

Change we can believe in: Comparing longitudinal network modelson consistency, interpretability and predictive power

Per Block, Johan Koskinen, James Hollway, Christian Steglich and Christoph Stadtfeld

The self-archived postprint version of this journal article is available at Linköping University Institutional Repository (DiVA):

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-142463>

N.B.: When citing this work, cite the original publication.

Block, P., Koskinen, J., Hollway, J., Steglich, C., Stadtfeld, C., (2017), Change we can believe in: Comparing longitudinal network modelson consistency, interpretability and predictive power, *Social Networks*. <https://doi.org/10.1016/j.socnet.2017.08.001>

Original publication available at:

<https://doi.org/10.1016/j.socnet.2017.08.001>

Copyright: Elsevier

<http://www.elsevier.com/>



Change we can believe in: Comparing Longitudinal Network Models on Consistency, Interpretability and Predictive Power^{*}

Per Block^{1†}, Johan Koskinen², James Hollway³,
Christian Steglich⁴ and Christoph Stadtfeld¹

¹ *Chair of Social Networks, ETH Zürich, Switzerland.*

² *The Mitchell Centre for SNA, and Social Statistics Discipline Area, University of Manchester, United Kingdom.*

³ *Department of International Relations/Political Science, Graduate Institute Geneva, Switzerland.*

⁴ *The Institute for Analytical Sociology, Linköping University, Sweden AND Department of Sociology/ICS, University of Groningen, The Netherlands.*

Please cite as

Block, P., Koskinen, J., Hollway, J., Steglich, C., and Stadtfeld, C. (2017, *forthcoming*). Change we can believe in: Comparing Longitudinal Network Models on Consistency, Interpretability and Predictive Power. *Social Networks*.

Abstract

While several models for analysing longitudinal network data have been proposed, their main differences, especially regarding the treatment of time, have not been discussed extensively in the literature. However, differences in treatment of time strongly impact the conclusions that can be drawn from data. In this article we compare auto-regressive network models using the example of TERGMs – a temporal extensions of ERGMs – and process-based models using SAOMs as an

^{*} The authors would like to thank the network groups at Nuffield College, Groningen University, ETH Zürich and University of Melbourne for their valuable comments and feedback to this work.

[†] corresponding author: per.block@gess.ethz.ch

example. We conclude that the TERGM has, in contrast to the ERGM, no consistent interpretation on tie-level probabilities, as well as no consistent interpretation on processes of network change. Further, parameters in the TERGM are strongly dependent on the interval length between two time-points. Neither limitation is true for process-based network models such as the SAOM. Finally, both compared models perform poorly in out-of-sample prediction compared to trivial predictive models.

1. Introduction

The study of social networks is increasingly concerned with modelling network change over time, as longitudinal analysis is usually better equipped for finding explanations and testing theories about the evolution of networks as well as the impact its structure has on constituent nodes (e.g. Steglich et al. 2010). Network analysis over time commonly uses network panel data: a network structure among the same set of nodes that is observed at two or more time points. By now (this is written in early 2017), several statistical approaches are available to analyse such data sets. The most widely used are the stochastic actor-oriented model (SAOM; Snijders, van de Bunt and Steglich 2010) and several extensions to the exponential random graph model (ERGM; Lusher, Koskinen and Robins 2013). These models and variations may appear almost indistinguishable to scientists interested in applying inferential methods to network panel data. However, they rest on quite different statistical assumptions that strongly affect the kind of inference one can draw from the estimated model parameters and, thus, the kind of questions that can be answered with each method.

While statistical models can be compared on many dimensions, we mainly focus in this article on differences in how they treat time. In particular, we discuss the difference between discrete-time, auto-regressive models and continuous-time, process-based models. Due to its increased use (e.g. in McFarland *et al.* 2014) and recent claims about its advantage relative to other models for network panel data (Desmarais and Cranmer 2012; Leifeld and Cranmer

2016), we choose the TERGM (or temporal ERGM) for this comparison case to represent auto-regressive models¹. The continuous-time model we discuss for comparison is the SAOM. Note that for ERGMs both continuous-time and auto-regressive extensions have been proposed – we focus on the latter group. The purpose is to compare the principles of auto-regressive and continuous-time network models and not the relative merits of either particular model – the two cases can be seen as representations of their respective model classes. This article highlights, by way of illustration, the most important differences in assumptions and their interpretive implications between these approaches and thus facilitates the applied researcher’s decision which to use in their own research.

1.1.Dimensions of comparison

When comparing statistical models it is tempting to ask which model is “better”. However, “better” implicates at least two quite different dimensions: explanation and prediction. On the one hand, it has been argued that accurate prediction is a chief criterion of a “good” model (Friedman 1953; Jasso 1988). Intuitively a “good” model should be able to extrapolate accurately into the future, which can be tested for a single dataset by simple out-of-sample prediction. At the same time, the criterion for what should be predicted correctly in a model with dependent data (such as networks) is not trivial, as a network is more than just an series of independent tie observations but also the structures that these ties form (see discussion in Section 5).

On the other hand, it has been argued that the endeavour of social science is not to predict, but to explain and understand the world (Hedström 2005; Elster 2007). Models with absurd assumptions or intractable algorithms can generate fairly accurate predictions, but teach us little

¹ We point out commonalities and differences between the TERGM (as defined in Desmarais and Cranmer 2012; Leifeld and Cranmer 2016) and other longitudinal variants of the ERGM where appropriate.

about the world. Social mechanisms, by contrast, can help us explain the social world and inform our understanding of our own and others' behaviour, but their concatenation in complex ways means that only in the simplest of systems can we expect this to result in accurate prediction at a micro-level. Indeed, even models with poor predictive power can generate valuable insights (see also Epstein 2008). In this line of reasoning, a good model is characterised by reasonable assumptions, as well as by clear interpretability of parameters in light of social mechanisms.

In this paper, we do not necessarily advocate for one or the other position, but investigate how different model assumptions make them applicable to different questions and thus to different empirical problems. As such, we elaborate what conclusions can be drawn from estimated parameters using the SAOM or the TERGM.

The remainder of the article is organised as follows. We first introduce the two different longitudinal/temporal network models (Section 2), and highlight their main features from a statistical point of view. The first main distinguishing feature of the model that is discussed concerns whether it is actor-oriented or tie-oriented (Section 3). Subsequently, the treatment of time is examined. Focus is on the interpretation of parameters and model consistency with regards to the differences between auto-regressive compared to process-based modelling (Section 4). The different treatment of time and how that influences parameters is shown in an empirical example. Finally, we demonstrate that both models perform poorly in out-of-sample prediction (Section 5) across two datasets, suggesting that we need to be careful as to the purposes of longitudinal network research.

2. The Models

A social network needs to be understood as a system of interdependent units. Whether one is interested in the details of network dependencies or just needs to control for them,

research on networks requires statistical tools that can adequately deal with this challenge. The model families that most explicitly deal with dependencies for such inferential-statistical analysis of social network data are *exponential random graph models* (ERGMs; Frank & Strauss, 1986; Pattison & Wasserman, 1999; Snijders, Pattison, Robins & Handcock, 2006; Lusher, Koskinen & Robins, 2013) and *stochastic actor-oriented models* for network evolution (SAOMs; Snijders 2001, 2005; Snijders, van de Bunt & Steglich, 2010)².

2.1. Exponential random graph models

ERGMs were originally formulated for cross-sectional data, i.e., a single observation of a network. The guiding idea behind the model family is to express the probability of observing a given network as a function of subgraphs in this network (called statistics denoted $z(x)$ on the network x), e.g. the reciprocated dyad, or the transitive triplet. These subgraphs express local dependencies between tie variables (reciprocity and transitive clustering, respectively). At the heart of the ERGM lies a linear predictor that weighs the prevalence of statistics in the network by the parameter vector θ :

$$\sum_k \theta_k z_k(x)$$

What is considered local differs between model specifications, with the general rule that specifications including more complex subgraphs instantiate more dependency (Pattison & Snijders 2013). Model parameters θ_k can be interpreted as expressing, on the tie-level, the probability of observing a specific tie, given the rest of the graph, or on the network-level, as

² There is a host of models that allow for dependent network ties (such as the p2 model, van Duijn *et al.* 2004; and an ever-expanding class of latent variable models, see for example the review by Salter-Townshend *et al.* 2012) that we do not discuss here.

indicating tendencies of a graph to exhibit certain sub-structures relative to what would be expected from a model not containing this parameter (this is discussed further in Section 4).

Longitudinal variants of the ERGM come in two forms, the continuous-time and the discrete-time version. The first, called longitudinal exponential random graph models (LERGMs; Snijders & Koskinen 2013; Koskinen, Caimo & Lomi 2015), is a longitudinal, continuous-time form of the ERGM, in the sense that changes to the network are modelled using the conditional probabilities of the ERGM and the process has the cross-sectional ERGM as its limiting distribution. In its treatment of time, the LERGM is identical to the SAOM, thus, we do not focus on the LERGM in this article – the interested reader can generalise from our discussion.

The most prominent discrete-time variant of the ERGM is the temporal exponential random graph model (TERGM; Robins and Pattison 2001, Hanneke, Fu and Xing 2010; Desmarais and Cranmer 2012)³. The model is based on the idea of panel regression. In a sequence of observations, lagged earlier observations or derived information thereof can be used as predictors for later observations. In other words, some of the statistics $z(x)$ are direct functions of an earlier realisation of the network. In its most basic form, the TERGM is a conditional ERGM with an earlier observation of the network occurring among the predictors. It is this basic TERGM (as presented in Desmarais and Cranmer 2012; Leifeld and Cranmer 2016) that we focus on in this article. While other statistics of a previous network realisation (e.g. past two paths) can be included in the model as predictors (e.g. to model transitivity over time), this does not change the fundamental challenges of parameter interpretability or time dependence of parameters modelling dependence as discussed in Section 4; consequently, we

³ Given the limited space in one article, we do not discuss other discrete-time models, such as the StERGM (Krivitsky and Handcock 2014), even though they deserve a similar comparison elsewhere that might give different results.

only deal with these extended specifications, when necessary, in footnotes. The interested reader can generalise⁴.

A different formulation of the basic TERGM does not use a previous time-point as a predictor, but instead models a “dyadic stability” parameter (Leifeld and Cranmer 2016), which models how many ties and non-ties remain constant between two observations. However, as we show in Appendix A.1.1 these two formulations are mathematically equivalent and we use the more intuitive version using the previous time-point as a predictor throughout the article for our discussion. As it is just an ERGM with a previous time-point as a dyadic covariate, the TERGM can be estimated with any software that can estimate ERGMs (see e.g. Hunter *et al.* 2008; Wang *et al.* 2014).

2.2. Stochastic actor-oriented models

The guiding idea behind SAOMs is the integration of statistical models that can account for network dependence with theoretical models of action that view social change as emanating from individual actors. Starting from the idea of modelling change, SAOMs are continuous-time models that connect two observations of a network through an unobserved sequence of smallest possible changes, called *mini-steps*. In these *mini-steps*, first an actor in the network is chosen to make a tie change according to the so-called *rate-function*⁵. Second, this actor considers which (if any) of its outgoing ties it will change, with its decision being based on a multinomial logit that uses the so-called *objective function*. Similar to the ERGM, the *objective*

⁴ It should be noted that the TERGM might only include transformations of an earlier network as predictors of the network, as presented in Hanneke, Fu & Xing (2010). In this case, all dependence between ties is assumed to be captured by the previous time-point; thus, ties in the network under analysis are assumed to be independent. This allows estimation as a simple logistic regression. However, a previous article (Lerner *et al.* 2013) has already treated this specific model in comparison to the process-based SAOM, concluding that this simplifying assumption leads to reasonable model fit in case observations are temporally very close, i.e. do not differ on a lot of tie variables. Consequently, in this article we focus on model specifications in which previous observations, as well as contemporaneous dependence terms are used to model the network.

⁵ In most applications of SAOMs the rate-function is constant, meaning that an actor is chosen at random.

function is a linear predictor that depends on statistics $s_i(x)$ as “seen” from the perspective of actor i and a statistical parameter β :

$$\sum_k \beta_k s_{i,k}(x)$$

Network dependencies are modelled within this *objective function* expressed as statistics and, thus, unfold over time, because the decision of which tie to update can depend on its embedding in network structures (such as reciprocation or transitive embedding). Model parameters can be interpreted in light of these *mini-steps*, indicating whether actors preferentially form ties to be embedded in certain configurations. Large values of the objective functions are associated with changes that are more likely.

SAOMs can be fitted to data by means of maximum-likelihood (Snijders, Koskinen & Schweinberger 2010), Bayesian (Koskinen & Snijders 2007), or method-of-moments estimation (Snijders 2001), each implemented in the RSiena software (Ripley *et al.* 2016). Details of the differences between estimation procedures are immaterial for the arguments presented here unless pointed out explicitly though properties of predictions and efficiency differ between estimation techniques.

3. Actor vs. tie based modelling

The principal difference between any model from the ERGM family and the SAOM family discussed in the literature is that the former is “tie-oriented” while the latter is “actor-oriented” (Block, Stadtfeld & Snijders 2016). In a very general sense, this means that the locus of modelling differs between the models. The former models whether a tie is likely to exist depending on how it is embedded in substructures in the network. This is reflected by the common interpretation of ERGM parameters, which provide the likelihood to observe a tie, conditional on the rest of the network. The latter models whether an actor is more likely to form

or maintain a tie depending on its embedding in substructures in the network from the actor's perspective. The SAOM is explicitly defined on the micro-level to allow for modelling change from an actor's perspective.

While these differences might seem trivial at first, there are two important implications. First, the different loci of modelling entail different dependence assumptions. In ERGMs, each tie is evaluated on its own for how it is embedded in substructures, but in SAOMs these substructures and ties are always identified and chosen from the perspective of a particular actor. This implies that a decision to, say, create one tie is simultaneously a decision against creating or deleting another at that time, inducing generally higher-level dependence between ties in the SAOM (which might or might not be an appropriate assumption about real-world-processes). Naturally, whether you interpret this 'decision' in behavioural terms or merely as formal way of expressing that ties are evaluated with reference to how they are embedded in one actor's local neighbourhood, is up to the researcher. Second, taking an actor's perspective allows model specifications that are closer to social theory, as ties in different positions in the same structure can be guided by different model parameters. However, this usually comes at the cost of model parsimony.

Block *et al.* (2016) discuss the above mentioned differences in detail with extensive illustrations and provide guidelines as to which model might be more appropriate for which research questions when only taking tie- and actor-orientation into account.

4. Process-based vs. auto-regressive modelling

The more important division among the discussed longitudinal network models for the comparison at hand is how they treat time. As argued in Section 2, continuous-time and discrete-time models differ fundamentally: the former models a process whereas the latter models a cross-sectional observation using a previous time-point as a predictor. The

conceptualisation of time upon which a model is based strongly impacts how parameters can be interpreted and, accordingly, which kind of research question can be answered.

The difference between process-based and auto-regressive models focussing on networks is presented in Section 4.1. Subsequently, we discuss any potential micro-level interpretations of parameters in connection to underlying social mechanisms (Section 4.2) and the dependence of parameter sizes on time elapsed between two observations (Section 4.3).

4.1. Models of change and models of structure

Continuous- and discrete-time models have different objects of inference and thus answer different types of questions. Continuous-time models like the SAOM answer, broadly speaking, questions about change such as: “According to which regularities does the network evolve from time t_{m-1} to t_m ?”. Conversely, discrete-time models like the basic TERGM answer questions about structure such as: “What regularities does the network at time t_m exhibit, taking into account knowledge about t_{m-1} ?”. Note that this means that the TERGM, despite including a previous realisation of the network, cannot make inference about change. Figure 1 provides an example that illustrates why it cannot.

Figure 1 around here

Consider modelling the two depicted networks $x_a(t_1)$ and $x_b(t_1)$ using a TERGM with three parameters. We include $x(t_0)$ as a dyadic covariate, a density parameter and a transitive triplets parameter. Recall that the probability to observe either network $x_a(t_1)$ or $x_b(t_1)$ only depends on the sufficient statistics of the networks:

$$p_{\theta}(x_*(t_1)|x(t_0)) = \frac{\exp(\sum_k \theta_k z_k(x_*(t_1)))}{\kappa}$$

As both, $x_a(t_1)$ and $x_b(t_1)$, have identical statistics (6 ties, 3 ties stable from $x(t_0)$ and 1 transitive triplet), they have the same probability to be observed for any combination of parameters and, consequently, the two hypothetical models have the same Maximum Likelihood Estimate. As the transitive triplets parameter in either model will be identical, we can see that it does not relate to *changes* between two time-points at all, but rather to a higher prevalence of a specific *structure* (the transitive triplet), given the other model parameters. Therefore, the notion that the TERGM models change because a previous time-point is included in the model is wrong.

For the SAOM though, estimating the evolution from $x(t_0)$ to $x_a(t_1)$ will give different results than estimating from $x(t_0)$ to $x_b(t_1)$, as the most likely chain of mini-steps that connects $x(t_0)$ to the respective network at the second time-point differs markedly. In one chain the transitive triplet is stable and the three non-embedded ties are broken and three ties are established in a different location, while in the other chain the transitive triplet must be broken and a new transitive triplet established in a different location. Consequently, the parameter estimates for transitive triplets differ considerably, with the transitive triplets parameter being larger in the case when the triplet emerges from network change⁶.

This illustration shows the fundamental difference between these models. In the SAOM dependence between ties unfolds over time – the embedding of ties guides the *changes* that actors make. In the TERGM, modelling of time is decoupled from the modelling of dependence; one parameter models similarity between a network and the same network at a previous time-point (the auto-regressive stability term), with other parameters modelling the prevalence of

⁶ Note that Maximum Likelihood estimation is necessary for this example calculation.

specific *structures* in this network without taking the past into account. These differences translate into how the model parameters can be interpreted.

4.2. Parameter interpretation

Interpretation of the SAOM

The natural interpretation of parameters in the SAOM (and similarly of other process-based models like the LERGM) is on the micro-level and follows directly from the formulation of the model. Even though the passing of time is explicitly modelled in the SAOM in the rate-function, generally it is the changes individuals make to their network themselves, expressed in the parameters β_k , that are of interest to researchers⁷.

A parameter's direction indicates whether actors in a (hypothetical) mini-step make choices that increase or decrease the statistic associated with the parameter. Referring to the underlying multinomial choice function, parameter sizes can be translated to (conditional) odds ratios. For example, when an actor has the opportunity to make a change and networks $x^{\pm ij}$ and $x^{\pm ik}$ are two possible outcomes referring to changing either the tie to j or to k , then the relative probability to choose $x^{\pm ij}$ over $x^{\pm ik}$ can be calculated by

$$\frac{p(x \rightsquigarrow x^{\pm ij})}{p(x \rightsquigarrow x^{\pm ik})} = \frac{\exp(\sum_k \beta_k s_{i,k}(x^{\pm ij}))}{\exp(\sum_k \beta_k s_{i,k}(x^{\pm ik}))}$$

In case $x^{\pm ij}$ and $x^{\pm ik}$ are identical on all dimensions but, for example, choosing $x^{\pm ij}$ would result in a reciprocated tie whereas $x^{\pm ik}$ would not, all terms but the difference in reciprocity would cancel out and

⁷ Early continuous-time models for networks that are not actor-oriented and focus explicitly on the rate at which a tie changes are the independent arcs, the reciprocity models (Wasserman, 1980; Leenders, 1995) and the pioneering work of Holland and Leinhardt (1977).

$$\frac{p(x \rightsquigarrow x^{\pm ij})}{p(x \rightsquigarrow x^{\pm ik})} = \exp(\beta_{reciprocity})$$

Because the interpretation of SAOM parameters is on the level of the mini-step, it is important that the modelled real-world process is reasonably approximated by a series of mini-steps. Under this assumption, the interpretation of SAOM parameters allows direct inference on whether particular social mechanisms, once translated into a network statistic, underlie the evolution of a network between multiple time-points.

Interpretation of the ERGM

To interpret the TERGM, it is instructive to review how parameters of the regular (cross-sectional) ERGM can be interpreted. For the ERGM it is common to interpret the model in terms of a micro-process (Lusher *et al.* 2013, Ch 3). That is, one can interpret the structural features of the network as resulting from biases towards particular types of tie-configurations in a hypothetical network formation process. For example, if there is a bias towards reciprocation, then whenever a tie-variable is being ‘re-evaluated’, it is more likely to remain or be created if it is or would be reciprocated. The underlying model casts the network as the result of a process of local evaluations or updates of ties, where a randomly chosen tie variable is considered for update and set to be present with probability

$$p(x \rightsquigarrow x^{+ij}) = \frac{\exp(\sum_k \theta_k z_k(x^{+ij}))}{\exp(\sum_k \theta_k z_k(x^{+ij})) + \exp(\sum_k \theta_k z_k(x^{-ij}))},$$

where $x^{+ij}(x^{-ij})$ is the network x where the tie from i to j is present (absent).

To accommodate this interpretation we must assume that the process is homogeneous and has continued for a sufficiently long time such that the initial state is irrelevant. In other words, to make this micro-level interpretation for a regular ERGM, one is forced to assume that the observed network is in an *equilibrium* state.

If a researcher is not willing to make this equilibrium assumption, the regular ERGM can still be interpreted in terms of over- or under-representation of specific substructures within a network, compared to what would be expected by chance. For example, a positive reciprocity parameter can be interpreted as a network having more reciprocated dyads compared to what would be expected at random, *controlling* for all other included statistics in the model. This second interpretation is equivalent to viewing ERGM as a log-linear model for dependent binary variables (Frank and Strauss, 1986). The dual interpretation option on the micro-level or on the network-level of the ERGM is a useful model feature because it extends the scenarios when ERGM interpretation is plausible. Is this the same for the TERGM?

Interpretation of the TERGM

The network-level interpretation of the basic TERGM is a model for a network at time t_m where, in addition to the dependencies that a regular ERGM can incorporate, one can also test whether ties that were present at t_{m-1} are more likely to be present at t_m . In this way, the interpretation does not differ from a model where we predict one network from another, e.g. predicting treaties between countries from their geographic proximity. In the TERGM, however, the additional network is the same set of relationships, only at a previous time point. Thus, the interpretation of, for example, a positive transitivity parameter in a TERGM would be that there is more transitivity than randomly expected, controlling for how the network looked in the past.

However, by controlling for a past version of the network, the relative importance of transitivity needs to be interpreted *net of* whichever mechanisms created and changed the network to its previous state. Following from the discussion in Section 4.1, the model for the network at $x(t_m)$ is agnostic about every aspect of the structure of $x(t_{m-1})$ other than the presence of individual tie-variables. Thus, if $x(t_{m-1})$ (the past) contains many of the same

dependencies as $x(t_m)$ (the present), the size of the parameters modelling these dependencies will reflect this. For example, an interpretation of a positive transitivity parameter might be that there is a tendency towards transitivity net of the past network *and* net of the tendency of ties to be transitive in the past. It is important to keep in mind that the TERGM's network-level interpretation of parameters does not pertain to change between two time-points, but to the structure of the second time point, controlling for the first.

Further, and worth *special emphasis*, is that a micro-level interpretation of the TERGM similar to the outlined interpretation of the regular ERGM is generally not possible. This is because the micro-level interpretation of the ERGM requires us to make the assumption that the network generating process is in equilibrium or, in other words, independent of its initial (past) state. A contradiction in the TERGM becomes clear here: To give the TERGM a process-interpretation would require that the network $x(t_m)$ is simultaneously in equilibrium (independent of the past) *and* dependent on the past network $x(t_{m-1})$. This logic is obviously inconsistent and thus a micro-level interpretation of the TERGM is impossible.

This shows that if we are interested in understanding how networks evolve and how ties come about on a micro-level, the TERGM is not a useful model, as parameters estimated from it only allow interpretation on the network level and cannot be interpreted as governing a process over time.

4.3. Dependence of parameters on time

To illustrate a further difference between auto-regressive and process-based models, we can draw on past statistical literature (summarised in e.g. Voelkle et al. 2012). This discusses that not only the size of the “stability-parameter” of auto-regressive models, but also most other parameters of these models depend on the time elapsed between two observations of the same system, in our case a network. This is closely related to the issue of interpretability discussed

in the previous section: that the parameters of a structural network need to be interpreted *net of* network features present at a previous time-point. In the remainder of this section, time-dependence is discussed referring first to parameters connected to time and, subsequently, to all other parameters.

Both the SAOM and the TERGM include parameters directly connected to how much time (meaning how many tie-changes) has passed between two observations. In the TERGM this is the “tie-stability” parameter or a transformation thereof; in the SAOM this is the rate-parameter⁸. A larger SAOM rate-parameter means more time has passed, while a smaller TERGM stability-parameter means the second network looks less like the first network, which allows the interpretation that more time has passed. How about other model parameters?

For the SAOM, more time/more change between observations results in more (simulated) instances in which an actor makes a tie-change and these changes are what is modelled. Thus, for example, a positive reciprocation parameter means that newly formed ties are more reciprocated than expected by chance and/or existing reciprocated ties are less likely to be broken. This means for more change between observations, we base the model parameters on more (hypothetical) decisions. Independence of parameter sizes on time elapsed is a consequence of modelling actors’ choices conditionally independent of the rate of change in the SAOM.

In contrast, TERGM parameter estimates are *not* independent of the amount of change between observations. Assume, for example, that we have an observed network $x(t_0)$ that exhibits a strong tendency towards reciprocation of ties, and some process by which ties change but the density and level of reciprocation remain constant. If this process starts from network

⁸ Assuming that we have a model in which the rate of change is constant between all actors.

$x(t_0)$ we can, after some time has passed, observe later realisations of the network $x(t_1)$, $x(t_2)$ etc.

If between $x(t_1)$ and $x(t_0)$ little change has happened, in an analysis of $x(t_1)|x(t_0)$ we will not only observe a strong stability parameter, but also a rather weak reciprocity parameter, as including $x(t_0)$ as a predictor, which is strongly reciprocated, will already explain a lot of the reciprocity in the network under analysis $x(t_1)$. However, the further we move away from $x(t_0)$ in time (read: changes in the network) the less reciprocation in network $x(t_s)$ will be explained by the persistence of reciprocity induced by the network $x(t_0)$ and the larger an estimated reciprocity parameter will be. This is because the reciprocity parameter will model the amount of reciprocation *not captured* by the stability parameter. As the amount of reciprocation not captured by tie-stability depends on the time elapsed, the reciprocity parameter itself will also depend on the time elapsed, where a longer inter-observation time will result in a stronger reciprocity parameter. As the same logic can be applied to all parameters in a TERGM, each one potentially depends on the time elapsed between two observations, *even if* the underlying process is time-homogeneous. The longer the time between the two observations, the less the second network will be explained by the first one and the more structural parameters will seem to matter.

The dependence of TERGM parameters on time also points to a more general observation: networks can only follow an assumed TERGM at specific points in time that are determined by a constant distance between observations. If the network at t_1 follows a TERGM conditional on t_0 , then a network at some later point e.g. t_2 does not follow any TERGM conditional on t_0 . This is because the network at t_1 enters the normalising constant, which means one cannot easily rewrite $p(x(t_2)|x(t_0))$ as a TERGM. Consequently one cannot, strictly speaking, estimate a TERGM for observations at t_{1+s} and t_0 (this is shown in detail in Appendix A.1.2), as t_{1+s} does not follow a TERGM distribution based on t_0 .

4.4. Empirical Illustration

The dependence of parameter size on time for the two models can be demonstrated using simulation-based analyses. One analysis shows the parameter consistency of the SAOM; the other analysis shows that parameters of a TERGM model are affected by the duration of the assumed process. Detailed description of the performed experiment can be found in Appendix A.2.1, in the current section only the intuition of the analyses and the results are presented.

The analysis proceeds in three steps. First, we estimate models based on empirical data to obtain realistic model parameters. The chosen data is a longitudinal friendship network collected in a school cohort in Glasgow in the 1990's from which we use two waves. It is well known and has been used previously in statistical analyses of network panel data (e.g. Steglich *et al.* 2010). Model specifications are based on recent literature and include terms that are equally available in RSiena and statnet, the software used for this comparison. The estimation of the same data is independently performed using a SAOM and a TERGM. For each of them a set of parameters is obtained. The estimated parameters are shown in Table 1 for either model. While the model specifications are not identical (they cannot be, see Block *et al.* (2016) and the discussion in Section 5), they include the same number of parameters that express the same types of interdependencies, i.e., fulfil the same functional role. The results are in line with what is known about networks of similar type. As these parameters are only estimated to obtain realistic parameter values for simulation, they will not be discussed further.

Table 1 around here

Second, those estimated model parameters are used to simulate 100 replicates of 10 waves of data each for SAOMs and TERGMs starting from the respective previous time-point. In both cases it is made sure that the generated sequences are similar in their amount of change to assure comparability. Note that two series of networks are simulated: one using the SAOM to generate a series of data and one using the TERGM to generate a series of data.

Third, we take a pair of two simulated networks that are between one and ten simulated periods apart and re-estimate models with these data. Each model only re-estimates the data that was produced with the same simulation method. Thus, the SAOM estimates data produced with a SAOM, and the TERGM estimates data produced with a TERGM. This is because we are interested in consistency of the models with regards to stability of parameters and want to exclude all other possible factors that can lead to diverging results.

Our theoretical discussions suggest that re-estimated parameters should be consistent with the simulation parameters in case of the SAOM, regardless of how far the observed networks are apart temporally, but that the distance between networks does affect the size of TERGM parameters.

The results of the re-estimation are shown in Figure 2. For each parameter, the red triangle is the original parameter used to simulate data, while the coloured points are the median re-estimated parameters after 1, 2, ..., 10 periods, respectively. Bars surrounding the points indicate the 90% of the range of parameter estimates in the repetitions. The estimated parameters of the SAOM show some stochastic variation over the different period lengths, but no systematic bias in one or the other direction. In contrast, however, most mean parameter estimates for the TERGM show a systematic and substantial bias in parameter estimates dependent on the number of simulated periods that lie between the two networks under analysis.

Further, after some periods the range of estimated parameters does not include the original data generating parameter anymore for four out of seven parameters, i.e. the data generating parameter is outside of the 90% interval of recovered parameters. However, not all parameters change with period length. This means that in an empirical analysis, depending on the time lag chosen, results will lead to quite different conclusions, in particular when comparing the relative strength of different parameters.

Figure 2 around here

This analysis shows that if we assume that some observed data are a result of a process that works in continuous time, an analysis of these data with a continuous-time model (SAOM) does not need to consider how far apart the network observations are, as parameter estimates are independent of elapsed time. Should we, however, assume the TERGM as a “data-generating process”, recording the network at the “correct” time-points is of crucial importance, as the estimated parameters are strongly biased otherwise. Arguably it is a herculean task to conclusively show that countries, for example, form treaties based on which treaties existed 1 year ago, but not 2 years ago (or 5 years, or 6 months...). However, without strong assumptions about this, parameter interpretation is ambiguous, as analysis of different time-points will lead to different results.

5. Model performance for tie prediction

After outlining issues of time-dependence of parameters that affect interpretation in the previous section, we now turn to discussing predictive power of the two models under analysis. We first discuss prediction of out-of-sample dependence structures more generally with the conclusion that a model specification should be available that gives a reasonable out-of-sample fit for either model. Then we test recent claims that TERGM provides greater predictive power on the tie-level and thus should be preferred for empirical analysis (Leifeld and Cranmer 2016). Our conclusion is that both models are weak predictively and that models for tie dependence are, for principled theoretical and statistical reasons, not suitable for tie-level predictions.

Cross-validation is a powerful technique for assessing the fit of a particular model to data. It is already common practice to evaluate model fit for ERGM (Hunter, Goodreau and Handcock 2008; Robins, Pattison, and Woolcock 2005) and SAOM (Schweinberger 2012; Lospinoso 2012) through simulating replicate data from the fitted model and comparing those data to the observations. For longitudinal models this can also be done in the form of predictive

distributions for out-of sample data (Koskinen and Snijders 2007). For dependent data, especially network data, this begs the question whether a model should perform well in terms of predicting dependence between ties at a future time point, or the presence of individual ties. If tie prediction is perfect, i.e., 100% accurate, it logically implies perfect prediction of all network dependencies. However, this implication breaks down in the absence of perfect tie prediction. What's more, reasonable prediction in terms of tie dependence need not derive from a model's high accuracy at predicting individual ties.

Overall, there seems to be a consensus in the field for the idea that the inferential task of network analysis is to uncover dependencies between network ties; in this view, out-of-sample prediction should be able to predict dependence in un-modelled observations rather than particular observations. The issue is well known more generally in statistics and the parallel to the so called Hamill (2001) forecast - where a forecast is “wrong” but “correct on average” - is particularly relevant. If you calibrate your model ‘marginally’ each prediction of a collection of events is unrealistic but you get the occurrence of each event correct on average; if you calibrate your model dynamically, each predicted outcome is realistic and whether you get the occurrence of each event correctly is not of interest (Gneiting *et al.* 2007). However, there are also some researchers that advocate for tie-prediction as a model-performance criterion.

When comparing model performance based on *a priori* model specification choices, we must be aware that this choice will likely influence the conclusions on which model performs better. Further, since they are quite different models, it is difficult to translate model specifications between the SAOM and (T)ERGM for comparing predictive power. Because one is actor-oriented and the other is tie-oriented, the SAOM parameter ‘transitive ties’ does not necessarily model the same as a ‘transitivity’ parameter in an (T)ERGM because the dependence implied differs (see Block et al. 2016). Therefore, model specifications would have to be selected independently, making impartial model comparison difficult due to the practically

infinite ways each model can be specified. Thus, we focus here more on the principle considerations of model performance, with special regards to some misconceptions in the literature.

Although assessing predictive power is an inherently practical task, we can derive two expectations on a purely theoretical basis (demonstrated empirically later). First, we can expect both models to perform poorly in tie-level prediction. Second, for the TERGM the inclusion of network terms in the model does not improve the expected performance of the predictive model, unless we take some information about the future into account.

5.1. Theoretical expectations

Both models typically analyse relatively sparse networks of medium size between a few dozen and a few hundred actors. For illustration purposes, let us assume a network of 100 nodes with 500 ties, which results in an average degree of 5. If we assume constant density and a turnover of half the ties (equalling a Jaccard network stability index of 0.33, see Ripley *et al.* 2016), the predictive task of a model would be to find the 250 ties that will be deleted and another 250 ties that will be created. The probability of an empty model to correctly predict a dropped tie will be 50% (250 out of 500), but the probability to correctly predict a tie that newly comes into existence is only about 2.5% (250 out of 9400). As we know, statistical models in the social sciences have notoriously low predictive power, even within sample, as exemplified by typically low R^2 values. Thus, even if we have an unusually good model, we will have difficulties to achieve predictions of tie creations that allow any degree of certainty. Based on probably any model, the answer to the question “will tie XYZ be formed in the future, given that it is absent now?” will always be “probably not (i.e. with a model-based probability of less than 50%)”. Thus, network based predictions about which countries will be the next to form a trade agreement, for example, will usually be bad. We show this in the empirical part by

demonstrating that the trivial prediction “the network in the future will look exactly like the network today” greatly and consistently outperforms either model in terms of both, precision (indicating false positives) and recall (indicating false negatives).

The second point mentioned above is more difficult to explain intuitively. It states that a basic TERGM, i.e. modelling tie dependence in addition to tie stability, does not in principle perform better than a logistic regression model that omits all structural network effects in terms of tie-level prediction. This is because the model is agnostic to where the modelled dependence structures are; the model parameter only ensures that a certain number of e.g. reciprocated dyads exist, not that they exist in the right place. Thus, we generally do not improve link prediction unless the inclusion of additional statistics puts further constraints on where these structures are. This can be illustrated on a simple example illustrated in Figure 3.

Figure 3 around here

Assumes that we estimate a TERGM including a density, a reciprocity and a tie stability term of $x(t_1)$ based on $x(t_0)$ (the training data) and use the estimated parameters to predict any of the networks $x_*(t_2)$ with $x(t_1)$ as a dyadic covariate (the hypothetical test data). It is clear that the process depicted in Figure 3 is time-homogeneous and that all networks have exactly the same statistics (5 ties, 1 reciprocated tie, 2 ties that existed in the previous wave). As a measure of tie-level prediction error we use the sum of squared residuals

$$D(p, x(t_{2,\cdot})) = \sum_{i,j} (\pi_{ij}^p - x_{ij,t})^2 ,$$

where p is a probability model (in our case the TERGM) and π_{ij}^p is the probability to observe the tie x_{ij} under this model.

Now an important observation is that $x_a(t_2)$, $x_b(t_2)$ and $x_c(t_2)$ are all equally likely under the TERGM. In other words, we have no reason to believe that either network will be

formed more likely than another, given our model. However, the tie-level prediction error between the three networks varies greatly. Compared to a logit model that only includes a density parameter and the previous time-point (i.e. the TERGM from above without structural effects), sometimes the predictive performance of the TERGM is better (e.g. $x_c(t_2)$), sometimes the logit performs better (e.g. $x_a(t_2)$) and sometimes performance is similar (e.g. $x_b(t_2)$). This is because some ties (e.g. from the upper right to the lower right node) are more likely to exist under the model, even though they were not there at the previous time point ($x_c(t_2)$). Thus, if a tie at a later time-point happens to be in this place, the TERGM prediction will be better than a logit model. However, under the TERGM, the network $x_a(t_2)$ is just as likely as $x_c(t_2)$, but here the newly emerged ties are in a location in which there is less predicted weight on this specific location (from the upper middle to the upper right node) – making the prediction of the TERGM worse than the prediction of the logit model. While sometimes one and sometimes the other model performs better, it is usually not possible to know beforehand whether in a particular case the TERGM will improve tie-level prediction. This is because the realisations of either network is equally likely under the model that the researcher chose – the TERGM⁹. This is closely linked to the observations in Section 4.1; the TERGM is only concerned with sufficient statistics, not with the location of where these ties are.

Readers familiar with multi-level modelling will see parallels in the outlined reasoning (see e.g. Snijders & Bosker 2012). In multilevel models, the inclusion of group level random terms cannot improve prediction of individual outcomes unless other explanatory variables with group-level variance are included in the model. This discussion illustrates that models designed to deal with interdependence between observations are not geared towards predicting (future)

⁹ It should be noted that this limitation does not hold true for process-based models (such as the SAOM), as these consider the embedding of ties in the updating process. This is because the probability to observe a certain network is not a function of network configurations, but of how likely the process that connects two time-points is. Thus, for example, reciprocated ties are more likely to remain than non-reciprocated ties.

outcomes of individual variable values, as this prediction, to make use of the dependence terms, requires knowledge of the (future) outcomes of other variable values. Since we cannot see partially into the future to gain some observations to anchor our inference about what value other, dependent observations may take, we are restricted to inference about the pattern those dependencies have and are likely to take.

5.2. Empirical Illustration

In this second empirical illustration of the article, we demonstrate that (these) network models that aim at modelling dependence structures in social networks are not well suited for predicting future outcomes. As above, we only present the intuition of the analyses and the results here; for detailed description of the experiment, refer to Appendix A2.2.

The empirical illustration is straightforward. We take three waves of empirical data from two different subject fields. First, we take one cohort from the ASSIST data (Steglich *et al.* 2012), a friendship networks among 80 adolescents, recorded at 3 time-points one year apart, which is representative of data typically analysed in the context of friendship studies. The network density is approximately constant over the three waves. The second dataset is composed of countries' bilateral fisheries treaties as used in Hollway and Koskinen (2016). Ties exist where a treaty concerning the allocation of shared fish stocks or access to fish stocks within the jurisdiction of only one of the parties has been concluded. While time-stamped data is available, we artificially create 3 waves of data in such a way that there is similar amount of change between waves 1 and 2 compared to between waves 2 and 3. The network evolution mainly consists of new treaties being formed and only few broken, i.e. the density increases over time. This is typical of many political science/international relations datasets.

Table 2 and 3 around here

Using these datasets, we fit a SAOM and a TERGM to the first two waves of the data, with model specifications mirroring specifications in current literature. Results of these estimations can be found in table 2 and 3. Based on the estimated models we simulate 1000 networks starting from wave 2 and evaluate the predictive power with regards to wave 3 based on the two criteria *precision* and *recall*. Precision measures the number of correctly predicted ties (simulated ties that are present in the empirical wave 3) over the total number of ties predicted by the simulation. Recall measures the number of correctly predicted ties over the total number of ties observed in wave 3. Perfect prediction would result in both indices being 1; a low value indicates poor prediction. We compare the distribution of precision and recall in 1000 simulated networks for the SAOM and the TERGM to the prediction and recall of a trivial model that assumes that the network in wave 3 will just be the same as the network in wave 2. This is known as the persistence technique in weather forecasting (the weather tomorrow will be like the weather today) – a usually not very accurate and certainly not very useful (yet simple) technique that is often considered as a baseline prediction. A further comparison is the simple logistic regression that excludes all dependence terms, but maintains the information on all covariates in the model.

Figure 4 around here

Results are presented in Figure 4. The y-axis denotes the value for precision and recall of the 1000 predictions for either model, the different models are stringed on the x-axis. It can be seen that neither model predicts better than the simple persistence model, suggesting that neither should be used for tie-level prediction. In fact, both models are, on average only marginally better than the logistic regression that does not use any tie-dependence for prediction and the 95% confidence intervals overlap in almost all cases for the different models. These results are in line with our intuition that predicting rare events, such as the creation of new ties in a sparse network is extremely difficult in a regression framework. In this light, we view any differences

in prediction between the TERGM and the SAOM as irrelevant. A further discussion of the results, including the differences between the logit prediction and the TERGM prediction can be found in Appendix A.2.2.

6. Discussion and Conclusion

Several approaches to analysing longitudinal network data have been proposed in the literature on statistical network modelling recently. In this article, we compared two of these models that are being applied by practical researchers – the SAOM and the TERGM. The former is a process-based, continuous-time model in which dependence unfolds over time, while in the latter, network dependence is modelled within an observation and decoupled from time. This means that the SAOM, and other process-based models, analyse *change* and network *evolution*. The TERGM, and models that treat time and tie dependence similarly, do not analyse change in the included dependence parameters, but can only show that some *structures* exist more than expected by chance, *controlling* for the past where the past carries some of the dependencies. Parameter sizes thus have no consistent micro-level interpretation. This paper shows that including multiple waves of network data in a statistical model is not sufficient to constitute a longitudinal network model allowing inference about change processes.

One implication from this discussion is that, when working with more than two data points, the TERGM (implicitly) requires equidistant observations. Otherwise, the interpretation of lagged predictors across data periods becomes rather incoherent. Deciding whether equidistant observations constitute network data where, between waves, the same amount of time has passed, or whether the same amount of network change has materialised is a further complication.

When considering predictive capabilities, the discussions in this paper suggest that (i) evaluating discrete-time statistical network models that include dependence terms within the

second time-point based on tie-level predictive power is debatable, as they generally do not improve over non-network models¹⁰. As an illustration of the modest predictive performance we show that (ii) both discussed network models perform poorly when used for tie-level prediction. We think this warrants the conclusion that predictive power is not a useful criterion for discriminating between these types of model. If one believes that prediction is a chief objective of modelling, one might look to a different strand of literature on link-prediction, originating in the physical sciences (see Lü & Zhao 2011) and not aiming at *explaining* processes of change. Today, experts in machine learning can efficiently solve such prediction tasks, but high predictive accuracy comes at the cost of little insight (Athey, 2017; Breiman, 2001).

Our conclusion that TERGM parameters strongly depend on the length of the interval between two waves have an interesting connection to the findings of Shalizi and Rinaldo (2013). They show that ERGM parameters for two networks of different size have to be different for most cases relevant to empirical researchers. These findings were previously suggested by e.g. Hunter and Handcock (2006) and equally apply to TERGMs and SAOMs. We add to these findings of non-scalability in terms of system size to by a temporal dimension. In a concise way, our findings show that the SAOM is temporally scalable, while the TERGM is not.

In this article we mainly focus on principle considerations of process-based and autoregressive models; thus, we neglect some further distinguishing features of these models. While not directly related to the theoretical foundations of the model that we focussed on here, there are further points that will be especially relevant to applied researchers. Future research should address these issues in more detail, for example how easily constraints imposed by the data can

¹⁰ These conclusions apply to the models as currently used in the literature; in case dependence terms that explicitly model the types of ties that are created or maintained are included, the conclusion are likely to differ.

be incorporated. Changing composition of a network happens regularly, especially in political sciences, where political actors are newly formed, split or dissolve between observations. For process-oriented models in which the passing of time is considered, changing composition can be taken into account by creating a new actor at some exogenously defined time-point during the modelled process. Discrete-time models currently lack a comparable solution.

This is tightly linked to modelling of multiple dependent variables, for example modelling a network and actor attributes interdependently, or modelling multiple relations between the same or interlocking set(s) of actors. A continuous-time approach can model the reciprocal influences between these dependent variables naturally, as the idea of mini-steps allows an understanding of how change in one variable can lead to change in another, potentially creating feedback-mechanisms. In discrete-time models it is not possible to take changes that occur between observations into account when considering how multiple variables mutually affect one another. Discrete-time models have difficulty instantiating *feedback processes* happening in-between observation moments, thereby missing a lot of what makes network research interesting and worth the effort.

Finally, some notes on processes that directly violate the assumption of stepwise updating of a network are warranted. This is the case, for example, when a group of actors decides jointly and interdependently to form ties amongst themselves or to a third party. Continuous-time models that are based on tie changes one at a time, like the SAOM, have no current implementation that allows this (it should be noted that the framework of the SAOM does not forbid such an extension, though). However, the TERGM has no clear advantage here either. The only dependencies that the TERGM affords are the standard ERGM dependencies and these are not well-equipped to explain ties forming as a result of coordinated action but would, in terms of interpretation, have to be interpreted in terms of how the ties formed conditionally on the others (like the SAOM, the ERGM does not generally forbid such extensions). Thus,

while the TERGM does not explicitly disallow coordinated action, it also has no way of accounting for coordinated action.

6.1. Advice for model selection

As outlined in the introduction, model performance in general can be evaluated on the basis of explanatory or predictive power. Recent model comparisons had considered relative out-of-sample predictive power a valid criterion of comparison (Leifeld & Cranmer 2016). However, such comparisons ignore the fact that both models suffer – much like most statistical models in use for studying social phenomena – from their inability to predict rare events such as the creation of new ties in a sparse network. Because a trivial persistence model consistently outperformed both longitudinal network models, we do not believe predictive power on the tie-level a useful criterion for model comparison. This empirical problem is in addition to the theoretical issues with evaluating models based on their predictive capabilities, as outlined in the introduction¹¹. This leaves explanatory power as a criterion for model selection.

Here the advice is very clear and follows directly from the discussion in Section 4. Is a researcher interested in explaining the evolution, i.e. change in a network between two time-points, or interested in explaining the structure of an observed network? In the former case we believe we have made a compelling case that process-based models, such as the SAOM (or the LERGM) are preferable, given that they directly model a process, which results in both consistent model parameters independent of the duration of the underlying process and a meaningful micro-level interpretation, allowing direct inference on underlying social mechanisms.

¹¹ However, this position might require further treatment if model specifications and empirical cases are found for which good predictive power of either model is achieved. In this case, comparing these models to further ones, for example different latent space models that makes use of a previous time-point and p2 models, is useful.

If a researcher is interested in explaining the structure of a network, the cross-sectional class of ERGMs is a well-established starting point. If additionally the previous states of the same network should be taken into account, the TERGM might be used, but its limitations in parameter interpretability and consistency should be considered. When a researcher is interested in whether modelled network dependencies are present in a network beyond dependencies that are solely a relic from the past, one could for example compare results of a TERGM and a cross-sectional ERGM of the second network.

Our critique of the TERGM's parameter interpretability and consistency does not mean that discrete-time network models in general should have no place whatsoever in the researcher's toolbox. In clearly round-based network-evolution processes, for example experimental research in lab-settings, or when knowledge about network ties is made public only at specific, equidistant times, treating this as evolving in discrete time may be warranted. However, for these types of data models like the proposed TERGM-variant by Hanneke, Fu & Xing (2010), might be used, where whether a tie exists in this round is *only* dependent on structures of the network in previous rounds, thus marrying time and tie dependence in a (maybe) more meaningful way.

6.2. Conclusion

Overall, the discussion in this paper shows that there is a mismatch between some of the research questions that are typically posited for longitudinal networks and the kind of substantive research questions that can be answered using a discrete-time model (such as TERGM). If our goal is to understand change in the world around us, we strongly advocate for statistical models that *model the processes* of change we observe in the world around us. This allows a direct test of theories in the social sciences and inference on how networks unfold over time. This is especially important, given the focus on using social mechanisms to explain

observed social phenomena. Whether these processes are best represented by actor-based models (such as the SAOM) or tie-based models (such as the LERGM) is a different matter. In case a researcher assumes a completely different process, not based on sequential, myopic changes, but based on for example co-ordinated or strategic action, we strongly advocate a principled approach that involves thinking about the assumed process and which established model represents this process best. There is, of course, always the possibility to develop a model tailored to specific applications and theoretical assumptions. We are looking forward to future developments that advance our understanding of network change processes of various kinds.

7. References

- Athey, Susan (2017). Beyond prediction: Using big data for policy problems. *Science* 355, 483–485.
- Block, P., Stadtfeld, C., & Snijders, T. A. B. (2016). Forms of Dependence: Comparing SAOMs and ERGMs from Basic Principles. *Sociological Methods & Research*, In Press.
- Breiman, Leo (2001). Statistical Modeling: The Two Cultures. *Statistical Science* 16(3), 199–231.
- Desmarais, B. A., & Cranmer, S. J. (2012). Micro-Level Interpretation of Exponential Random Graph Models with Application to Estuary Networks. *Policy Studies Journal*, 40(3), 402–434.
- Duijn, M. A. J. van, Snijders, T. A. B., & Zijlstra, B. J. (2004). p2: a random effects model with covariates for directed graphs. *Statistica Neerlandica*, 58(2), 234–254.
- Elster, J. (2007). Explaining social behavior: More nuts and bolts for the social sciences. Cambridge, UK: Cambridge University Press.
- Epstein, J. M. (2008). Why Model?. *Journal of Artificial Societies and Social Simulation*, 11(4), 12.

- Frank, O., & Strauss, D. (1986). Markov Graphs. *Journal of the American Statistical Association*, 81(395), 832–842.
- Friedman, M. (1953). The methodology of positive economics. In Friedman, M. (Ed) *Essays in positive economics* (pp. 210-244). Chicago: University of Chicago Press.
- Gneiting, T., Balabdaoui, F., & Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 69(2), 243–268.
- Hamill, T. M. (2001). Interpretation of Rank Histograms for Verifying Ensemble Forecasts. *Monthly Weather Review*, 129(3), 550–560.
- Hanneke, S., Fu, W., & Xing, E. P. (2010). Discrete temporal models of social networks. *Electronic Journal of Statistics*, 4, 585–605.
- Hedström, P. (2005). *Dissecting the social: On the principles of analytical sociology*. Cambridge, UK: Cambridge University Press.
- Holland, P. W., & Leinhardt, S. (1977). A dynamic model for social networks. *Journal of Mathematical Sociology*, 5(1), 5-20.
- Hollway, J., & Koskinen, J. (2016). Multilevel embeddedness: The case of the global fisheries governance complex. *Social Networks*, 44, 281–294.
- Hunter, D. R., Goodreau, S. M., & Handcock, M. S. (2008). Goodness of Fit of Social Network Models. *Journal of the American Statistical Association*, 103(481), 248–258.
- Hunter, D. R., & Handcock, M. S. (2006). Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics*, 15(3), 565–583.
- Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., & Morris, M. (2008). ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks. *Journal of Statistical Software*, 24(3).
- Jasso, G. (1988). Principles of Theoretical Analysis. *Sociological Theory*, 6(1), 1–20.

- Koskinen, J. H., & Snijders, T. A. B. (2007). Bayesian inference for dynamic social network data. *Journal of Statistical Planning and Inference*, 137(12), 3930–3938.
- Koskinen, J., Caimo, A., & Lomi, A. (2015). Simultaneous modeling of initial conditions and time heterogeneity in dynamic networks: An application to Foreign Direct Investments. *Network Science*, 3(1), 58–77.
- 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.
- Leenders, R. T. a. J. (1995). Models for network dynamics: A Markovian framework. *The Journal of Mathematical Sociology*, 20(1), 1–21.
- Leifeld, P., & Cranmer, S. J. (2016). *A Theoretical and Empirical Comparison of the Temporal Exponential Random Graph Model and the Stochastic Actor-Oriented Model*. working paper: <https://arxiv.org/pdf/1506.06696.pdf>.
- Lerner, J., Indlekofer, N., Nick, B., & Brandes, U. (2013). Conditional independence in dynamic networks. *Journal of Mathematical Psychology*, 57(6), 275–283.
- Lospinoso, J. A. (2012). *Statistical Methods for Social Network Dynamics*. Oxford: DPhil Thesis, University of Oxford.
- Lü, L., & Zhou, T. (2011). Link prediction in complex networks: A survey. *Physica A: Statistical Mechanics and Its Applications*, 390(6), 1150–1170.
- Lusher, D., Koskinen, J., & Robins, G. (2013). *Exponential Random Graph Models for Social Networks: Theories, Methods and Applications*. Cambridge, UK.: Cambridge University Press.
- McFarland, D. A., Moody, J., Diehl, D., Smith, J. A., & Thomas, R. J. (2014). Network Ecology and Adolescent Social Structure. *American Sociological Review*, 79(6), 1088–1121.
- Pattison, P. E., & Snijders, T. A. B. (2013). Modeling Social Networks: Next Steps. In D. Lusher, J. Koskinen, & G. L. Robins (Eds.), *Exponential random graph models for*

social networks: Theories, methods and applications (pp. 287–301). Cambridge, UK: Cambridge University Press.

Pattison, P., & Wasserman, S. (1999). Logit models and logistic regressions for social networks: {II}. Multivariate relations. *British Journal of Mathematical and Statistical Psychology*, 52(2), 169–193.

Ripley, R. M., Snijders, T. A. B., Boda, Z., Vörös, A., & Preciado, P. (2016). *Manual for RSiena*. Oxford: University of Oxford, Department of Statistics; Nuffield College.

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., & Woolcock, J. (2005). Small and Other Worlds: Global Network Structures from Local Processes. *American Journal of Sociology*, 110(4), 894–936.

Salter-Townshend, M., White, A., Gollini, I., & Murphy, T. B. (2012). Review of statistical network analysis: Models, algorithms, and software. *Statistical Analysis and Data Mining*, 5(4), 243–264.

Shalizi, C. R., & Rinaldo, A. (2013). Consistency under sampling of exponential random graph models. *Annals of Statistics*, 41(2), 508–535.

Schweinberger, M. (2012). Statistical modelling of network panel data: Goodness of fit. *British Journal of Mathematical and Statistical Psychology*, 65(2), 263–281.

Snijders, T. A. B. (2005). Models for Longitudinal Network Data. In P. Carrington, J. Scott, & S. Wasserman (Eds.), *Models and methods in social network analysis* (pp. 215–247). New York: Cambridge University Press.

Snijders, T. A. B. (2001). The Statistical Evaluation of Social Network Dynamics. *Sociological Methodology*, 31(1), 1–33.

Snijders, T. A. B., & Bosker, R. J. (2012). Multilevel analysis: An introduction to basic and advanced multilevel modeling. London: Sage.

- Snijders, T. A. B. and Koskinen, J. (2013). Longitudinal Models. In Lusher, D., Koskinen, J., and Robins, G. (Eds.), *Exponential random graph models for social networks: Theories, methods and applications*, pp. 130–140. Cambridge: Cambridge University Press.
- Snijders, T. A. B., Koskinen, J., & Schweinberger, M. (2010). Maximum Likelihood Estimation for Social Network Dynamics. *Annals of Applied Statistics*, 4(2), 567–588.
- Snijders, T. A. B., Pattison, P., Robins, G., & Handcock, M. S. (2006). New specifications for exponential random graph models. *Sociological Methodology*, 36(1), 99–153.
- Snijders, T. A. B., van de Bunt, G., & Steglich, C. E. G. (2010). Introduction to Stochastic Actor-based Models for Network Dynamics. *Social Networks*, 32(1), 44–60.
- Steglich, C. E. G., Snijders, T. A. B., & Pearson, M. (2010). Dynamic Networks And Behavior: Separating Selection From Influence. *Sociological Methodology*, 40, 329–393.
- Steglich, C. E. G., Sinclair, P., Holliday, J., & Moore, L. (2012). Actor-based analysis of peer influence in A Stop Smoking In Schools Trial (ASSIST). *Social Networks*, 34(3), 359–369.
- Voelkle, M. C., Oud, J. H. L., Davidov, E., & Schmidt, P. (2012). An SEM approach to continuous time modeling of panel data: relating authoritarianism and anomia. *Psychological Methods*, 17(2), 176–92.
- Wang, P., Robins, G. L., Pattison, P. E., & Koskinen, J. H. (2014). *MPNet*. Melbourne: Melbourne School of Psychological Sciences.
- Wasserman, S. (1980). Analyzing Social Networks as Stochastic Processes. *Journal of the American Statistical Association*, 75(370), 280–294.

8. Tables and figures

SAOM		TERGM	
Effect	<i>param.</i>	<i>param.</i>	Effect
Outdegree	-2.43	-5.38	Edges
Reciprocity	1.35	1.41	Reciprocity
GWESP transitive	0.91	0.40	GWESP transitive
GWESP cyclic	0.34	0.04	GWESP cyclic
Indegree Popularity (sqrt)	0.02	-0.38	Indegree GW
Two-Paths (OutPop) (sqrt)	-0.45	-0.09	Two-paths GW
Outdegree Activity (sqrt)	-0.09	0.10	Outdegree GW
Same Sex	1.18	1.23	Same Sex
Rate	2.63	4.65	Tie-Stability

Table 1: Parameters used for forwards Simulation of the SAOM and TERGM

SAOM				TERGM			
	<i>est</i>		<i>s.e.</i>	<i>est.</i>		<i>s.e.</i>	
Rate	1.12	***	(0.28)	8.23	***	(1.10)	Stability
Density	-1.97	**	(0.66)	-15.76	***	(3.30)	Density
GWESP	1.13	**	(0.38)	0.73	***	(0.20)	Alternating Triads
4-Cycles	-0.11		(0.11)	0.01		(0.07)	4-Cycles
Indegree Popularity (sqrt)	0.87	*	(0.31)	-0.27		(0.27)	Alternating In-stars
Contiguity	2.05	***	(0.38)	2.35	***	(0.35)	Contiguity
Same IGOS	0.03		(0.02)	0.03		(0.02)	Same IGOS
Trade (log)	-0.16		(0.10)	-0.20	*	(0.08)	Trade (log)
GDP	2.47	*	(1.19)	2.44	**	(0.77)	GDP
GDP Simimlarity	-0.90		(1.04)	1.59		(0.95)	GDP Difference
GDP PC	-0.25		(0.20)	-0.04		(0.12)	GDP PC
GDP PC Similarity	-0.63		(0.71)	0.14		(0.18)	GDP PC Difference

Table 2: Estimated Parameters used for out-of-sample prediction for the fish data

SAOM				TERGM			
	<i>est</i>		<i>s.e.</i>	<i>est.</i>		<i>s.e.</i>	
Rate	10.02	***	(1.20)	1.90	***	(0.14)	Stability
Density	-2.20	***	(0.33)	-5.49	***	(0.66)	Density
Reciprocity	2.60	***	(0.28)	2.36	***	(0.30)	Reciprocity
GWESP trans	1.89	***	(0.16)	1.35	***	(0.10)	Alternating Trans. Trip
GWESP trans * Recip.	-1.12	***	(0.33)	-0.42	***	(0.09)	Alternating Cyclic. Trip
Indegree Popularity (sqrt)	-0.06		(0.16)	0.25		(0.32)	Alternating In-stars
Outdeg. Popularity (sqrt)	-0.50	**	(0.17)	1.18	**	(0.35)	Alternating Two-Paths
Sex Alter	0.13		(0.14)	-0.51	*	(0.20)	Sex Alter
Sex Ego	-0.10		(0.16)	-0.81	**	(0.27)	Sex Ego
Same Sex	0.46	**	(0.15)	1.40	***	(0.32)	Sex Interaction

Table 3: Estimated Parameters used for out-of-sample prediction for the ASSIST data

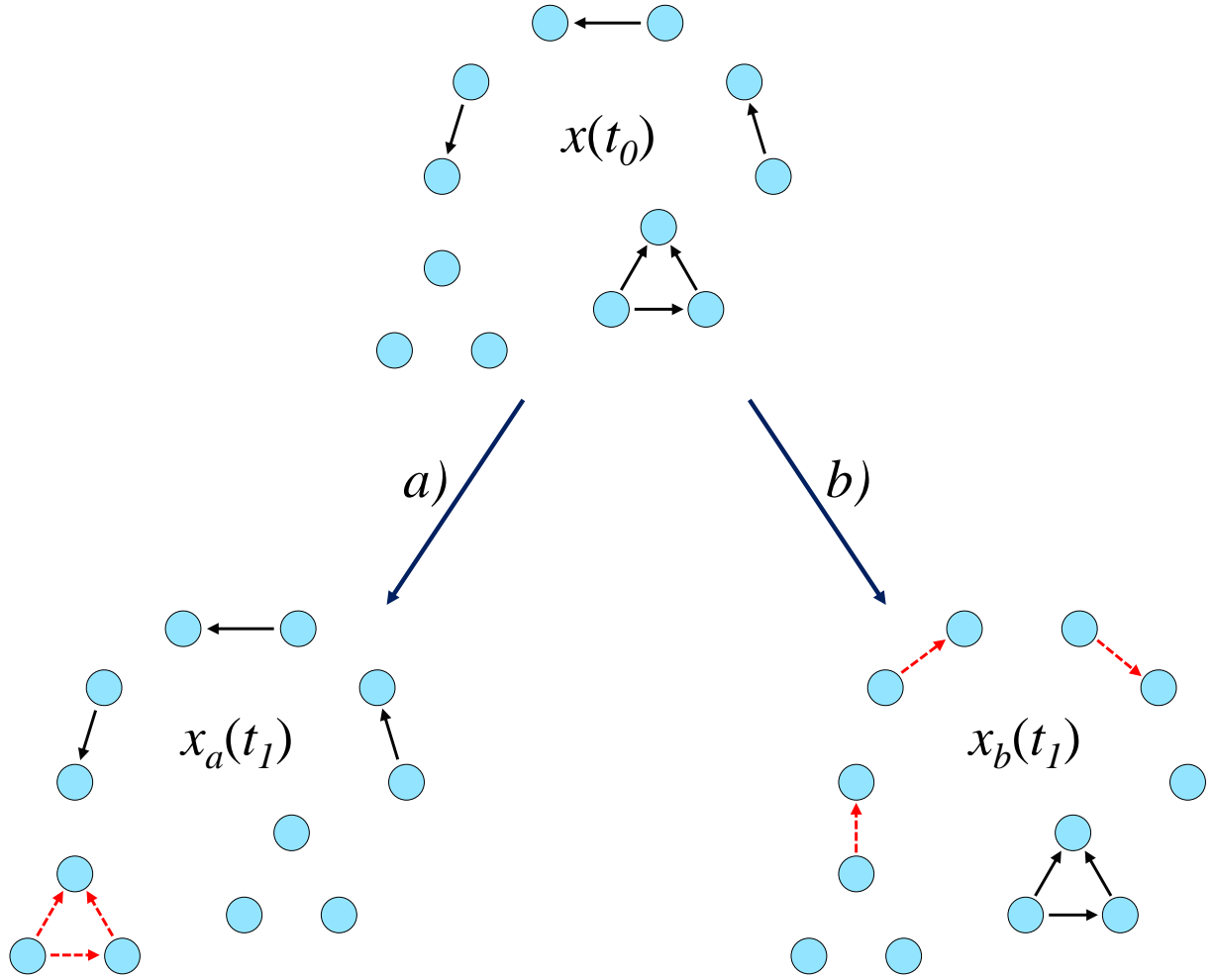


Figure 1: Two potential network sequences

Notes: Example of two network realisations $x_a(t_1)$ and $x_b(t_1)$ that are equally likely under the TERGM but imply different change processes for the SAOM given $x(t_0)$. Either network contains 6 ties, 1 transitive triplet and 3 ties change. Solid black arrows indicate stable ties, dotted red arrows indicate changing ties.

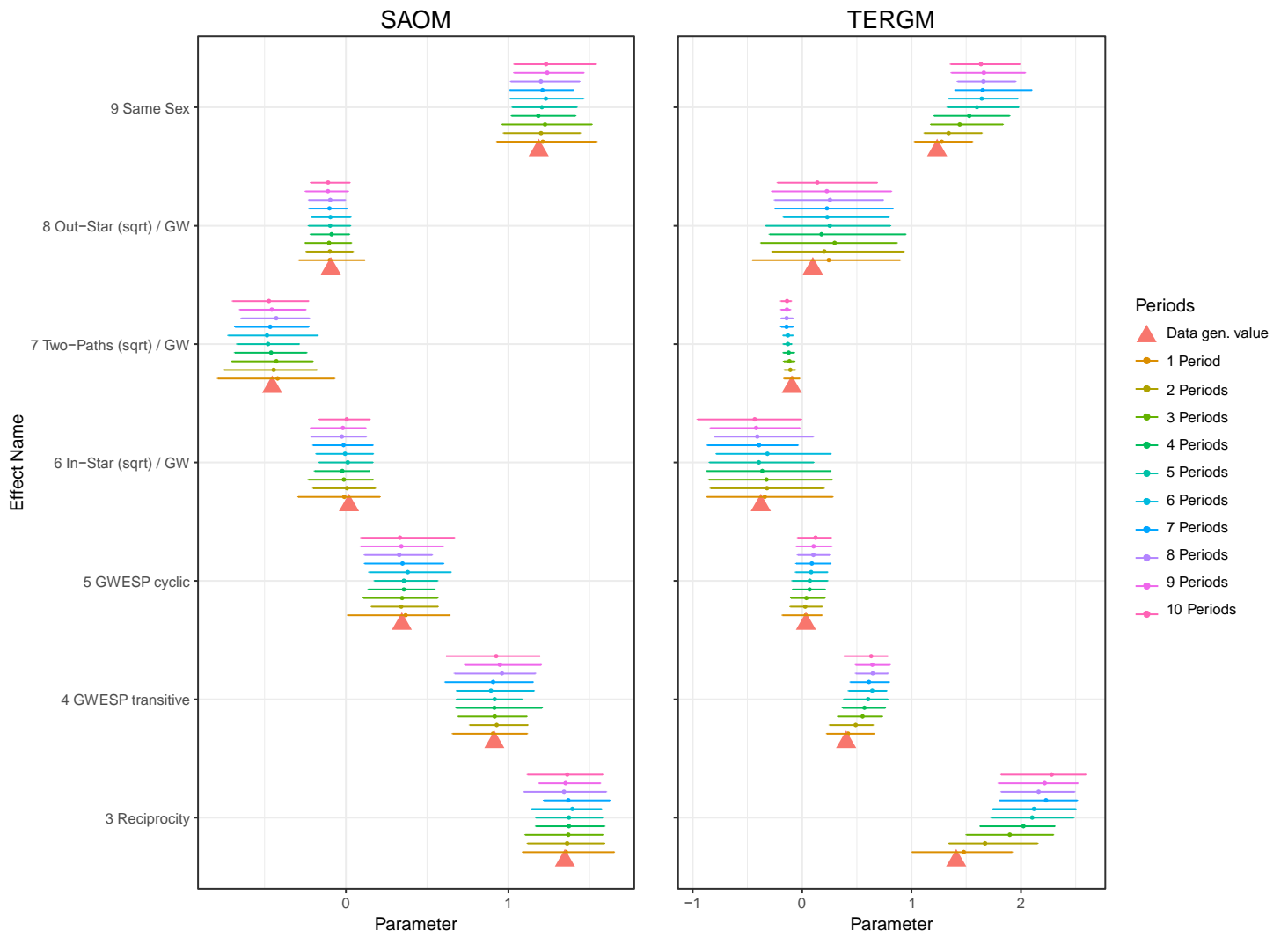


Figure 2: Estimated Parameters and 90% range of re-estimation of the Glasgow data.

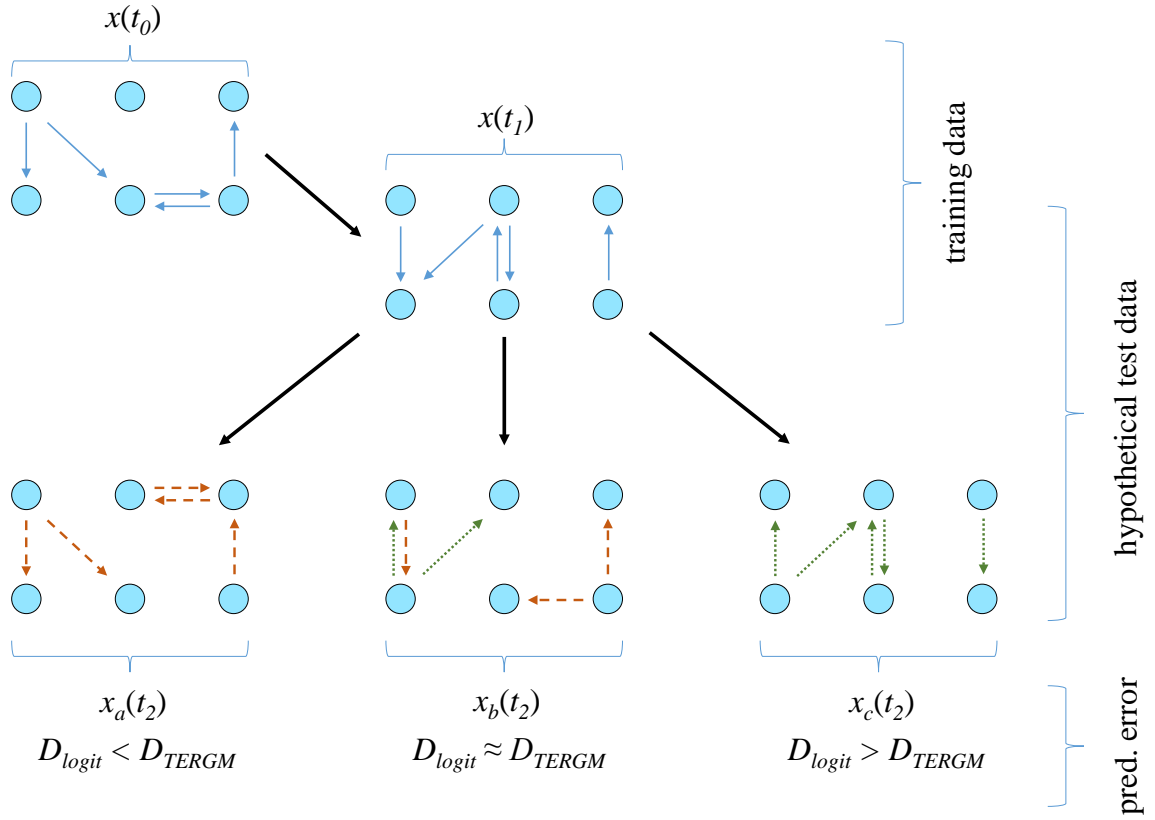


Figure 3: Potential series of networks at 3 time-points.

Notes: The figure depicts the hypothetical scenario in which a network model is estimated for the evolution from t_0 to t_1 and these parameters are used to predict the network at a future time-point t_2 . Each of the three potential networks $x_*(t_2)$ are equally likely under a TERGM as well as under a logit model, but the tie-level prediction error D differs by network realisation. Brown dashed arrows depict ties that are more likely under the logit model, green dotted arrows depict ties that are more likely under the TERGM.

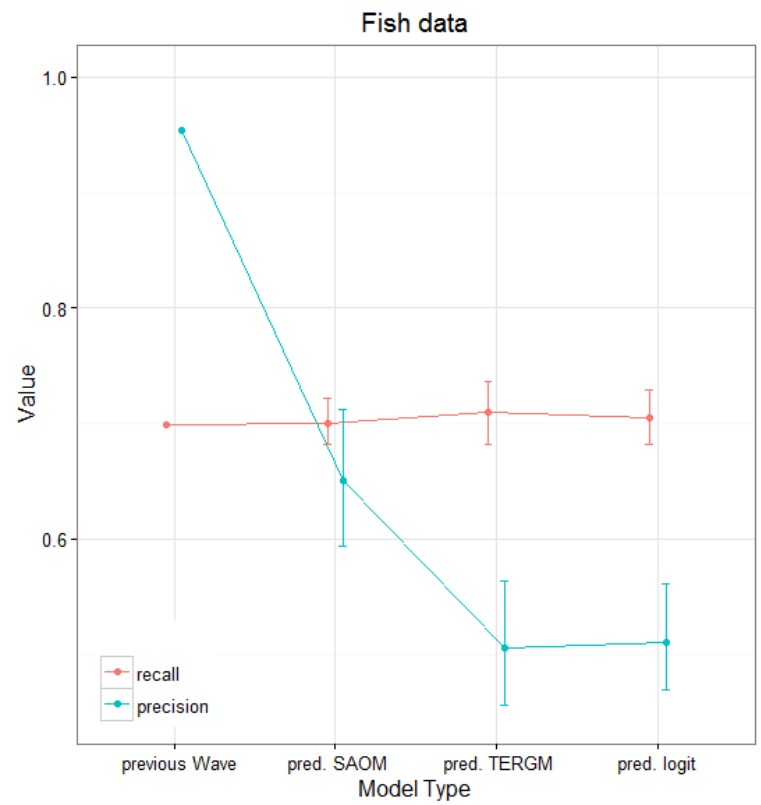
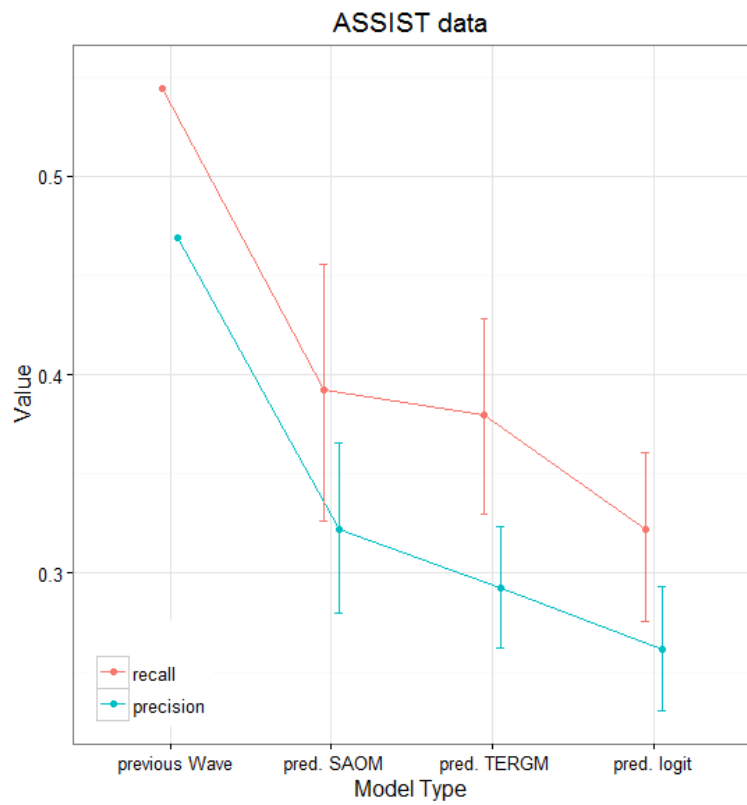


Figure 4: Predictive performance for out-of-sample tests of SAOM and TERGM for ASSIST data and Fish data.

A.1. Appendix: Mathematical elaborations

A.1.1 Equivalence of TERGM formulations

In this section we show that the TERGM including a “dyadic stability” term is identical to a TERGM that includes the previous time-point as a dyadic covariate, as proposed in Section 2.1. We assume that the basic form of a TERGM includes the dyadic stability term as the statistic (as proposed in Leifeld and Cranmer 2016)

$$d(x(t_m), x(t_{m-1})) = \sum_{i,j} [x_{ij}(t_m)x_{ij}(t_{m-1}) + (1 - x_{ij}(t_m))(1 - x_{ij}(t_{m-1}))],$$

together with the ERGM terms $z(x(t_m))$, with density parameter θ_L and stability parameter θ_D , where $x(t_m)$ is the network at time m and x_{ij} denotes the tie variable from i to j . This model is mathematically identical to estimating an ERGM with statistics $z(x(t_m))$ and using $x(t_{m-1})$ as a dyadic covariate with statistic

$$d(x(t_m), x(t_{m-1})) = \sum_{i,j} x_{ij}(t_m)x_{ij}(t_{m-1}),$$

and density and stability parameters $\tilde{\theta}_L = \theta_L - \theta_D$ and $\tilde{\theta}_D = 2\theta_D$, respectively. Consequently, the formulation of a TERGM that includes a “dyadic stability” term can be interpreted as a cross-sectional ERGM for $x(t_m)$ with $x(t_{m-1})$ as a dyadic covariate network.

A.1.2 Consistency of ERGM with lagged covariate network

Assume that we have time-points t_0, t_1, \dots, t_M and the network at these points follows conditional ERG models, $p_\theta(x(t_m)|x(t_{m-1}))$, where dependence between waves is captured by a stability term. The pmf for the fully observed sequence is

$$\begin{aligned}
p_{\theta}(x(t_1), x(t_2), \dots, x(t_M) | x(t_0)) &= \prod_{m=1}^M p_{\theta}(x(t_m) | x(t_{m-1})) \\
&= \prod_{m=1}^M \frac{\exp\{\theta^T z(x(t_m)) + \theta_D d(x(t_m), x(t_{m-1}))\}}{\sum_{y \in \mathcal{X}} \exp\{\theta^T z(y) + \theta_D d(y, x(t_{m-1}))\}},
\end{aligned}$$

where θ_D is a stability parameter as in A.1.1. To gain further insight into what properties this conditional model has in terms of panel data, consider what distribution data have at each point in time. The distribution of $X(t_M)$ for any t_M (given $x(t_0)$) is given by the mixture distribution

$$p_{\theta}(x(t_M) | x(t_0)) = \sum_{(x(t_1), x(t_2), \dots, x(t_{M-1})) \in \mathcal{X}^{M-1}} p_{\theta}(x(t_1), x(t_2), \dots, x(t_M) | x(t_0)),$$

where the sum is taken over all sequences of graphs. Thus, while $p_{\theta}(x(t_m) | x(t_{m-1}))$ follows an ERG model conditionally, a graph $x(t_m)$ only follows an ERG model conditionally on $x(t_0)$ if $\theta_D = 0$. If $\theta_D \neq 0$, the intermediate networks enter in the normalising constant of the pmf of $x(t_m)$.

For example, the log-odds for a reciprocated dyad (1,1) against an asymmetric dyad (0,1) at t_1 for a reciprocity model with parameters for edges θ_L , reciprocation ρ , and stability θ_D is $\theta_L + \rho$ irrespective of whether the corresponding dyad at t_0 was equal to (0,1) or (0,0). The ratio of these log-odds (reciprocated divided by asymmetric) for t_0 equals (0,1) and t_0 equals (0,0), and therefore must equal 1 for the TERGM. Figure A1 plots this odds-ratio under the distribution (based on the exact calculation) at time-point t_2 conditional on the network at time t_0 . The odds-ratio is the same as for a TERGM, i.e. is equal to 1, only when $\theta_D = 0$. The interpretation is that stability amplifies reciprocity and will lead to exaggerated effects at $x(t_k)$. Further, it shows that $p_{\theta}(x(t_2) | x(t_0))$ does not follow a TERGM for any stability parameter not equal to 0.

It is also straightforward to show that the limiting distribution for an inhomogenous Bernoulli graph with stability parameter θ_D is a homogenous Bernoulli graph with

$$\Pr(X_{ij} = 1) = \frac{e^{\theta_L}(1 + e^{\theta_L + \theta_D})}{1 + e^{\theta_L}(2 + e^{\theta_L + \theta_D})}$$

The interpretation is that all tie-variables have the same probability but if $\theta_D > 0$ this probability is greater than $\frac{e^{\theta_L}}{(1+e^{\theta_L})}$.

Figure A1 around here

While the limiting behaviour might not be of great interest for panel-ERGM, it is important to carefully consider how the observed panel wave time points were chosen. If one assumes that the networks follow conditional ERG models for the discrete time-points t_0, t_1, \dots, t_M , then one cannot fit the model to t_0, t_2, \dots, t_M nor to $t_0, t_{0.5}, t_1, \dots, t_M$, should more data become available. The time between observations not only affects similarity between the networks, as captured by $d(\cdot, \cdot)$, but the networks are only ERG models at very specific time-points. To assure that ERGM dependencies can be modelled at any time-point changes need to be modelled using a process in continuous-time. An example could be modelling time between changes as a Poisson process and model changes according to their conditional probabilities according to the ERGM as in Snijders and Koskinen (2013).

A.2. Appendix: Description of experiments

A.2.1 Time-dependence of parameters

Section 4.3 analysed whether parameters of the TERGM and the SAOM are dependent on the process time between two subsequent observations. It was argued theoretically that a

dependence on process time is expected in autoregressive models but not in process-based models. This appendix details the simulation-based analysis strategy that was applied.

The analysis worked in three steps. First, we *estimated* TERGM and SAOM models from empirical data. Second, the estimated models were used to *simulate* sequences from the SAOM and from the TERGM. Third, we *re-estimated* SAOM and TERGM parameters from the simulated sequences and compared ten cases: one case where the length of the simulated process time is equal to the process time of the original estimated data, and nine cases where the process time was increased by a factor of two to ten as compared to the first case.

A.2.1.1 Step 1: Estimation of a SAOM and a TERGM from empirical data

To estimate parameters of a SAOM and a TERGM we used a sequence of two networks that are based on empirical “Glasgow data” collected by Mitchell and West (1996). The original data set consists of a sequence of three friendship networks between 160 school children (84 of which are male; mean network density = 2.1%; original Jaccard index between waves 1 and 2 = 46.3%). The data have been used in a number of longitudinal network studies (e.g., Steglich et al. 2010, Stadtfeld & Lomi 2014) where further descriptions can be found. It is publically available on the SIENA website (<http://www.stats.ox.ac.uk/~snijders/siena/>). The data set was simplified in a number of ways. First, we only used a subset of 129 individuals that are present through all data collection waves to avoid missing data. Second, we artificially added and dropped ties from the second data collection wave so that both waves had the same density (2.7 percent) and a Jaccard distance of 0.75. These two adjustments made sure that the simulation models in step 2 would not easily suffer from near-degeneracy problems. The density of the second network was set by randomly adding ties that exclusively existed in wave 1 to the (sparser) second network until the density was the same. Then the Jaccard index was set to approximately 0.75 by dropping ties $i \rightarrow j$ in the second network that were not present in the first wave network and simultaneously adding ties $k \rightarrow l$ that were present in the first wave network

but not in the second wave network. This replacement was repeated until the Jaccard index was above 0,75. It is important to note that the artificial changes to the original networks were made solely to achieve parameters that produce stable network sequences. Results in this experiment for either model were not negatively influenced by these changes but are instead more reliable as we rule out unrelated problems such as model degeneracy.

In both models we used a similar parameter specification. The basic endogenous network effects outdegree (density) and reciprocity were included. We further added geometrically weighted versions of a transitive triad and a three-cycle structure, and counts of two paths, in-stars and out-stars (Snijders *et al.* 2006). The latter three effects were included as geometrically weighted versions in case of the TERGM and as square-rooted versions in the SAOM as geometrically weighted versions of effects are not available in the RSiena software. In the SAOM, to model two paths the Outdegree-popularity effect was chosen. The selection of these effects aimed at producing models that do not easily suffer from degeneracy when simulated. A gender homophily parameter was added as the only covariate-related effect. The SAOM and TERGM parameters were estimated from the two data waves described above. The estimates are found in Table 1.

A.2.1.2 Step 2: Simulation of network sequences using the SAOM and TERGM

From both empirically informed models, 100 sequences of networks were simulated using the estimated parameters in Table 1. The SAOM is a process model and can thus be straightforwardly simulated. The forward simulation of the SAOM was started from the empirically observed first network of the Glasgow data. A sequence of random mini-steps was generated based on the model specification using the RSiena software (Ripley et al., 2016, section 9). To generate networks in a distance of two to ten, the rate parameter was multiplied by the corresponding factor.

The ERGM is not a model for change processes but a probability model for networks (and so is the TERGM variant). The probability distribution of networks can be approximated with an MCMC simulation algorithm that traverses the space of all possible networks, given this algorithm runs for a very long time. These chains can be used to sample networks from the ERGM distribution, which fulfils the function of simulating into the future with a TERGM. To reduce auto correlation between subsequently sampled networks and a dependence on the starting network we choose a burn-in parameter of 100,000 and a thinning parameter of 100,000. We used the simulate function of the statnet package in R (Hunter *et al.* 2008). MCMC simulations, and details about burn-in and thinning are further explicated in Lusher et al. (2013, chapter 12).

From both the SAOM and the TERGM, ten waves of data were simulated with 100 observations each. To simulate the 100 networks at the second time-point, in the SAOM the original network was used as a starting point, in the TERGM the original network was used as a dyadic covariate. Subsequently, in case of the TERGM the k -th network in the m -th wave was sampled by using a k -th sampled network of wave $(m-1)$ as a dyadic covariate (the persistence covariate).

A.2.1.3 Step 3: Re-estimation from simulated data

In a final step, we re-estimated 100 models from simulated networks that for each of the periods that are one to ten waves apart. For each re-estimation we used the first data collection wave as a starting point (SAOM), or a dyadic covariate (TERGM), respectively. The model specifications were exactly the same as the ones in Step 1 (estimation of a SAOM and TERGM from empirical data). The distribution of the estimates¹⁴ that we generated in Step 3 were

¹⁴ Especially in the later waves, some estimations did not converge, however, their number was small at each time-points (<10). Thus, each estimates are based on at least 90 estimates results.

compared with the empirical parameters of Step 1 that were used in the network simulation to test whether parameters are consistent within a model. Results are shown in Figure 2 and discussed in section 4.4. As outlined there, the SAOM parameters are consistent and independent of the amount of time passed between waves, while in the TERGM the parameter size of several parameters clearly depends on the amount of time passed. This is true even if we assume a TERGM process underlying the evolution of the data, as in this experiment.

A.2.2 Predictive capabilities of the SAOM and the TERGM

In the second empirical illustration, we tested in how far the SAOM and the TERGM are capable to predict future network realizations (out-of-sample prediction). We compared their predictive power to a simple persistency model that predicts the future by assuming no change from the past, and a logit model that takes no network dependencies into account.

A.2.2.1 Data

The first dataset we use comes from the *A Stop Smoking In Schools Trial (ASSIST)*. This large dataset was collected in the context of a school-based intervention study. We use a cohort from one school of this dataset for which panel network data was collected. The cohort contains 80 adolescents (41% girls) aged 12-13 in the first wave. Three waves of data one year apart were gathered. Friendship data was collected using a name generator allowing up to 6 nominations per individual; further, many demographic and behavioural variables of the participants were collected; only sex is used in this empirical analyses. This data has been used in multiple empirical studies to date (e.g. Steglich *et al.* 2012) and appears representative of friendship networks among adolescents. The reason to choose this specific school cohort from the larger dataset is that (i) the network size is small enough to allow straightforward analysis, but (ii) large enough for this illustration. Further, (iii) the Jaccard index, indicating stability

between consecutive wave for wave 1 and wave 2 is identical to the Jaccard index for wave 2 and wave 3 (0.35), which is a useful feature when assessing predictive power.

The second dataset we use combines bilateral fisheries agreements retrieved from two sources, ECOLEX (2011) and the International Environmental Agreement (IEA) database of Mitchell (2013), as outlined in Hollway and Koskinen (2016). This results in a comprehensive archive of bilateral fisheries treaties, in which start and end date of each recorded treaty are known. We further use a set of exogenous predictors inspired by the analyses of Kinne (2014), to obtain a reasonable model specification. The monadic covariates included are GDP and GDP per capita, with the data being recovered from the World Bank. As dyadic covariates we include trade between countries (Barbieri and Keshk 2012), contiguity (Stinnet *et al.* 2002), and number of shared intergovernmental organizations (IGOs; Pevehouse, Nordstrom, and Warnke 2004).

To use this data for analysis with both, TERGMs and SAOMs, network panels have to be created. An undirected, binary tie in the network at any of the time-points is present if at least one treaty between two countries was existent then. To create the panels, two questions need to be addressed; which countries to include and when to set the time-points to create the panel data. To the former, we include all countries that had a fisheries treaty at any time-point in our dataset, which amounts to a total of 97 countries. To the latter, we are interested in finding a period in which tie formation can be assumed to be (more or less) time-homogeneous over three waves. Further, as the network tends to increase in density, we want two periods in which the network changes the same number of ties, which, we argue, is closest to equidistant observations (where distance is measured by how much happened, rather than by how much time has passed). These considerations lead to using the networks in 1968, 1977 and 1984. The changes in the number of ties are 98 and 100, respectively. The covariates used in the analyses are measured at the same time-point, except for information on IGOS, which is available for the years 1965 and 1975.

A.2.2.2 Model estimation

For the adolescent friendship data, the TERGM was estimated using MPNet (Wang *et al.* 2014) with the network at the second time point as the dependent variable and the first wave as a dyadic covariate. The SAOM was estimated using RSiena (Ripley *et al.* 2016), analysing the evolution from the first to the second wave. The model specification for the SAOM was taken from recent literature on this dataset and the specification for the TERGM was chosen to closely mirror this by effects that fulfil the same function. As the included parameters reflect those in Appendix A.2.1, they are not further discussed here. Results from the model estimations can be found in Table 3. They show no major surprises, but, in line with previous studies, find tendencies towards clustering, reciprocation and sex homophily, but against cyclic triplets / reciprocation in triplets. Estimates for two-paths and main effects of sex are inconsistent across models. However, for the effects related to sex this is mainly due to the covariate sex not being centred in the TERGM analysis.

For the international treaty data, both models were estimated using the period from 1968 to 1977. The TERGM was estimated as an undirected network using MPNet with the network of 1977 as the dependent network and the network of 1968 as a dyadic covariate. The SAOM was estimated using RSiena, analysing the evolution between these networks. There are multiple options to estimate undirected networks in RSiena; we chose “option 3”, which assumes a two-step process of unilateral proposal by one actor and acceptance or rejection by the other actor for tie creation; tie deletion is a unilateral decision. Details on RSiena for undirected networks are found in Snijders and Pickup (2017).

Model specification was inspired by Kinne (2014) as well as by Hollway and Koskinen (2016), to reflect contemporary model specification. As structural effects both models include a transitivity parameter, modelled as Alternating Triads in the TERGM and as geometrically weighted edgewise shared partners (GWESP) in the SAOM, a parameter for preferential

attachment, modelled in the TERGM by alternating stars and in the SAOM as a square-rooted Indegree Popularity parameter, and a four-cycle parameter. As dyadic covariates, contiguity, number of shared IGOS and trade (as log of trade volume between countries) are included. For the monadic covariates GDP and GDP per capita the main effect and a dyadic similarity effect are included. The Results of the estimations can be found in Table 2. The model shows that there is a tendency towards clustering and forming treaties with contiguous countries, as well as a tendency of high GDP countries to be involved in more treaties. Results on preferential attachment are additionally positive on the SAOM, while trade between countries is significant only in the TERGM.

A.2.2.3 Generating out-of-sample predictions

The estimated parameters of both datasets for both models were used to generate out-of-sample predictions that could be compared to the respective third wave of data. For the SAOM, once again RSiena was used, with its standard simulation routine that took all previously estimated parameters, and the network at the second time-point for either dataset as a starting value. For the TERGM, the simulation routine implemented in MPNet was used, with the previously estimated TERGM parameters and the second network as a dyadic covariate. For each model and dataset 1000 simulated networks were produced. Additionally, predictions from a logit-model were simulated to have a further comparison case. As the TERGM without any structural effects is a logit-model, we equally used MPNet for this simulation, retaining all covariates and the previous time-point in the model, but excluding structural effects (using, of course, re-estimated parameters from analyses excluding structural effects). 1000 networks for either dataset were simulated using this method.

A.2.2.4 Results

We applied two criteria to assess predictive power. *Precision* evaluates the ratio of correctly predicted ties over all ties predicted to be present in the out-of-sample network, *recall* evaluates the ratio of correctly predicted ties over all ties observed in the out-of-sample network. A perfect prediction model will have values of 1 for both precision and recall. Trivial prediction models can typically easily maximize one of the two parameters. For example, predicting all ties will be present in a network leads to perfect recall but the precision will be poor. Precision and recall were evaluated for each of the 1000 simulations and the mean values of each combination of method and data calculated. The predictions for each method were further compared to the persistence method, which proposes that the networks at the third time-point looks exactly like the network at the second time-point.

Results can be found in Figure 4. For the ASSIST data on the left hand side it can be seen that the persistence model vastly outperforms both the SAOM and the TERGM. The differences between the SAOM and TERGM are rather small. Given that they could result from the model specification to favour one model over the other and the plotted 95% confidence intervals overlap, we disregard them here and only conclude that neither model has noteworthy predictive power.

For the data on international fisheries treaties the picture is somewhat different but the conclusions are very similar. Given that the network mainly grows between all waves, both the SAOM and the TERGM perform just as well in terms of recall compared to the persistence model (however, following the principle of Ockham's razor the persistence model would still be preferable, as it predicts equally well with a much simpler model). This is because all models retain all or most of the ties present in wave 2, which is in accordance with the data and directly results in a good recall. However, the ties that are newly predicted to be formed in each model are very inaccurate, shown by the low precision value of either model. The differences between

the TERGM and SAOM here would likely require a deeper treatment of the process of undirected tie formation in the SAOM; however, as our goal is only to show that neither model can predict the future with any accuracy, we abstain from speculation about this difference. Overall, we believe that these results clearly indicate that neither of the models has a reasonable capability to predict the future that would warrant to use it as a predictive model, at least using a standard model specification as found in the literature.

Finally, some notes on the differences in predictive power between the TERGM and the simple logistic model are in order. The extent to which the former can improve on the latter depends on how ‘structured’ the first time-point is in a non-trivial way, i.e. putting constraints on where structures can be. The argument in Section 5 still holds true, in the sense that the ties ‘do not care’ where they end up in the network as long as the overall statistics match the observed network. As is evident from Figure 4, there is a very slight difference for the ASSIST data set and none at all for the fisheries data set.

A.3. Appendix: Additional references in the appendix

Barbieri, K., & Keshk, O. (2012). Correlates of War Project Trade Data Set Codebook, Version 3.0. <http://correlatesofwar.org>.

ECOLEX. (2011). The Gateway to Environmental Law. <http://www.ecolex.org/start.php>.

Kinne, B. J. (2014). Dependent Diplomacy: Signaling, strategy, and prestige in the diplomatic network. *International Studies Quarterly*, 58(2), 247–259.

Mitchell, R. B. (2013). International Environmental Agreements Database Project (Version 2013.2). University of Oregon.

Michell, L. & West, P. (1996) Peer pressure to smoke: the meaning depends on the method. *Health Education Research*, 11, 39-49.

Pevehouse, J., Nordstrom, T., & Warnke, K. (2004). The Correlates of War 2 International Governmental Organizations Data Version 2.0. *Conflict Management and Peace Science*, 21(2), 101–119.

Snijders, T. A. B., & Pickup, M. (2017). Stochastic Actor-Oriented Models for Network Dynamics. In J. N. Victor, A. H. Montgomery, & M. Lubell (Eds.), *Oxford Handbook of Political Networks*. Oxford University Press.

Lomi, A., & Stadtfeld, C. (2014). Social networks and social settings: Developing a coevolutionary view. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 66(1), 395-415.

Stinnett, D. M., Tir, J., Diehl, P. F., Schafer, P., & Gochman, C. (2002). The Correlates of War (Cow) Project Direct Contiguity Data, Version 3.0. *Conflict Management and Peace Science*, 19(2), 59–67.

A.4 Figures used in the Appendix

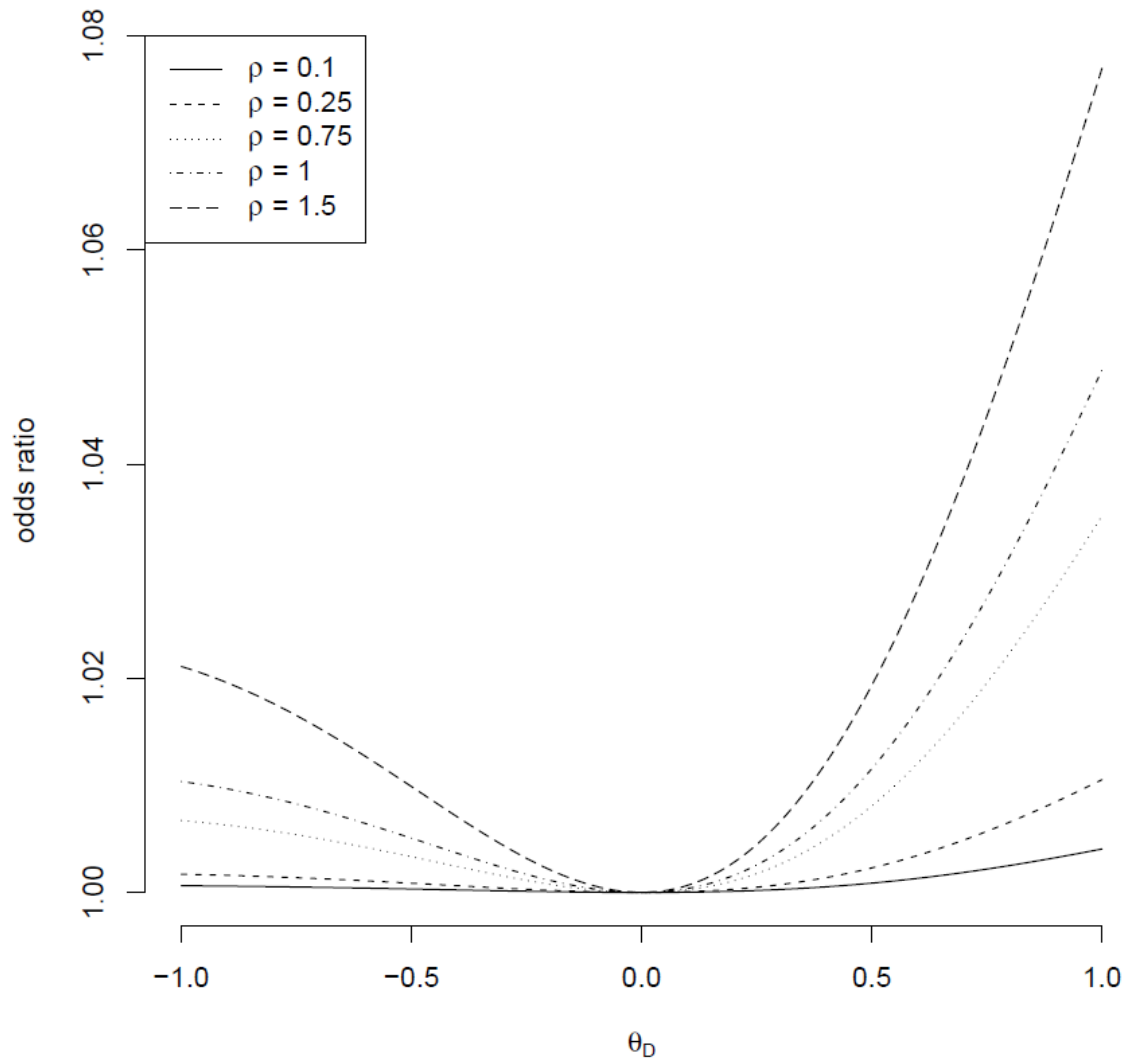


Figure A1: Odds ratio of the realisations in the example dyad for different values of θ_D and ρ .