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

Christian S. Schmid[1], Ted Hsuan Yun Chen[2], and Bruce A. Desmarais[3]

[1]*Department of Statistics, The Pennsylvania State University, University Park, PA. Email: schmid@psu.edu*
[2]*Faculty of Social Sciences, University of Helsinki, Finland. Email: ted.hsuanyun.chen@gmail.com*
[3]*Department of Political Science, The Pennsylvania State University, University Park, PA. Email: bdesmarais@psu.edu*

1   c-ERGM Estimation

The normalizing constant in Equation 1 of the main text is intractable. For example, in the simple case of adding three cases to a network in which six cases already exists—like that depicted in Figure 4 of the main text—there are 16,777,216 unique configurations of C_t that could be observed. The typical Supreme Court term involves adding hundreds of cases to a network that already includes thousands of previous cases. This means that straightforward methods of maximum likelihood estimation (MLE) are infeasible with the c-ERGM.

The common alternative relies on Monte Carlo methods to approximate the normalizing constant by simulating a large set of networks (Hunter and Handcock 2006; Hummel, Hunter, and Handcock 2012). The resulting estimator, the Monte Carlo MLE (MCMLE), is approximately consistent, meaning that it converges to the MLE as the sample size, i.e., the number of simulated networks, increases. However, one drawback is that with the number of nodes in the Supreme Court citation network being in the order of $10,000$, obtaining the MCMLE is computationally expensive (Schmid and Desmarais 2017) and the success of the algorithm heavily relies on the starting parameter vector θ_0, which is ideally chosen in the proximity of the unknown MLE (Hummel, Hunter, and Handcock 2012). The prevailing choice for θ_0 is the maximum pseudo-likelihood estimation (MPLE) (Strauss and Ikeda 1990), a fast estimation method that is defined as maximizing the log product of the conditional probability of each citation (and non-citation), conditional on the other elements of the observed citation network. The joint probability of all citations is replaced by the product over conditional probabilities, which, as we demonstrated in the main text, assume a logit form. The MPLE is simple to obtain, but does not guarantee a starting value close to the MLE (Schmid and Hunter 2020).

The MCMLE of networks up until the 90s was obtainable in a reasonable time frame starting at the MPLE and sampling $10,000$ networks to approximate the normalizing constant. However, the estimation of most networks in the 90s with the MPLE as starting values was not feasible in an reasonable time frame anymore. Instead, we improved the choice of starting value θ_0 by fixing it at the MCMLE of the previous term $t-1$ and successfully obtained the MCMLE of the network at term t. But even this approach started to fail for networks around the turn of the millennium. Neither the MPLE nor the MCMLE of previous terms as starting values led to successful estimation, and neither did the Stepping algorithm (Hummel, Hunter, and Handcock 2012). The MCMLE for these large citation networks was obtained by setting the starting value according to an novel approach introduced by Schmid and Hunter (2020). This method is based on the fact that the MLE of exponential family distributions is solely a function of the vector of sufficient statistics $h(C_t, C_{<t})$, meaning that the MLE of two networks A and B is equal if $h(A) = h(B)$. However, the MPLE of

networks with the same sufficient statistics is not necessarily the same. Instead of starting the MCMLE algorithm at the network's MPLE, Schmid and Hunter (2020) propose searching for a new network C_t^* on the same nodes as the observed network that satisfies $h(C_t, C_{<t}) = h(C_t^*, C_{<t})$, and has a weak dependence structure among unfixed ties. Such a network can be found using simulated annealing algorithms (Kirkpatrick, Gelatt, and Vecchi 1983). For networks with a weak dependence structure among unfixed ties, the MPLE is similar to the MLE, in addition, the same sufficient statistic between C_t^* and the observed network C_t guarantees the same MLE between these two networks. This makes the MPLE of C_t^* an effective starting value for the MCMLE algorithm. Since for some networks the MCMLE was only obtainable using simulated annealing method to find a starting value, the final results in the paper have all been estimated using simulated annealing. The simulated annealing algorithm for finding an improved starting value for c-ERGMs was implemented in the **cERGM**-package for **R** (R Core Team 2020) and can be found at http://github.com/schmid86/cERGM.

2 Goodness-of-Fit

We evaluate the goodness-of-fit of the model following Hunter, Goodreau, and Handcock (2008) by examining the distribution of four hyper statistics, e.g., the out- and indegree distribution and the distribution for two different edgewise shared partners statistics. OTP stands for *outgoing two-paths* and refers to the number of cases r that are cited by case i and that cite case j, while j is also directly cited by i. The second ESP statistic is the OSP specification that has been introduced in section 4.1.2 in the paper. Figure 1 visualizes the goodness-of-fit results for the citation network for the 1950 (top) and 2015 (bottom) term. The solid black line indicates the statistic's distribution in the Supreme Court citation network of that given term and the boxplots depict the statistic's distribution of 1000 networks that have been simulated from the ERGM defined by the MCMLE. This means that in the ideal case the solid black line passes through ever single boxplot.

We see that our models do a good job capturing the out and indegree distribution of the citation network, since the black line falls almost exclusively within the ranges spanned by the boxplots. For the ESP distributions we can observe that the number of ties with $r = 0$ shared partners is captured well for both the OTP as well as for the OSP statistic. However, the model overestimates the number of $r = 1$ shared partners and then, especially in the 2015 term network, underestimates the number of ties with more than $r = 1$ shared partners.

3 Checking for Model Degeneracy

A common challenge when fitting ERGMs is model degeneracy. Model degeneracy occurs when the probability distribution defined by the parameter vector does not predominantly yield networks with similar statistics as the observed network. Generally, model degeneracy results in simulated networks with no ties or all possible ties. In a non-degenerate model the statistics of the networks that were simulated from the probability distribution defined by the MCMLE fall in the proximity of the observed network's statistics. Figures 2 and 3 depict trace and density plots for the dependence terms in the 1950 and 2015 term citation network. The histograms on the left visualize a statistic's density from 1000 simulated networks, while the right side shows the statistic's trace plot of the same 1000 networks. The solid black line indicates the statistic's value in the actual citation network. Both figures indicate that this model is non-degenerate and that the simulated network's statistics fall almost evenly around the observed statistic. The density and trace plots for the ERGM of the terms not depicted provide similar results.

Supplementary Material

For supplementary material accompanying this paper, please visit https://doi.org/10.1017/pan.xxxx.xx.

References

Hummel, R. M., D. R. Hunter, and M. S. Handcock. 2012. "Improving simulation-based algorithms for fitting ERGMs." *Journal of Computational and Graphical Statistics* 21 (4): 920–939.

Hunter, D. R., S. M. Goodreau, and M. S. Handcock. 2008. "Goodness of fit of social network models." *Journal of the American Statistical Association* 103 (481): 248–258.

Hunter, D. R., and M. S. Handcock. 2006. "Inference in curved exponential family models for networks." *Journal of Computational and Graphical Statistics* 15 (3): 565–583.

Kirkpatrick, S., C. D. Gelatt, and M. P. Vecchi. 1983. "Optimization by Simulated Annealing." *Science* 220 (4598): 671–680. ISSN: 0036-8075. https://doi.org/10.1126/science.220.4598.671.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Schmid, C. S., and B. A. Desmarais. 2017. "Exponential Random Graph Models with Big Networks: Maximum Pseudolikelihood Estimation and the Parametric Bootstrap." *IEEE International Conference on Big Data, http://ieeexplore.ieee.org/document/8257919/,* 116–121.

Schmid, C. S., and D. R. Hunter. 2020. "Revisiting pseudolikelihood estimation for ERGMs." *Working Paper, The Pennsylvania State University, Department of Statistics,* no. 1.

Shalizi, C. R., and A. Rinaldo. 2013. "Consistency under sampling of exponential random graph models." *Annals of statistics* 41 (2): 508.

Strauss, D., and M. Ikeda. 1990. "Pseudolikelihood estimation for social networks." *Journal of the American Statistical Association* 85 (409): 204–212.
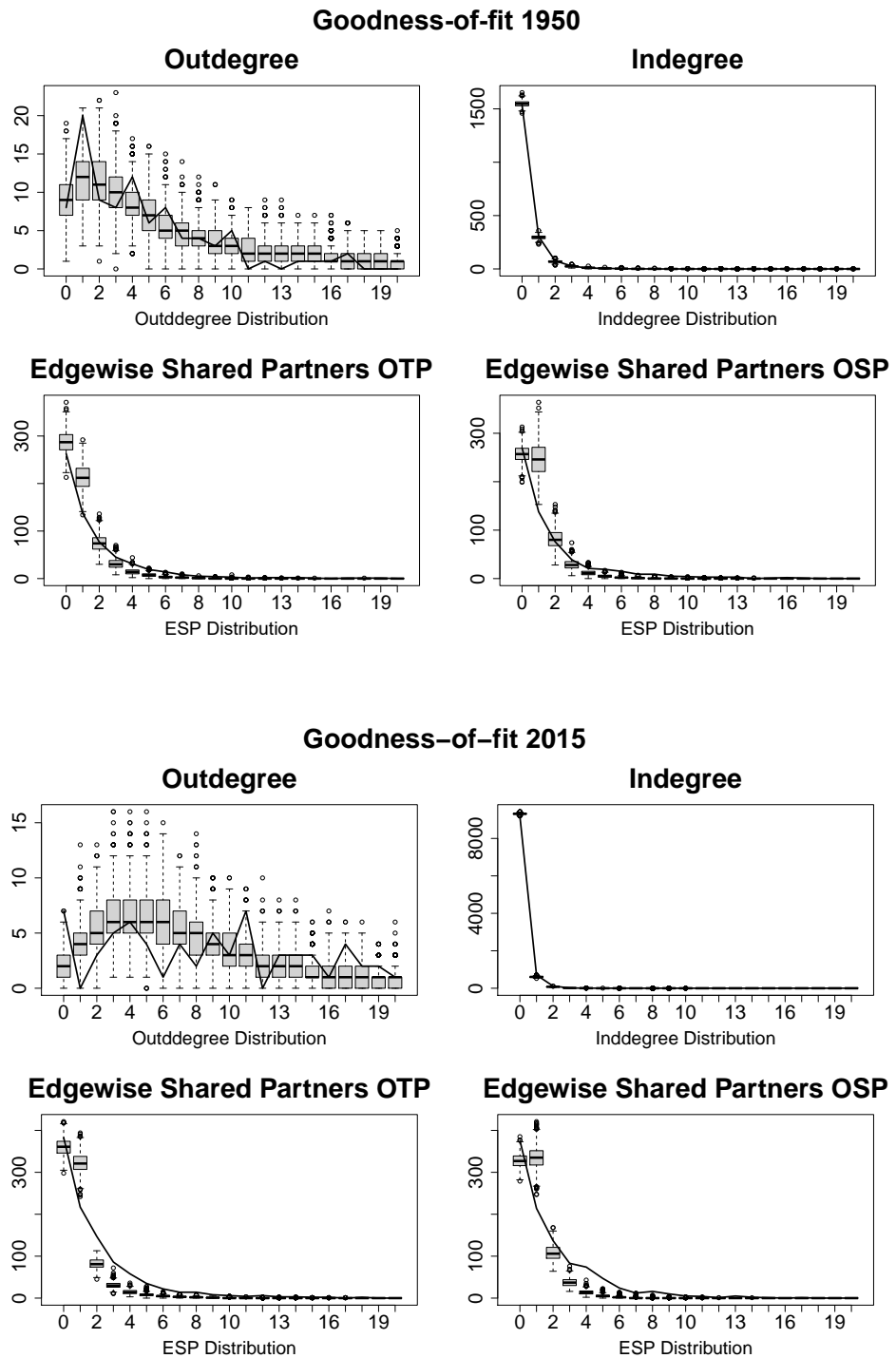
Goodness-of-fit 1950

Outdegree

Indegree

Edgewise Shared Partners OTP

Edgewise Shared Partners OSP

Goodness−of−fit 2015

Outdegree

Indegree

Edgewise Shared Partners OTP

Edgewise Shared Partners OSP

Figure 1. Goodness-of-fit diagnostic for the 1950 network (top) and the 2015 network (bottom).
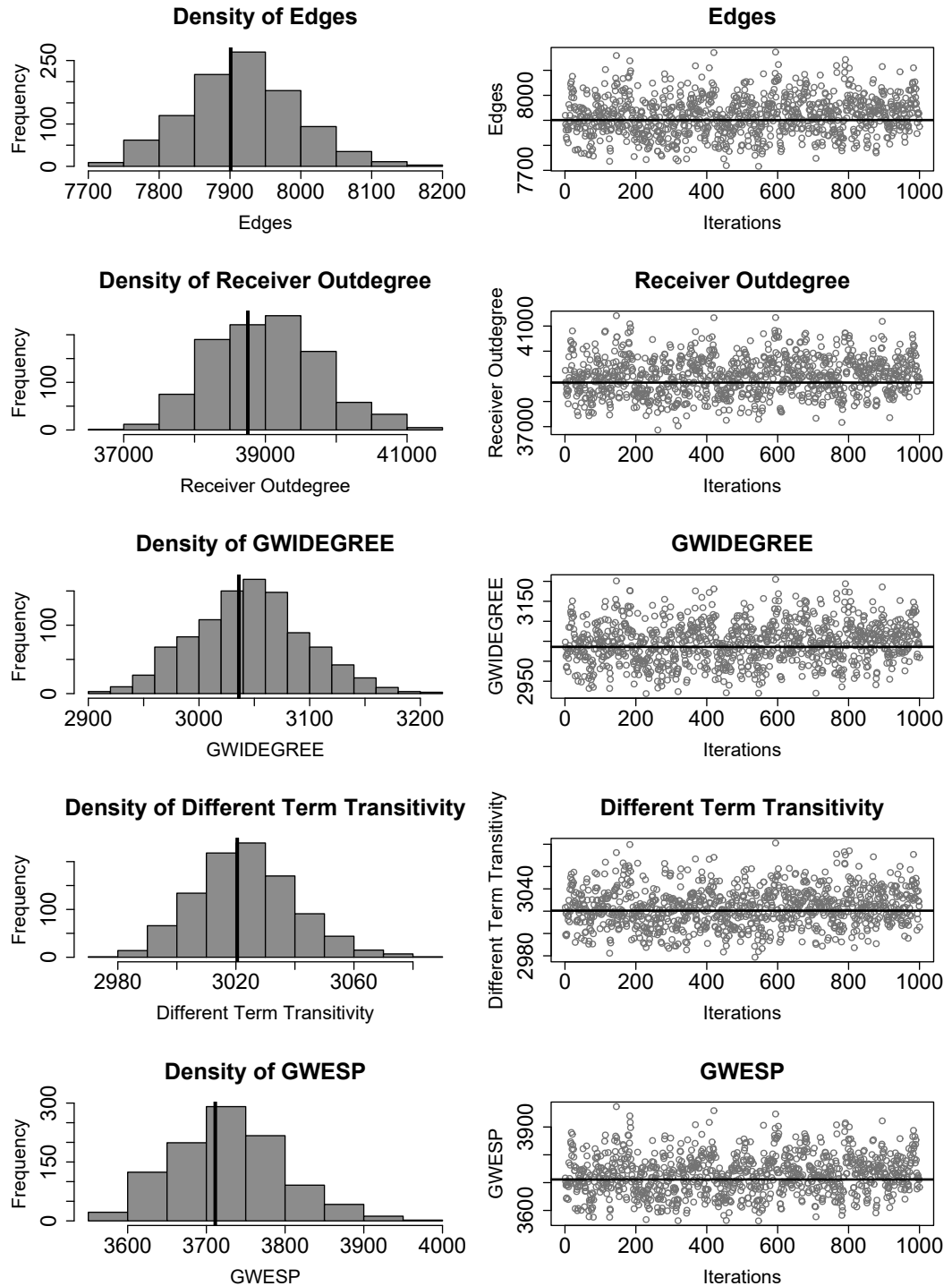
Degeneracy Check 1950

Figure 2. Density and trace plots for the dependency terms of the 1950 term citation network.
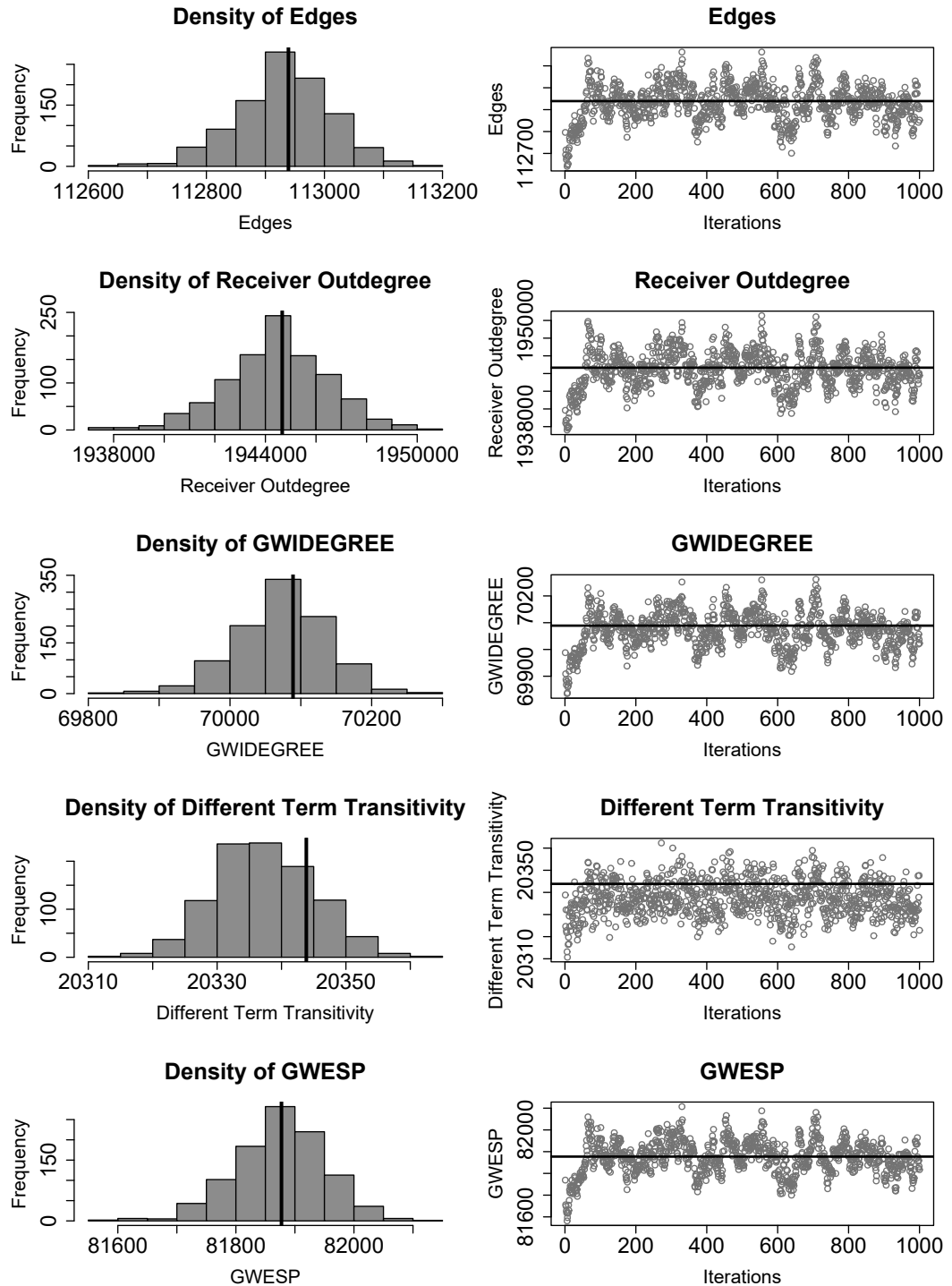
Degeneracy Check 2015

Figure 3. Density and trace plots for the dependency terms of the 2015 term citation network.