

Subject: Manuscript Revisions for *Political Analysis*

Manuscript ID: PA-2020-075

Manuscript Title: Generative Dynamics of Supreme Court Citations: Analysis with a New Statistical Network Model

Date: August 25, 2020

We are grateful for the opportunity to revise and resubmit our manuscript. In this memo we have separated the reviewers' comments into separate criticisms and suggestions. Under each comment, we describe how we have revised the manuscript in response to the feedback provided. We generally agree with the criticisms offered, and think that our manuscript has improved substantially as a result of incorporating this feedback.

Reviewer: 1

R1.1: *In terms of estimation, it reads as if the smaller networks used MCMLE and the later the simulated annealing approach. Is that right? If so, does the simulated approach arrive at the same values for the smaller year networks as the MCMLE? Showing that might help validate the approach for the bigger/later year networks. At the very least it might mitigate my next point on the overtime changes.*

Addressed: All networks were estimated using MCMLE, however, for smaller networks we used the network's MPLE as the algorithm's starting value, while for larger networks the starting value was obtained using simulated annealing, since the MPLE as starting value did not lead to converged MCMLEs anymore. While in theory the algorithm should converge to the MLE regardless of the starting value, in practice this is not the case [1].

In order to be consistent with the used estimation approach, we estimate the MCMLE now for all networks using simulated annealing starting value. The results have not changed noticeably.

The text in the appendix has been updated accordingly.

R1.2: *For at least a few of the variables — reciprocity, gwidegree, and Justices in majority — we see a substantial change in recent periods. Does this have anything to do with the change in estimation strategy? How convinced can we be that these fairly big changes are substantively driven?*

Addressed: We have re-estimated all networks using the simulated annealing approach and the results look very similar to the initially submitted results. This is not too surprising, since the MCMLE algorithm only reports a successful estimation if it has converged to the unique MLE. However, with a poorly chosen starting value the algorithm might fail to converge towards the MLE. This is what happened for some large networks using the MPLE.

R1.3: *Relatedly, why isn't overtime change in the covariates discussed? The authors instead choose to make broad statements: "in/consistently significant" over the period of study. This seems to be a wasted opportunity and contrary to their motivation (a la Shalizi and Rinaldo 2013). If the authors want to make a statement about whether the covariate generally holds, then why not run a single model across all the years? Otherwise, wouldn't it make sense to discuss these over time changes, at least when they appear to be quite meaningful? Not to belabor the point, but the motivation for the exogenous and endogenous terms are nicely situated in the literature, but the discussion of the results are quite anticlimactic. It merely notes that some results are generally in line with expectations while others are*

not (the ideology finding being terribly interesting and surprising!). I'd like to see a more substantive engagement with the results, particularly any trends / disruptions in the series of coefficients. This has the potential to be a much bigger contribution to the judicial literature than the authors acknowledge.

Addressed: We thank the reviewer for pushing us to expand on our results; we believe the extended discussion greatly improves the paper. We changed the results section in two main ways (pp. 13-14).

First, we situated our more unexpected findings (i.e. negative outdegree count and ideological distance) in the judicial politics and legal studies literatures. For these findings, we offer explanations that comport with the field's understanding of justice behavior. The hub effect is negative because outdegree count reflects case complexity, and justices avoid citing these opinions in anticipation of them being overruled. Judges are more likely to cite ideologically distant cases because they increasingly attempt to provide more balanced opinions.

Second, we extend this discussion to account for temporal dynamics. Using these same effects, we illustrate how our approach of fitting term-by-term models allows us to better capture the complexities that arise from both the temporal context (e.g. chief justice characteristics) and from the network growth characteristic of citation networks.

R1.4: *While this is a well-executed paper (and visually beautiful), it's not clear where the contribution is meant to be. Is this paper an introduction of a new model or an application of the ERGM to citation data? If this is a new network model that would merit its own name, c-ERGM, what other networks might this model apply to? Does it really make sense to have a model that just applies to this one judicial network...? I suppose this model works for any citations — e.g., academic papers — so maybe discussing a bit of that literature and data would help. In short, the change to the ERGM is slight, even when combining that with the necessary simulated estimation approach, so telling us why these changes are so useful beyond this network could drive home the methodological contribution.*

Addressed:

R1.5: *Alternatively, one could think about this paper more in terms of its substantive contributions. The front end is deeply ingrained in the citations and judicial literatures, as are the data, of course, but little attention is given in the end to what we learn from this paper. It seems like the authors decided not to engage either frame fully, which leaves the contribution lacking on both. Having said that, I think one could go either way with this paper, and I liked a lot about it on both fronts.*

Addressed:

Reviewer: 2

R2.1: *I would really like to see discussion near the end of section 4.1 potentially expanded and/or cleaned up considerably to make some things more clear.*

Addressed:

R2.2: *Additionally, implementation details are almost wholly relegated to the appendix; while this is largely appropriate, there is not even much information in the appendix, so I would recommend that one or both of the estimation details in the main paper and the appendix are expanded. What software was used? Do the authors generate novel convenience software to handle the specifics of the cERGM model and/or the computational overhead they discuss in the appendix?*

Addressed:

R2.3: *These issues that I mention are particularly important if the authors intend for this model to be used by the judicial politics community rather than only network modelers in other contexts where the cERGM structure makes sense; a bit more practical information would be necessary I think for this to travel within the judicial politics community.*

Addressed:

R2.4: *In many places the authors do a very good job of explaining the technical side of what they are doing precisely, as well as providing easy to digest intuitive explanations. However, one place where we seem to somewhat lack both clarity as well as thoroughness and precision is unfortunately the section where it was arguably needed the most: the section on dependence terms. For example, the second paragraph of this section warns that "Unfortunately, adding the triangle, out-2-star, or in-2-star statistic causes model degeneracy," as far as I can tell, only for degeneracy as a result of the in-2-star statistic is a plan to guard against degeneracy addressed at all (surrounding equation (2)). I think more information about whether and how degeneracy resulting from the other statistics can be guarded against would be a good addition. Additionally, the authors fix the decay parameters for gwidegree and gwesp to 1 and 0.25 respectively with no discussion of the justification for those specific values; perhaps setting lambda for gwidegree to unity makes intuitive sense, but what drives the decision for gwesp? Considerably more guidance for applied researchers is needed in this section if the authors intend this model to be used by judicial politics researchers, which seems to be the case considering their discussion in the introduction. This would be a very good advance to the state of the art in citation studies in judicial politics, but the end of section 4.1 is not clear and thorough enough in my opinion to be useful to most applied judicial scholars whose work could be improved by the authors' methodological advance.*

Addressed:

R2.5: *As a small comment, in relation to the degeneracy discussion, I think footnote 4 would be best served by not being a footnote. After your discussion of degeneracy I think the typical reader will be wondering, "Did this end up being a problem for their application?" The answer is, "No, we checked that and it was fine," and I just think that as it's such a natural thing for a reader to want to know after you spent multiple full paragraphs discussing the issue, this sentence deserves its place in the main text.*

Addressed:

R2.6: *I think the manuscript would be greatly improved as well by pointing practitioners toward implementation software, even if this is a very short, simple statement. For example, if the answer for that is, "We used software X which is often used for ERGMs, and our custom calls to that software to handle the cERGM and large network we analyze can be found in the replication archive," great. If the authors have written convenience wrappers for existing software to handle these issues, or wholly novel software, even better; advertise it both for your own sake and for the sake of the substantive scholars who want to use your method because they care about modeling their data in the most appropriate way.*

Addressed:

R2.7: *I think the manuscript would be greatly improved as well by pointing practitioners toward implementation software. A point which may be useful to the authors in future work (though I'm not sure whether they find it has great utility for this methods piece): The authors find that there are not reliable effects for the ideological variables, and thus wonder, "whether, and if so, how, citations are shaped by ideological factors." I think one natural response could be that the ideological effects may actually be*

found within the dependence effects they investigate in some way, and I wonder if there are creative ways within the context of their model to test for that.

Addressed:

References

- [1] Ruth M Hummel, David R Hunter, and Mark S Handcock. “Improving simulation-based algorithms for fitting ERGMs”. In: *Journal of Computational and Graphical Statistics* 21.4 (2012), pp. 920–939.