

Appendix IV

Testing for Multicollinearity in ERGMs

Table A4.1: Results from calculating the variance inflation factor (VIF) for the model terms in the local, regional, and multi-level ERGMs

Model variable (ERGM term)	VIF: Local ERGM	VIF: Regional ERGM	VIF: Multi- level ERGM
Shared partners (gwesp)	6.43	5.10	5.26
Anti-preferential attachment (gwdegree)	12.89	3.39	1.28
Within-level homophily (nodematch)	--	--	3.11
Within-country homophily (nodematch)	45.59	1.26	2.47
Level activity: local (nodefactor)	--	--	6.75
Country activity: Costa Rica (nodefactor)	5.77	1.25	2.41
Country activity: El Salvador (nodefactor)	8.12	1.13	2.83
Country activity: Guatemala (nodefactor)	5.85	1.07	2.16
Country activity: Honduras (nodefactor)	8.08	1.03	3.12
Country activity: Nicaragua (nodefactor)	4.04	1.05	1.70
Country activity: Panama (nodefactor)	3.97	1.06	1.67
Reported ties on website? (nodefactor)	13.03	5.70	7.69
Isolates	6.66	3.18	--

We calculated the VIF for each ERGM following the methodology developed by Scott Duxbury, a Ph.D. student at the Ohio State University. The formulation of the “vif-ergm” calculation is publically available on Github,¹ and is implemented as a function in the R Environment for Statistical Computing. In a paper posted to the Open Science Framework,²

Mr. Duxbury elaborates on the development of the VIF calculation for ERGMs, and discusses how to interpret the results. He determined any value above 20 is considered to be evidence of potentially problematic collinearity, and any value above 100 is representative of extreme collinearity. In the event of extreme collinearity in an ERGM, even the parameter estimates of terms that do not exhibit a large degree of correlation may be essentially random. We performed the calculation of the VIF for ERGMs exactly as described by Mr. Duxbury, with the only exception being we ran 10,000 simulations instead of using the function's default value of 1,000 simulations.

There is no evidence of extreme collinearity in the three models presented in this paper. Only in one model, specifically the local network ERGM, does a term possess a VIF greater than 20. The VIF score for within-country homophily is 45.6 for the local ERGM, and the presence of two other terms with relatively elevated VIF scores – anti-preferential attachment (12.9) and the “reported ties” variable (13.0) – are correlated with within-country homophily. It is encouraging that further testing, including re-estimating the model, indicated the parameter estimates and standard errors for these three terms were stable and, therefore, multicollinearity in the model may not be highly problematic. However, the presence of even potentially problematic levels of multicollinearity here means the model coefficients must be interpreted with caution, and in this situation it is especially important not to overlook the descriptive statistics assessing open and closed structures.

In the ERGMs for the regional and multi-level networks, all of the VIF scores for the model terms were below 20, indicating low levels of collinearity. In both models the “reported ties” variable displayed the highest VIF score, indicating this variable tends to be the most highly correlated, relatively speaking, with other variables in the model. In the regional model the “reported ties” variable (5.7) is most highly correlated with the shared partners variable (5.1), followed by the anti-preferential attachment variable (3.4). Again, this is due to the fact that actors contributing ties to the network are also more likely to be partners, and to have more partners. In the multi-level model the “reported ties” variable (7.7) is most highly correlated with the “level activity” of local actors (6.8), followed by the shared partners variable (5.3). Similar to the regional model, this is due to the fact that actors contributing ties to the network are more likely to be partners, and less likely to be local actors (who were less likely to have organizational websites and/or report collaborative ties with other actors). However, in the multi-level model there is little to no correlation between anti-preferential attachment and the “reported ties” variable.

While these overall results suggest multicollinearity is not a huge problem given the rule of thumb proposed by Mr. Duxbury, there are coefficients that rank higher than others in terms of VIF. Given this does potentially pose a problem, caution is warranted when interpreting the ERGM coefficients, especially for the suspect variables.

¹ Duxbury, S. 2018. R Function to Detect Multicollinearity in ERGM. <https://github.com/sduxbury/vif-ergm>

² Duxbury, S. 2018. Diagnosing Multicollinearity in Exponential Random Graph Models. Forthcoming in *Sociological Methods and Research*. <https://osf.io/yqdxh/>