

INVITED COMMENTARY

The stochastic actor-oriented model is a theory as much as it is a method and must be subject to theory tests

Philip Leifeld^{1,*}  and Skyler J. Cranmer²

¹Department of Government, University of Essex, Colchester, CO4 3SQ, UK and ²Department of Political Science, The Ohio State University, Columbus, OH, USA

*Corresponding author. Email: Philip.Leifeld@essex.ac.uk

Action Editor: Ulrik Brandes

Keywords: stochastic actor-oriented model; SAOM; SIENA; temporal exponential random graph model; TERGM; ERGM

Leifeld & Cranmer (2019) compare the stochastic actor-oriented model (SAOM) and the temporal exponential random graph model (TERGM) theoretically, using simulations, and using an empirical illustration. Block et al. (2022) argue that an error in the empirical illustration towards the end of the article invalidates the main contributions of the article. The main contributions are a set of theoretical differences between the two models and a proposed methodology for model comparison based on out-of-sample prediction. Block et al. do not challenge the theoretical comparison or simulation framework. Yet they argue that network models should not be compared using tie prediction because the inferential goal is to model endogenous network processes, rather than the location of ties. This logical leap from an error in the empirical illustration to discarding the article's main findings and the impossibility of model comparison using dyadic prediction is unsupported by evidence, and it distracts from the main discussion: Does the SAOM contain theory, and how can its inherent theory be tested in any given application? Here, we will refocus the discussion on this central question, which we believe must be urgently addressed.

1. The SAOM has a theory hardwired into its core, and it should be part of theory testing

Theories must be falsifiable to hold scientific value. They are subjected to hypothesis tests using statistical methods. The SAOM is not only a statistical method. It also contains non-trivial theoretical elements. We described these theoretical elements in our 2019 article in comparison with the TERGM and found that the TERGM was lighter on the theoretical side. For example, the SAOM contains a temporal updating process in which actors make one-period look-ahead decisions on rewiring their ties, one by one in a sequential manner and with full information of the remaining network. This should exclude cases like email communication, where actors can send messages to multiple recipients simultaneously; power grids, in which the nodes do not possess agency; large networks where the knowledge of all remaining $n^2 - n$ dyads would pose an impossible cognitive task to the actors, and many other scenarios in which this sociological theory is likely at odds with reality.

As the SAOM has sociological theory hardwired into its core, it only permits a test of the user-specified theory *given* that its other theoretical core assumptions are true. But they may not be true in all given applications. They may sometimes hold fully, sometimes partially, and sometimes not at all. The fact that this is not factored into the testing methodology is a cause for concern

and needs to be addressed because science can only progress if theory is falsifiable. The SAOM, however, contains a non-falsifiable component—unless it is subjected to a testing methodology that removes the substantive model assumptions from the test. The only way in which this can be achieved is through predictive modeling of network waves and comparison with the held-out observations, the very methodology that is being criticized by the creators of the SAOM here and elsewhere, based on an error in an illustrative empirical example.

But the error in the example is unrelated to the argument. In Section 2.2, Block et al. examine the replication code and specification used in the empirical illustration. They find that three of the thirteen model terms in the TERGM are contaminated by observations originating from the time step that is held out for out-of-sample assessment of model fit. This circularity indeed invalidates the model comparison in the empirical example. We agree with everything written in Section 2.2 and apologize to readers of the original article for the oversight. Yet, the error is in the empirical example that was used for illustration, not in the approach per se. The conclusion that “using tie-level predictive capabilities as indicators for comparison between these model classes is unsuitable” does not follow logically from the error in the empirical example. In fact, any model that estimates coefficients can be used for out-of-sample prediction, and it makes sense to do so when assessing model performance (see, e.g., Cranmer & Desmarais, 2017 for a detailed discussion).

2. Attempts by Block et al. to fix the example fail

In Section 2.3, Block et al. attempt to fix the error in the model specification by proposing a new model specification, which makes any differences in fit disappear and trumps prior model fit for both the SAOM and the TERGM. This attempt is beside the point because it does not fix the original problem. Rather than being faithful to the given model specification in their re-analysis, they suggest a new set of model terms that would conveniently lead to similar fit of the two models. Of the thirteen original model terms, six were dropped and replaced by new terms, thus entirely changing the specification for this comparison. But choosing a new specification cannot inform us of whether there are any differences in fit. Apparently, the desired result guides the modeling choices. The proper way to replicate an analysis would be to stick to the original specification and show how the differences in fit go away (or do not go away) if the coding error is fixed. This aspiration is not met, and as a consequence, the authors’ conclusion that there is no difference in model fit between the SAOM and the TERGM is not supported by the evidence provided. In fact, when done properly, model estimation fails due to degeneracy, a problem plaguing poor theoretical specifications (Schweinberger et al., 2020). The question then becomes why the original theory seems feasible and estimable with the SAOM while it is apparently deemed so unrealistic that estimation becomes impossible by the TERGM. It seems to suggest that the auxiliary theory hardwired into the SAOM accounts for this difference, which leads back to the point that theory must be fully falsifiable to be of scientific value.

Model fit is a function of the *theoretical specification* of the chosen model (i.e., the set of model terms), the *parameters* that are attached to the theoretical specification (i.e., the estimated coefficients for the model terms given the observed data), and the theoretical elements hardwired into the technique (i.e., how the nodes are chosen to update their tie profile, the time horizon with which they evaluate their objective function, whether they make sequential or simultaneous choices, etc.—these are additional theoretical elements that may differ across *models*). In short:

Model fit = f (choice of specification; choice of parameters; hardwired model assumptions | data)

In the original article, we argued that if one keeps the specification constant and the parameters are estimated given the specification and the data (thus eliminating parameters from the

equation), then any remaining differences in model fit one may find must be due to differences in inherent model assumptions. We pointed out several theoretical differences in model assumptions associated with the choice of the SAOM or TERGM. Thus,

$$\Delta \text{ Model fit} = f(\Delta \text{ hardwired model assumptions} \mid \text{data; choice of specification})$$

Any manipulations of the givens, such as the data or the specification, invalidate the research design. The task is to compare models given the specification, not to find a specification until one model fits better or until both models fit equally well.

3. Predictive fit is the only way to evaluate the fit of the theoretical component of the SAOM

In contrast, in applications of the SAOM or the TERGM, applied researchers usually want to test theories on network formation, using a dataset as an empirical realization of an underlying population process. They specify this theory by selecting model terms, possibly including covariates, and test the theory by estimating parameters and uncertainty measures for the model terms.

But in the SAOM, additional theory on the network formation process is added by the model itself. It is impossible to separate the two parts of the theoretical process—the one specified by the user and the one inherent in the model—during estimation because they interact with each other. As it is impossible to switch these theoretical assumptions of the SAOM on or off using parameters, part of the variance that ought to be explained by the specification provided by the analyst is explained by the SAOM process. Conveniently, this is often the part of the variance where the provided specification alone fits poorly. This is why one can find cases where the same theoretical specification can be estimated using a SAOM but not using a TERGM. The endogenous version of the original specification of the failed illustration in the 2019 article is an example.

The SAOM theory can compensate for poor model specification through its built-in theoretical updating process. Any estimated coefficients and uncertainty measures are net of the assumed SAOM process. If the SAOM theory is in good keeping with the respective population process, this poses no problem and yields useful parameters for the user-supplied theory. If the inherent SAOM theory is not how the network forms, the hypothesis test of the remaining parameters becomes increasingly meaningless the more the SAOM theory and the population process diverge.

Unfortunately, the SAOM does not contain a test of how useful its built-in theory is, given any population process. It does permit theory testing and comparison *given* its core assumptions, but the question remains as to whether those assumptions are met by a given case.

The SAOM theory can only be factored out of the equation by making model comparisons independent of model choice. The only feasible way of achieving this goal, at least approximately, is to evaluate the usefulness of a theory by asking how well it predicts independent data from the same population process that did not contribute to the estimation of parameters. Such data can usually not be found concurrently, and slicing the network cross-sectionally would violate the independence assumption. Hence, predictive model selection through prediction of future states of the network and assessment of the quality of these predictions is the only way in which this can be achieved. Predicting a panel wave as a test set that was not part of the training set does not violate conditional independence, and it eliminates any factors pertaining to the model itself and the scaling of the parameters: As the TERGM is lighter on theory—it assumes merely a long-running equilibrium of ties between the fixed set of nodes along with an exponential-family probability mass function—any differences in predictive fit with a given specification should be attributable only to differences in model assumptions and a small random error because the held-out panel wave is a draw from the population process (though this remaining error could be minimized

by predicting multiple held-out time steps similarly to cross-validation—see, for example Smyth (2000) for cross-validation and model selection for non-nested models with different likelihood functions).

Model terms in the SAOM are formulated from the perspective of an actor, while they are formulated in a tie-based fashion in the TERGM, resulting in minimally different notation (see Equations (7) and (8) in our 2019 article for an example). Theoretically, they express the same ideas from the perspective of the applied researcher. Block et al. (2019) show that they lead to different distributions of the corresponding subgraph counts in the simulated networks. However, this is not a difference in the theoretical specification per se but rather in the theory inherent in the SAOM as opposed to the TERGM. Either way, the analyst would test for a theoretical building block like reciprocity or triadic closure, but the SAOM and TERGM would translate the respective mechanism into a different distribution as a function of their inherent theoretical model differences. This makes it so important to isolate these inherent theoretical differences while assessing the value of an estimated model.

Simulation evidence in our 2019 article shows that the results one gets indeed depend on model choice and, therefore, on the inherent theory built into the SAOM: Simulations with data that were generated by adhering to the data-generating process of the SAOM found a better predictive model fit for the SAOM in recovering the specification used to generate the data, while simulations with data that were generated by adhering to the data-generating process of the TERGM found a better predictive model fit for the TERGM in recovering the specification used to generate the data.

The claim by Block et al. that their finding “leaves the paper with ambivalent results for [...] the theoretical discussion and simulation study” is therefore questionable, and the conclusion that “there is no basis for suggesting that the TERGM is superior to the SAOM” does not extend to cases in which the data-generating process in the population that generated the observed network is closer to the TERGM than the SAOM, as demonstrated in the simulations. (As an aside, our 2019 article contains no claim that one model is generally superior to the other model.)

4. Tie location must be an essential part of theory evaluation

The final point of contention concerns the value and applicability of dyadic prediction. Block et al. posit that the location of ties in the network is irrelevant, especially to the SAOM. Taken to the extreme, this claim would mean that the labeling of nodes in the network can be discarded and that it is only endogenous processes that should matter for explaining network formation. Consequently, in their view, only endogenous properties like the shared partner distribution or degree distribution should be used to evaluate model fit when comparing predictions with observations.

Restricting the value of theory in this way would denigrate the value of any exogenously imposed structure on the network. If only endogeneity mattered for network formation, nobody should ever control for exogenous node variables, such as age, location, income—or one’s past interactions. This view is oddly detached from theoretical questions many social scientists try to answer with network models. For example, it should matter in explaining friendship networks where and how far apart any two individuals live and whether they had prior opportunities for becoming acquainted (such as through primary school co-attendance). Similarly, for modeling international alliances, it would be at odds with theories in international relations if nation states were not anchored in regions or prior conflicts and if their natural resources or military capabilities were ignored. Contextual information about nodes structures the network formation process and its resulting observable network topology. Including structure through covariates often makes

the difference between a poorly theorized (i.e., degenerate) model and an estimable model, especially with ERGMs as they are not supported by a full-fledged inherent actor-based theory. Why should one disregard such information when assessing how well a model fits and instead focus solely on the endogenous properties?

In the dynamic network context, such a loss of nodal identity would be especially unrealistic: In most empirical contexts, one should expect that past interactions structure current interactions in some way. Assuming non-inertia would destroy the linkages between consecutive network panel waves. Omitting past information from an explanation would amount to nodes completely reinventing their tie profiles in each round irrespective of history. But is collective amnesia of nodes realistic? Do all nodes suffer from memory loss? Are international conflicts and alliances independent of world history other than through endogenous rewiring? We believe not.

Most of the time network modeling is based on a blend of endogenous processes and exogenous variables, sometimes including the past location of nodes or ties in the network, which provide structure to the network topology. As we argued in our 2019 article, both criteria—distributions of endogenous properties and the location of ties in the network as co-determined by endogenous and exogenous processes and information—are important aspects of model fit. Ignoring one of them risks an incomplete picture of model fit.

Only by using prediction, including endogenous and dyad-based assessment of the prediction quality, can one assess the predictive value of a theory when method and theory are inseparably intertwined. While a non-predictive measure of fit would be preferable, this would be necessarily confounded by the theory component inherent in the model. While a model choice based on theory alone—or “explanation” as Block et al. (2019) put it—may sound tempting, in practice it can be hard to evaluate hypothetically how well the built-in theory of the SAOM is met by the population process. At any rate, one should not exclude tie prediction as an “unsuitable” criterion just to mask poor tie-based model fit, and we thus continue to encourage the use of tie prediction.

Competing interests. None.

References

- Block, P., Stadtfeld, C. & Snijders, T. A. B. (2019). Forms of dependence: Comparing SAOMs and ERGMs from basic principles. *Sociological Methods & Research*, 48(1), 202–239.
- Block, P., Hollway, J., Stadtfeld, C., Koskinen, J. & Snijders, T. A. B. (2022). Circular specifications and “predicting” after peeking into the future: Flaws in the empirical SAOM – TERGM comparison of Leifeld & Cranmer. *Network Science*, 10(1), 2–13.
- Cranmer, S. J. & Desmarais, B. A. (2017). What can we learn from predictive modeling? *Political Analysis*, 25(2), 145–166.
- Leifeld, P. & Cranmer, S. J. (2019). A theoretical and empirical comparison of the temporal exponential random graph model and the stochastic actor-oriented model. *Network Science* 7(1), 20–51.
- Schweinberger, M., Krivitsky, P. N., Butts, C. T. & Stewart, J. R. (2020). Exponential-family models of random graphs: Inference in finite, super and infinite population scenarios. *Statistical Science*, 35(4), 627–662.
- Smyth, P. (2000). Model selection for probabilistic clustering using cross-validated likelihood. *Statistics and Computing*, 10(1), 63–72.

Cite this article: Leifeld P. and Cranmer S. J. The stochastic actor-oriented model is a theory as much as it is a method and must be subject to theory tests. *Network Science* <https://doi.org/10.1017/nws.2022.7>