Network Science **10** (1): 3–14, 2022
doi:10.1017/nws.2022.6

CAMBRIDGE UNIVERSITY PRESS

COMMENTARY

Circular specifications and "predicting" with information from the future: Errors in the empirical SAOM–TERGM comparison of Leifeld & Cranmer

Per Block[1*], James Hollway[2], Christoph Stadtfeld[3], Johan Koskinen[4] and Tom Snijders[5,6]

[1]Department of Sociology, Leverhulme Centre for Demographic Science, and Nuffield College, University of Oxford, Oxford, UK, [2]The Graduate Institute Geneva, Geneva, Switzerland, [3]Chair of Social Networks, ETH Zürich, Zürich, Switzerland, [4]Melbourne School of Psychological Sciences, University of Melbourne, Melbourne, Australia, [5]Nuffield College, University of Oxford, Oxford, UK and [6]Department of Sociology, University of Groningen, Groningen, The Netherlands
*Corresponding author. Email: per.block@sociology.ox.ac.uk

Abstract

We review the empirical comparison of Stochastic Actor-oriented Models (SAOMs) and Temporal Exponential Random Graph Models (TERGMs) by Leifeld & Cranmer in this journal [Network Science 7(1):20–51, 2019]. When specifying their TERGM, they use exogenous nodal attributes calculated from the outcome networks' observed degrees instead of endogenous ERGM equivalents of structural effects as used in the SAOM. This turns the modeled endogeneity into circularity and obtained results are tautological. In consequence, their out-of-sample predictions using TERGMs are based on out-of-sample information and thereby predict the future using observations from the future. Thus, their analysis rests on erroneous model specifications that invalidate the article's conclusions. Finally, beyond these specific points, we argue that their evaluation metric—tie-level predictive accuracy—is unsuited for the task of comparing model performance.

Keywords: SAOM; ERGM; TERGM; statistical modeling; dynamic networks; inferential network analysis; longitudinal networks

1. Introduction

The past couple of decades have seen important advances in statistical network modeling, including the proliferation of statistical network models for making inference on a range of questions across the social and natural sciences. Applied researchers and statisticians alike are interested in identifying the appropriate modeling solution for a given research problem. In a recent article in this journal, "*A theoretical and empirical comparison of the temporal exponential random graph model and the stochastic actor-oriented model*", Leifeld & Cranmer (2019a) seek to contribute to this goal.[1] They compare two statistical methods for the analysis of longitudinal network data: the Stochastic Actor-oriented Model (SAOM) and a variant of a Temporal Exponential Random Graph Model (TERGM).[2] In this article Leifeld & Cranmer (2019a) present a theoretical discussion, a simulation study, and an empirical application. While the theoretical and simulation sections are presented as indeterminate—"it is usually not possible to choose the SAOM or the TERGM [...] purely on theoretical grounds" (p. 46, see also p. 20)—their empirical comparison provides them with conclusive results: "when considering out-of-sample predictive performance, the TERGM outperformed the SAOM by a substantial margin" and "even on the type of data for which the SAOM was designed and which the developers of the SAOM use as an expository case, the TERGM out-performed the SAOM" (p. 46).

This is a strong claim, especially as Leifeld & Cranmer (2019a) recommend researchers "not to put too much stock into the *a priori* selection of a model" and instead "contrast the out-of-sample (or in-sample) predictive performance of the two models." Indeed, again based on the empirical results, they claim that the SAOM would have to have "its updating assumptions met with a high degree of precision [. . .] for that specific model to outperform the more general TERGM" (p. 46). Though in their abstract, Leifeld & Cranmer (2019a) state they "do not aim to make a general claim about either being superior to the other across all specifications," their recommendations in the conclusion are clear.

Leifeld & Cranmer (2019a) provide access to all replication materials for the published article (Leifeld & Cranmer, 2019b). We explore a number of surprising findings: why one model outperformed the other by substantial margins, even though both model the same structures in the analyzed network; why Leifeld & Cranmer (2019a) had obtained empirical results suggesting that there is no evidence for transitive closure operating between waves; and why some coefficients seemed too large to be non-degenerate, in contrast with the vast established literature on model specification in ERGMs (e.g., Snijders et al., 2006; Robins et al., 2007).

All these observations are explained by the fact that Leifeld & Cranmer (2019a) replaced endogenous degree effects, typically used in ERGMS and also in the SAOM, by nodal covariates calculated from the observed outcome network. By doing so, instead of explaining differences in nodal degrees as an emerging popularity-based process, they explain the degrees of a node with fixed node covariates that are informed from the empirical data used as the dependent variable (i.e., a transformation the dependent variable is used among the explanatory (independent) variables). Furthermore, when they use this model for out-of-sample predictions, they use degree information *from the future* to predict future nodal degrees. The incorrect model specification and the use of future information to predict the future are entirely responsible for the "substantial margin" between these models in the empirical comparison since they do not repeat this error in the SAOM specification. Thus, the SAOM and the TERGM estimated by Leifeld & Cranmer (2019a) do not have "the same specification" (p. 42). When this is corrected, there is little to discriminate the models in terms of performance. The other surprising findings are similarly explained. Jointly, this invalidates Leifeld & Cranmer's (2019a) comparative results.

In the next section, we present the error in Leifeld & Cranmer's (2019a) specification of the TERGM and explain its consequences. We show that a conventional model specification does not yield the differing results between the SAOM and the TERGM as presented in this article. Next, we illustrate how the specification error is carried over into the out-of-sample prediction. Again, we show that a conventional model does not yield differing out-of-sample predictions. Finally, we outline why tie-level predictive performance is a poor metric to discriminate between the compared models, as the TERGM explicitly models the location of ties, while the SAOM only models their embedding in structures.

2. Empirical model specification

In this section, we first introduce the ERGM and TERGM. Second, we highlight the error in the specification by reference to the code and explain its consequences. Third, we show that analysis in line with the appropriate ERG specification does not support Leifeld & Cranmer's (2019a) conclusions and that there is no difference in out-of-sample performance.

2.1 ERGMs and TERGMs

The TERGM has an Exponential Random Graph Model (ERGM) at its core. As an exponential family model, the ERGM comes with well-known properties (Wasserman & Pattison, 1996; Robins et al., 2007; Lusher et al., 2013). The ERGM defines the probability to observe a network based on endogenous network terms and exogenous covariates. The most commonly used statistics are counts of substructures, for example, the number of reciprocated ties or the number of

https://doi.org/10.1017/nws.2022.6 Published online by Cambridge University Press

in-stars of some order, which may be combined with nodal or dyadic attributes. The statistics that are typically used are based on principled assumptions about the dependencies among the ties (Frank & Strauss, 1986; Snijders et al., 2006). The expected prevalence of these statistics is determined by a statistical parameter. The probability to observe the realization x of a network is given by

$$p_{\mathrm{ERGM}}(X = x) = \kappa^{-1} \exp\left(\sum_k \theta_k s_k(x)\right),$$

where X is the random network state, θ is a statistical parameter, and $s(x)$ is a vector of statistics describing the network; κ is a normalizing constant.

For later use, we elaborate the difference between endogenous and exogenous effects. We represent the network by its adjacency matrix $x = (x_{ij})$, where x_{ij} is the indicator of the tie from node i to node j. An effect is exogenous if it is a linear function of the x_{ij}, and endogenous if it is not. Dependence considerations for networks (see the cited ERGM literature) lead to effects depending on subgraph counts, which are equivalent to products of tie indicators. Examples are counts of two-stars, representing degree variability, and counts of triangles or other triadic configurations, representing transitivity. Endogenous effects imply emergence in the evolution of networks and imply dependence between the tie indicator variables, which is essential for representing a network by a statistical model. In contrast, the impact of nodal and dyadic covariates is usually represented by exogenous effects.

The TERGM is a model for network panel data where the t'th network observation is modeled by an ERGM in which functions of the preceding network observations can enter as exogenous variables. Various ways to specify this have been proposed (Robins & Pattison, 2001; Hanneke et al., 2010; Desmarais & Cranmer, 2012; Krivitsky & Handcock, 2014). The TERGM for two waves at times t-1 and t can be represented by the conditional probability function

$$\begin{aligned} p_{\mathrm{TERGM}}(X(t) = x(t) | X(t-1) = x(t-1)) \\ = \kappa^{-1} \exp\left(\sum_k \theta_k s_k(x(t)) + \sum_h \theta_h z_h(x(t), x(t-1))\right) \end{aligned} \quad (1)$$

where $x(t)$ and $x(t\text{-}1)$ are the realizations, $X(t)$ and $X(t\text{-}1)$ are the random networks, $s(x(t))$ is a vector of statistics for network $x(t)$ like above, and $z(x(t), x(t\text{-}1))$ is a vector of statistics of both networks. For the parameter θ, we denote by θ_k those pertaining to $s(x(t))$ and by θ_h those pertaining to $z(x(t), x(t\text{-}1))$. The difference with the standard ERGM lies in the extra statistics $z_h(x(t), x(t\text{-}1))$ that are memory terms depending on the networks at time t as well as time t–1. These are usually exogenous effects, that is, linear functions of the tie statistics $x_{ij}(t)$. A basic example of such a statistic representing the match between the two consecutive observations is the number of identical tie variables,

$$z_h(x(t), x(t-1)) = \sum_{ij} \left(1 - \left|x_{ij}(t) - x_{ij}(t-1)\right|\right),$$

called the dyadic stability term by Leifeld & Cranmer (2019a); it can be rewritten as a linear function of $x_{ij}(t)$, modeling inertia.

2.2 The model specification in Leifeld & Cranmer's (2019a) TERGM

We now evaluate Leifeld & Cranmer's (2019a) model specification based on the replication script empirical.R (Leifeld & Cranmer, 2019b). Leifeld & Cranmer (2019a) use the btergm package for all analyses, which uses functionality for the established "ergm" and "RSiena" package in R (Hunter et al., 2008; Ripley et al., 2021). Leifeld & Cranmer (2019a) state that they use "the same specification" (p. 42) in their TERGM as in the SAOM and give the statistics used in Equations (15)–(21)

https://doi.org/10.1017/nws.2022.6 Published online by Cambridge University Press

in their article. The ERGM equivalents of the SAOM terms "in-degree popularity," "out-degree popularity," and "out-degree activity" (Equations (19)–(21)) are "in-2-stars," "mixed 2-paths," and "out-2-stars" (Frank & Strauss, 1986). The replication script allows to identify how they have translated Equations (19)–(21) used as endogenous effects in the SAOM to statistics used in the TERGM.

The specification can be found on lines 243–247 of the empirical script:

```
tergm.0.firstthree <- mtergm(friendship[1:3] ~ edges + mutual
  + ttriple + transitiveties + ctriple + nodeicov("idegsqrt")
  + nodeicov("odegsqrt") + nodeocov("odegsqrt") + nodeofactor("sex")
  + nodeifactor("sex") + nodematch("sex") + edgecov(primary)
  + memory("stability"), control = control.ergm(MCMC.samplesize = 5000,
  MCMC.interval = 3000))
```

The three terms `nodeicov("idegsqrt") + nodeicov("odegsqrt") + nodeocov ("odegsqrt")` are meant to correspond to the endogenous effects "in-degree popularity," "out-degree popularity," and "out-degree activity", in the SAOM. However, rather than endogenous network terms, "`idegsqrt`'' and "`odegsqrt`" refer to *exogenous* vertex attributes representing a square root transformation of nodes' indegree and outdegree in the dependent variable, respectively, as defined in lines 222–230 of the script:

```
for (i in 1:length(friendship)) {
  s <- adjust(sex, friendship[[i]])
  friendship[[i]] <- network(friendship[[i]])
  friendship[[i]] <- set.vertex.attribute(friendship[[i]], "sex", s)
  idegsqrt <- sqrt(degree(friendship[[i]], cmode = "indegree"))
  friendship[[i]] <- set.vertex.attribute(friendship[[i]], "idegsqrt",
  idegsqrt)
  odegsqrt <- sqrt(degree(friendship[[i]], cmode = "outdegree"))
  friendship[[i]] <- set.vertex.attribute(friendship[[i]], "odegsqrt",
  odegsqrt)
}
```

This means that the terms `nodeicov("idegsqrt") + nodeicov("odegsqrt") + nodeocov("odegsqrt")` model (i) the tendency of nodes with high *observed* indegree to receive ties, (ii) the tendency of nodes with high *observed* outdegree to receive ties, and (iii) the tendency of nodes with high *observed* outdegree to send ties, respectively. The first and third of these express, respectively, that the indegrees predicted in the model are similar to observed indegrees, and that outdegrees predicted in the model are similar to observed outdegrees. These are tautological terms, by definition receiving high parameter estimates in networks. The second term expresses that modeled indegrees are similar to observed outdegrees. This is a circular term at the level of networks, but not a direct tautology, and the resulting parameter estimate is small. An intuitive way of conceiving of what is happening is that a crucial summary—vertex degrees—of the observed dependent network is used to predict this same feature of the dependent network.

This specification means that the model is not of the form (1), but of the form

$$\begin{aligned} p_{\text{TERGM; LC}}(X(t) = x(t) | X(t-1) = x(t-1); x_{\text{obs}}) \\ = \kappa^{-1} \exp\left(\sum_k \theta_k s_k(x(t)) + \sum_h \theta_h z_h(x(t), x(t-1)) \right. \\ \left. + \sum_l \phi_l u_l(x(t), x_{\text{obs}}(t)) \right). \end{aligned} \tag{2}$$

https://doi.org/10.1017/nws.2022.6 Published online by Cambridge University Press

Added to the model (1) is a further set of statistics, here denoted $u_l(x(t), x_{\text{obs}}(t))$, that is dependent on a particular realization $x_{obs}(t)$ of the networks. A positive parameter associated with these statistics $u_l(x(t), x_{\text{obs}}(t))$, expressing an aspect of the similarity between $x(t)$ and $x_{obs}(t)$, implies that any realization of the network that is closer to the actual observation according to this match has a higher probability to be observed. Thus, the appropriate endogenous specification (1) is turned into circularity. This gives no indication about the salient features of a network, but only artificially improves the fit of the model to this particular dataset without uncovering any dependencies. Leifeld & Cranmer's (2019a) analysis uses three such terms, depending on the observed indegree and outdegree sequences. These are exogenous terms treating the observed network as a covariate for itself. Taken to the extreme, in this spirit we could include the entire observed network as a dyadic covariate among the predictors of the network—this would lead to perfect fit of the model, but would have no explanatory value whatsoever.

Moreover, as the degree sequence is used in the estimation of parameters, it also impacts the parameter estimates of other statistics directly through the strong correlations between statistics in network models. Thus, not only are the parameters ϕ_l associated with statistics $u_l(x(t), x_{\text{obs}}(t))$ tautological, but also all other model parameters θ_k and θ_h will be systematically distorted. In the analysis by Leifeld & Cranmer (2019a), this can be seen in the astonishing finding that there is no support for transitivity in friendship dynamics. As we show in the following subsection, we find clear support for transitivity when replacing the circular exogenous terms with conventional endogenous terms, in line with the very vast literature on friendship networks.

2.3 Replication with correct ERGM terms

In this section, we replicate the analysis by Leifeld & Cranmer (2019a) using conventional ERGM terms instead of the artificially exogenous variables[3]. A model that would be the direct analogue of the SAOM specification has the three star-statistics, transitive and cyclic triples, all of which are sufficient statistics in a Markov graph derived from first principles and assumptions about dependence among the ties (Frank & Strauss, 1986). This Markov model cannot be estimated, something which is to be expected for most networks[4]. While the pseudo-maximum likelihood estimation (MLE) can be determined—that is, pseudo-MLE will produce parameter estimates—these estimates correspond to a degenerate model[5].

Following standard practice (Snijders et al., 2006; Hunter & Handcock, 2006; Lusher et al., 2013), we have to replace the star statistics and the triadic statistics with non-degenerate ones, in particular their equivalent, geometrically weighted statistics, derived from another set of principled dependence assumptions (Robins et al., 2009), in order to specify a model that can be estimated. We substitute the terms `nodeicov("idegsqrt")`, `nodeicov("odegsqrt")`, and `nodeocov("odegsqrt")` with the ERGM terms that model degree dispersion, in particular geometrically weighted in-stars, two-paths, and geometrically weighted out-stars (statnet terms `gwidegree`, `twopath`, and `gwodegree`). Further, we substitute the triadic terms transitive triplets (`ttriple` and `ctriple`) with their geometrically weighted versions (`dgwesp(type = "OTP")` and `dgwesp(type = "ITP")`).

In the SAOM analysis, we also substitute the terms for transitive and cyclic triplets with the geometrically weighted versions. This is done for two reasons. First, even though model terms in ERGM and SAOM analyses are necessarily different in their dependence assumptions (Block et al., 2019), using geometrically weighted versions in both makes the results more comparable. Second, the specification in the tutorial article from 2010 is not "canonical" as claimed by Leifeld & Cranmer (2019a) (p. 42), and so a more contemporary specification should be used for both models.

The results from both models, presented in Table 1, are remarkably similar. The "Transitive ties" parameter and the "GWESP Transitive" parameter need to be interpreted together, as both

https://doi.org/10.1017/nws.2022.6 Published online by Cambridge University Press

Table 1. Results of TERGM and SAOM analysis using standard endogenous ERGM terms

SAOM analysis				TERGM analysis			
	est.		*s.e.*	*est.*		*s.e.*	
Rate period 1	8.38	°	(1.55)	0.73	°	(0.08)	Memory
Rate period 2	8.66	°	(1.39)				
Outdegree (density)	−1.84	°	(0.66)	−2.98	°	(0.43)	Edges
Reciprocity	1.61	***	(0.28)	2.03	***	(0.37)	Reciprocity
Transitive ties	−0.16		(0.33)	−2.16	***	(0.46)	Transitive ties
GWESP transitive	1.76	***	(0.34)	2.84	***	(0.34)	GWESP transitive
GWESP cyclic	−0.27		(0.27)	−0.52	**	(0.12)	GWESP cyclic
Indegree-popularity (sqrt)	−0.29		(0.27)	1.07		(0.70)	GW Indegree
Outdegree-popularity	−0.10		(0.07)	−0.06		(0.03)	Two-paths
Outdegree-activity (sqrt)	0.01		(0.12)	−0.65		(0.52)	GW outdegree
Same primary class	0.39	*	(0.20)	0.44	*	(0.17)	Same primary class
Boy alter	−0.12		(0.17)	−0.06		(0.15)	Boy alter
Boy ego	0.37	*	(0.19)	0.25	*	(0.13)	Boy ego
Same sex	0.71	**	(0.19)	0.53	**	(0.15)	Same sex

Notes: All analyses performed using standard best practises. Significance levels: * = 0.05; ** = 0.01; *** = 0.001; ° = not tested.

model the same tendency of transitive closure, but with different functional forms. The combination of parameters shows that both models find a strong tendency toward friendships being transitive. No endogenous sorting of degrees is found, while the impact of the exogenous covariates "Same primary class" and sex is in the same direction. The main difference is that the "GWESP cyclic" term is significant in the TERGM analysis but not in the SAOM analysis. This will be related to the different formulations of the various GWESP and transitivity terms in this ERGM-type and SAOM-type model as outlined in Block et al. (2019), but some differences are to be expected because the models are different (see literature cited in footnote 1). We conclude that the substantive insights we can draw from either model are not very different, but that understanding how these differences come about is a more complex task than attributing this to simple differences in model fit.

2.4 Out-of-sample prediction

Out-of-sample analysis plays a major role in Leifeld & Cranmer's (2019a) results and final recommendations. In out-of-sample analysis, a dataset is split into training data and test data. Leifeld & Cranmer (2019a) choose the first three waves of the *Knecht network data* as training data and the test data is the fourth wave. Out-of-sample analysis estimates a model using the training data, after which data beyond the training data are simulated based on those estimates and compared to the test data. Importantly, the model should use no information from the test data to generate the predictions. However, the "out-of-sample" prediction using Leifeld & Cranmer's (2019a) TERGM specification violates this principle.

We saw above that in the estimation of parameters for each wave the empirically observed degree sequences for waves at the end of the period were extracted and used as exogenous nodal covariates in the model of change for the period. The same was done to generate likely outcomes of wave 4 (out-of-sample prediction).

One of the wrapper functions of the btergm package is the "`gof`" function that facilitates simulating out-of-sample predictions for a specified model. This "`gof`" function is prominently applied

https://doi.org/10.1017/nws.2022.6 Published online by Cambridge University Press

in section 4 of this article for the out-of-sample comparison. The earlier error is repeated here in lines 254–260:

```
tergm.0.oos <- gof(tergm.0.firstthree, nsim = nsim,
  target = friendship[[4]], formula = friendship[3:4] ~ edges + mutual
  + ttriple + transitiveties + ctriple + nodeicov("idegsqrt")
  + nodeicov("odegsqrt") + nodeocov("odegsqrt") + nodeofactor("sex")
  + nodeifactor("sex") + nodematch("sex") + edgecov(primary)
  + memory("stability"), statistics = c(esp, dsp, ideg, geodesic, rocpr),
  parallel = parallel, ncpus = ncpus)
```

The crucial part here is "`formula = friendship[3:4] ~ ...`", which tells the algorithm to use the "covariates" stored in the network object `friendship[[4]]` in the simulations. However, these covariates include square root transformed information about observed indegrees and outdegrees *at wave 4*. The "`gof`" function as used here thus simulates a model using future information (the future in- and out-degree of nodes). The same circularity as before is even more evident here; part of the test data is used to predict the test data. This mistake appears to be encoded in the `btergm` package that is used in Leifeld & Cranmer's (2019a) analysis.

This analysis cannot then be regarded as out-of-sample testing. This circularity has major consequences for the article results. The comparison case, that is, the out-of-sample predictions of the SAOM, makes no use of any information from the fourth wave, biasing the comparison toward the TERGM.

2.5 Replication with correctly specified ERGM terms

It would in principle be possible that a properly specified TERGM could still perform substantially better in out-of-sample prediction, even if the substantive conclusions do not differ much. To test this, we used the TERGM estimates obtained as outlined above.

The results using the identical metrics for comparison as employed by Leifeld & Cranmer (2019a) are presented in Figure 1. The GOFs for the auxiliary statistics "Edge-wise shared partners," "Dyad-wise shared partners," "Indegree distribution," and "Geodesic distances" show no clear trend favoring either model. The ROC and PR curves are, for what they are worth, very similar between the SAOM and the re-specified TERGM. In sum, these analyses indicate no gain in predictive value between a correctly specified TERGM and the SAOM corroborating earlier findings and theoretical understanding of such processes (Block et al., 2018).[6]

In sum, it was on the basis of a supposed *performance margin* that Leifeld & Cranmer (2019a) recommend researchers estimate both models and use the wrapper functions provided in btergm to identify the one with the better predictive fit. However, we have seen that this comparison based on the btergm functionality and thus this conclusion is fundamentally undermined by their repeated error.

3. Using tie-level predictive accuracy as a comparison criterion

Up to this point, we focused on the particular problems of the empirical comparison in Leifeld & Cranmer's (2019a) article. In this section, we go beyond that article and discuss using tie-level predictive accuracy as a goodness-of-fit metric in such model comparisons. This discussion is based on differences between models like the TERGM that have an auto-regressive component to explain network structure, on the one hand, and models that treat network structure as emergent as the SAOM or ERGM, on the other hand.

The defining feature of statistical network models such as the SAOM or the ERGM is that structures in a network are conceptualized and modeled as the outcome of emergent processes. This means that ties mutually influence each other's emergence and disappearance. Examples for such

https://doi.org/10.1017/nws.2022.6 Published online by Cambridge University Press

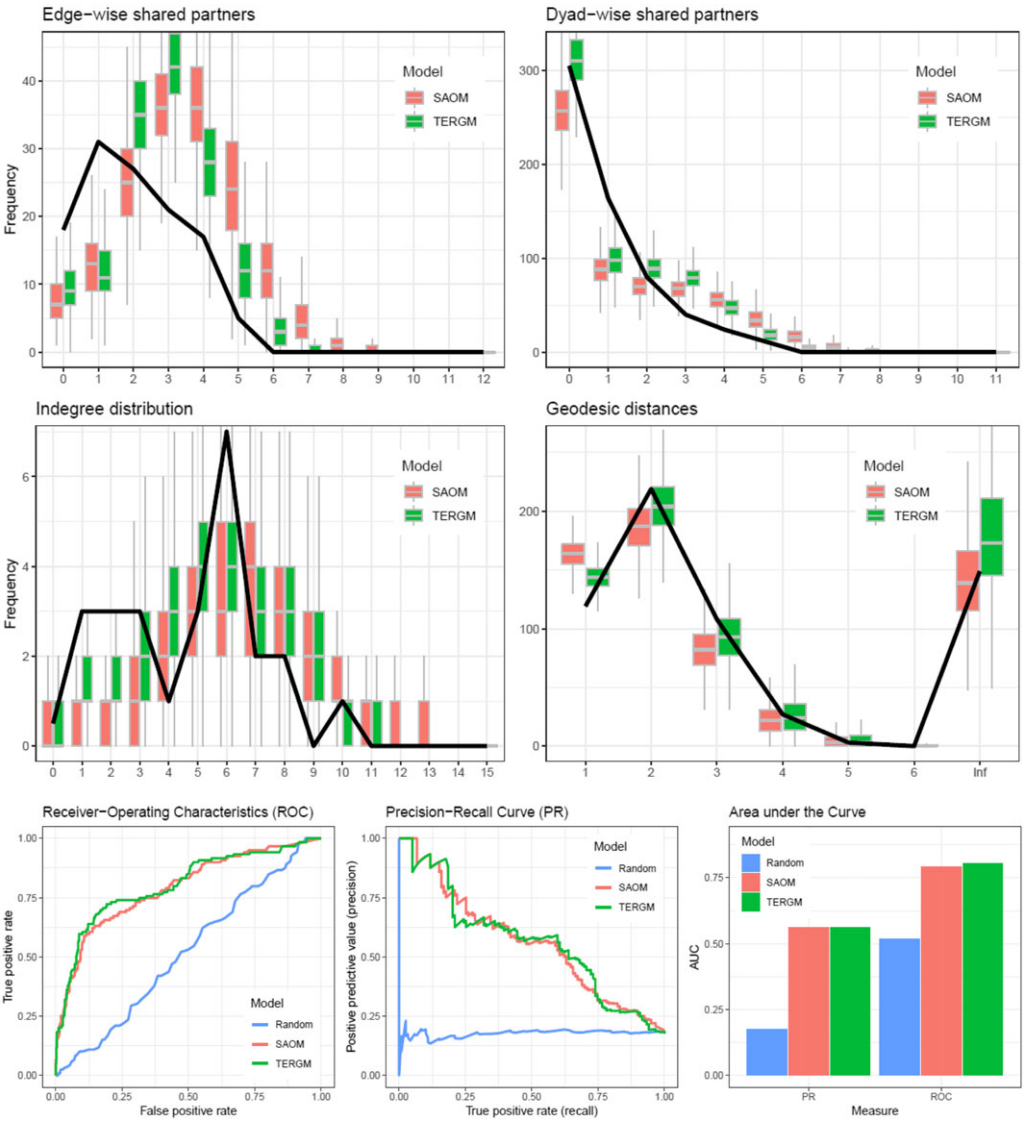

Figure 1. Replication of Figures 4 and 5 from Leifeld & Cranmer's (2019a) article with ERGM terms in the model that do not use future information. This shows that the claim that there is a substantial margin in performance differences is false.

substructures that tend to stabilize the existence of many types of ties include reciprocation, transitive closure, and degree centralization. In these cases, ties are more likely if they are mutual, embedded in transitive groups, or sent/received by actors that are particularly active/popular. Because these structure emerge endogenously, it is not possible to distinguish between ties as explanations or outcomes *per se* (or assigning them either "independent" or "dependent" variable status). Each tie can be both, influenced in its emergence by other ties and, at the same time, influencing the emergence of other ties—even though longitudinal data sometimes allow to assign such roles in particular instances. In this modeling framework, the exact location of ties that emerge or disappear is irrelevant. The focus is firmly on emergent structures of ties, and we are agnostic to where these structures are located in the network. Thus, clustering in networks, reciprocation, or degree centralization are understood to emerge *endogenously* and therefore may appear anywhere

https://doi.org/10.1017/nws.2022.6 Published online by Cambridge University Press

in a network, not in some *exogenously* given location. This is true for both the ERGM and the SAOM.

The TERGM differs from models that are based on endogenous parameters by the additional inclusion of auto-regressive parameters (or by fully substituting endogenous parameters with auto-regressive parameters, see Hanneke et al., 2010). These auto-regressive parameters $z(x(t), x(t\text{-}1))$ appear related to endogenous parameters, but differ in that they predict network structure by configurations in a past realization of the network. Examples include delayed reciprocity or delayed transitivity (Robins & Pattison, 2001; Hanneke et al., 2010), that is when a tie at t_2 is predicted by an incoming tie at t_1, or when a tie at t_2 closes a two-path that existed at t_1, respectively. Degree persistence is a further example.[7] Here, a high popularity/activity at t_1 of an actor predicts incoming and outgoing ties at t_2. These parameters model structure not as the outcome of *endogenous* processes, but as resulting from *exogenous* predictors that are calculated from past realizations of the network. This means the endogenous processes modeled in ERGMs and SAOMs are *exogenized* in the auto-regressive parameters of the TERGM; for example, reciprocated ties are modeled as mutually stabilizing each other in ERGM and SAOM, but when modeled with auto-regressive parameters it is modeled as past ties $j \rightarrow i$ predicting current ties $i \rightarrow j$. This means that auto-regressive parameters predict the *location* of ties based on the past, that is, where exactly in the network ties can be expected. They do not model the prevalence of structures in the modeled network. In the example of using a past tie $j \rightarrow i$ to predict a current tie $i \rightarrow j$, the term does not care if the tie $j \rightarrow i$ still exists. In TERGM specifications such as introduced by Hanneke et al. (2010), only auto-regressive parameters are used, which means that the prevalence of structures of interest is not modeled *at all* in the network under analysis.[8] This means there is a fundamental difference in the way ties are explained between endogenous parameters and auto-regressive parameters, even if they appear similar on the surface. Auto-regressive parameters model the *location* of ties, while endogenous parameters model the prevalence of *structures* in a network. Consequently, parameters of the two types of models are interpreted differently.

This difference in modeling approaches—modeling structure as emergent vs. as predicted by exogenous variables—has direct implications for which fit criteria are meaningful when assessing model fit of different model types. Auto-regressive parameters optimize the correct location of ties in their estimation. In comparison, structural parameters do not. Thus, when comparing which model type performs better based on tie-prediction it might favor the model type that explicitly optimizes this feature in the estimation process. Naturally, the converse is also true. Comparing these model types based on the prevalence of (modeled) substructures is questionable when ERGMs or SAOMs explicitly model this characteristic, while the TERGM of Hanneke et al. (2010) with only auto-regressive parameters does not model this feature. To find an analogy from a different area of network research, this issue is similar to some approaches that compare community detection algorithms on the basis of modularity. Those community detection algorithms that maximize modularity tend to fare better on such comparisons than algorithms that do not maximize it. As such, modularity can only be a meaningful comparison metric if all or no algorithm under comparison use modularity on their approach.

It is not immediately obvious how these issues translate to the performance of different models in out-of-sample predictive accuracy as also used by Leifeld & Cranmer (2019a). However, we believe it might be worth exploring this further before drawing conclusions when using tie prediction as a comparison metric, or when assessing fit in models whose primary goal is uncovering structure, much more than it is predicting tie location.

4. Conclusion

This article is a post-publication review of the recent article by Leifeld & Cranmer (2019a) in *Networks Science* followed by a discussion of more general points pertaining to model comparison.

https://doi.org/10.1017/nws.2022.6 Published online by Cambridge University Press

The error of the original authors, which is repeated both in the specification of the model and also in the out-of-sample analysis, is the use of structural information of the outcome network about nodal degrees as exogenous predictors. This is most clearly (and consequentially for the conclusions of this article) made for the out-of-sample analysis. Leifeld & Cranmer (2019a) calculate (a square root transformation of) nodes' indegrees and outdegrees in the test data and generate sample networks based on these data that they then compare with the same test data.

This has several consequences. First, the out-of-sample comparison conducted in this article is undermined by the tautological specification. We found that using correctly specified ERGM terms resulted in performance no better than that of an analogous SAOM, corroborating previous analyses that came to the same conclusion of no substantial differences (Block et al., 2018). This leaves the article with ambivalent results for the empirical section, as well as from the theoretical discussion, which has been covered in the previous literature more thoroughly, and ambivalent results from the simulation study. At any rate, there is no basis for suggesting that the TERGM is superior to the SAOM.

Second, parameter estimates from TERGMs that use the Leifeld & Cranmer (2019a) specification are contaminated by the inclusion of exogenous structural covariates. The resulting parameter estimates do not say anything about self-organizing tendencies of the network, neither about dependence between ties nor about degree centralization of the network. If used, they must logically have positive parameter estimates. Furthermore, the inclusion of these statistics strongly affects other parameter estimates in the model, due to high collinearities between modeled statistics. The differing results found by Leifeld & Cranmer (2019a) are a direct consequence of the circularity introduced into the model specification. One of these surprising results is the absence of transitive clustering in Leifeld & Cranmer's (2019a) results. These differences in parameter estimates disappeared when the model was specified with degree terms modeled endogenously. As there does not appear to be any general performance improvement from a properly specified TERGM over the SAOM, we believe we should return the question of which model is better motivated theoretically.

Third, using tie-level predictive capabilities as indicators for comparison between these model classes is unsuitable. The auto-regressive parameters in a TERGM are used to predict the location of ties in the dependent network, while models such as the SAOM are mainly interested in the embedding in structures. In summary, we do not advocate for such assessment for the kinds of highly interdependent data as present in network studies, not least because these types of model that rely on the generation of stationary stochastic processes generally fare worse than trivial prediction models in out-of-sample predictions (Block et al., 2018)[9].

Acknowledgments. We thank Christian Steglich, Viviana Amati and Felix Schönenberger for feedback. Closely related issues about statistical network modelling were brought up at the annual Duisterbelt meetings; we thank all participants that contributed to these discussions. Per Block is supported by the Leverhulme Centre for Demographic Science.

Competing interests. None.

Notes

1 Many points that concern the theoretical and principled differences between the models Leifeld & Cranmer (2019a) treat have been discussed: Schaefer & Marcum (2018) generally discuss SAOMs and (S)(T)ERGMs. Block et al. (2019) focus on fundamental dependence assumptions between the models, showing that it is not possible to formulate "equivalent" models, even if the parameter names might suggest otherwise. Block et al. (2018) discuss what Leifeld & Cranmer (2019a) call the "data-generating process" (DGP), concluding that the TERGM DGP is a purely technical solution to obtain samples under the model that has no coherent interpretation about a network evolution.

2 A different version of the TERGM is the STERGM of Krivitsky & Handcock (2014) which is implemented in "tergm" (Krivitsky & Handcock, 2016).

3 We perform this replication analysis with the code provided by Leifeld & Cranmer (2019a), thus leaving all other modeling and software choices intact.

https://doi.org/10.1017/nws.2022.6 Published online by Cambridge University Press

4 Issues with Markov models have been extensively treated in Strauss (1986); Jonasson (1999); Snijders (2002); Handcock (2003); and Schweinberger (2020). How these degeneracies are alleviated by different dependence assumptions is covered, for example, in Snijders et al. (2006) and Schweinberger (2011).
5 A model with such parameters predicts either near-empty and near-complete networks, neither of which shares any similarities with observed data. As such, the model is useless for predictions.
6 While it would certainly be possible to optimize the model specification of both the TERGM and the SAOM to improve on these metrics, especially as the geometrically weighted terms have an internal parameter that can be adjusted, since Leifeld & Cranmer's (2019a) purpose was comparison rather than optimization, we only compare similarly specified models here.
7 Degree persistence is close to Leifeld & Cranmer's (2019a) analysis, had they used observed degree in the explanatory network instead of the dependent network in their model.
8 It should be noted that such specifications that exogenize all structural parameters have very similar results to continuous-time models such as the SAOM when inter-observation times are very short and start to perform worse the longer the observations are apart, as shown by Lerner et al. (2013). Thus, for very short inter-observation times they might be equally suited while having much lower computational costs.
9 See also Snijders (2010) on marginalization of ERGM and further on the consequence on link pre-diction and residuals for ERGM (Koskinen et al., 2018).

References

Block, P., Koskinen, J., Hollway, J., Steglich, C., & Stadtfeld, C. (2018). Change we can believe in: Comparing longitudinal network models on consistency, interpretability and predictive power. *Social Networks*, 52, 180–191. http://doi.org/10.1016/j.socnet.2017.08.001

Block, P., Stadtfeld, C., & Snijders, T. A. (2019). Forms of dependence: Comparing SAOMs and ERGMs from basic principles. *Sociological Methods & Research*, 48(1), 202–239.

Desmarais, B. A., & Cranmer, S. J. (2012). Statistical mechanics of networks: Estimation and uncertainty. *Physica A: Statistical Mechanics and Its Applications*, 391(4), 1865–1876.

Frank, O., & Strauss, D. (1986). Markov Graphs. *Journal of the American Statistical Association*, 81, 832–842.

Handcock, M. (2003). Assessing degeneracy in statistical models of social networks. Tech. rep., Center for Statistics and the Social Sciences, University of Washington, http://www.csss.washington.edu/Papers.

Hanneke, S., Fu, W., & Xing, E. P. (2010). Discrete temporal models of social networks. *Electronic Journal of Statistics*, 4, 585–605.

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

Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., & Morris, M. (2008b). ergm: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software*, 24(3), nihpa54860.

Jonasson, J. (1999). The random triangle model. *Journal of Applied Probability*, 36, 852–876.

Knecht, A. (2008). Friendship selection and friends' influence: Dynamics of networks and actor attributes in early adolescence. (Ph.D. dissertation). University of Utrecht, Utrecht.

Koskinen, J., Wang, P., Robins, G., & Pattison, P. (2018). Outliers and influential observations in exponential random graph models. *Psychometrika*, 83(4), 809–830.

Krivitsky, P. N., & Handcock, M. S. (2014). A separable model for dynamic networks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1), 29–46.

Krivitsky, P. N., & Handcock, M. S. (2016). tergm: Fit, simulate and diagnose models for network evolution based on exponential-family random graph models. *The Statnet Project (http://www.statnet.org). R package version, 3*(0).

Leifeld, P., & Cranmer, S. J. (2019a). A theoretical and empirical comparison of the temporal exponential random graph model and the stochastic actor-oriented model. *Network Science*, 7(1), 20–51. https://doi.org/10.1017/nws.2018.26

Leifeld, P., & Cranmer, S. J. (2019b). Replication data for: a theoretical and empirical comparison of the temporal exponential random graph model and the stochastic actor-oriented model. *Harvard Dataverse*, V1, https://doi.org/10.7910/DVN/NEM2XU

Lerner, J., Indlekofer, N., Nick, B., & Brandes, U. (2013). Conditional independence in dynamic networks. *Journal of Mathematical Psychology*, 57(6), 275–283.

Lusher, D., Koskinen, J., & Robins, G. (2013). Exponential random graph models for social networks. Cambridge: Cambridge University Press.

Ripley, R., Snijders, T. A. B., Boda, Z., Vörös, A., & Preciado, P. (2021). *Manual for SIENA version 4.0*. Oxford: University of Oxford, Department of Statistics, http://www.stats.ox.ac.uk/ snijders/siena/.

Robins, G., & Pattison, P. (2001) Random graph models for temporal processes in social networks. *Journal of Mathematical Sociology*, 25(1), 5–41.

Robins, G., Pattison, P., & Wang, P. (2009). Closure, connectivity and degree distributions: Exponential random graph (p*) models for directed social networks. *Social Networks*, 31(2), 105–117.

https://doi.org/10.1017/nws.2022.6 Published online by Cambridge University Press

Robins, G., Snijders, T., Wang, P., Handcock, M., & Pattison, P. (2007). Recent developments in exponential random graph (p*) models for social networks. *Social networks*, 29(2), 192–215.
Schaefer, D. R. & Marcum, C. S. (2018). Modeling network dynamics. https://doi.org/10.31235/osf.io/6rm9q
Schweinberger, M. (2011). Instability, sensitivity, and degeneracy of discrete exponential families. *Journal of the American Statistical Association*, 106(496), 1361–1370.
Schweinberger, M., Krivitsky, P. N., Butts, C. T., & Stewart, J. (2020). Exponential-family models of random graphs: inference in finite-, super-, and infinite population scenarios. *Statistical Science* (forthcoming).
Snijders, T. A. B. (2002). Markov chain Monte Carlo Estimation of exponential random graph models. *Journal of Social Structure*, 3, 1–40.
Snijders, T. A. B. (2010). Conditional marginalization for exponential random graph models. *Journal of Mathematical Sociology*, 34, 239–252.
Snijders, T. A. B., Pattison, P. E., Robins, G. L., & Handcock, M. S. (2006). New specifications for exponential random graph models. *Sociological Methodology*, 36, 99–153.
Strauss, D. (1986). On a general class of models for interaction. *SIAM Review*, 28, 513–527.
Wasserman, S., & Pattison, P. (1996). Logit models and logistic regressions for social networks: {I}. An introduction to Markov graphs and p*. Psychometrika, 61(3), 401–425.

Cite this article: Block P., Hollway J., Stadtfeld C., Koskinen J. and Snijders T. (2022). Circular specifications and "predicting" with information from the future: Errors in the empirical SAOM–TERGM comparison of Leifeld & Cranmer. *Network Science* **10**, 3–14. https://doi.org/10.1017/nws.2022.6