UNIVERSITY OF CALIFORNIA, LOS ANGELES   **UCLA**

BERKELEY • DAVIS • IRVINE • LOS ANGELES • RIVERSIDE • SAN DIEGO • SAN FRANCISCO   SANTA BARBARA • SANTA CRUZ

DUNCAN A. CLARK
DEPARTMENT OF STATISTICS
8125 MATHEMATICAL SCIENCES BUILDING
LOS ANGELES, CALIFORNIA 90095-1554

August 30, 2021

Jouni Kuha
Joint Editor
Journal of the Royal Statistical Society Series A.

Dear Professor Kuha,

I have submitted via the web-portal revisions to the manuscript titled "Comparing the Real-World Performance of Exponential-family Random Graph Models and Latent Order Logistic Models" for consideration for the *Journal of the Royal Statistical Society Series A*.

We thank you for the thoughtful reviews by yourself, the Associate Editor and two reviewers. Their overriding theme was that the paper presentation and writing needed to improved. We have used the reviews to rewrite the paper, addressing each comment individually. We believe the paper is much stronger for this effort. In particular, the rewrite has dramatically improved the structure and writing of the paper. The coherence of the presented analyses and their descriptions is greatly improved.

Our detailed responses to the reviews are below, following comments of reviewers in *italic* text. I also detail the corresponding revisions made to the paper itself.

Thank you for your time and consideration.

Sincerely,

Duncan A. Clark

<u>Joint Editor's Comments:</u>

The reviewers have recommended some major revisions to your manuscript. Therefore, I am inviting you to respond to their comments and to resubmit your manuscript accordingly.

Many of my comments overlap with ones from the referees and the Associate Editor. In the revision, you should of course address all of their comments, including the ones not touched upon here.

The motivation of several of my comments is that for this journal the paper should be, as far as possible, accessible also to interested readers who are not experts in network analysis. This means, for example, explaining some things more carefully than you would in a specialist journal and avoiding or explaining network analysis jargon where necessary. When preparing the revision, please consider the text also from this point of view, even beyond my specific suggestions below.

1. *As noted also by the reviewers, the text is currently a mess and in need of thorough copyediting. This applies to the main text (including careless editing, ungrammatical expressions, bad punctuation, undefined acronyms etc.), tables and figures, and bibliography. Some of my comments mention instances of this, but they are not meant to be comprehensive.*

   We have extensively rewritten all of the text, especially explaining jargon and improving the quality of the text overall. These changes are detailed in the responses below and those to the Associate Editor and reviewers.

2. *When revising the paper, please aim to do this without increasing its length. The revised paper should be no longer than 30 pages in the current formatting.*

   The revised text is now 25 pages long. The body of the text has decreased in length, though the network descriptions have increased in length and have been moved to a 9 page supplement.

3. *Sections 2 and 3 are currently somewhat fragmentary. I would suggest that you combine them into one section on the definitions and theory of these models and make it more self-contained and coherent.*

   We have complete rewritten Sections 2 and 3, combined them and provided more explanation of the LOLOG model, together with a brief ERGM introduction. In the interest of brevity, the ERGM section is significantly shorter.

   For example, this could include the following:
   (a) Describe both LOLOG models and ERGMs in roughly the same way, in matching notation and covering the same topics (with perhaps some more detail on LOLOG, but not taking ERGM as known to the extent you do now). This will make it easier to follow the discussion of the comparisons between them.

   This has been done (Section 2.1)

DUNCAN A. CLARK · DEPARTMENT OF STATISTICS · 8125 MATHEMATICAL SCIENCES BUILDING

(b) In this discussion, clearly separate model specification, interpretation, estimation, explicit comparisons between the two models, and other such topics.

This has been done explicitly in Sections 2.1 – 2.5.

(c) Please give some more explanation of the logic of LOLOG. At first reading it seems confusing to see a time variable t and references to time ordering, given that most network data have no such observed ordering. If I understand this correctly, this is a hypothetical (latent) device which is used define the model, and the distribution of the observed network is regarded as an average over all the possible orderings. This is clear enough on second reading of the paper, but saying it even more clearly would be useful, especially since the paper would serve as an introduction to LOLOG to many readers.

We have rewritten the logic for LOLOG starting in paragraph 3 of Section 2.1, emphasizing the latent time dimension interpretation.

(d) Give some more explanation (and examples) of what the graph statistics g may be. Are all the covariates in your examples (and more generally) of this kind?

We have done this in paragraphs 1 and 2 of Section 2.1. We have also cited Morris et al 2008, which is a paper length treatment of the types of graph statistics. All the covariates/graph statistics in our examples are of the kind typically used (and described in the cited Morris et al (2008).

(e) Say something about the interpretation of the parameters theta in both models. Without this, it is harder to understand what to make of the examples that come later, and the comparisons between them.

This is now done in the new Section 2.2

(f) p 5: "… the expectations and standard errors of the MLE and MOM parameter distributions do not exist either" – so where do the estimated standard errors in your results come from?

Whilst our statement is true theoretically, it perhaps does not add to the main arguments of the paper. We have removed this paragraph and simply state that we report approximate standard errors derived from MCMC-estimated Fisher information matrices.

(g) I would suggest that you explain the goodness of fit method that you use (now on p11) already in this section, giving also some more explanation of what it does.

We included the goodness of fit in Section 2.5 of the newly combined Sections 2 and 3.

4. *The formatting and labelling of the tables and figures should be improved:*

Duncan A. Clark · Department of Statistics · 8125 Mathematical Sciences Building

(a) Make the captions more self-contained and self-explanatory.
(b) Explain what the numbers shown in the tables are.
(c) Put the explanation of the significance stars in a footnote of a table rather than caption.
(d) Remove the underscores in "University_2005" etc.

We have updated the tables and figures as suggested. Captions now describe and explain results contained in the tables.

(e) In the figures, overrule the ggplot2 defaults on labels of the plots and legend texts.

We have amended the plots so as not to rely on colour and, overidden defaults to match the fonts in the body.

5. *Please give better explanation of the details of the example in Section 5.*

(a) Please give a much clearer explanation of what the results actually mean, in terms of substantive research questions about interactions of people in offices. How are the estimated coefficients interpreted, for LOLOG and ERGM? How do these interpretations compare with each other? You now start to say something about this in Section 5.3, but this is still obscure and a little too late.

We have added a description of the terms in the model in in Section 4.1, where they are first used. All terms we use are described in detail in the cited Morris et al (2008). We have kept the descriptions short for brevity. We also include a summary of the original paper's (ERGM) qualitative conclusions in Section 4.

We have rewritten Section 4.3 to says how the interpretations compare, although we consider goodness of fit first so that we can assess if the fits are valid to interpret.

(b) Where appropriate, use less unexplained network analysis language; e.g. the "nodes" can be "workers" or "persons", and "edges" and "dyads" could be "social interactions" or something like that.

We have done this throughout the paper.

(c) Explain what the covariates (GWESP, in and out stars, but also team match, metric distance etc.) mean.

This has been done in the introduction to Section 4 in the context of the analysis of the first case-study of networks of daily social interactions and elsewhere the first time the terms appear.

(d) p 10: "Table 5 shows the fitted model where the nodes are added in the order of their average usefulness, as reported by the other nodes". What does this mean? Does it mean that you used some non-uniform choice for p(s) for these estimates?

DUNCAN A. CLARK · DEPARTMENT OF STATISTICS · 8125 MATHEMATICAL SCIENCES BUILDING

We have removed the results for the Office layout network using edge ordering by usefulness, and reduced the commentary on this issue. While this is of interest it does not address the primary point of the paper.

Answering the direct question: Yes, we used the non-uniform choice for *p(s)* that placed probability 1 on the usefulness ordering. While this is extreme, such fits can help allow the sensitivity to the choice of *p(s)* to be assessed.

(e) Give a clearer explanation of the goodness of fit procedure (in combination with its general explanation, see 2(g) above). What does "the goodness of fit on the in-, outdegree and ESP distributions" (p 11) mean?

We have rewritten the explanation and included it earlier (in Sections 4.2).

(f) The six goodness of fit plots take up quite a lot of space in the main text. Perhaps consider putting some of them in the appendix.

We have moved the ESP and out-degree plots to a new appendix

(g) p 16-17: The language of "performance in fit" and "simulation" is a slightly confusing way to refer to estimation times. Here it will also help if the methods of estimation for both models have been more explicitly explained before (see 2(a) above).

The methods of estimation are now described earlier (in Section 2.3). We have amended the computational comparison wording to remove performance and refer explicitly to time (The end of Section 4.1).

6. *Please summarise briefly in the main text what criteria you used to decide on the questions in Table 8. In particular, discuss this for question 8.*

We decided that tabular form was to restrictive to discuss the questions. We have removed Table 8 and placed its contents at the top of Section 5. This enabled us to expand on the questions/criteria and provide overall motivation at the beginning of the section. After each question in the list we describe the criteria we used to include the question. This was a very helpful suggestion.

7. *For each of the examples in the Appendix, please give also a brief summary (couple of sentences) of the substantive research questions and conclusions, in the language of the application. The current descriptions are a bit lifeless, and give the impression that the purpose of social network analysis is to fit network models rather than to use them to answer substantive questions.*

We have added more detailed description to the summaries. We focussed on the substantive questions and conclusions in the language of the relevant application.

DUNCAN A. CLARK · DEPARTMENT OF STATISTICS · 8125 MATHEMATICAL SCIENCES BUILDING

8. *Please provide example code for fitting the two kinds of models. This is particularly useful in a comparative paper like this. It can be included as supplementary materials or even (if short and simple) in the appendix. Also, in the "Access to data and computer code" statement on the manuscript submission system (see also the instructions to the authors) provide links to the seven examples for which the data are publicly available (as mentioned on p 6).*

   The complete set of code for the analysis of each member of the ensemble is quite substantial. In many cases multiple models were fit and goodness-of-fit assessed. We have spent considerable effort in developed a GitHub site with the publicly available data and an archive of code to model it. We believe this is an important contribution of the paper to the readers as they can benefit from our experience. It is linked here and described at the end of the discussion section. The site is likely more accessible than an appendix or supplementary materials. If the journal has a preferred method of making these resources available, we would be happy to use them.

9. *Other comments:*

   (a) *In the title, I suggest you add "...for social network analysis", or words to that effect.*

   We have amended the title as suggested.

   (b) *"dyad" and "dyad independence" are never really defined*

   We have defined "dyad" when the networks are defined and removed the term "dyad independence", replacing it with direct text.

   (c) *p 4, l 13: what is s? [element of script-s]*

   In our overall rewrite, this is defined above equation (3). It is an edge order permutation.

   (d) *p 4: what is MOM? PMF?*

   These have now been defined.

   (e) *p 5 l 25: "heavy lifting" is a bit too casual*

   This text has been removed in the rewrite.

   (f) *Include the references to the studies in Table 1. You can then delete Table 10. Also, include the references in the subsection headings of the appendix.*

   We have amended the text and tables as suggested. We have removed the covariates column to accommodate the citations.

   (g) *p 8, l 39: This paragraph (which is about relative advantages of LOLOG vs. ERGM rather than construction of your ensemble) does not seem to belong to this place in the text.*

DUNCAN A. CLARK · DEPARTMENT OF STATISTICS · 8125 MATHEMATICAL SCIENCES BUILDING

We have placed a modified version of this paragraph into the results section.

(h) p 10: "when substituting the GWESP term for a triangle term which is not possible with ERGM" – I think this says the opposite of what you mean it to say.

We have mended the sentence for clarity.

(i) p 16: "role", not "roll"

This has been fixed.

(j) "[!h]" in Table 8, "?? (Add)" on p 21, "?? (Dis)" on p 24, the first two entries in the References – please see comment 1 above.

These broken references have been corrected.

(k) Table 9: State also in the table (caption or column heading) that 1-8 refer to the numbers in Table 8.

This has been corrected in the updated Section. 5.

(l) p 19: Delete the lonely subsection number and label for 6.1.

This has been removed.

Reviewers' Comments to Author

Associate Editor:

Comments to the Author:

This paper performs a comparative analysis between the more commonly used ERGM models and the LOGLOG models on a large set of networks. While the objective is formulated well and the study clearly justified, the structure and the writing of the paper should be significantly improved by:

We have carefully rewritten the paper to improve the clarity and structure. These changes are detailed in the responses below and those to the other reviewers.

- *Correcting the many many typos, unusual formulations and wrong inclusion of references in the text (many instances needing \citep rather than \cite)*

We have rewritten the text to improve the clarity and correct the typos.

- *Making a careful selection of the results: many tables and figures are included and not really commented on or thoroughly explained. Make sure that the captions of the selected tables and figures are informative.*

  We have rewritten the captions for all figures and tables, expanding them. We have moved many goodness-of-fit figures to Appendix A. We have added explanations for the included statistical summaries.

- *Table 9: I personally could not follow the results presented in this table. I would also like to see more quantitative summaries of the overall results on the ensemble to see whether there is support for the claims put forward by the authors in the comparison between the methods.*

  We embedded Table 9 (now Table 7) in the new Section 5, to make it clearer. We have now included a list of the column headings with more detailed explanation, as well as providing a column proportion. We have also removed Table 8 and placed its contents at the top of Section 5. This enabled us to expand on the questions/criteria and provide overall motivation for the original Table 9 at the beginning of the section. After each question in the list we describe the criteria we used to include the question.

- *Generally clarifying, and where possible *quantifying*, the comparison between the two methods based on the evidence gathered from the empirical analysis. If needed here, perhaps as a final section, you could also include additional examples – not included in your ensemble, but adding insight into the comparison between the two methods - such as those mentioned by one of the reviewers*

  We have added references to the relevant columns in the results table in the text, which points to the proportion of the networks in the ensemble that satisfy the relevant question. We would like to include an additional example but are constrained by the length of the manuscript.

  We have also keen to find ways to quantify the comparison. A possible tool are the goodness-of-fit metrics of Hunter, Goodreau and Handcock (2008). However, scientists that model social network data typically have multiple objectives with some models more suited to some of those objectives rather than others. Our constructed rubric allows an assessment of the models that is closer to scientist's criterion. We have added a brief discussion of this at the top of Section 5 and expanded on the comparative nature of the criteria used in (the new) Table 7.

Referee 1

Comments to the Author

Because my comments are quite long, I have attached them as a text file. But in summary, there is really only one point that I think needs addressing: a potential misinterpretation of GWDEG parameter in the discussion of the models for the Sailer & McCulloh networks, as described in the attachment.

The other points raised in my "general comments" section may (or may not) be of interest to the authors and

inform a revision (if any is felt necessary). And the "minor points" section at the end is just a few typographical errors etc. which I noticed while reading the manuscript.

Comments from Referee 1's text file:

In the following, citations are to references in the manuscript, with any additional references not in the manuscript at the end of this report.

Summary:

This manuscript compares the well-known exponential-family random graph model (ERGM) with the recently proposed latent order logistic (LOLOG) model (Fellows, 2018b) on a (substantial: 35 end up included in the final analysis) set of networks with ERGM models published in the Social Networks journal. The ERGMs are reproduced (where possible) and then equivalent, and sometimes improved (in terms of goodness-of-fit) LOLOG models, are also estimated.

A short description of the LOLOG model and its relation to ERGM is provided, briefly recapitulating the original descriptions in Fellows (2018b).

The case of the networks from one particular paper (Sailer & McCulloh 2012) is explored in detail, while the rest are summarized in a table with some details in the Appendix.

It is shown that, generally, the LOLOG is qualitatively in agreement with the ERGM. Often the LOLOG is able to fit with simpler terms such as star and triangle, which usually result in neardegeneracy in the ERGM, requiring the "alternating" or "geometrically weighted" model terms instead. Further, the LOLOG tends to be easier (and faster) to fit, requiring less "tuning" than ERGM (where, as mentioned, simple terms like triangles often lead to degeneracy, and decay parameters for geometrically-weighted terms need to be manually tuned, or estimated by curved ERGMs, the latter often being very slow or hard to converge).

The ERGMs are estimated with the well-known statnet ergm R package, and the LOLOG models with the R package lolog provided by the the creator of the model (Fellows, 2018a).

General comments:

The manuscript is easy to read and clearly is the result of significant effort in collecting and reestimating (in ERGM and LOLOG) models for the ensemble of networks. It makes a significant and novel contribution in taking the "data-centric" perspective (given the data, what modeling approach is best?) rather than the usual perspective of a "methodology" paper (given our model, what data can we fit with it?), showing the usefulness of the LOLOG model for situations in which ERGM would be the conventional choice. Another interesting contribution this manuscript makes is the suggestion (p.6 lines 30-35) of a reason as to why LOLOG does not suffer a similar degeneracy problem to ERGM. This was not clear (at least to me) from the Fellows (2018b) paper (unlike for example Fellows & Handcock (2017) where phase transitions are explicitly removed by the "tapered" Gibbs measure).

The argument that this population of networks is biased towards ERGM and hence provides stronger evidence for the value of LOLOG models (a conservative indicator that LOLOG is useful) is a good one,

DUNCAN A. CLARK · DEPARTMENT OF STATISTICS · 8125 MATHEMATICAL SCIENCES BUILDING

given the the "file drawer" effect of only models with good ERGM fits likely to published in Social Networks, and the occasional (or perhaps frequent) difficulty of estimating ERGM models leading researchers to omit terms they might have wanted or considering fewer models. This problem has been discussed also by (for example) Martin (2020) [and his statistics textbook he cites therein].

We agree, as this logic was a primary motivator to consider the ensemble we did.

The only weakness in this argument is if we consider the possibility that LOLOG is a model that works well exactly when ERGM works well (and indeed theoretically this is not so ridiculous, given they are equivalent in the dyad independent case). In which case, this sample is not quite such a conservative indicator of the utility of LOLOG.

That said, we know that this is not the case, given Fellows (2018b) describes LOLOG as "complementary to the popular ERGM framework", and states that there are simple LOLOG models where an ERGM requires prohibitively many terms, and also vice versa. Something like this is also shown in this manuscript, with not all networks being able to be fit better by LOLOG than ERGM, and varying results as to which model class (including nonzero counts for "Both" and "Neither") is "more useful subjectively".

Although I had read the Fellows (2018b) preprint, and was aware of at least one use of the LOLOG model which was useful in a case where an ERGM was not able to be fit, I had not experimented with the LOLOG model myself, until reading this manuscript as a reviewer. This motivated me to try the lolog R package (Fellows 2018a) on some networks which are difficult to fit with ERGM.

The very first example I tried did indeed demonstrate the usefulness of LOLOG: the E. coli transcriptional regulation network (Salgado et al, 2001; Shen-Orr et al, 2002) available in the ergm package (network ecoli1 in data(ecoli)) and for which simple models (inherently directed network converted to undirected; no triangles or other transitivity parameter) are estimated in Saul & Filkov (2007); Hummel et al. (2012). Estimating ERGMs for this network left in its original state as directed, and especially if triangles or alternating or geometrically-weighted versions of transitivity are used, had apparently proved impossible with statnet or PNet/MPNet, and the only such ERGM I am aware of is in Stivala & Lomi (2020), requiring the use of more recently published methods (Byshkin et al., 2018; Stivala, Robins, & Lomi, 2020), which are based on a version of "persistent contrastive divergence" and do not have the provable statistical properties of MoM or MCMLE for example. I was able to easily obtain a LOLOG model of this network including the twoPath and triangles terms (not geometrically-weighted) and ingwdegree and out-gwedegree (both with decay 0.2 arbitrarily), showing qualitative agreement with the (not exactly the same) models in Stivala & Lomi (2020), with similar goodness-of-fit. Importantly, however, the LOLOG was able to estimate a model with the triangles term, which ERGM could not (requiring the "alternating" transitive triangles statistic), facilitiating interpretation in this context as a "motif" (although "motifs" are usually considered induced subgraphs, so the relevant test is GoF on the 030T triad, which is good). I was also easily able to estimate a similar LOLOG model for the other network (yeast transcriptional regulation) in that preprint.

We find this to be very encouraging, though due to lack of space and scope of our study do not comment on this aspect. The enhanced "motif context" interpretation provided by the ability to fit triangle terms to large networks, may be very useful in future applied studies.

DUNCAN A. CLARK · DEPARTMENT OF STATISTICS · 8125 MATHEMATICAL SCIENCES BUILDING

Although excluded from consideration from the network ensemble due to the time period chosen (1979 - 2016) and only journal Social Networks, there has been progress in new methods for ERGM estimation for very large networks, and it is now possible to estimate ERGM parameters for a network with over one million nodes, e.g. the ERGM for an online social network estimated in Stivala, Robins, & Lomi (2020). Albeit the ability to estimate ERGM parameters for such large networks relies on the methods discussed above, which achieve this speed and scale precisely by avoiding the expensive simulation to find equilibrium of Markov chain for ERGM, and hence there is as yet no goodness-of-fit test possible for models of such large networks.

By taking advantage of local dependence and latent block structure, hierarchical ERGMs have also been estimated for a network with over 10 000 nodes (Babkin et al., 2020). The latter method is implemented in the R package hergm (Schweinberger & Luna, 2018) and allows the conventional simulation-based goodness-of-fit tests.

These advances in ERGM estimation may somewhat qualify the claim (p. 19, lines 48-50) regarding LOLOG "increasing the maximum size of networks that can be feasibly analysed." Although perhaps LOLOG, with its advantage over ERGM of the ease of simulation from the model, is a better way forward here. It might be interesting to see how well it can scale in practise.

The ability of the LOLOG for large models is perhaps not its primary utility. The point we were trying to make is that applied researchers likely will find it easier to fit a LOLOG to a say 10,000 nodes network, than the corresponding ERGM, due to the lack of degeneracy and speed of estimation. We have amended the sentence to reference Stivala, Robins, & Lomi (2020), and make this point clearer.

Another aspect of LOLOG, which is used in the online social network example (the one related to hamsters and their owners) in Fellows (2018b), but not used in this manuscript except in one example in the Appendix (due to the selection criteria of the network ensemble) is the edge ordering mechanism. Although it may indeed be "at odds with some researchers' philosophical beliefs" (p. 20 line 31) about networks, the ability to explicitly model edge ordering (and preferential attachment) could be highly relevant (and not so at odds with beliefs about social networks) in modeling citation networks, for example, where ERGMs are sometimes used, and have to include some mechanism to account for citation temporal direction (e.g. McLevey et al., 2018). It seems to me that LOLOGs could be highly useful in such cases, where there definitely is an edge ordering in the data.

This is a very important point we did not emphasise in the paper. We have now rewritten the last paragraph in the discussion to reflect these points (and be less negative on edge order process modeling). We have also cited McLevey et al., 2018, if that is acceptable.

Major (i.e. substantive) points [there is only one]:

In the second-last paragraph (lines 42-49) on p. 16, concerning geometrically-weighted degree (GWDEG) term being negative and significant in an ERGM, it says this is "at odds with the LOLOG model". However if I understand correctly, a LOLOG model with a GWDEG was never fit, so this refers to a comparison with a LOLOG model containing the in and out 2 and 3 star terms described in the previous paragraph (lines 28-40). The out-2-star term was found to be positive and significant, the the out-3-star term negative and significant. This is interpreted as a "tendency for super daily interactors". (No similar tendency was found significant for corresponding in-(2,3)-stars).

Duncan A. Clark · Department of Statistics · 8125 Mathematical Sciences Building

This finding (positive 2-star and negative 3-star) is a fairly typical result in ERGMs (I am assuming the same applies to LOLOG, although perhaps without, at least to the same extent, near-degeneracy problems as a motivation for geometrically-weighted statistics) indicating a tendency towards centralization ("super daily interactors" here), and indeed the alternation of the signs as higher order terms are added (edge, 2-star, 3-star,...) is a motivation for the alternatingk-[in|out]-stars and gw[io]degree parameters (see e.g. pp. 65-67 of Koskinen & Daraganova, 2013).

However the interpretation of GWDEG (gw[io]degree) in statnet is (confusingly: see Levy (2016); Levy et al. (2016)) that a negative parameter estimate indicates a tendency towards centralization, and positive a tendency against centralization (Hunter, 2007). This is opposite (i.e. signs swapped) to the interpretation of the alt-k-star parameter - and so in fact this would seem to be consistent with the LOLOG out-(2,3)-star interpretation above. Which is reassuring (rather than "concerning"), although the point that the LOLOG allows a better fitting model (or at least more easily to estimate a model that fits well, perhaps not requiring the geometrically-weighted parameters usually required in ERGM to avoid near-degeneracy) stands.

This careful analysis is exactly correct: indeed, we have interpreted the sign as opposite of what it should be. We have amended the wording in the text and verified that the statnet ergm GWDEG terms should be interpretated as you suggest, as they include 1-degree as the first term in the sum, the 2-degree term therefore receives a negative sign, resulting in the opposite interpretation of the GWDEG parameter to the combination of 2- and 3-star parameters. We have updated the wording in the case study to this effect.

Minor points:

Some typos or missing references, easily fixed e.g.:

We have carefully edited the paper to correct these. Thank you for these improvements.

"Adolescent Health Survey." and "Pajek Datasets" at top of References (p. 27) and some corresponding '??' missing references (e.g. pp. 21,24), arXiv identifier missing from the Fellows (2018b) reference, etc.

Reference to Table 5.2 (p. 11) should presumably be to Table 7.

"Matching floors play a roll..." (p. 16) should obviously be "play a role..."

One or two LaTeX errors e.g. "[!h]" appearing before Table 8 caption.

Sailer & McCulloch (2012) should be Sailer & McCulloh (2012) [only two "c"s in McCulloh].

References (in addition to those already in the manuscript references): Babkin, S., Stewart, J., Long, X., & Schweinberger, M. (2020). Large-scale estimation of random graph models with local dependence. Computational Statistics & Data Analysis, 152, 107029.

Byshkin, M., Stivala, A., Mira, A., Robins, G., & Lomi, A. (2018). Fast maximum likelihood estimation via equilibrium expectation for large network data. Scientific reports, 8(1), 1-11.

Duncan A. Clark · Department of Statistics · 8125 Mathematical Sciences Building

Hummel, R. M., Hunter, D. R., & Handcock, M. S. (2012). Improving simulation-based algorithms for fitting ERGMs. Journal of Computational and Graphical Statistics, 21(4), 920939.

Hunter, D. R. (2007). Curved exponential family models for social networks. Social Networks, 29(2), 216-230.

Koskinen, J., & Daraganova, G. (2013). Exponential random graph model fundamentals. Ch. 3 (pp. 49-76) in Lusher D, Koskinen J, Robins G, editors. Exponential random graph models for social networks. Structural Analysis in the Social Sciences. Cambridge University Press: New York.

Martin, J. L. (2020). Comment on "Geodesic Cycle Length Distributions in Delusional and Other Social Networks". Journal of Social Structure 21(1):77-93. doi:10.21307/joss-2020-003

McLevey, J., Graham, A. V., McIlroy-Young, R., Browne, P., & Plaisance, K. S. (2018). Interdisciplinarity and insularity in the diffusion of knowledge: an analysis of disciplinary boundaries between philosophy of science and the sciences. Scientometrics, 117(1), 331-349.

Levy, M. A. (2016). gwdegree: Improving interpretation of geometrically-weighted degree estimates in exponential random graph models. Journal of Open Source Software, 1(3), 36.

Levy, M., Lubell, M., Leifeld, P., & Cranmer, S. (2016). Interpretation of GW-Degree Estimates in ERGMs. Political Networks 2016 (Conference poster). https://doi.org/10.6084/m9.figshare.3465020

Saul, Z. M., & Filkov, V. (2007). Exploring biological network structure using exponential random graph models. Bioinformatics, 23(19), 2604-2611.

Salgado et al (2001), RegulonDB (version 3.2): Transcriptional Regulation and Operon Organization in Escherichia Coli K-12,Nucleic Acids Research, 29(1):72-74.

Schweinberger, M., & Luna, P. (2018). HERGM: Hierarchical exponential-family random graph models. Journal of Statistical Software, 85(1).

Shen-Orr et al (2002), Network Motifs in the Transcriptional Regulation Network of Escerichia Coli, Nature Genetics, 31(1): 64-68. Stivala, A., & Lomi, A. (2020). Testing biological network motif significance with exponential random graph models. arXiv preprint arXiv:2001.11125.

Stivala, A., Robins, G., & Lomi, A. (2020). Exponential random graph model parameter estimation for very large directed networks. PloS one, 15(1), e0227804.

Referee 2

The authors compare the fitting of a new latent order logistic (lolog) generative graph model with the fitting of the established ERGM on the data sets from the articles published in the journal Social Networks. Even though the new model has its benefits and the goal of the study is well justified, I find the results not conviencing and not well presented.

1. In column 6 of Table 9, a large majority of marks is represented by `Yes'. However, in Section 5.2 the authors write "we note that the LOLOG model did not seem to help improve the fit for any of the networks in question" and in Section 7 "Goodness of fit of LOLOG models also compare favourably with the ERGMs, with little drop in quality".

The Section 5.2 quotation truncates the remainder of the sentence: "In particular we note that the LOLOG model did not seem to help improve the fit for any of the networks in question here, when using the same terms as the ERGM required to be non degenerate."

We have rewritten this sentence to be clearer. We do note later in Section 5.2 that including star terms in the LOLOG model, which are usually degenerate with ERGM, does improve the fit. We have also amended the Table 9 column labels to include additional explanation to be clear that the ERGM GOF is compared to the best fitting LOLOG model, that we were able to fit, which is permitted to included terms usually degenerate in ERGM, for example, triangles and stars. The primary point is that, while LOLOG are competitive with the best ERGM using terms choose n for the best ERGM, they often produce a better fit as they can use terms that an ERGM cannot. LOLOG models should increase the scope of terms that researchers use, as they can focus on their representation of the underlying social processes being less restricted by computational and class specific representation issues.

2. The shades of the boxes in Figs. 1-6 are the same, so I was not able to understand the implications of these figures and had to rely only on the text.

We have improved the figures to be interpretable without colour.

3. There are too many typos. I'll give some example but there are many more. A paper should be prepared much more carefully for a high profile journal such as JRSS.

We have completely rewritten the text for clarity. We have carefully edited the paper to improve this aspect.

p.1,l.45 "a social network a collection of fixed nodes" ->"a social networks represented by a collection of fixed nodes"? BTW, why nodes need to be fixed, in particular, if we consider dynamic or growing networks?

p.2,l.49 "inverse problem' -> "inverse problem"

p.5,l.16 Newton Raphson approach -> Newton-Raphson approach

p.6,l.15 "the the MCMC" -> "the MCMC"

p.6,l.19 "an dyad ordering" -> "a dyad ordering"

p.6,l.45 Social network journal -> {\it Social Networks} journal

p.9,l.24 "form a single published paper" -> "from a single published paper"?

Duncan A. Clark   ·   Department of Statistics   ·   8125 Mathematical Sciences Building

We have corrected the above errors.