17,500
 346 datasets comprised of Bernoulli networks (i.e., an ERGM model only defined
 347 by the *edgcounts* sufficient statistic). In this case, we drew different sets of
 348 sample sizes: for each of $\{5, 10, 15, 20, 30, 50, 100\}$ we generated 2,500 datasets
 349 using the Bernoulli model with *edgcount* parameter uniformly distributed in
 350 the range $[-2, -0.1] \cup [0.1, 2]$. We then estimated the models using MC-MLE
 351 and MLE, as implemented in the *ergm* and *ergmito* R packages respectively,
 352 and calculated the type I error rates, using a misspecified model, this is, fitting
 353 ERGMs including a *transitive triads* count statistic.

354 As shown in table Table 3, when models are fit to datasets with samples sizes
 355 of 15 or less networks, the MLE estimates had smaller type I error rates, that
 356 were much closer to the nominal 5% level, especially in small sample sizes. With
 357 datasets that had samples sizes of 20 or more networks, there was no statistically

Figure 9: Empirical power by proportion of networks of size five in the sample (color coded) and sample size (rows).



significant difference in type 1 error rates between the two methods.

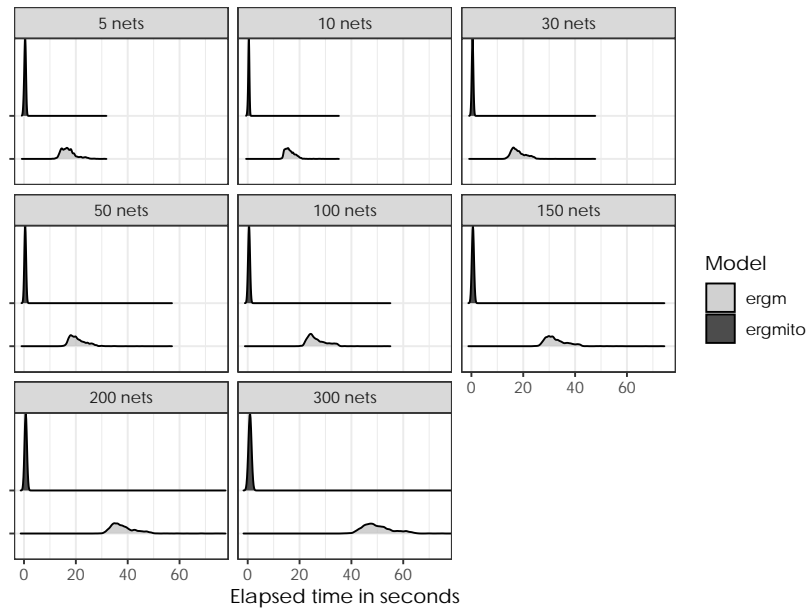
Sample size	N. Simulations	P(Type I error)		chi2
		MC-MLE	MLE	
5	2,189	0.084	0.057	11.71 ***
10	2,330	0.070	0.045	12.46 ***
15	2,395	0.084	0.066	5.55 *
20	2,430	0.074	0.060	3.58
30	2,460	0.057	0.052	0.67
50	2,495	0.046	0.044	0.17
100	2,499	0.048	0.048	0.00

Table 3: Empirical Type I error rates. The χ^2 statistic is from a 2-sample test for equality of proportions, and the significance levels are given by *** $p < 0.001$, ** $p < 0.01$, and * $p < 0.05$. The lack of fitted samples in some levels is due to failure of the estimation method.

6. Discussion

In this paper we revisit and extend Exponential Family Random Graph Models (ERGM) for the case of small networks. Given the interest in testing hypotheses about small networks in the literature, but limited application of statistical models to small network data [6, 11, 21, 26, 33, and others], we proposed

Figure 10: Distribution of elapsed time (in seconds) for the estimation process for MC-MLE (ERGM) versus MLE (using ERGMito). Overall, the MLE implementation is orders of magnitude faster compared to the time required by the MC-MLE implementation to do the parameter estimation.



364 a special case of ERGMs for small networks, called *ERGMitos*. An appeal-
 365 ing feature of *ERGMitos* is that they are estimated using MLE directly, rather
 366 than estimation via MCMC methods, as is traditionally done with ERGMs
 367 for larger networks. This approach provides a couple of important benefits
 368 for small network data: (1) it avoids at least partially, if not completely, the
 369 model near-degeneracy problems that affects the MC-MLE methods; and (2)
 370 its use of pooled estimates is valuable when datasets are comprised of small
 371 networks, as is often the case with research on multiple families, small teams, or
 372 ego-networks. Our simulation study finds that *ERGMitos* are easily estimated,
 373 and that this approach can improve accuracy and efficiency for modeling small
 374 networks compared to MC-MLE estimation approaches implemented in other
 375 ERGM packages.

376 Another major benefit of using the full likelihood directly – when feasible –
 377 is that it allows researchers tremendous flexibility in terms of constructing and
 378 estimating new models. In terms of estimation, for example, the ability to eas-
 379 ily calculate the likelihood allows researchers to make use of standard tools and
 380 techniques for ML estimation and MCMC estimation. This is important, be-
 381 cause current techniques for intractable models (e.g., auxiliary variable MCMC,
 382 MC-MLE, and noise-contrastive estimation), while effective, are not necessarily
 383 straightforward to implement for non-experts. This places a high barrier to en-
 384 try for researchers who would like to develop statistical models that go beyond
 385 the standard packages.

386 As a simple example, take recent work on multilevel network models [24]. In
 387 that work, constructing a multilevel Bayesian model of a sample of networks –
 388 while conceptually straightforward – required the development of custom code
 389 and algorithms to implement. By contrast, having the full likelihood available
 390 in R means that the same models studied in that paper (assuming small-N net-
 391 works) can be constructed and estimated as easily as any other non-intractable
 392 model for which we can calculate the likelihood, using the full range of tra-
 393 ditional tools and algorithms for ML and Bayesian estimation. This frees re-
 394 searchers from focusing only on models that are implementable in packages like

statnet, and allows greater freedom to think about ways that models for graphs can be modified and incorporated into other statistical models. In addition, being able to estimate gradients opens the possibility of estimating models using modern Bayesian algorithms like Hamiltonian Monte Carlo (HMC) and stochastic gradient langevin dynamics (SGLD), which may offer advantages in terms of speed or scalability, respectively.

The development and evaluation of *ERGMitos* in this simulation study also brings up topics for future work. One is the evaluation of model goodness-of-fit, and identifying statistics that are most important to evaluate with small networks, and that are reasonable to expect in a model that would suggest a “good fit”. Because *ERGMitos* enable a rather simple way of conducting simulation studies (relative to traditional ERGMs), this will facilitate this work in future. Another topic to explore in future work is the value of *ERGMitos* for estimating ERGMs for very large networks, by drawing *samples* of local network structures from a large graph. There is ongoing work extending ERGMs to very large networks [27, 28], and *ERGMitos* may be a valuable and *efficient* approach for fitting these pooled models to a large sample (e.g., in the order of the thousands) of small local network structures, with 5 or 6 nodes, drawn from a large network.

Overall, *ERGMitos* provide a promising extension to the ERGM framework for the analysis of small social networks, which can generate a richer understanding of the local social processes that give rise to the formation of networks in small social groups, and processes that are the local social building blocks of larger social structures.

7. Acknowledgements

BLIND REVIEW

References

- [1] E. Bott and E.B. Spillius. *Family and Social Network*. Families & Marriage. Routledge, 2001. ISBN: 9780415264174. URL: <https://books.google.com/books?id=XVwMqHb56TsC>.

- 425 [2] Carter T. Butts. *sna: Tools for Social Network Analysis*. R package version
426 2.4. 2016. URL: <https://CRAN.R-project.org/package=sna>.
- 427 [3] Dorothy R. Carter et al. "Social network approaches to leadership: An
428 integrative conceptual review." In: *Journal of Applied Psychology* 100.3
429 (May 2015), pp. 597–622. ISSN: 1939-1854. DOI: [10.1037/a0038922](https://doi.org/10.1037/a0038922). URL:
430 <http://doi.apa.org/getdoi.cfm?doi=10.1037/a0038922>.
- 431 [4] N. Crossley et al. *Social Network Analysis for Ego-Nets: Social Network
432 Analysis for Actor-Centred Networks*. SAGE Publications, 2015. ISBN:
433 9781473927339. URL: <https://books.google.com/books?id=v0IdCAAQBAJ>.
- 434 [5] Dirk Eddelbuettel and Romain François. "Rcpp: Seamless R and C++
435 Integration". In: *Journal of Statistical Software* 40.8 (2011), pp. 1–18. DOI:
436 [10.18637/jss.v040.i08](https://doi.org/10.18637/jss.v040.i08). URL: <http://www.jstatsoft.org/v40/i08/>.
- 437 [6] O Frank and David Strauss. "Markov graphs". In: *Journal of the American
438 Statistical Association* 81.395 (1986), pp. 832–842. DOI: [10.2307/2289017](https://doi.org/10.2307/2289017).
439 URL: [http://amstat.tandfonline.com/doi/abs/10.1080/01621459.](http://amstat.tandfonline.com/doi/abs/10.1080/01621459.1986.10478342)
440 [1986.10478342](http://amstat.tandfonline.com/doi/abs/10.1080/01621459.1986.10478342).
- 441 [7] Jeff Gill and Gary King. "What to Do When Your Hessian is Not In-
442 vertible: Alternatives to Model Respecification in Nonlinear Estimation".
443 In: *Sociological Methods & Research* 33.1 (2004), pp. 54–87. DOI: [10.1177/](https://doi.org/10.1177/0049124103262681)
444 [0049124103262681](https://doi.org/10.1177/0049124103262681). URL: <https://doi.org/10.1177/0049124103262681>.
- 445 [8] Mark S. Handcock. "Assessing Degeneracy in Statistical Models of So-
446 cial Networks". In: *Working Paper No. 39* 76.39 (2003), pp. 33–50. ISSN:
447 1936900X. DOI: [10.1.1.81.5086](https://doi.org/10.1.1.81.5086). URL: [http://citeseerx.ist.psu.](http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.81.5086)
448 [edu/viewdoc/summary?doi=10.1.1.81.5086](http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.81.5086).
- 449 [9] Mark S. Handcock et al. *ergm: Fit, Simulate and Diagnose Exponential-
450 Family Models for Networks*. R package version 3.9.4. The Statnet Project
451 (<http://www.statnet.org>). 2018. URL: [https://CRAN.R-project.org/](https://CRAN.R-project.org/package=ergm)
452 [package=ergm](https://CRAN.R-project.org/package=ergm).
- 453 [10] Kaisa Henttonen. "Exploring social networks on the team level-A review
454 of the empirical literature". In: *Journal of Engineering and Technology
455 Management* 27.1 (2010), pp. 74–109. ISSN: 0923-4748. DOI: [https://](https://doi.org/10.1016/j.jengtecman.2010.03.005)
456 doi.org/10.1016/j.jengtecman.2010.03.005.
- 457 [11] Paul W. Holland and Samuel Leinhardt. "An exponential family of prob-
458 ability distributions for directed graphs". In: *Journal of the American
459 Statistical Association* 76.373 (1981), pp. 33–50. DOI: [10.2307/2287037](https://doi.org/10.2307/2287037).
- 460 [12] David R Hunter, Steven M Goodreau, and Mark S Handcock. "Goodness
461 of Fit of Social Network Models". In: *Journal of the American Statistical
462 Association* 103.481 (2008), pp. 248–258. DOI: [10.1198/016214507000000446](https://doi.org/10.1198/016214507000000446).
463 eprint: <https://doi.org/10.1198/016214507000000446>. URL: [https://](https://doi.org/10.1198/016214507000000446)
464 doi.org/10.1198/016214507000000446.

- [13] David R. Hunter, Pavel N. Krivitsky, and Michael Schweinberger. “Computational Statistical Methods for Social Network Models”. In: *Journal of Computational and Graphical Statistics* 21.4 (2012). PMID: 23828720, pp. 856–882. DOI: [10.1080/10618600.2012.732921](https://doi.org/10.1080/10618600.2012.732921). eprint: <https://doi.org/10.1080/10618600.2012.732921>. URL: <https://doi.org/10.1080/10618600.2012.732921>.
- [14] David R. Hunter et al. “ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks”. In: *Journal of Statistical Software* 24.3 (2008), pp. 1–29.
- [15] Pavel N. Krivitsky, Mark S. Handcock, and Martina Morris. “Adjusting for network size and composition effects in exponential-family random graph models”. In: *Statistical Methodology* 8.4 (July 2011), pp. 319–339. ISSN: 15723127. DOI: [10.1016/j.stamet.2011.01.005](https://doi.org/10.1016/j.stamet.2011.01.005). URL: <https://linkinghub.elsevier.com/retrieve/pii/S1572312711000086>.
- [16] Philip Leifeld. “texreg: Conversion of Statistical Model Output in R to L^AT_EX and HTML Tables”. In: *Journal of Statistical Software* 55.8 (2013), p. 24. URL: <http://www.jstatsoft.org/v55/i08/>.
- [17] D. Lusher, J. Koskinen, and G. Robins. *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications*. Structural Analysis in the Social Sciences. Cambridge University Press, 2012. ISBN: 9781139851039.
- [18] J. Møller et al. “An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants”. In: *Biometrika* 93.2 (June 2006), pp. 451–458. ISSN: 0006-3444. DOI: [10.1093/biomet/93.2.451](https://doi.org/10.1093/biomet/93.2.451). eprint: <http://oup.prod.sis.lan/biomet/article-pdf/93/2/451/590012/932451.pdf>. URL: <https://doi.org/10.1093/biomet/93.2.451>.
- [19] Jaewoo Park and Murali Haran. “Bayesian Inference in the Presence of Intractable Normalizing Functions”. In: *Journal of the American Statistical Association* 113.523 (2018), pp. 1372–1390. DOI: [10.1080/01621459.2018.1448824](https://doi.org/10.1080/01621459.2018.1448824). eprint: <https://doi.org/10.1080/01621459.2018.1448824>. URL: <https://doi.org/10.1080/01621459.2018.1448824>.
- [20] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2018. URL: <https://www.R-project.org/>.
- [21] Garry Robins et al. “An introduction to exponential random graph (p*) models for social networks”. In: *Social Networks* 29.2 (2007), pp. 173–191. DOI: [10.1016/j.socnet.2006.08.002](https://doi.org/10.1016/j.socnet.2006.08.002).
- [22] Michael Schweinberger and Mark S. Handcock. “Local dependence in random graph models: characterization, properties and statistical inference”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 77.3 (June 2015), pp. 647–676. ISSN: 13697412. DOI: [10.1111/rssb.12081](https://doi.org/10.1111/rssb.12081). URL: <http://doi.wiley.com/10.1111/rssb.12081>.

- [23] Cosma Rohilla Shalizi and Alessandro Rinaldo. “Consistency under sampling of exponential random graph models”. In: *Ann. Statist.* 41.2 (Apr. 2013), pp. 508–535. DOI: [10.1214/12-AOS1044](https://doi.org/10.1214/12-AOS1044). URL: <https://doi.org/10.1214/12-AOS1044>.
- [24] Andrew J Slaughter and Laura M Koehly. “Multilevel models for social networks: hierarchical Bayesian approaches to exponential random graph modeling”. In: *Social Networks* 44 (2016), pp. 334–345. DOI: [10.1016/j.socnet.2015.11.002](https://doi.org/10.1016/j.socnet.2015.11.002).
- [25] Tom A. B. Snijders. “Statistical Models for Social Networks”. In: *Annual Review of Sociology* 37.1 (2011), pp. 131–153. DOI: [10.1146/annurev.soc.012809.102709](https://doi.org/10.1146/annurev.soc.012809.102709).
- [26] Tom A B Snijders et al. “New specifications for exponential random graph models”. In: *Sociological Methodology* 36.1 (Dec. 2006), pp. 99–153. ISSN: 0081-1750. DOI: [10.1111/j.1467-9531.2006.00176.x](https://doi.org/10.1111/j.1467-9531.2006.00176.x). URL: <http://www.jstor.org/stable/25046693><http://smx.sagepub.com/lookup/doi/10.1111/j.1467-9531.2006.00176.x>.
- [27] Alex D. Stivala et al. “Snowball sampling for estimating exponential random graph models for large networks”. In: *Social Networks* 47 (2016), pp. 167–188. ISSN: 0378-8733. DOI: <https://doi.org/10.1016/j.socnet.2015.11.003>. URL: <http://www.sciencedirect.com/science/article/pii/S0378873315000878>.
- [28] Alex Stivala, Garry Robins, and Alessandro Lomi. *Exponential random graph model parameter estimation for very large directed networks*. 2019. arXiv: [1904.08063](https://arxiv.org/abs/1904.08063) [stat.ME].
- [29] David Strauss. “On a General Class of Models for Interaction”. In: *SIAM Review* 28.4 (Dec. 1986), pp. 513–527. ISSN: 0036-1445. DOI: [10.1137/1028156](https://doi.org/10.1137/1028156). URL: <http://epubs.siam.org/doi/10.1137/1028156>.
- [30] George Vega Yon. *ergmito: Exponential Random Graph Models for Small Networks*. R package version 0.1-0. 2018. URL: <https://github.com/muriteams/ergmito>.
- [31] George Vega Yon and Paul Marjoram. “slurmR: A lightweight wrapper for HPC with Slurm”. In: *The Journal of Open Source Software* 4.39 (July 2019). DOI: [10.21105/joss.01493](https://doi.org/10.21105/joss.01493). URL: <https://doi.org/10.21105/joss.01493>.
- [32] W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. Fourth. ISBN 0-387-95457-0. New York: Springer, 2002. URL: <http://www.stats.ox.ac.uk/pub/MASS4>.
- [33] Stanley Wasserman and Philippa Pattison. “Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p*”. In: *Psychometrika* 61.3 (1996), pp. 401–425. DOI: [10.1007/BF02294547](https://doi.org/10.1007/BF02294547).
- [34] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. ISBN: 978-3-319-24277-4. URL: <http://ggplot2.org>.

551 [35] Achim Zeileis and Torsten Hothorn. “Diagnostic Checking in Regression
552 Relationships”. In: *R News* 2.3 (2002), pp. 7–10. URL: [https://CRAN.R-](https://CRAN.R-project.org/doc/Rnews/)
553 [project.org/doc/Rnews/](https://CRAN.R-project.org/doc/Rnews/).

554 Appendix A. MLE

555 The estimation process of *ERGMITOS* (as pooled models of small networks)
556 is done entirely on R using the `ergmito` R package. While a significant amount
557 of the implementation of the methods described here was done using `Rcpp` [5],
558 a core component of the package is based on `statnet`’s `ergm` R package, and in
559 particular, in the function `ergm.allstats` which does exhaustive enumeration
560 of statistics in a compact way. In general, the estimation process for any list of
561 networks is as follows:

- 562 1. Analyze the model to be estimated: Extract the networks from the left-
563 hand-side as specified in the `ergm` package, and calculate the exact statis-
564 tics using the `ergm.allstats` function.
- 565 2. With the full enumeration of statistics, build the joint likelihood function
566 of the model in a compact form (i.e., using the weights instead of the full
567 enumeration of the support of the model). This improves speed when it
568 comes to evaluating the log-likelihood function.
- 569 3. Because we are dealing with exact statistics, it is also possible to calculate
570 the exact gradient function. We compute the gradient as follows:

$$\sum_p \nabla l_p(\theta) = s(y_p, X_p)^t - \frac{Q_p^t (W_p^t \circ \exp\{Q_p \theta\})}{\kappa_p} \quad (\text{A.1})$$

571 Where $s(y, X)$ is a vector of observed sufficient statistics (usually called
572 target statistics), Q is a matrix of sufficient statistics, in particular, the
573 isomorphic sufficient statistics associated with the model, and W is a
574 vector of frequency weights.

575 These first three steps carry the most part of the computing time.

- 576 4. Finally, the joint log-likelihood is maximized using the BFGS algorithm
577 implemented in the `optim` function in the `stats` package.

578 The final set of estimates is analyzed separately by another program included
579 in the package. The next section describes the evaluation steps followed by
580 *ERGMITO*.

581 **Appendix B. Evaluation of estimates**

582 After the optimization procedure finalizes, the *ergmito* package performs a
583 series of tests checking the quality of the estimates. In particular, we conduct
584 the following evaluations after every call to the main optimization function:

- 585 1. Since the BFGS algorithm, as implemented in the *optim* function from the
586 stats R package, only works with real numbers, we check whether the log-
587 likelihood function increases with changes in the attained value in cases
588 when the theoretical maxima lies in $\pm\infty$. To do this, we increase the
589 magnitude of each estimate by 1.5 and check the value of the log-likelihood
590 function with that change. If an increase in that value is observed, we
591 assume that the maxima for the given parameter equals $\text{sign}(\hat{\theta}_i) \times \infty$.
- 592 2. If all parameters turn out to be $\pm\infty$ after this check, the function will send
593 a warning message to the user and the function returns without computing
594 the variance-covariance matrix.
- 595 3. If, on the other hand, a fraction of the parameters were switched to $\pm\infty$,
596 the function recalculates the Hessian and the log-likelihood using the value
597 $\text{sign}(\hat{\theta}_i) \times 10^5$. This is done instead of using ∞ , because in most cases using
598 infinite will result in the function being undefined. Again, the function
599 will warn users about this issue.
- 600 4. Finally, using the current state of the Hessian matrix, the function will
601 attempt to compute the variance-covariance matrix by inverting the neg-
602 ative of the Hessian. If an error is caught, then the generalized inverse is
603 returned instead [7].

604 The possible return codes are:

605 **00** *optim* converged, no issues reported.

606 **01** optim converged, but the Hessian is not p.s.d.

607 **10** optim did not converged, but the estimates look OK.

608 **11** optim did not converged, and the Hessian is not p.s.d.

609 **20** A subset of the parameters estimates was replaced with +/-Inf.

610 **21** A subset of the parameters estimates was replaced with +/-Inf, and the
611 Hessian matrix is not p.s.d.

612 **30** All parameters went to +/-Inf suggesting that the MLE may not exists.