

PENNSYLVANIA STATE UNIVERSITY

PLSC 597E

Replication Exercise: *Self-Organizing Policy Networks: Risk, Partner, Selection, and Cooperation in Estuaries*

Joshua Snoke

CITATION

Berardo, Ramiro, and John T. Scholz. "Self-Organizing Policy Networks: Risk, Partner, Selection, and Cooperation in Estuaries." *American Journal of Political Science* 54, no. 3 (2010):632-649.

DATA AND CODE

<https://github.com/jsnoke/Replication-and-Extension>

1 BACKGROUND

My replication study uses the paper "Self-Organizing Policy Networks: Risk, Partner, Selection, and Cooperation in Estuaries." by Berardo and Ramiro. This paper is of interest due to its use of two waves of longitudinal data with missingness in the second wave. The original paper did nothing beyond simple modal imputation or listwise deletion to account for this missingness.

The paper seeks to understand what attributes or network structures may cause nodes to create ties with one another. In this case, they consider water Estuaries and directed ties represent one group going to another group for advice. They also wish to model the level of trust between groups, measured at both time points. To do this they include a behavioral model to estimate different attribute effects on trust and simultaneously estimate these parameters alongside the network model. The two equations for the SAO model are as follows

$$f^{net}(x_{ij}) = \sum_k \beta_k^{net} s_{ijk}^{net}(x_{ij}) + \beta_k e_{ik}(x_i) + \beta_k a_{jk}(x_j) + \beta_k d_{ijk}(x_{ij}) + \epsilon^{net} \quad (1.1)$$

$$f^{beh}(x) = \sum_k \beta_k e_{ik}(x_i) + \beta_k a_{jk}(x_j) + \epsilon^{beh} \quad (1.2)$$

where s_{ijk} represent network attributes, e_{ik} represent ego (node forming tie) attributes, a_{jk} represent alter (node being tied with) attributes, and d_{ijk} represent similarity attributes.

I have obtained the complete data from the authors website, and have replicated most of the summary statistics and the Stochastic Actor Oriented model from the original paper. I have yet to be able to run the ERGM model or the MCMC method described due to a couple of issues with coding and missing data. The rest of the replication was successful, except there appears to be a slight discrepancy between the published result and the available data. Primarily this shows up in the second wave of the network data, where the data shows more missing nodes and different network statistics than reported.

2 REPLICATION

This study replicated the simple network summary statistics and the Stochastic Actor Oriented model estimates across two waves of networks for cooperation in Estuaries. Extension involves ERGMs and considering alternative methods to deal with missing data prevalent in both the second wave and covariate data.

2.1 NETWORK VISUALS

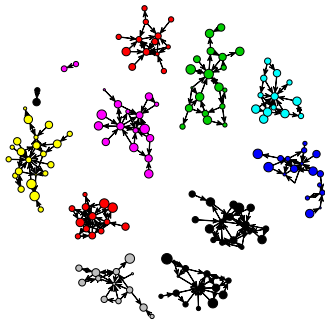


Figure 2.1: Wave 1 network

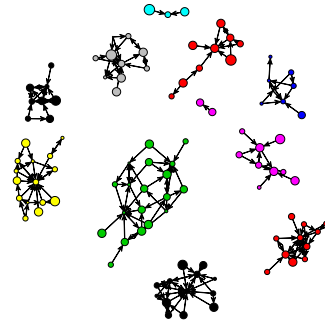


Figure 2.2: Wave 2 network

First, we simply visualize the first wave of the network data. We see as described and plotted in the paper that the network is directed and comprised of ten distinct communities. The nodes are colored by their Estuary group, and the node size is based on a covariate

measured at each wave of how “trustworthy” a node was. Interestingly, there are also a number of isolates (not shown) which are not mentioned in the paper. In wave 2 we see that the network is substantially smaller, and the missingness is not even across Estuaries. This will lead to the extension of interest, namely imputing data.

2.2 NETWORK EXPLORATION

To further explore the network, we consider some network statistics. This is not part of the replication, as these exact statistics were not covered in the original paper, but it will help to better understand the data. Table 2.1 gives results across the two waves for reciprocity, transitivity, betweenness centrality, and eigenvector centrality.

Table 2.1: Network Statistics

Statistic	Wave 1	Wave 2
Reciprocity	0.217	0.259
Transitivity	0.295	0.351
Mean Normalized Betweenness Centrality	0.047	0.095
Mean Eigenvector Centrality	0.027	0.052

We see that all statistics increase for the second wave, which is interesting given its smaller number of nodes. Reciprocity and transitivity tell us that in the second wave, nodes are slightly more likely to reciprocate ties and to tie with a third node if already tied to a second node which is tied to the third. Neither of the centrality measures are particularly high in either network. Figures 2.3 and 2.4 give histograms of the centralities across all nodes for the two waves. We see both have fairly long tail distributions.

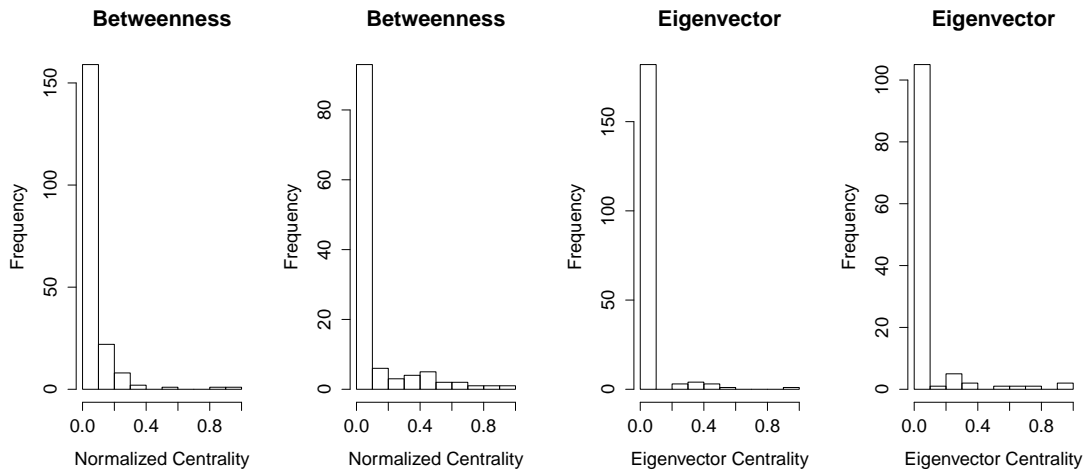


Figure 2.3: Betweenness Centrality Dist. Figure 2.4: Eigenvector Centrality Dist.

2.3 DESCRIPTIVE STATISTICS

Next, following Table A2 in the reference paper, we replicate the descriptive statistics for the two waves and the covariates. There were a number of difference between the replicated statistics and those given in the paper. These are bolded in the tables below.

Table 2.2: Covariate Statistics

Variable	Mean	SD	Min	Max
Trust - wave 1	5.85	2.00	0	10
Trust - wave 2	6.04	2.01	0	10
Prodevelopment beliefs	5.44	1.15	1	7
Government actor	0.49	0.50	0	1

First, we see that the general covariate descriptive statistics in table 2.2 are almost all exactly the same as published.

Table 2.3: Network Statistics

Statistics	1999 (t_1)	2001 (t_2)
Number of nodes	194	118
Number of networks	10	10
Average node in-degree	1.43	1.69
Number of total ties	277	193
Number of mutual dyads	30	28
Numer of asymmetric dyads	217	143
Number of transitive triplets	72	51
Number of out 2-stars	205	116
Number of in 2-stars	533	398
Number of 2-path	505	318
Missing data fraction	0.00	0.39

The network statistics given in table 2.3 on the other hand have a number of differences (bolded). First, the number of missing data points in the second wave is incorrect. 74 nodes are missing rather than 69, which also leads to all the network statistics being off. Secondly, there are a couple statistics in the first wave which are off, but I was not able to find exact definitions of how the original authors calculated these totals (transitive triplets, out 2-stars, in 2-stars), so it could be that I am simply calculating these differently. Lastly, it does not show up in these statistics, but a few nodes in the second wave have self-loops which by definition should be impossible. This leads me to believe the second data wave became somehow corrupted. This would explain the differences.

As exepcted based on the previous table, the changes in links shown in table 2.4 are slightly off from the published numbers. Curiously, the 1 to 0 and 1 to 1 changes were not corrupted. In the original paper, this table included some distance statistics but these were undefined and I was unable to replicate them.

Table 2.4: Change in Links

Links	1999 to 2001
$0 \rightarrow 0$	35814
$0 \rightarrow 1$	152
$1 \rightarrow 0$	108
$1 \rightarrow 1$	89
Missing	1473
<i>note: undefined distances also given</i>	

Table 2.5: Change in Trust

Actor changes	1999 to 2001
Constant	37
Up	46
Down	42
Total increase	86
Total decrease	89
Missing	69

Finally in table 2.5 we look at the change in the trust covariate, since it was also measured at both waves. Here, the number of dropouts in the second wave is correct (69) unlike the network data. The number of nodes increasing or decreasing is off, only by 1, from the published results.

2.4 STOCHASTIC ACTOR ORIENTED MODEL

The paper relies on one primary stochastic actor oriented (SAO) model. This is longitudinal and simultaneously models (i) a network equation for the choices made by nodes to form edges from the first wave to the second and (ii) a behavioral equation for the trust between nodes. The predictors in these models can be network attributes, actor or partner (referred to as ego and alter) covariates, or dyadic similarities across covariates. To fit this model requires a Markov Chain Monte Carlo (MCMC) estimation, switching between the equations to optimize across all parameters. Accordingly, results obtained here for replication will not be exact but should be similar to the published results. Models were estimated using the *RSiena* package in *R*.

As these are simulated results the coefficients and standard errors will not be exactly the same, but generally they are very close to the published estimates. We can get a better idea also by looking at which variables are determined as significant. We see that there is one extra variable in our replication which is significant at the 0.01 level. Overall the smaller difference in model outcomes is good considering the data used for replication is not exactly the same as reported (based on the descriptive statistics). Note that p-values are obtained from running five simulations and performing a meta-analysis to determine

significance among any simulation. They do not state in the published paper how they obtain p-values.

2.5 ERGM MODELS

Discussion of ERGM models fit in paper, which are not replicable.

3 EXTENSION

Two primary extensions.

1. Respecifying ERGM models in order to allow for estimating of ERGM models.
2. Considering different imputations.

3.1 RESPECIFICATION OF ERGM MODELS

- Set constraints for outdegree following stated survey design.
- Use previous wave as predictor of ties.
- Implement a "build-up" modeling approach. First fitting an ERGM with appropriate structural statistics and then fitting successive models that add layers of the authors hypotheses. The goal will be to assess at what layer the model is best.

3.2 DATA IMPUTATION MODELS

- Modal imputation.
- Carry forward imputation.
- MPLE imputation - as a multiple imputation task
- MPLE imputation - as a classification task

3.3 RESULTS - SAO MODELS AND NEW ERGMs WITH IMPUTATION

4 CONCLUSIONS

Table 2.6: SAO Model

Network Model - Partner Selection		
Variable	Coefficients	Standard Error
Average choice per actor	4.85	0.49
Network structures		
Outdegree	-2.24	0.14
Indegree	0.21	0.02
Reciprocity	0.73	0.23
Transitive triplets	0.16	0.10
Ego attributes		
Generalized trust	-0.04	0.05
Prodevelopment beliefs	0.003	0.06
Government actor	0.003	0.14
Alter attributes		
Generalized trust	0.006	0.05
Prodevelopment beliefs	0.02	0.06
Government actor	0.22	0.13
Dyadic Similarity		
Generalized trust	0.16	0.82
Prodevelopment beliefs	0.61	0.46
Government actor	0.19	0.13
Behavior Model - Partner Trust		
Variable	Coefficients	Standard Error
Average choices per actor	6.85	1.20
Ego effects		
Trust in 1999	0.04	0.05
Trust tendency	-0.006	0.02
Prodevelopment beliefs	-0.06	0.05
Government actor	0.10	0.10
Alter effects		
Influence of Alters' trust	4.12	2.22

Table 2.7: Significant Coefficients at 0.01 level

Published results	<i>Outdegree, Indegree, Reciprocity, Influence of Alters' trust</i>
Replicated results	<i>Outdegree, Indegree, Reciprocity,</i> <i>(Alter) Government Actor, Influence of Alters' trust</i>