Computational Statistics and Data Analysis 152 (2020) 107052

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Varying-coefficient models for dynamic networks[☆]

Jihui Lee[a,*], Gen Li[b], James D. Wilson[c]

[a] Department of Population Health Sciences, Weill Cornell Medicine, New York, NY, USA
[b] Department of Biostatistics, Columbia University, New York, NY, USA
[c] Department of Mathematics and Statistics, University of San Francisco, San Francisco, CA, USA

ARTICLE INFO

Article history:
Received 27 March 2018
Received in revised form 16 July 2020
Accepted 16 July 2020
Available online 24 July 2020

Keywords:
Exponential random graph model
Temporal graphs
Basis spline
Pseudo likelihood
Penalized logistic regression

ABSTRACT

Dynamic networks are commonly used to model relational data that are observed over time. Statistical models for such data should capture both the temporal variation of the relational system as well as the structural dependencies within each network. As a consequence, effectively making inference on dynamic networks is a computationally challenging task, and many models are intractable even for moderately sized systems. In light of these challenges, a family of dynamic network models known as varying-coefficient exponential random graph models (VCERGMs) is proposed to characterize the evolution of network topology through smoothly varying parameters. The VCERGM provides an interpretable dynamic network model that enables the inference of temporal heterogeneity in dynamic networks. Estimation of the VCERGM is achieved via maximum pseudo-likelihood techniques, thereby providing a computationally tractable strategy for statistical inference of complex dynamic networks. Furthermore, a bootstrap hypothesis testing framework is presented for testing the temporal heterogeneity of an observed dynamic network sequence. Application to the U.S. Senate co-voting network and comprehensive simulation studies both reveal that the VCERGM provides relevant and interpretable patterns and has significant advantages over existing methods.
© 2020 Elsevier B.V. All rights reserved.

1. Introduction

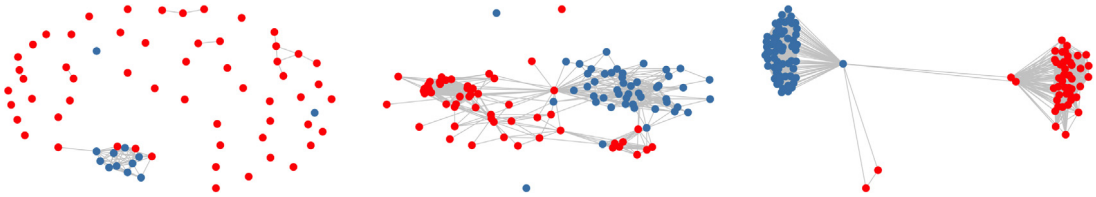
Networks have been extensively used to explore, model, and analyze the relational structure of individual units, or actors, in a complex system. In a network model, nodes represent the actors of the system, and edges are placed between nodes if the corresponding actors share a relationship. In many applications, the relationships among the actors of a modeled system change over time, necessitating the use of dynamic networks. Two diverse examples, which we analyze later in our application study, include the Congressional co-voting networks in Fig. 1 and resting state brain connectivity networks in Fig. 2. A prominent way to analyze relational network systems is through the use of probabilistic models, or graphical models, which describe the generative mechanism of an observed network. Although there is a rich body of literature on graphical models for static networks (Fienberg, 2012; Goldenberg et al., 2010), the development of interpretable and computationally tractable models for dynamic networks is in its early stages.

An important feature of dynamic networks that needs to be captured in any statistical model is the extent to which its local and global features change through time. We refer to this property as *temporal heterogeneity*. Heterogeneity directly affects the underlying process that best describes the formation of networks. In parametric models, heterogeneity may

☆ The R code for the analyses of this article is publicly available at https://www.github.com/jihuilee/vcergm.

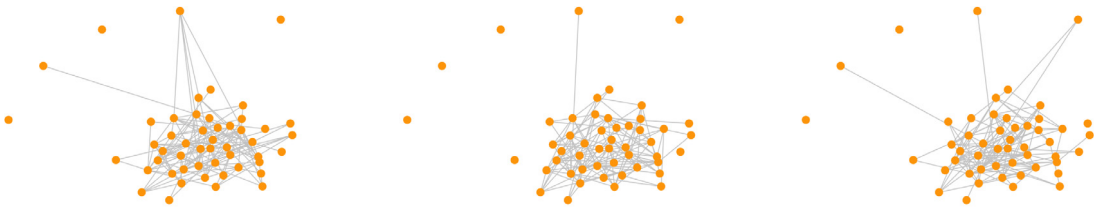
* Correspondence to: 402 East 67th Street, New York, NY 10065, USA.
E-mail address: jil2043@med.cornell.edu (J. Lee).

https://doi.org/10.1016/j.csda.2020.107052
0167-9473/© 2020 Elsevier B.V. All rights reserved.



(a) 40th Congress (b) 70th Congress (c) 113th Congress

Fig. 1. U.S. Senate co-voting network: Co-voting networks of U.S. senators in Congress 40, 70 and 113. Red nodes represent Republican Senators and blue nodes represent Democratic Senators. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



(a) Time 10 (b) Time 20 (c) Time 47

Fig. 2. Resting state fMRI network: Resting state fMRI network at observed times 10, 20 and 47. Each node represents a brain region. The top 10% of partial correlation between regions forms an edge.

result in significant changes in parameters that characterize the observed network. Consider the U.S. Senate co-voting network shown in Fig. 1. One can readily observe an evolution of the network to form distinct clusters of Republicans and Democrats by the 113th Congress. This configuration is in stark contrast with the sparse, seemingly random configuration formed in the 40th Congress. On the other hand, the resting state functional magnetic resonance imaging (fMRI) network shown in Fig. 2 remains fairly stable through time with only minor local changes in edge formation. These contrasting examples exemplify the need to explicitly model the heterogeneity of a network. We further analyze these dynamic networks in Section 6 and Appendix F.

In this paper, we propose a probabilistic model for dynamic networks called the varying-coefficient exponential random graph model (VCERGM). The model parameterizes time-varying topological features of dynamic networks in continuous time. Our model builds on two major statistical methodologies. One is the exponential family of random graph models (Holland and Leinhardt, 1981; Wasserman and Pattison, 1996) that characterizes the marginal effect of local and global network features on the likelihood of the network. The other major component is a varying-coefficient specification (Hastie and Tibshirani, 1993), which flexibly models the changes of effect parameters over time. The VCERGM characterizes the temporal heterogeneity of dynamic networks by modeling the parameter associated with each topological feature as a smooth function of time.

One prominent advantage of the VCERGM is its interpretability. By quantifying temporal heterogeneity of a network via fluctuating parameters, we are able to analyze key properties of the local and global features of a dynamic network. In addition to serving as a means to test for heterogeneity of a dynamic sequence, our method can also be directly used for interpolation of missing networks or edges. For networks at unobserved time points, our method provides robust estimates that reflect the structure of the unobserved networks without being strongly influenced by outliers in the sequence. Furthermore, estimation of the VCERGM can be done with a computationally scalable maximum pseudo-likelihood estimation (MPLE) approach, enabling efficient inference for large dynamic networks.

There are several related dynamic network models that have been investigated. The exponential random graph model (ERGM) is a family of probability distributions on unweighted static network. The ERGM has been adapted to dynamic networks in the pivotal work of Hanneke et al. (2010). The method is called the temporal exponential random graph model (TERGM). The TERGM models the difference in topological features between every two consecutive networks in a similar fashion to the ERGM. However, it ignores the heterogeneity of the differences, and cannot fully capture the time-varying patterns of the network structure. We investigate the situations under which the TERGM degenerates to a collection of independent and identically distributed ERGMs in Appendix A.

The TERGM has been further investigated in many different perspectives. Guo et al. (2007) devised the hidden TERGM, which utilizes a hidden Markov process to express the nature of rewiring networks and model a time-specific network topology. Krivitsky and Handcock (2014) generalized the TERGM to the separable TERGM (STERGM). The STERGM models the formation and dissolution of networks by separately parameterizing prevalence and duration of fluctuations. It allows time-varying overall rate of tie formation and dissolution while retaining homogeneous parameters for the other terms.

Almquist and Butts (2014) relaxed the temporal Markov and fixed vertex set assumption of Hanneke et al. (2010) and demonstrated a parametric model for temporal networks via dynamic network logistic regression.

The VCERGM, like the TERGM and ERGM, is a network-centric model. In other words, the VCERGM does not explicitly model individual changes in a network; rather, the focus is to characterize motifs or subgraph properties. An alternative to this modeling approach is what is known as ego-centric analysis, where individual change is the focus. Perhaps the most popular ego-centric modeling strategy is the stochastic actor-oriented model (SAOM) (Snijders, 2001). It provides an alternative to dyadic models and instead is a localized actor-based model, which characterizes network evolution as a consequence of each actors' connectivity. Even if the SAOM considers the fluctuation between two time points, it does not provide explicit form to parameterize the fluctuation in network topology. Sarkar and Moore (2005) and Sewell and Chen (2015) generalized the latent space model developed by Hoff et al. (2002) to dynamic networks. Unlike our current model, latent space models characterize the dynamics of network structure through random effects in a latent space.

Another recent ego-centric model is that proposed in Hoff et al. (2015), where dynamic networks are modeled using multilinear tensor regression. This work adapted autoregressive models to dictate temporal dependence in a sequence of networks, and like the SAOM, proposed an actor-based dependence structure between edges in each network. It directly models the temporal heterogeneity but may not be adequate for larger networks due to its computational complexity. Kolar et al. (2010) consider capturing time-varying attributes of dynamic networks and parametrizes the evolving relationship of each edge between nodes as a smooth function of time. Along with kernel smoothing approach, the ℓ_1-regularization is utilized to ensure the smoothness. The parameters in the model provide a valuable intuition in understanding the topological change of each edge, but fitting this model for larger networks can be computationally expensive considering the number of parameters.

As an alternative, the proposed model exploits a varying-coefficient framework to model the temporal heterogeneity of topological features. The varying-coefficient framework is a family of semi-parametric models, where the coefficient of a parametric model evolves with some characteristics in a nonparametric fashion. It was first developed to model non-linear effects of covariates on real-valued response variables (Hastie and Tibshirani, 1993). Later it was extended to the dynamic generalized linear models (Hoover et al., 1998; Zhang et al., 2015). The varying-coefficient models extend the classic parametric models to understand the dynamic pattern of temporally evolving structure (Fan and Zhang, 2008). A detailed review of varying-coefficient models and their applications are provided in Fan and Zhang (2008). In our proposed model, we model the coefficients of the topological features in the ERGM as a function of time. As a result, the varying coefficients effectively capture the dynamic pattern of the network structure. To our best knowledge, the VCERGM is the first attempt to generalize the idea to dynamic networks.

2. Model

We begin by describing the exponential family of random graph models (ERGMs) and their temporal extension, the TERGM, since our proposed model is closely related to these specifications. We then introduce our proposed model the VCERGM.

2.1. Temporal exponential random graph models

Let the $n \times n$ random matrix X represent an unweighted network with n vertices, whose (i, j)th entry X_{ij} is an indicator that specifies whether or not node i and node j are connected by an edge. Self-loops are not allowed, and thus the diagonal elements of X are all zero. Let $\mathcal{X}$ denote the family of all $n \times n$ unweighted networks so that $X \in \mathcal{X}$. The ERGM is a probability distribution that characterizes the likelihood of X via a function of network statistics $\mathbf{h} : \mathcal{X} \to \mathbb{R}^p$ that describe the topological structure of X.

Given $\mathbf{h}$, the ERGM models X as a binary random matrix generated from the following probability mass function

$$\mathbb{P}(X = x \mid \boldsymbol{\phi}) = \frac{\exp\{\boldsymbol{\phi}^T \mathbf{h}(x)\}}{\sum_{z \in \mathcal{X}} \exp\{\boldsymbol{\phi}^T \mathbf{h}(z)\}}, \tag{1}$$

where $\boldsymbol{\phi} \in \mathbb{R}^p$ parameterizes the influence of the network statistics $\mathbf{h}(X)$ on the likelihood of X. The coefficient corresponding to the number of triangles in an undirected network, for example, characterizes how the number of triangles changes the likelihood of a network with n nodes. Positive coefficients suggest that networks with a higher number of triangles are more likely to occur than networks with lower number of triangles, and reflects clustering in the observed network.

The ERGM has been successfully applied in a wide variety of fields, ranging from social networks to brain connectivity networks (Goodreau et al., 2009; Simpson et al., 2011; Székely et al., 2016). Recent tutorials of exponential random graph models and their applications are provided in Cranmer and Desmarais (2011), Fellows and Handcock (2012) and Robins et al. (2007). Despite their popularity, an important obstacle that arises in discrete exponential family model specification is the problem of degeneracy, a condition under which only a few network configurations – usually very sparse and very dense networks – have high probability mass (Handcock et al., 2003; Rinaldo et al., 2009; Schweinberger, 2011). The issue of degeneracy strongly influences the effectiveness of estimation algorithms, which often rely on Markov chain

Monte Carlo simulation. In the case that nearly empty (or nearly complete) networks are most probable, estimation via MCMC will fail to converge to consistent parameter estimates. A common strategy to help mediate the degeneracy problem is to use geometrically weighted network statistics, which downweight higher order statistics and reduces the computational complexity of subgraph counting (Snijders et al., 2006; Hunter et al., 2008a; Wyatt et al., 2010). Furthermore, the generalized exponential random graph model for networks with continuous-valued edge weights has been shown to avoid likelihood degeneracy in common specifications (Wilson et al., 2017; Bhamidi et al., 2018).

We now describe the TERGM, a generalization of the ERGM that enables statistical inference of dynamic networks (Desmarais and Cranmer, 2012; Hanneke et al., 2010). Consider a temporally ordered sequence of networks $\mathbf{X} = \{X_1, X_2, \ldots, X_T\}$ that is observed at T discrete and non-overlapping time periods, where each graph $X_t \in \mathcal{X}$ from $\mathbf{X}$ is unweighted, and observed for the set of vertices $[n] = \{1, \ldots, n\}$. The TERGM is a generative model for $\mathbf{X}$ that characterizes the conditional probability of X_t given $\mathbf{X}_t^- = \{X_s : s = 1, \ldots, t-1\}$ via an exponential family of probability distributions. Under the first order TERGM, $\mathbf{X}$ exhibits a one-step Markov dependence between sequential networks as follows:

$$\mathbb{P}(X_t = x_t \mid \mathbf{X}_\mathbf{t}^- = \mathbf{x}_\mathbf{t}^-) = \mathbb{P}(X_t = x_t \mid X_{t-1} = x_{t-1}). \tag{2}$$

Under (2), one can fully specify the joint probability mass function of $\mathbf{X}$ by parameterizing the one-step transitions from X_{t-1} to X_t. One models these dependencies using a function of transition statistics $\mathbf{g} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}^p$. These statistics represent the temporal potential over cliques across two sequential networks and can represent, for example, the change in the clustering or the change in overall connectivity between each pair of networks. For a chosen $\mathbf{g}$, the first-order TERGM specifies the likelihood of $X_t \mid X_{t-1}$ for $t = 2, \ldots, T$ as

$$\mathbb{P}(X_t = x_t \mid \mathbf{X}_t^-; \boldsymbol{\phi}) = \mathbb{P}(X_t = x_t \mid X_{t-1} = x_{t-1}; \boldsymbol{\phi}) = \frac{\exp\{\boldsymbol{\phi}^T \ \mathbf{g}(x_t, x_{t-1})\}}{\sum_{z \in \mathcal{X}} \exp\{\boldsymbol{\phi}^T \ \mathbf{g}(z, x_{t-1})\}}, \tag{3}$$

where $\boldsymbol{\phi} \in \mathbb{R}^p$ parameterizes the influence of the transition statistics $\mathbf{g}(X_t, X_{t-1})$ on the conditional likelihood of X_t given X_{t-1}. Suppose that the marginal distribution $\mathbb{P}(X_1 = x_1 \mid \boldsymbol{\phi})$ is specified. The TERGM characterizes the joint distribution of the dynamic sequence $\mathbf{X}$ by

$$\mathbb{P}(\mathbf{X} = \mathbf{x} \mid \boldsymbol{\phi}) = \mathbb{P}(X_1 = x_1 \mid \boldsymbol{\phi}) \prod_{t=2}^{T} \mathbb{P}(X_t = x_t \mid X_{t-1} = x_{t-1}, \boldsymbol{\phi}). \tag{4}$$

We note that in general if one is able to specify appropriate transition statistics, then the TERGM in (3) and (4) is readily generalized to higher-order Markov dependency. Bootstrapped maximum-pseudolikelihood techniques are typically used for estimation of the TERGM (Desmarais and Cranmer, 2012). Estimation can be carried out using the `btergm` software package (Leifeld et al., 2018).

2.2. Varying-coefficient exponential random graph models

Let $\mathbf{X} = \{X_t : 0 \le t \le T\}$ be a stochastic sequence of temporally ordered networks observed continuously up to some time $T > 0$. At each time point t, $X_t \in \mathcal{X}$ represents an unweighted, directed or undirected network with time-invariant size n. Our goal is to provide a dynamic network model for $\mathbf{X}$ that directly accounts for the temporal heterogeneity of its local and global network structure.

The VCERGM consists of two components — (i) an ERGM representation for the marginal likelihood of each observed network, and (ii) the coupling of networks over time via a varying-coefficient model, where the coefficients at time t parameterize the marginal likelihood of the network X_t. We first specify a set of functions $\mathbf{h}(x_t) : \mathcal{X} \to \mathbb{R}^p$ for $t \in [0, T]$, which quantify the p topological features of network x_t with n nodes. Given $\mathbf{h}(x_t)$ and the coefficient vector $\boldsymbol{\phi}(t) = (\phi_1(t), \ldots, \phi_p(t))^T \in \mathbb{R}^p$, the marginal likelihood of X_t at time t has an ERGM representation given by

$$\mathbb{P}(X_t = x_t \mid \boldsymbol{\phi}(t)) = \frac{\exp\{\boldsymbol{\phi}(t)^T \mathbf{h}(x_t)\}}{\sum_{z \in \mathcal{X}} \exp\{\boldsymbol{\phi}(t)^T \mathbf{h}(z)\}}, x_t \in \mathcal{X}. \tag{5}$$

A large collection of topological features can be used in the VCERGM. Traditionally, the network statistics are raw counts of different features in an observed network, such as the number of edges (edge density), the number of triangles, or the number of reciprocal edges in a directed network.

The coefficients $\boldsymbol{\phi}(t)$ in model (5) characterize the influence of the corresponding network statistics on determining the network structure. By evaluating the coefficient at time point $t \in [0, T]$, we can write the marginal distribution of a graph X_t as described in model (5). When a dynamic network evolves gradually over time, it is reasonable to believe the coefficients will also change gradually. In such a case, $\boldsymbol{\phi}(t)$ can be represented by smooth functions of t with continuous second order derivatives over $[0, T]$ (Ramsay, 2006). In the special case where all the separate functions in $\boldsymbol{\phi}(t)$ are constant, the generative models underlying the dynamic networks are identical over time and the VCERGM reduces to a family of marginally identically distributed ERGMs. In Section 4, we introduce a formal hypothesis testing procedure to test the temporal heterogeneity of the coefficients.

2.3. Generalization to higher order varying-coefficient exponential random graph models

The VCERGM, in general, can be used to model the parameters describing the smooth transitions between consecutive networks in time. Model (5) investigates the dynamics of coefficients for marginal network statistics without accounting for temporal dependency. However, this model can readily be extended to networks with a Markov dependency like that described by the TERGM. For any non-negative integer q, one can incorporate an order q Markov dependency in the VCERGM. We denote such a model as a VCERGM of order q (VCERGM(q)). We refer to model (5) as the varying-coefficient exponential random graph model of order 0 (VCERGM(0)). For $q \geq 1$, one must specify summary statistics that couple the dependence among q observed networks in the sequence. For example, when $q = 1$ we can model the one step transition between X_{t-1} and X_t using a suite of statistics $\mathbf{h}_1(x_t, x_{t-1})$ as

$$\mathbb{P}(X_t = x_t \mid \boldsymbol{\phi}_1(t), x_{t-1}) = \frac{\exp\{\boldsymbol{\phi}_1^T(t)\,\mathbf{h}_1(x_t, x_{t-1})\}}{\sum_{z\in\mathcal{X}} \exp\{\boldsymbol{\phi}_1(t)^T\ \mathbf{h}_1(z, x_{t-1})\}}, x_t \in \mathcal{X}. \tag{6}$$

Here, $\mathbf{h}_1$ is the *temporal potential* over cliques across two time-adjacent networks. For examples of transition statistics $\mathbf{h}_1$, see Hanneke et al. (2010). In model (6), $\boldsymbol{\phi}_1(t) = \{\phi_{1k}(t), k = 1, \ldots, p\}$ can be modeled as smooth functions that describe the impact of the one-step transition statistics from x_{t-1} to x_t. Therefore, model (6) effectively captures the *rate of change* of the temporal potential between sequential graphs rather than the rate of change of the marginal features as done in this work. Considering the higher order dependency, one could use the VCERGM to predict the network structure in the future based on the Markovian framework. Like the TERGM, we can generalize the VCERGM to a higher order Markov dependency, say order $q > 1$, by specifying appropriate transition statistics $\mathbf{h}_q(x_t, x_{t-1}, \ldots, x_{t-q})$.

In general, the VCERGM(q) characterizes the impact of the changes of transition between q consecutive networks. Due to the Markov properties, the VCERGM with lags can be used for prediction. Furthermore, since coefficients are smooth functions through time, one can readily interpolate for unobserved networks. Notably, the TERGM of order q is a special case of the VCERGM(q) where $\boldsymbol{\phi}_q(t) \equiv \boldsymbol{\phi}_q$. This requirement greatly restricts the family of dynamic networks that can be modeled through the TERGM. By allowing smooth fluctuations, the VCERGM models the effects of temporal heterogeneity more efficiently.

3. Estimation

3.1. Spline-based representation of time-varying coefficients

Without any constraint, the collection of coefficients $\{\boldsymbol{\phi}(t) : 0 \leq t \leq T\}$ contain an infinite number of parameters, making inference on (5) intractable. To address this problem, we approximately represent these smooth functions as a linear combination of basis functions. Possible strategies of defining basis functions include piecewise polynomials (De Boor et al., 1978), Fourier series (Konidaris et al., 2011) and wavelets (Daubechies et al., 1992). For inferential purposes, we employ basis splines (b-splines) (De Boor et al., 1978; Eilers and Marx, 1996) as a way to reduce the dimensionality of estimation. B-splines are commonly used due to its flexibility in incorporating smoothing constraints.

In particular, we first specify a collection of basis functions $B_1(t), \ldots, B_q(t)$, $0 \leq t \leq T$, and then approximate $\phi_k(t)$ by a linear combination of these functions

$$\phi_k(t) = \sum_{\ell=1}^{q} \Phi_{k\ell}\, B_\ell(t),$$

where $\Phi_{k\ell}$ quantifies the contribution of the ℓth basis function on $\phi_k(t)$. Let $\boldsymbol{\Phi} = \{\Phi_{k\ell};\ k = 1, \ldots, p,\ \ell = 1, \ldots, q\}$ denote the $p \times q$ basis coefficient matrix and let $\mathbf{B}(t) = (B_1(t), \ldots, B_q(t))^T$ be the length q vector of basis functions. We can represent the coefficients $\boldsymbol{\phi}(t)$ as

$$\boldsymbol{\phi}(t) = \boldsymbol{\Phi}\mathbf{B}(t). \tag{7}$$

The set of q basis functions represents the smoothness of $\boldsymbol{\phi}(t)$, and the coefficient matrix $\boldsymbol{\Phi}$ determines the shape and trajectory of the fluctuations through time. Under the basis representation in (7), the distribution of $\mathbf{X}_t$ in (5) is fully specified by the pq parameters in the coefficient matrix $\boldsymbol{\Phi}$.

3.2. Fast estimation via maximum pseudo likelihood

For an observed dynamic sequence of unweighted graphs $\mathbf{x} = \{x_s \in \mathcal{X} : s = t_1, \ldots, t_K, t_j < t_{j+1} \in [0, T]\}$, our goal is to estimate the coefficients $\{\boldsymbol{\phi}(t) : 0 \leq t \leq T\}$ given the sequence $\mathbf{x}$. Let $\mathbf{B}_s = \{B_{s,\ell};\ \ell = 1, \ldots, q\}$ be a vector of length q of which elements are the basis functions evaluated at time s. By applying the basis representation in (7), we denote $\boldsymbol{\phi}_s = \boldsymbol{\Phi}\mathbf{B}_s$ as the smooth function $\boldsymbol{\phi}(\cdot)$ evaluated at time s. Therefore, this estimation reduces to the task of estimating the $p \times q$ coefficient matrix $\boldsymbol{\Phi}$. A major obstacle in obtaining the maximum likelihood estimators of the parameters in Model (5), similar to that of fitting an ERGM, is that calculation of the normalizing constant in the denominator is computationally intractable. Although numerical approaches such as the Markov chain Monte Carlo method can be used

to estimate $\boldsymbol{\Phi}$ for small networks (Hunter and Handcock, 2006; Wilson et al., 2017), the computational cost is prohibitive for moderate to large networks, let alone a sequence of networks. To alleviate the computational complexity, we exploit a maximum pseudo-likelihood approach, originally adapted for fitting the ERGM (Strauss and Ikeda, 1990; Van Duijn et al., 2009; Wasserman and Pattison, 1996). We show that the maximum pseudo-likelihood estimator (MPLE) for the VCERGM can be efficiently obtained via maximum likelihood estimation of a logistic regression model. Below we describe the estimation procedure in more detail.

For each observed time point $s = t_1, \ldots, t_K$, let X_{ij}^s denote the binary random variable that describes whether or not there is an edge between node i and node j at time s. Furthermore, let $\mathbf{X}_{-(ij)}^s$ be the collection of $\binom{n}{2} - 1$ binary random variables that describe whether or not there is an edge between all other pairs of nodes other than the node pair i and j. For each observed time point $s = t_1, \ldots, t_K$, assume the conditional independence between edges. The marginal pseudo-likelihood function of $\boldsymbol{\Phi}$ given x_s at time s is defined as

$$\mathrm{PL}(\boldsymbol{\Phi}|x_s) = \prod_{i,j\in[n]} \mathbb{P}(X_{ij}^s = x_{ij}^s | \mathbf{X}_{-(ij)}^s = \mathbf{x}_{-(ij)}^s). \tag{8}$$

Subsequently, the marginally independent composite pseudo likelihood of model (5) is

$$\mathrm{PL}(\boldsymbol{\Phi}|\mathbf{x}) = \prod_{s=t_1}^{t_K} \prod_{i,j\in[n]} \mathbb{P}(X_{ij}^s = x_{ij}^s | \mathbf{X}_{-(ij)}^s = \mathbf{x}_{-(ij)}^s).$$

The MPLE $\widehat{\boldsymbol{\Phi}}$ is obtained by maximizing $\mathrm{PL}(\boldsymbol{\Phi}|\mathbf{x})$. The pseudo-likelihood approach used for estimation and hypothesis testing treats pairs of edges as pairwise independent. As the temporal dependence is parametrized by the coefficient $\boldsymbol{\phi}(t)$, the composite pseudo-likelihood function can be written as a product of marginal pseudo-likelihood functions at the observed time points $t_1, \ldots, t_k$.

Let $x_{s,ij}^+$ denote the realization of x_s with x_{ij}^s set to 1 and let $x_{s,ij}^-$ be the realization of x_s with $x_{ij}^s = 0$. Define $\boldsymbol{\Delta}_{ij}^s = \mathbf{h}(x_{s,ij}^+) - \mathbf{h}(x_{s,ij}^-)$ as the vector describing the element-wise difference in the network statistics when x_{ij}^s changes from 0 to 1. One can readily show that for each $s = t_1, \ldots, t_K$, the following relationship holds for all $i, j \in [n]$:

$$\begin{aligned}\mathrm{logit}\left\{\mathbb{P}(X_{ij}^s = 1|\mathbf{X}_{-(ij)}^s = \mathbf{x}_{-(ij)}^s)\right\} &= \log\left\{\frac{\mathbb{P}(X_{ij}^s = 1|\mathbf{X}_{-(ij)}^s = \mathbf{x}_{-(ij)}^s)}{\mathbb{P}(X_{ij}^s = 0|\mathbf{X}_{-(ij)}^s = \mathbf{x}_{-(ij)}^s)}\right\} \\ &= \log\left[\exp\{\boldsymbol{\phi}(t)^T(\mathbf{h}(x_{s,ij}^+) - \mathbf{h}(x_{s,ij}^-))\}\right] \\ &= \boldsymbol{\phi}_s^T \boldsymbol{\Delta}_{ij}^s. \end{aligned} \tag{9}$$

Let $Y_{ij}^s = \mathrm{logit}\{\mathbb{P}(X_{ij}^s = 1|\mathbf{X}_{-(ij)}^s = \mathbf{x}_{-(ij)}^s)\}$ and let $\mathbf{Y}_s = (Y_{11}^s, Y_{12}^s, \ldots, Y_{nn}^s)^T$. Similarly, define $\boldsymbol{\Delta}_s = (\boldsymbol{\Delta}_{11}^s, \boldsymbol{\Delta}_{12}^s, \ldots \boldsymbol{\Delta}_{nn}^s)$ as the $p \times \binom{n}{2}$ matrix whose rth row contains the change in the rth network statistic when each edge changes from 0 to 1. Let $\mathrm{vec}(X)$ be the operator that stacks the columns of X into a column vector and let $\otimes$ represent the Kronecker product operator. Combining (7) and (9) yields

$$\mathbf{Y}_s = \boldsymbol{\Delta}_s^T \boldsymbol{\Phi} \mathbf{B}_s = (\mathbf{B}_s \otimes \boldsymbol{\Delta}_s)^T \ \mathrm{vec}(\boldsymbol{\Phi}), s = t_1, \ldots, t_K. \tag{10}$$

Let $\mathbf{Y} = (\mathbf{Y}_{t_1}, \ldots, \mathbf{Y}_{t_K})^T$ and define the $pq \times K\binom{n}{2}$ design matrix $\mathbf{H}$ as

$$\mathbf{H} = \begin{pmatrix} \mathbf{B}_{t_1} \otimes \boldsymbol{\Delta}_{t_1} \\ \vdots \\ \mathbf{B}_{t_K} \otimes \boldsymbol{\Delta}_{t_K} \end{pmatrix}.$$

The relationship in (10) connects to a logistic regression where $\mathbf{H}$ represents a design matrix with its coefficient $\mathrm{vec}(\boldsymbol{\Phi})$. In Strauss and Ikeda (1990), it was shown that maximizing the pseudo-likelihood $\mathrm{PL}(\boldsymbol{\Phi}|x_s)$ in (8) is equivalent to finding the maximum likelihood estimator (MLE) of $\boldsymbol{\Phi}$ in the logistic regression model given in (9) with independent entries X_{ij}^s. Dependency among nodes in a network can be indirectly modeled by conditioning on the rest of the edges. As a result, the assumption of independent data points is not required. We expand this estimation strategy to temporal networks. It follows from the independence of $\mathbf{X}_s$ and $\mathbf{X}_{s'}$ for $s \neq s'$ that maximizing $\mathrm{PL}(\boldsymbol{\Phi}|\mathbf{x})$ is equivalent to calculating the MLE of $\boldsymbol{\Phi}$ in the logistic regression model $\mathbf{Y} = \mathbf{H}^T\mathrm{vec}(\boldsymbol{\Phi})$ treating $\{X_{ij}^s : i, j \in [n], s = t_1, \ldots, t_K\}$ as mutually independent variables. Correlation between neighboring time points is not explicitly specified, but the joint pseudo likelihood of dynamic networks is defined by multiplying the pseudo likelihood for each observed time point. Qu and Li (2006) showed that the estimation procedure for varying-coefficient models based on the penalized spline and quadratic inference function directly incorporates the correlation across time without further specifying a nuisance parameter associated with the correlation.

This maximum pseudo likelihood approach can be also applied to the VCERGM(q) in an analogous fashion. For simplicity, we consider the VCERGM(1) with one-step Markov dependence. The conditional likelihood of graph X_t given X_{t-1} is specified as (5). For a collection of basis functions $B_1(t), \ldots, B_q(t)$, $0 \le t \le T$, we approximate $\phi_{1k}(t)$ by a linear

combination of these functions as $\phi_{1k}(t) = \sum_{\ell=1}^{q} \Phi_{1k\ell} B_\ell(t)$ and represent the coefficient $\boldsymbol{\phi}_{1k}(t)$ as a $\boldsymbol{\phi}_{1k}(t) = \boldsymbol{\Phi}_1 \mathbf{B}(t)$, with a $p \times q$ matrix of basis coefficients $\boldsymbol{\Phi}_1$. The composite pseudo-likelihood functions for VCERGM(1) can be expressed as

$$\begin{aligned}\mathrm{PL}(\boldsymbol{\Phi}|\mathbf{x}) = &\prod_{i,j\in[n]} \mathbb{P}(X_{ij}^{t_1} = x_{ij}^{t_1} | \mathbf{X}_{-(ij)}^{t_1} = \mathbf{x}_{-(ij)}^{t_1}) \\ &\times \prod_{s=t_2}^{t_K} \prod_{i,j\in[n]} \mathbb{P}(X_{ij}^{s} = x_{ij}^{s} | \mathbf{X}_{-(ij)}^{s} = \mathbf{x}_{-(ij)}^{s},\ X_{s-1} = x_{s-1}).\end{aligned}$$

The likelihood of X_{t_1} has an ERGM representation and thus the marginal pseudo-likelihood at time t_1 is unconditional; it has the same form as (8). Let $\boldsymbol{\Delta}_{ij}^{s} = h(x_{s,ij}^{+}, x_{s-1}) - h(x_{s,ij}^{-}, x_{s-1})$ denote the vector of differences in the transition statistics when $x_{s,ij}$ changes from 0 to 1. Then the relationship described in (9) can be similarly applied to VCERGM(1). A logistic regression model like (10) is used for parameter estimation and thus the remainder of the estimation steps described above remain the same.

3.3. Penalized logistic regression

To obtain smooth estimates of the time-varying coefficients $\boldsymbol{\phi}(t)$, we further consider a roughness penalty on the coefficients of the basis functions (see Hastie and Tibshirani, 1993; Eilers and Marx, 1996; Hoover et al., 1998, for example). A commonly used penalty, which we use throughout this paper, is the integrated squared second derivative defined for kth row of $\boldsymbol{\Phi}$, denoted as $\boldsymbol{\Phi}_{(k)}$, as

$$\mathcal{P}(\boldsymbol{\Phi}_{(k)}) = \int \{D^2\phi_k(u)\}^2\, du = \boldsymbol{\Phi}_{(k)}^T\, \boldsymbol{\Omega}\, \boldsymbol{\Phi}_{(k)},$$

where a smoothness matrix $\boldsymbol{\Omega}$ in this case can be specified as

$$\boldsymbol{\Omega} = \left\{ \Omega_{ij} = \int \{D^2 B_i(u)\}\{D^2 B_j(u)\}\, du\,;\ i, j = 1, \ldots, q \right\}.$$

For networks observed at discrete time points $t_1, \ldots, t_K$, the (i, j)th element of $\boldsymbol{\Omega}$ is

$$\Omega_{ij} = \sum_{s=t_1}^{t_K} \{D^2 B_i(s)\}\{D^2 B_j(s)\}, \quad i, j = 1, \ldots, q.$$

For more examples of possible penalties, see the Chapter 5 in Ramsay (2006). As the same collection of basis functions are used to express $\phi_k(t)$, $k = 1, \ldots, p$, via basis representation, we impose the same $\boldsymbol{\Omega}$ on all $\phi_k(t)$. Consequently, we add the penalty term $\mathcal{P}_{\boldsymbol{\Omega}}(\boldsymbol{\Phi})$ to the logistic log pseudo likelihood function where $\mathcal{P}_{\boldsymbol{\Omega}}(\boldsymbol{\Phi})$ is defined as

$$\mathcal{P}_{\boldsymbol{\Omega}}(\boldsymbol{\Phi}) = \sum_{k=1}^{p} \boldsymbol{\Phi}_{(k)}^T\, \boldsymbol{\Omega}\, \boldsymbol{\Phi}_{(k)} = \mathrm{vec}(\boldsymbol{\Phi})^T (\boldsymbol{\Omega} \otimes \mathbf{I}_p) \mathrm{vec}(\boldsymbol{\Phi}).$$

Similar with $\mathbf{Y}$, let $\tilde{\mathbf{x}}$ denote a vector that stacks all edges from networks at $t_1, \ldots, t_K$. That is, $\tilde{\mathbf{x}} = \{x_{ij}^{s} : i, j \in [n], s = t_1, \ldots, t_K\}$. We calculate the penalized pseudo-likelihood estimator $\widehat{\boldsymbol{\Phi}}_{\Omega}$ by maximizing the following penalized log likelihood with tuning parameter λ:

$$\tilde{\mathbf{x}}^T \mathbf{H}^T \mathrm{vec}(\boldsymbol{\Phi}) - \mathbf{1}^T \log[1 + \exp\{\mathbf{H}^T \mathrm{vec}(\boldsymbol{\Phi})\}] - \lambda \mathcal{P}_{\boldsymbol{\Omega}}(\boldsymbol{\Phi}). \tag{11}$$

To fit (11), we implement an iteratively reweighted least squares (IRLS) algorithm. A detailed description of this procedure is available in Appendix B.

4. Testing for heterogeneity

A key assumption of the VCERGM is that the effects of a specified collection of statistics vary through time. This assumption reflects heterogeneity in an observed sequence of graphs $\mathbf{x}$ and provides intuition as to whether or not summaries of $\mathbf{x}$ can be treated in aggregate. One can formally test for heterogeneity in $\mathbf{x}$ using bootstrap inference the observed sequence of networks. Bootstrap inference has recently received a lot of attention for uncertainty quantification of network summaries. For example, Snijders and Borgatti (1999) applied a bootstrap to assess the variance of network statistics in an observed network and demonstrated how to test differences between networks using t-test comparisons of calculated network statistics. Recently, Akcora et al. (2019) developed a bootstrap inference strategy for uncertainty quantification of estimators for node features in large networks. Here, we apply bootstrap inference on an observed sequence of networks to test for temporal heterogeneity using a likelihood ratio test. Our approach is closely related to the recent model selection strategy introduced in Chen and Onnela (2019).

We begin with a null hypothesis that $\mathbf{x}$ is homogeneous, namely that the coefficients $\boldsymbol{\phi}(t)$ under model (5) are fixed as constants over time. This serves as the null model, under which the VCERGM(0) is equivalent to fitting independent

and identically distributed ERGMs. Let $\phi_1^0, \ldots, \phi_p^0$ be the estimates under the time-invariant model fitting. With fixed constants $\phi_1^0, \ldots, \phi_p^0$, the null hypothesis corresponding to a homogeneous sequence of graphs can be written as

$$\mathrm{H}_0 : \phi_1(t) = \phi_1^0, \ \ldots, \ \phi_p(t) = \phi_p^0. \tag{12}$$

With basis spline (b-spline) setup of basis expansion (De Boor et al., 1978; Eilers and Marx, 1996), the basis functions satisfy $\sum_{l=1}^{q} B_l(t) = 1$ for all $t \in [0, T]$. As a result, any $\phi_k(t)$ is uniquely expressed as a linear combination of b-spline basis functions and setting the function $\phi_k(t) = \phi_k^0$ is equivalent to writing $\Phi_{k\ell} = \phi_k^0$ for all $\ell = 1, \ldots, q$. In other words, the null hypothesis in (12) can be expressed more succinctly as

$$\mathrm{H}_0 : \boldsymbol{\Phi} = \boldsymbol{\Phi}^0 = (\phi_1^0, \ldots, \phi_p^0)^T \times \mathbf{1}_q^T,$$

where $\mathbf{1}_q$ is length q vector of 1's. Such simplification is applicable for spline basis functions and we have implemented a hypothesis test with spline basis functions. The condition $\sum_{l=1}^{q} B_l(t) = 1$ for all $t \in [0, T]$ does not necessarily hold when other types of basis functions are used. Hypothesis test under other basis function specification remains to be explored in the future. The coefficients under the null hypothesis are the restricted form of the VCERGM where the basis coefficients for each network statistic are constants for all q basis functions.

The likelihood ratio test (LRT) is commonly used for conducting the test for heterogeneity in varying-coefficient models (Cai et al., 2000; Fan et al., 2001; Fan and Zhang, 2000, 2008). Due to the dependence between entries in each graph, we utilize a pseudo likelihood ratio test (pLRT) (Staicu et al., 2014). As previously emphasized in Section 3, the joint pseudo likelihood consists of the distribution of X_{ij}^s given the rest of the data $\mathbf{X}_{-(ij)}^s$ for all $i, j \in [n], s = t_1, \ldots, t_K$. Furthermore, maximizing the pseudo likelihood simplifies the estimation process as fitting a logistic regression. Namely, with observed networks $\mathbf{x} = \{x_s : s = t_1, \ldots, t_K\}$ with n nodes, the pLRT compares the pseudo log likelihood function below under the null and alternative hypotheses:

$$\begin{aligned}\log \mathrm{PL}(\boldsymbol{\Phi}|\mathbf{x}) &= \sum_{s=t_1}^{t_K} \sum_{i,j\in[n]} \log\{\mathbb{P}(X_{ij}^s = x_{ij}^s | \mathbf{X}_{-(ij)}^s = \mathbf{x}_{-(ij)}^s)\} \\ &= \sum_{s=t_1}^{t_K} \sum_{i,j\in[n]} \Big[x_{ij}^s \mathbf{B}_s^T \boldsymbol{\Phi}^T \boldsymbol{\Delta}_{ij}^s - \log\{1 + \exp(\mathbf{B}_s^T \boldsymbol{\Phi}^T \boldsymbol{\Delta}_{ij}^s)\}\Big].\end{aligned}$$

Let $\hat{\boldsymbol{\Phi}}_{\mathrm{H0}}$ and $\hat{\boldsymbol{\Phi}}_{\mathrm{H1}}$ be the estimates of $\boldsymbol{\Phi}$ under the null and alternative hypotheses. The estimate $\hat{\boldsymbol{\Phi}}_{\mathrm{H1}}$ can be calculated by fitting the VCERGM specified in (5) and $\hat{\boldsymbol{\Phi}}_{\mathrm{H0}}$ is the estimate from the VERCM with a restriction of constant basis coefficients. Accordingly, let $\log \mathrm{PL}(\hat{\boldsymbol{\Phi}}_{\mathrm{H0}}|\mathbf{x})$ and $\log \mathrm{PL}(\hat{\boldsymbol{\Phi}}_{\mathrm{H1}}|\mathbf{x})$ denote the pseudo log likelihood functions under the null and alternative, respectively. Then, the test statistic is

$$\begin{aligned}T &= 2\{\log \mathrm{PL}(\hat{\boldsymbol{\Phi}}_{\mathrm{H1}}|\mathbf{x}) - \log \mathrm{PL}(\hat{\boldsymbol{\Phi}}_{\mathrm{H0}}|\mathbf{x})\} \\ &= 2\sum_{s=t_1}^{t_K} \sum_{i,j\in[n]} \Big[x_{ij}^s B_s^T (\hat{\boldsymbol{\Phi}}_{\mathrm{H1}} - \hat{\boldsymbol{\Phi}}_{\mathrm{H0}})^T \boldsymbol{\Delta}_{ij}^s + \log\Big\{ \frac{1 + \exp(B_s^T \hat{\boldsymbol{\Phi}}_{\mathrm{H0}}^T \boldsymbol{\Delta}_{ij}^s)}{1 + \exp(B_s^T \hat{\boldsymbol{\Phi}}_{\mathrm{H1}}^T \boldsymbol{\Delta}_{ij}^s)} \Big\} \Big].\end{aligned} \tag{13}$$

We reject the null hypothesis when $T > C_\alpha$ where C_α is the critical value of the test with significance level α. We introduce an approach that involves generating bootstrap samples to construct the null distribution of T (Cai et al., 2000; Fan and Zhang, 2008; Huang et al., 2002). It is s preferable for moderate network size. Analogous to the work in De Brabanter et al. (2006), McLachlan (1987) and Tekle et al. (2016), the steps of obtaining the critical value C_α or calculating the p-value with parametric bootstrapping can be described as follows. For a large value of B, the test statistics (13) calculated based on B bootstrap samples successfully represent the null distribution of T.

1. Create B bootstrap samples. For each bootstrap, indexed by $b = 1, \ldots, B$, $\mathbf{x}^{*(b)} = \{x_s^{*(b)} : s = t_1, \ldots, t_K\}$ is a sample from $\mathbb{P}(\mathbf{X}|\hat{\boldsymbol{\Phi}}_{\mathrm{H0}})$.
2. For each bootstrap sample $\mathbf{x}^{*(b)}$, estimate $\boldsymbol{\Phi}$ under the null and alternative hypotheses and denote them as $\hat{\boldsymbol{\Phi}}_{\mathrm{H0}}^{*(b)}$ and $\hat{\boldsymbol{\Phi}}_{\mathrm{H1}}^{*(b)}$, respectively.
3. Calculate the test statistic for each bootstrap sample as
$$T^{*(b)} = 2\{\log \mathrm{PL}(\hat{\boldsymbol{\Phi}}_{\mathrm{H1}}^{*(b)}|\mathbf{x}^{*(b)}) - \log \mathrm{PL}(\hat{\boldsymbol{\Phi}}_{\mathrm{H0}}^{*(b)}|\mathbf{x}^{*(b)})\}, \quad b = 1, \ldots, B.$$
4. The critical value C_α is determined as the $(1-\alpha)$th quantile of $(T^{*(1)}, \ldots, T^{*(B)})$. The p-value is the proportion of times that the bootstrap test statistic values exceed the observed test statistic T. Define an indicator function $I(A)$ which takes a value of 1 if A is true and 0 otherwise. Then the p-value can be written as
$$p\text{-value} = \frac{\sum_{b=1}^{B} I(T < T^{*(b)})}{B}.$$

Table 1
Simulation results: Proportion of cases that we reject the null hypothesis out of 100 simulations at the significance level of $\alpha = 0.05$. Bootstrap samples of size $B = 1000$ and permuted samples of size $P = 1000$ are used to make a decision for hypothesis testing.

M	Bootstrap					Permutation				
	$K = 10$	30	50	70	100	$K = 10$	30	50	70	100
0	0.02	0.03	0.07	0.01	0	0.03	0.04	0.03	0	0.01
0.05	0.15	0.36	0.46	0.59	0.71	0.07	0.29	0.49	0.63	0.73
0.1	0.42	0.77	0.91	0.93	0.97	0.32	0.82	0.98	0.99	1
0.15	0.74	0.98	1	1	0.99	0.52	1	1	1	1
0.2	0.98	1	1	1	1	0.63	1	1	1	1
0.25	1	1	1	1	1	0.87	1	1	1	1
0.3	1	1	1	1	1	0.97	1	1	1	1

The p-value is then used to determine whether or not to reject the null hypothesis. For values below a specified significance value, α, one rejects the null hypothesis in (12) and decides that the sequence of networks does exhibit heterogeneity in its parameters. In our applications below, we choose $\alpha = 0.05$ when evaluating any hypothesis test.

5. Simulation study

The goal of our simulation study is two-fold: (i) to evaluate the power of the hypothesis testing procedure described in Section 4 and (ii) to assess the goodness of fit of the VCERGM on dynamic networks with various magnitudes of temporal heterogeneity. In Section 5.1, we evaluate the sensitivity of the hypothesis test in (12) for detecting temporal heterogeneity in a sequence of networks with fluctuating parameters using both the bootstrap and permutation procedures. Section 5.2 assesses the performance of the VCERGM under various varying-coefficient specifications. We compare the performance of the VCERGM with other competing methods. We further investigate how the VCERGM performs when the networks are observed at unequally spaced time points due to missing networks. We explore the performance of VCERGM when the network size is time-varying in Appendix D.

5.1. Power evaluation for testing heterogeneity

We first investigate the power of the hypothesis test for heterogeneity that we introduce in Section 4. To do so, we investigate both Type I and Type II errors of the test on dynamic networks over various magnitudes of temporal heterogeneity. We simulate 100 sequences of dynamic networks $\mathbf{x} = \{\mathbf{x}_1, \ldots, \mathbf{x}_{100}\}$, where each sequence $\mathbf{x}_w = \{x_{w,1}, \ldots, x_{w,K}\}$, contains K networks with 30 nodes observed at equally-spaced times $t_1, \ldots, t_K$ under the VCERGM that models the temporal contributions of the edge density statistic. We set the coefficient on the edge density term, $\phi(t)$, to be a sinusoidal curve with amplitude M and period T. In particular, we model

$$\phi(t) = M \sin\left(\frac{2\pi t}{T}\right), \quad t \in [0, T].$$

We vary the number of observed time points K from 10 to 100, and the amplitude M from 0 to 0.3 in increments of 0.05. In case that $M = 0$, we set $\phi(t) = 1$, $t \in [0, T]$ to represent an Erdős-Rényi model. For each value of K and M, we calculate the proportion of rejections at $\alpha = 0.05$ level out of the 100 simulated dynamic network sequences. Table 1 reports these proportions when using the bootstrap procedure as well as the permutation test. For the permutation test, instead of simulating networks from the estimated null, we simply permute the observed networks to break any time-varying pattern, and re-estimate the model under the null and the alternative and then calculate the test statistics. We learned that both testing strategies appear to be overly conservative. It is a valuable point we would like to address for future research.

When $M = 0$, $\phi(t)$ is a constant function and as a result the proportion of rejections in this case provides an estimate for the Type I error of each test. From Table 1, we see that both strategies obtain a Type I error at or below 0.05, as desired. For $M > 0$, the proportion of rejections provides an estimate of the power of the test. We see that for higher signal (larger M) and for a larger number of observed networks (larger K), we obtain a higher power, as expected. Across K, we see in general that the bootstrap procedure is consistently more powerful than the permutation procedure for each amplitude value M. For $M > 0.25$ the power of both tests reaches 1, indicating that heterogeneity is successfully identified by both tests. These results suggest that both tests are powerful for large enough signal size, and that the bootstrap procedure slightly outperforms the permutation procedure for small signal sizes (between $M = 0.05$ and 0.20).

5.2. Estimation performance

We now evaluate the performance of VCERGM to accurately estimate fluctuating parameters $\boldsymbol{\phi}(t)$, $t \in [0, T]$. We consider four different settings for $\boldsymbol{\phi}(t)$: (i) sinusoidal curve $\phi(t) = a \sin\{(t + b)/c\} + d$ of varying amplitude a;

Table 2
Simulation results with 30 nodes and (0, 1, 5, 10) missing networks: Mean and standard deviation of the integrated absolute errors (IAE) for each method.

	Missing	Edges			Reciprocity		
		ERGM	ERGM2	VCERGM	ERGM	ERGM2	VCERGM
Sinusoidal	0	18.63 (11.06)	11.79 (11.76)	**5.07** (1.05)	20.99 (11.11)	14.28 (12.24)	**6.54** (1.25)
	1		12.41 (11.79)	**5.35** (1.19)		14.63 (12.28)	**7.13** (1.32)
	5		12.92 (11.83)	**5.67** (1.37)		14.55 (12.25)	**7.6** (1.3)
	10		12.89 (13.76)	**5.44** (1.21)		13.87 (12.97)	**7.53** (1.33)
Quadratic	0	6.33 (0.74)	2.87 (1.08)	**2.86** (0.97)	8.58 (0.87)	**3.07** (1.1)	3.19 (0.83)
	1		2.9 (1.09)	**2.87** (0.99)		**3.16** (1.14)	3.22 (0.86)
	5		3.13 (1.11)	**2.98** (1)		**3.38** (1.27)	**3.38** (0.89)
	10		3.29 (1.28)	**3.05** (0.97)		3.56 (1.41)	**3.52** (0.93)
Erdős-Rényi	0	14.17 (2.5)	6.92 (3.8)	**5.93** (2.71)	15.91 (2.58)	7 (3.76)	**6.06** (2.67)
	1		7 (3.86)	**5.98** (2.76)		7.09 (3.83)	**6.1** (2.75)
	5		7.09 (3.79)	**6.06** (2.68)		7.14 (3.83)	**6.17** (2.69)
	10		7.11 (4.03)	**6.15** (2.83)		7.28 (3.93)	**6.35** (2.81)
Non-smooth	0	**12.62** (6.53)	32.32 (2.96)	31.27 (0.35)	**15.88** (6.61)	21.9 (4.46)	24.08 (0.93)
	1		32.25 (2.79)	**31.16** (0.36)		**21.86** (4.58)	22.65 (0.89)
	5		32.46 (3.47)	**31.34** (0.45)		**22.19** (5)	22.48 (0.88)
	10		32.63 (4.13)	**31.42** (0.44)		**22.91** (6.14)	23.46 (0.92)

(ii) quadratic curve $\phi(t) = a(t - T/2)^2 + b$ of varying strength a; (iii) dynamic Erdős-Rényi random graph with probability p of edges; and (iv) non-smooth (spiky) functions as a form of a sequence of random numbers with varying mean and standard deviation for normal distribution. For each setting of varying coefficients, we model the occurrence of graphs using the VCERGM with edge density and reciprocity statistics. We simulate 100 dynamic sequences of directed graphs $\{\mathbf{x}_1, \ldots, \mathbf{x}_{100}\}$ where each sequence $\mathbf{x}_w = \{x_{w,1}, \ldots, x_{w,50}\}$ is observed at $K = 50$ equally-spaced time points. We assume that the network size remains constant through time and consider estimation with networks of three different sizes $n = 30, 50, 100$. Furthermore, we repeat (i)–(iv) with 1, 5, and 10 randomly chosen networks removed from the time series to evaluate the performance on dynamic networks with observations missing at random.

5.2.1. Competing methods

For each simulated dynamic network, we compare the VCERGM with two other dynamic network models. First, we fit cross-sectional ERGMs, where the ERGM in model (1) is fit separately at each of the K observed time points. As an alternative competitive method, we also develop an ad hoc 2-step procedure, which adapts an ad hoc smoothing procedure after fitting cross-sectional ERGMs for observed networks. Namely, let $\hat{\phi}(t)$ denote the estimate of $\phi(t)$. The ad hoc smoothing mechanism aims to find a smooth function $f(t)$ that minimizes the penalized residual sum of squares (RSS)

$$\mathrm{RSS}(f, \lambda) = \sum_{s=t_1}^{t_K} \{\hat{\boldsymbol{\phi}}(s) - f(s)\}^2 + \lambda \int \{f''(t)\}^2 dt,$$

where λ is a tuning parameter that controls the amount of roughness. The generalized cross validation (GCV) is used to choose the tuning parameter λ (Golub et al., 1979).

5.2.2. Performance metrics

To assess the performance of each method, we calculate the integrated absolute error (IAE) of the estimated coefficient curves. It measures the sum of point-wise absolute difference between estimated curve $\hat{\phi}(t)$ and true curve $\phi(t)$ at observed time points $t_1, \ldots, t_K$, namely

$$\mathrm{IAE}(\phi(t), \hat{\phi}(t)) = \sum_{s=t_1}^{t_K} |\phi(s) - \hat{\phi}(s)|.$$

The mean and standard deviation (SD) are calculated to evaluate the performance of our proposed method compared to cross-sectional ERGMs and ad hoc 2-step procedure. We provide the summary of IAE for each method on dynamic networks with 30 nodes in Table 2 with (0, 1, 5, 10) missing networks. Settings for the results are (i) sinusoidal curves with $(a, b, c, d) = (1, 30, 5, 1)$ (edges) and $(a, b, c, d) = (0.6, 20, 3, 0.4)$ (reciprocity); (ii) quadratic curves with $(a, b) = (1/20^2, 0)$ (edges) and $(a, b) = (-1/25^2, 0.5)$ (reciprocity); (iii) Erdős-Rényi with $p_{edges} = 0.85$; (iv) a sequence of random numbers from $N(0, 1)$ (edges) and $N(1.5, 0.5)$ (reciprocity). The performances of cross-sectional ERGMs, ad hoc 2-step procedure, and VCERGM become more comparable with larger network size. For results of $n = 50$ and $n = 100$ case, see Tables C.5 and C.6 in Appendix C.

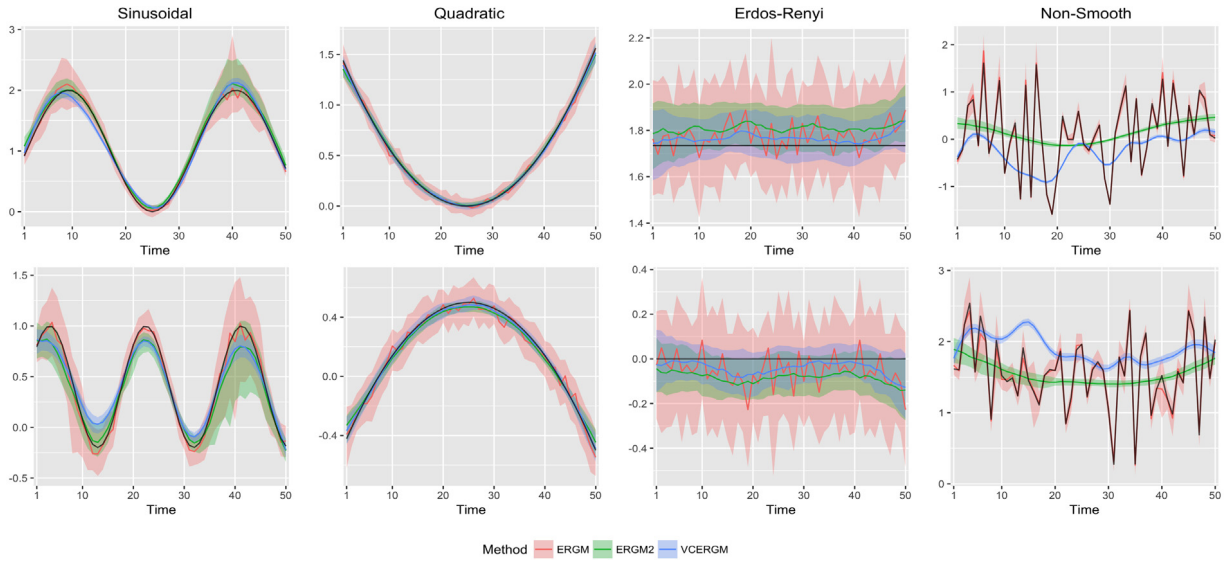


Fig. 3. Parameter estimates with 30 nodes: Estimated parameters for edges (top) and reciprocity (bottom). Black line is the true $\boldsymbol{\phi}(t)$. Red (ERGM) is for cross-sectional ERGMs, green (ERGM2) is for ad hoc 2-step procedure, and blue (VCERGM) is for VCERGM. For each method, solid line indicates the average of 100 estimated curves and the shaded band illustrates the first and third quantiles. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5.2.3. Results

We first investigate the results of the methods when there are no missing networks. These results are presented in Fig. 3. We find that cross-sectional ERGMs are more likely to introduce unexpected spikes or increased variability in estimating true $\boldsymbol{\phi}(t)$, compared to VCERGM. Overall, the VCERGM estimates deviate less from the true $\boldsymbol{\phi}(t)$ and has smaller variability compared to cross-sectional ERGMs and ad hoc 2-step procedure. In the first three functional types, the VCERGM outperforms the cross-sectional ERGMs and ad hoc 2-step procedure. In the case of non-smooth functions, the ad hoc 2-step procedure shows better performance than the VCERGM with respect to IAE. Both Table 2 and Fig. 3 indicate that the VCERGM potentially misses random deviations, which causes greater bias on average compared to cross-sectional ERGMs. Despite this, the true non-smooth $\boldsymbol{\phi}(t)$ is well captured by the VCERGM, and the variability of the VCERGM estimators are smaller than the coefficients for the cross-sectional ERGMs. The performance of ad hoc 2-step smoothing procedure is comparable with the VCERGM, but the VCERGM provides a more principled model for incorporating time-varying coefficients.

For all four functional types, the VCERGM is computationally more efficient than the cross-sectional ERGMs. We conduct an additional simulation study specifically tailored to compare the computing time between methods and the results are presented in Tables 3 and 4 (see Fig. 3).

When there exist missing networks, cross-sectional ERGMs are no longer available to provide the estimates at unobserved time points. Therefore, the IAE is calculated only for ad hoc 2-step procedure and VCERGM. Notably, the performance of the VCERGM remains stable across each number of missing networks. Cross-sectional ERGMs and the 2-step approach, on the other hand, suffer more than the VCERGM in the case of missing networks. Indeed, as shown in Table 2, the VCERGM outperforms these competitive methods in the case that observations are missing and is better able to capture the true coefficient curve in these cases.

In order to compare the computational efficiency, we vary the number of time points K and the number of nodes n and record the computing time for VCERGM and cross-sectional ERGMs. Table 3 summarizes the computing times of 100 simulated dynamic network sequences of 30 nodes and displays how computing time changes as the number of time points K changes. Similarly, Table 4 shows the distribution of computing times of 100 simulated dynamic network sequences for varying number of nodes n in temporal networks with 30 time points. The maximum pseudo-likelihood approach is used for both ERGM and VCERGM estimation.

According to Table 3, the VCERGM takes significantly less time than cross-sectional ERGMs to complete the parameter estimation. Even if both VCERGM and cross-sectional ERGMs show a linear increase in computing time, the rate of change is much smaller for VCERGM. Both methods entail K separate steps to construct design matrix and response vector at each time point, but the cross-sectional ERGMs require K separate MPLE steps while VCERGM only needs one estimation. In other words, the longer the time series of networks are, the more efficient VCERGM is compared to cross-sectional ERGMs.

Table 4 shows a consistently shorter computing time for VCERGM compared to cross-sectional ERGMs. For both VCERGM and cross-sectional ERGMs, the increment in computing times is almost linear up to temporal networks with 100 nodes. As the network size gets bigger than 100, the computing time increases exponentially, and the difference in

Table 3
Computing Time: Summary (Mean(SD)) of computing time (second) for dynamic networks with different number of time points K.

	Number of time points K				
	20	40	60	80	100
ERGM	1.35 (0.07)	2.64 (0.08)	3.83 (0.13)	5.08 (0.10)	6.35 (0.17)
VCERGM	0.95 (0.07)	1.83 (0.09)	2.69 (0.14)	3.41 (0.12)	4.37 (0.11)

Table 4
Computing Time: Summary (Mean(SD)) of computing time (second) for dynamic networks with different number of nodes n.

	Number of nodes n						
	20	40	60	80	100	200	500
ERGM	1.84 (0.11)	2.35 (0.12)	3.20 (0.09)	4.31 (0.12)	5.89 (0.18)	20.46 (0.59)	224.73 (12.74)
VCERGM	1.21 (0.08)	1.66 (0.06)	2.24 (0.07)	3.15 (0.09)	4.25 (0.14)	14.53 (0.48)	179.09 (11.11)

mean computing time between VCERGM and cross-sectional ERGMs for temporal networks with 500 nodes is close to 1 min.

6. Application to U.S. congressional co-voting behavior

We apply the VCERGM to analyze how the co-voting patterns among U.S. Senators have changed through time. We analyze the effects of political affiliation (Republican or Democrat) on the likelihood of the voting networks. We first test for temporal heterogeneity of any statistic included in the model using our bootstrap procedure from Section 4. We compare the VCERGM with cross-sectional ERGMs and the ad hoc 2-step procedure described in Section 5.2.1 We furthermore compare these with estimates from the TERGM model with the same specified statistics for comparison.

6.1. Data and model specification

This dynamic network that describes the co-voting patterns among U.S. Democrat and Republican Senators from 1867 (Congress 40) to 2015 (Congress 113). Three of the voting networks are shown in Fig. 1. This network was first investigated in Moody and Mucha (2013) and has been subsequently analyzed in Wilson et al. (2019). The network is based on the roll call voting data from http://voteview.com, which contains the voting decision of each Senator (yay, nay, or abstain) for every bill brought to Congress. We model the co-voting tendencies of the Senators using a dynamic network where nodes represent Senators and an edge is formed between two nodes if the two Senators vote concurringly (both yay or both nay) on at least 80% of the bills to which they were both present. We note that we exclude Independent Senators for this analysis.

As shown in Fig. 1, there are noticeable fluctuations in the co-voting network structure over time. Previous analyses in Moody and Mucha (2013) and Wilson et al. (2019) have identified significant changes in the community structure of the network over time, and that this community structure is closely associated with the political affiliation of the Senators. To model these fluctuations, we include a mixing matrix effect that counts the number of edges among Senators with the same affiliation as well as the number of edges between Senators with different affiliations. We also include the geometrically weighted edge shared partners (GWESP, decay = 1) and geometrically weighted dyadic shared partners (GWDSP, decay = 1). The GWESP statistic measures the extent to which two senators who are tied share connections with other senators. The GWDSP statistic captures the extent to which each pair of Senators shares similar ties.

6.2. Model estimates and analysis

The estimated parameters from (i) cross-sectional ERGMs (ERGM), (ii) ad hoc 2-step procedure (ERGM2), (iii) VCERGM, and (iv) TERGM are presented in Fig. 4. Notably, all five network statistics exhibit temporal heterogeneity. The permutation test p-value for testing heterogeneity is $< .001$. Like we found in the simulation results, the cross-sectional ERGMs exhibits spiky estimates, but the ad hoc smoothing recovers the lack of smoothness efficiently and produces similar estimates as the VCERGM. We note that the TERGM coefficient estimates from the `btergm` package are averages of the coefficient estimates from the cross-sectional ERGM. As explained in Desmarais and Cranmer (2012) and Leifeld et al. (2018), the reason for this averaging behavior of the TERGM is the `btergm` package identifies the maximum pseudo-likelihood estimator for the coefficients in the sequence of graphs, which turns out to be the same as taking the average coefficient estimate across independent ERGM models with the same coefficients. In comparing the VCERGM with competing methods, we see that the coefficient estimates from the VCERGM are smooth functions across time that simultaneously account for the heterogeneity of the coefficients as well as avoids outliers.

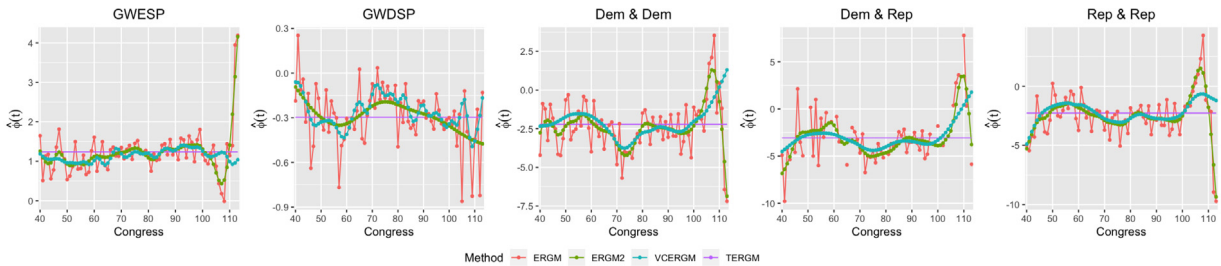


Fig. 4. Parameter estimates of political networks: Coefficient estimates for the VCERGM, cross-sectional ERGM (ERGM), the ad hoc smoothing approach to the ERGM (ERGM2), as well as the TERGM.

The political mixing matrix effect parameter estimate reveals an important trend in the political network. We see that the coefficient value for this term generally increases over time, suggesting that political affiliation has had a growing importance in co-voting behavior over time. This increasing pattern in the coefficient across affiliations reflects that the number of ties formed with the same political affiliation positively influences the likelihood. We notice that since the increase has been particularly evident since Congress 95, which matches the current theory of "political polarization" described in Moody and Mucha (2013).

Fig. 4 reveals that the GWESP coefficient remained positive and relatively stable over time until Congress 107. This trend suggests that the clustering of Senator votes positively influences the formation of the co-voting network. At Congress 107, the GWESP coefficient estimate greatly increased. This closely aligns with the notable polarization of Republicans and Democrats starting in the Clinton administration, which has also been noted in previous studies (Moody and Mucha, 2013). The GDWSP coefficient remained negative and relatively stable over time. This, in combination with the GWESP results, supports the claim that the connection of two Senators has a positive influence on having shared partners and suggests that their political alliances were formed over history. These findings suggest that the U.S. Senate transitioned from an "individual centric" network, where central figures influenced the voting habits of the Congress, to a "party centric" network in which political affiliation was the primary determinant of voting habits. This finding augments the empirical work in Moody and Mucha (2013).

6.3. Goodness of fit

We next assess the within-sample accuracy of the VCERGM through a goodness of fit study. To assess goodness of fit, we follow the strategy established in Hunter et al. (2008b), described as follows. First, for the tth network x_t in the observed sequence, we calculate the marginal coefficients $\boldsymbol{\phi}(t)$ from the results of the VCERGM. We then simulate 100 networks with parameters $\boldsymbol{\phi}(t)$ using Markov chain Monte Carlo as described in Hunter and Handcock (2006). We calculate a family a summary statistics for each of the simulated networks and compare the distribution of these statistics with the true observed value of the statistic for x_t. We did this comparison for each of the statistics in the model – triangle, two-star, and node-mix terms – as well as several statistics not included in the model, including an edges term and a geometrically weighted edge shared partners (GWESP) and geometrically weighted dyad shared partners (GWDSP). Our goodness of fit results are plotted in Fig. 5.

Fig. 5 does not appear to suggest any systematic bias in the simulated network statistics. Indeed, the dynamic trend of the simulated networks closely matches the trend of the statistics in the true dynamic sequence. To test this, we calculated the correlation between the median of network statistics from the simulated networks and the true observed value. The range of calculated correlations is between 0.76 and 0.98. This is particularly reassuring, as pseudo-likelihood estimation for static ERGMs may lead to biased parameter estimates for small networks (Strauss and Ikeda, 1990; Van Duijn et al., 2009). We note that there are some time points at which the simulated networks do not closely match the observed Republican to Republican mixing matrix effect and GWESP terms. These time points correspond to congresses where the smoothed coefficient estimates of the VCERGM tend to differ from the jumpy nature of the cross-sectional ERGM estimates as seen in Fig. 4.

7. Discussion

In this paper, we introduce varying-coefficient models for dynamic networks. In particular, we described the formulation and estimation of the VCERGM, a model that incorporates temporal changes in the coefficients of an exponential random graph family of models. We demonstrated the advantages of applying the VCERGM over competing methods through simulations and two dynamic network case studies. First, the VCERGM provides an intuitive explanation of how a network changes through time. Both the cross-sectional ERGMs and ad hoc 2-step procedure seemed to capture the temporal heterogeneity in a sense. However, by incorporating the temporal heterogeneity in the modeling step, the VCERGM provides a compact and meaningful model to formally explain the temporal structure of dynamic networks.

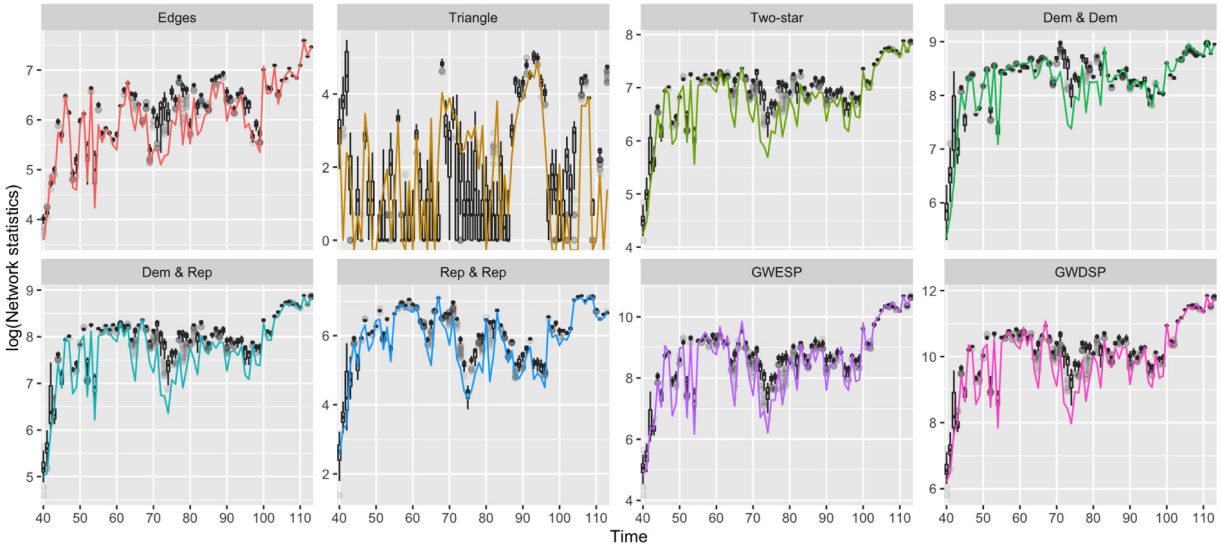


Fig. 5. Goodness of fit of the VCERGM on political networks: Estimated coefficients from the VCERGM were used to simulate 100 networks for each observed time point. The distribution of each network statistic (log transformed) is shown for each Congress. The log of true observed statistics of the co-voting network are shown with solid lines.

Second, the VCERGM is robust to perturbations in observed temporal data. By imposing smoothness on the coefficients, we are able to provide robust estimates that are resistant to outliers and noise. Third, the VCERGM enables interpolation for missing networks through time. In practice, one can only observe a finite number of networks in a dynamic sequence, which may be observed in unequally spaced time increments. Estimates of the coefficients to the VCERGM can be evaluated at any time point in the domain and immediately interpreted as the impact of network statistics at that time point. By presenting the results with unequally-spaced networks, we illustrated how the varying-coefficients through time can be useful especially in terms of interpolation.

Our work provides several avenues for future research. First, it is important to consider the evaluation of goodness of fit and model selection in a dynamic context. Through empirical exploration, we found that the network statistics used to fit a model are often highly correlated. For example, if there exists a triangle in a network, it is more likely to find two-stars in the network. Model identifiability should be investigated both in static ERGM models and the VCERGM to ensure appropriate model selection. For static ERGMs, one generally assesses goodness of fit through a comparison of quantitative summaries of simulated networks from the fitted model with the summaries of the observed network (Hunter et al., 2008b). However, for dynamic networks this type of goodness of fit comparison captures only the marginal aspects of the dynamic sequence. How exactly to assess the quality of a dynamic model is still an open problem. A second avenue for future work involves adapting the varying-coefficient framework introduced here to networks with weighted edges. To do this, one can extend the exponential models of networks for integer-valued weights from Krivitsky (2012) or to the models of networks for continuous-valued weights considered in Desmarais and Cranmer (2012), Wilson et al. (2017), Stillman et al. (2017) and Stillman et al. (2019).

We discussed a maximum pseudo-likelihood approach for parameter estimation. This strategy provides a computationally feasible approach to fitting dynamic networks with a large number of nodes or time steps, especially when compared to the typical simulation-based MCMC-MLE approach. There have been several studies exploring the relationship between the pseudo-likelihood and the likelihood of the ERGM, including Strauss and Ikeda (1990), Desmarais and Cranmer (2012), where the efficacy of MPLE was empirically compared to MCMC-MLE. More recently, Schmid and Desmarais (2017) compared the performance of MCMC-MLE and MPLE and empirically supported the accuracy of MPLE. Despite this, theoretical support for MPLE is still lacking and is an open area for future research.

In many dynamic networks, it is often of interest to identify change-points in the network, namely points in time where the network undergoes significant local or global structural change (Woodall et al., 2017; Bindu and Thilagam, 2016). It would be interesting to further analyze how to utilize dynamic network models like the VCERGM to identify such changes. The test for heterogeneity that we use in the paper may provide some idea of how to formally test for a change — through the identification of a change in network parameter. We plan to pursue this idea further in future research.

Acknowledgments

The fMRI data were provided [in part] by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support

the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University. Gen Li's research was partially supported by the Calderone Junior Faculty Award from the Mailman School of Public Health at Columbia University. James D. Wilson was partially supported by the National Science Foundation grant NSF DMS-1830547.

Appendix A. Stochastic equivalence under the difference statistic specification

Comparing the first-order TERGM with model (1), we see that the TERGM is closely related to the ERGM in that it characterizes the conditional distribution of X_t given X_{t-1} using an ERGM representation. Perhaps not surprisingly, these two models are much more closely related than this relationship.

Consider a simple univariate time series represented by the stochastic process $\mathbf{Z} = \{Z_1, \ldots, Z_T\}$ for $Z_t \in \mathbb{R}$. Without any other information about $\mathbf{Z}$, a natural non-parametric manner to investigate the rate of change in $\mathbf{Z}$ involves analyzing the difference between sequential observations Z_{t-1} and Z_t, namely analyzing $\Delta(Z_t) = Z_t - Z_{t-1}$. The analysis of $\Delta(Z_t)$ in univariate and multivariate time series is known as *differencing*, and is a common first step in the analysis of time series data (Brockwell and Davis, 2013). In the context of the TERGM, differencing corresponds to the analysis of *difference statistics*, where one specifies transition statistics of the form

$$\mathbf{g}(x_t, x_{t-1}) = \mathbf{h}(x_t) - \mathbf{h}(x_{t-1}), \text{ t = 2, ..., T}, \tag{A.1}$$

where $\mathbf{h} : \mathcal{X} \to \mathbb{R}^p$ is a topological summary of an input network with n vertices. Statistics of the form in (A.1) can capture, for example, the differences in the edge weight of the network from time $t-1$ to t, or the difference in the number of triangles from one network to the next. Although incorporating difference statistics in the TERGM is a natural first-step in the analysis of temporal networks, it turns out that doing so is equivalent to modeling each network $X_t \in \boldsymbol{X}$ as an independent realization from the *same* exponential family probability mass function. This is made precise in the next proposition.

Proposition 1. *Let $\boldsymbol{\mathcal{X}}$ denote the family of unweighted dynamic graph sequences on n vertices with $T \geq 1$ discrete observations. Suppose that $\boldsymbol{X} = \{X_1, \ldots, X_T\} \in \boldsymbol{\mathcal{X}}$ is generated under the TERGM in (3), where for $t = 2, \ldots, T$*

$$X_t \mid \boldsymbol{X}_t^- \sim \mathbb{P}(X_t = x_t \mid X_{t-1} = x_{t-1}; \boldsymbol{\phi}) = \frac{\exp\{\boldsymbol{\phi}^T \mathbf{g}(x_t, x_{t-1})\}}{\sum_{z \in \mathcal{X}} \exp\{\boldsymbol{\phi}^T \mathbf{g}(z, x_{t-1})\}}.$$

Suppose $\mathbf{g}(\cdot,\cdot) \in \mathbb{R}^p$ is a difference statistic of the form (A.1) where $\mathbf{g}(x, y) = \mathbf{h}(x) - \mathbf{h}(y)$ for some $\mathbf{h}(\cdot) \in \mathbb{R}^p$. Then for all $t \geq 2$, X_t is independent of $\boldsymbol{X}_t^-$ and can be generated as an independent realization of an ERGM with the following probability mass function

$$X_t \mid \boldsymbol{X}_t^- \sim \mathbb{P}(X_t = x \mid \boldsymbol{\phi}) = \frac{\exp\{\boldsymbol{\phi}^T \mathbf{h}(x)\}}{\sum_{z \in \mathcal{X}} \exp\{\boldsymbol{\phi}^T \mathbf{h}(z)\}}.$$

Proposition 1 reveals that under the difference statistic model specification, a dynamic network under the TERGM reduces to an independent and identically distributed sequence of graphs under a corresponding ERGM. Hence under this family of specifications, the TERGM does *not* capture temporal dependence in the underlying dynamic network sequence. Although in practice one may utilize statistics that are not of the form (A.1), this relatively simple example motivates further investigation between the relationship of the ERGM and the TERGM.

Appendix B. Iterative reweighted least squares (IRLS)

The penalized logistic regression problem for fitting a VCERGM is to maximize the following penalized log likelihood function:

$$\tilde{\mathbf{x}}^T \mathbf{H}^T \text{vec}(\boldsymbol{\Phi}) - \mathbf{1}^T \log[1 + \exp\{\mathbf{H}^T \text{vec}(\boldsymbol{\Phi})\}] - \lambda \mathcal{P}_{\boldsymbol{\Omega}}(\boldsymbol{\Phi}). \tag{B.1}$$

The tuning parameter λ controls the amount of roughness. We implement the iteratively reweighted least squares (IRLS) to fit the logistic regression with the penalty term. Consider a link function $g(\mu) = \log(\mu/(1-\mu))$ and a convex function $b(\eta) = \log(1 + e^\eta)$. The IRLS without penalty term updates Φ at the $(u+1)$th iteration

$$\text{vec}(\boldsymbol{\Phi}^{(u+1)}) = (\mathbf{H}^T \mathbf{W}^{(u)} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{W}^{(u)} \left\{ \mathbf{H}\text{vec}(\boldsymbol{\Phi}^{(u)}) + (\tilde{\mathbf{x}} - \boldsymbol{\mu}^{(u)}) \cdot g'(\boldsymbol{\mu}^{(u)}) \right\}, \tag{B.2}$$

where $\boldsymbol{\mu}^{(u)} = b'(\mathbf{H}\text{vec}(\boldsymbol{\Phi}^{(u)}))$ and $\mathbf{W}^{(u)}$ is a diagonal matrix with

$$\mathbf{W}_{(i,i)}^{(u)} = \frac{1}{b''(\mathbf{H}_{(i)}^T \text{vec}(\boldsymbol{\Phi}^{(u)}))} \frac{1}{\{g'(\boldsymbol{\mu}_i^{(u)})\}^2}, \qquad i = 1, 2, \ldots, (p \times q).$$

Table C.5
Simulation results with 50 nodes and (0, 1, 5, 10) missing networks: Mean and standard deviation of the integrated absolute errors (IAE) for each method.

	Missing	Edges			Reciprocity		
		ERGM	ERGM2	VCERGM	ERGM	ERGM2	VCERGM
Sinusoidal	0	7.84 (2.05)	4.24 (2.89)	**4.06** (2.8)	8.53 (0.99)	4.62 (1.23)	**4.35** (1.05)
	1		4.82 (2.89)	**4.45** (2.82)		5.05 (1.23)	**4.89** (1.06)
	5		5.58 (2.85)	**4.94** (2.8)		5.65 (1.2)	**5.46** (1)
	10		5.19 (2.77)	**4.43** (2.63)		5.42 (1.46)	**5.32** (1.1)
Quadratic	0	3.8 (0.62)	**1.84** (0.96)	1.89 (0.91)	5.06 (0.57)	**1.76** (0.63)	2.06 (0.46)
	1		**1.91** (0.93)	1.93 (0.89)		**1.82** (0.64)	2.08 (0.49)
	5		2.12 (0.87)	**2.05** (0.83)		**2.04** (0.69)	2.19 (0.53)
	10		2.2 (0.93)	**2.12** (0.82)		**2.09** (0.67)	2.31 (0.58)
Erdős-Rényi	0	8.4 (2.62)	4.28 (3.81)	**4.21** (3.7)	8.62 (1.43)	3.17 (1.94)	**3.14** (1.4)
	1		**4.22** (3.82)	**4.22** (3.7)		**3.15** (1.95)	3.16 (1.39)
	5		4.34 (3.75)	**4.3** (3.65)		**3.2** (2.01)	3.24 (1.45)
	10		**4.29** (3.58)	4.31 (3.48)		3.43 (2.06)	**3.38** (1.35)
Non-smooth	0	**5.77** (1.34)	30.92 (0.28)	30.21 (0.28)	**7.9** (1.95)	19.68 (0.74)	23.23 (1.01)
	1		30.98 (0.3)	**30.03** (0.29)		**19.64** (0.76)	21.91 (0.84)
	5		30.81 (0.25)	**30.12** (0.38)		**19.75** (0.79)	21.64 (0.74)
	10		30.62 (0.26)	**30.24** (0.39)		**19.85** (0.79)	23.01 (0.89)

Table C.6
Simulation results with 100 nodes and (0, 1, 5, 10) missing networks: Mean and standard deviation of the integrated absolute errors (IAE) for each method.

	Missing	Edges			Reciprocity		
		ERGM	ERGM2	VCERGM	ERGM	ERGM2	VCERGM
Sinusoidal	0	18.48 (20.76)	17.81 (21.23)	**17.74** (21.3)	8.44 (5.11)	7.61 (5.71)	**7.5** (5.75)
	1		18.4 (20.95)	**18.28** (21.04)		7.9 (5.5)	**7.85** (5.51)
	5		18.95 (20.55)	**18.81** (20.65)		8.26 (5.34)	**8.22** (5.38)
	10		18.3 (20.78)	**18.07** (20.97)		8.23 (5.45)	**8.07** (5.51)
Quadratic	0	9.14 (12.84)	**8.59** (13.16)	8.67 (13.12)	3.53 (1.95)	**2.44** (2.53)	2.56 (2.33)
	1		**8.61** (13.09)	8.66 (13.06)		**2.49** (2.52)	2.59 (2.33)
	5		**8.58** (12.96)	8.6 (12.95)		**2.59** (2.42)	2.68 (2.27)
	10		**8.63** (12.97)	8.65 (12.95)		**2.6** (2.42)	2.67 (2.26)
Erdős-Rényi	0	23.22 (27.46)	**22.32** (28.11)	22.47 (28.01)	3.61 (0.96)	**1.09** (0.53)	1.5 (0.4)
	1		**22.34** (28.11)	22.49 (28.02)		**1.05** (0.54)	1.49 (0.39)
	5		**22.36** (28.08)	22.55 (27.97)		**1.14** (0.52)	1.59 (0.38)
	10		**22.36** (28)	22.52 (27.87)		**1.3** (0.65)	1.73 (0.49)
Non-smooth	0	**10.95** (9.95)	32.92 (3.96)	30.6 (3.98)	**16.63** (15.89)	26.48 (10.91)	26.32 (9.67)
	1		33.05 (3.97)	**30.52** (4.01)		26.39 (10.8)	**25.66** (9.94)
	5		32.72 (3.92)	**31.16** (3.99)		26.75 (11.08)	**25.73** (10.39)
	10		32.69 (4.06)	**31.53** (4.02)		**26.76** (10.99)	27.15 (9.7)

With the penalty term $\mathcal{P}(\boldsymbol{\Phi})$, we only need to replace $\mathbf{H}^T\mathbf{W}^{(u)}\mathbf{H}$ by $\mathbf{H}^T\mathbf{W}^{(u)}\mathbf{H}+\lambda(\boldsymbol{\Omega}\otimes\mathbf{I}_p)$ in (B.2). The generalized cross validation (GCV) is used to choose the tuning parameter λ (Golub et al., 1979). Namely, the λ is a minimizer of $G(\lambda)$, which is defined as

$$G(\lambda)=\frac{1}{N}\|\tilde{\mathbf{x}}-\mathbf{H}(\mathbf{H}^T\mathbf{H}+N\lambda\boldsymbol{\Omega})^{-1}\mathbf{H}^T\tilde{\mathbf{x}}\|^2\Bigg/\left\{\frac{1}{N}\mathrm{tr}(I-\mathbf{H}(\mathbf{H}^T\mathbf{H}+N\lambda\boldsymbol{\Omega})^{-1}\mathbf{H}^T)\right\}^2,$$

where N is the number of rows in matrix $\mathbf{H}$.

Appendix C. Additional simulation results

Tables below show the mean and standard deviation of IAE associated with fitting ERGMs and VCERGMs to temporal networks of size 50 and 100 with 0, 1, 5, and 10 randomly missing networks. The results are from the settings (i) sinusoidal curves with $(a,b,c,d)=(1,30,5,1)$ (edges) and $(a,b,c,d)=(0.6,20,3,0.4)$ (reciprocity); (ii) quadratic curves with $(a,b)=(1/20^2,0)$ (edges) and $(a,b)=(-1/25^2,0.5)$ (reciprocity); (iii) Erdős-Rényi with $p_{edges}=0.85$; (iv) a sequence of random numbers from $N(0,1)$ (edges) and $N(1.5,0.5)$ (reciprocity).

Table D.7
Simulation results with time-varying network size and (0, 1, 5, 10) missing networks: Mean and standard deviation of the integrated absolute errors (IAE) for each method.

	Missing	Edges			Reciprocity		
		ERGM	ERGM2	VCERGM	ERGM	ERGM2	VCERGM
Sinusoidal	0	20.63 (12.37)	13.79 (13.14)	**4.99** (1.28)	22.77 (12.32)	16.54 (13.57)	**6.72** (1.37)
	1		14.43 (13.16)	**5.26** (1.37)		16.97 (13.43)	**7.5** (1.33)
	5		14.21 (13.21)	**5.32** (1.47)		16.3 (13.32)	**7.98** (1.41)
	10		14.86 (14.6)	**5.32** (1.18)		15.58 (12.83)	**7.81** (1.52)
Quadratic	0	6.48 (0.66)	2.72 (1.11)	**2.47** (0.74)	8.95 (1.02)	**3.06** (1.27)	3.14 (0.86)
	1		2.79 (1.07)	**2.56** (0.69)		**3.15** (1.35)	3.16 (0.87)
	5		3.1 (1.1)	**2.82** (0.68)		**3.23** (1.24)	3.33 (0.93)
	10		3.11 (1.1)	**2.94** (0.78)		**3.5** (1.49)	3.52 (1.02)
Erdős-Rényi	0	14.52 (1.67)	5.25 (2.43)	**4.72** (1.39)	16.24 (1.84)	5.48 (2.57)	**5.18** (1.52)
	1		5.26 (2.51)	**4.74** (1.42)		5.42 (2.47)	**5.21** (1.54)
	5		5.4 (2.44)	**4.9** (1.49)		5.54 (2.47)	**5.41** (1.59)
	10		5.51 (2.65)	**5.12** (1.6)		5.83 (2.59)	**5.71** (1.61)
Non-smooth	0	**11.36** (5.46)	31.84 (2.61)	31.84 (0.43)	**14.65** (5.52)	21.08 (3.98)	24.61 (1.25)
	1		31.81 (2.47)	**31.57** (0.41)		**21.01** (4.03)	23.35 (1.22)
	5		31.92 (3.11)	**31.74** (0.45)		**21.29** (4.42)	22.96 (1.23)
	10		**31.87** (3.46)	32.29 (0.58)		**21.76** (5.21)	23.82 (1.27)

Appendix D. Estimation for networks with time-varying network size

In dynamic networks, networks at different time points may have differing numbers of nodes, making it inappropriate to compare networks using un-normalized counts. Instead, one should standardize the network statistics to make them comparable over time. We propose to standardize network counts by its maximal possible value. By using density (proportion) instead of count, we can measure and compare the change in the ratio of certain network statistics when the number of nodes is time-varying. For a directed binary graph X_t with n_t nodes, for example, edge density and reciprocity can be defined as $\sum_{i\neq j} x_{ij}^t / \{n_t(n_t-1)\}$ and $\sum_{i<j} x_{ij}^t x_{ji}^t / \binom{n_t}{2}$, respectively. Fitting a VCERGM to temporal networks is to capture the evolution of connectivity pattern of overall relational data. Therefore, even if the network size is time-varying, using standardized statistics enables us to detect the overall pattern as well as maintain the smoothness assumption of $\boldsymbol{\phi}(t)$.

Let $\mathcal{X}_n$ denote the all obtainable networks with n nodes, and define a set of functions $\mathbf{h}(x_t, n_t) : \mathcal{X}_{n_t} \to \mathbb{R}^p$ for $t \in [0, T]$, which quantify the p topological features of network x_t with size n_t. Given $\mathbf{h}(x_t, n_t)$ and the coefficient vector $\boldsymbol{\phi}(t) = (\phi_1(t), \ldots, \phi_p(t))^T \in \mathbb{R}^p$, the marginal likelihood of X_t at time t has an ERGM representation given by

$$\mathbb{P}(X_t = x_t \mid \boldsymbol{\phi}(t)) = \frac{\exp\{\boldsymbol{\phi}(t)^T \mathbf{h}(x_t, n_t)\}}{\sum_{z\in\mathcal{X}_{n_t}} \exp\{\boldsymbol{\phi}(t)^T \mathbf{h}(z, n_t)\}}, x_t \in \mathcal{X}_{n_t}. \tag{D.1}$$

We randomly vary the network size over time, simulate the networks with time-varying network size and (0, 1, 5, 10) randomly missing networks, and estimate the parameters. The results are from the settings (i) sinusoidal curves with $(a, b, c, d) = (1, 30, 5, 1)$ (edges) and $(a, b, c, d) = (0.6, 20, 3, 0.4)$ (reciprocity); (ii) quadratic curves with $(a, b) = (1/20^2, 0)$ (edges) and $(a, b) = (-1/25^2, 0.5)$ (reciprocity); (iii) Erdős-Rényi with $p_{edges} = 0.85$; (iv) a sequence of random numbers from $N(0, 1)$ (edges) and $N(1.5, 0.5)$(reciprocity) (see Table D.7).

Appendix E. Estimation with different number of basis functions

We vary the number of basis functions (5, 7, 10, 12, 15, 17, 20) and compare the performance of VCERGM in estimating the smooth sinusoidal true $\boldsymbol{\phi}(t)$ from temporal networks with $K = 50$. There is no significant difference in estimation performance when the number of basis functions is greater than 10 (see Fig. E.6).

Appendix F. fMRI dataset

We next analyze the structure of brain connectivity in the data provided by the WU-Minn Consortium Human Connectome Project (HCP). The dataset is available at https://db.humanconnectome.org. See Van Essen et al. (2012) for an overview of data acquisition and analysis. The dataset includes the resting-state functional magnetic resonance imaging (rfMRI) of 500 subjects. For each subject, a 15-minute run of rfMRI is recorded. We set 47 local windows and calculate a precision matrix between 50 brain regions based on observations within each window. For a transition from precision matrices to a sequence of dynamic networks, we define the edge density of a network as the proportion of edges in the network. Once the edge density is specified, the threshold can be determined to form an edge between brain regions. With the edge density of 10%, for example, the greatest 10% of partial correlation values would form edges.

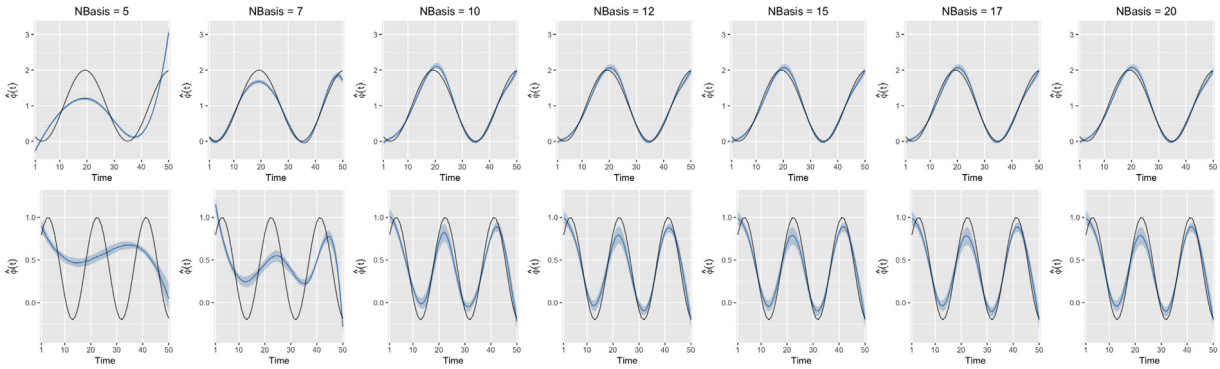


Fig. E.6. Estimation with different number of basis functions: Estimated parameters for edges (top) and reciprocity (bottom). Black line is the true $\boldsymbol{\phi}(t)$. The blue line indicates the average of 100 estimated curves and the shaded band illustrates the first and third quantiles. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

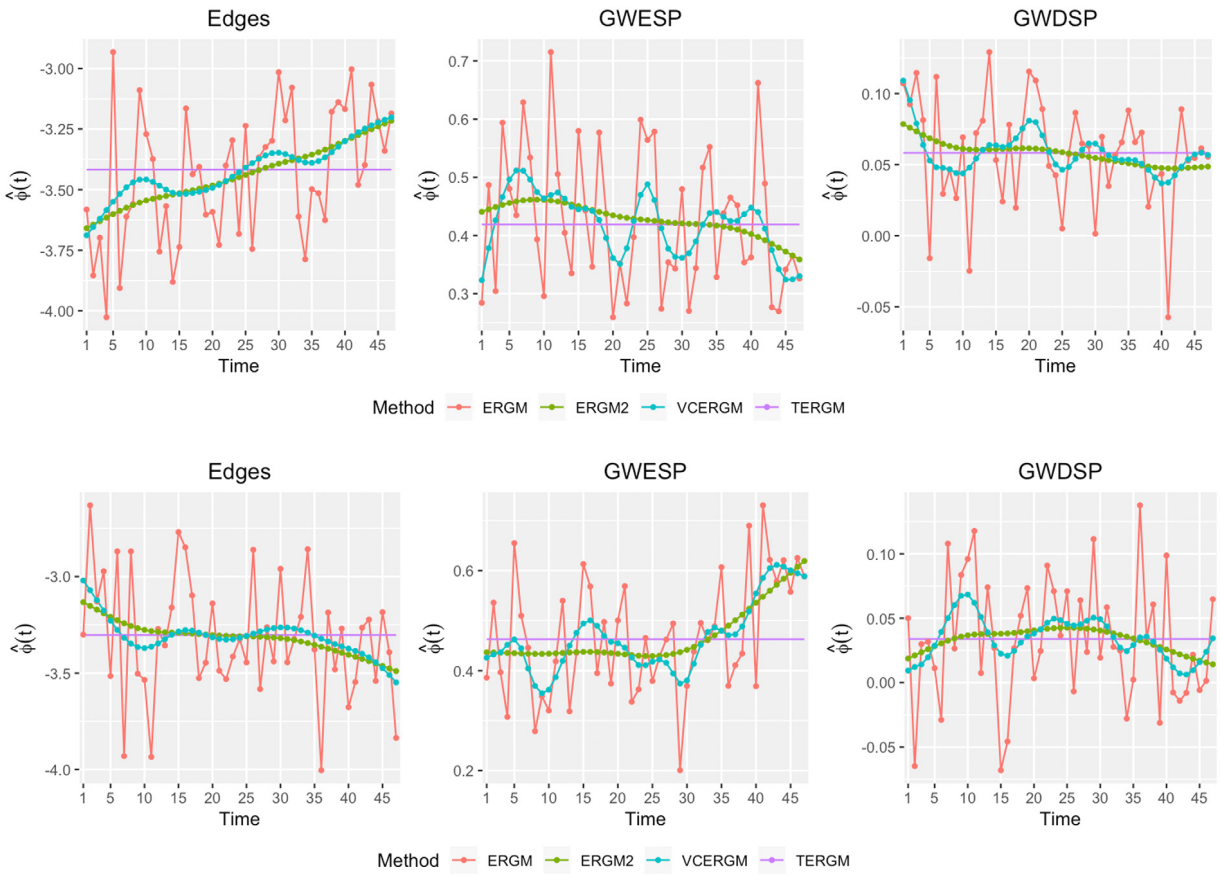


Fig. F.7. Parameter estimates of fMRI networks: Results of two randomly chosen individuals. For all three network statistics, one individual (first row; permutation test p-value $= 0.314$) displays slightly more fluctuations than the other individual (second row; permutation test p-value $= 0.681$). The ad hoc 2-step procedure and VCERGM show similar estimates.

Simpson et al. (2011, 2012) fit the ERGMs to brain networks and conducted extensive model selection. Their final model includes network statistics such as geometrically weighted edge-wise shared partner (GWESP) and geometrically weighted non-edge-wise shared partner (GWNSP). We keep our analysis simple for the sake of comparison of methods. We model our rfMRI networks with three network statistics: edges, triangle and two-star and compare i) cross-sectional ERGMs (ERGM), ii) ad hoc 2-step procedure (ERGM2) and iii) VCERGM. We leave the model selection for the VCERGM for future research.

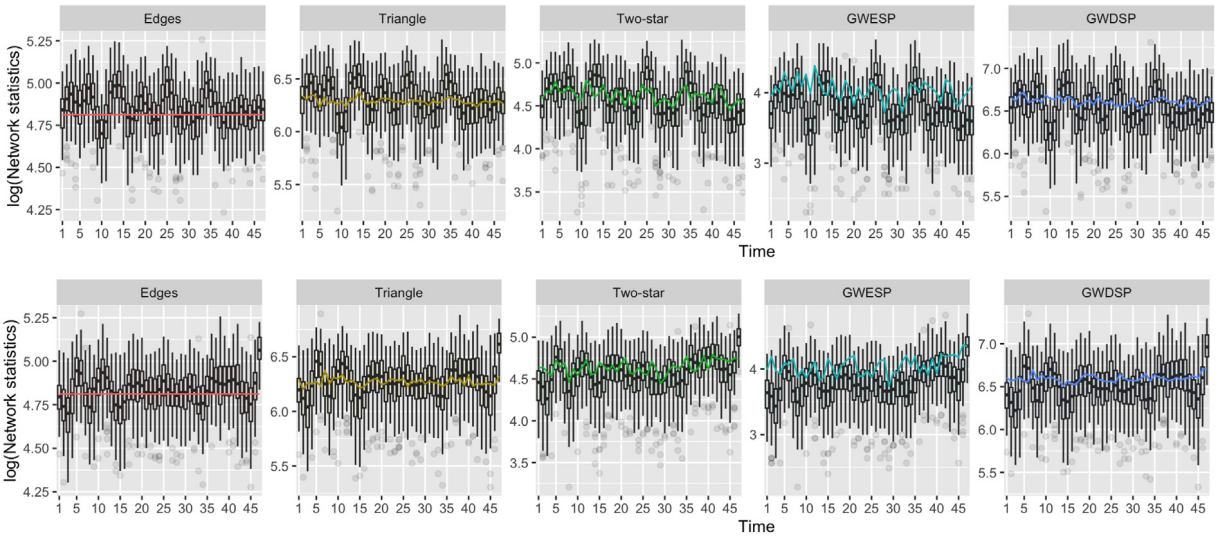


Fig. F.8. Goodness of fit of the VCERGM on fMRI networks: Results of two randomly chosen individuals. Estimated coefficients from the VCERGM were used to simulate 100 networks for each observed time point. The distribution of each network statistic (log transformed) is shown for each time point. The log of true observed statistics of the fMRI network is shown with solid lines.

Fig. F.7 shows the results of two individuals from this study. Computing time for cross-sectional ERGMs and VCERGM is about 1 s. As the data are the resting-state fMRI records, little fluctuation is expected in parameters over time. For both individuals, both ad hoc 2-step procedure and VCERGM provide estimates with a small range of fluctuation for all three network statistics. Overall, the ad hoc 2-step procedure and VCERGM provide relatively similar estimates, while both estimates cross the cross-sectional ERGM estimates. The estimates from cross-sectional ERGMs are extremely jagged that they may introduce inaccurate inference with regard to explaining the topological change in brain networks over time. The VCERGM not only produces fairly static estimates but also captures small variations through time more sensitively than ad hoc 2-step procedure. Therefore, even with relatively stable dynamic networks, the VCERGM performs consistently well. Fig. F.8 shows the goodness of fit plots for two individuals in Fig. F.7. Similar with co-voting network, there was no systematic bias in the simulated network statistics.

References

Akcora, C., Gel, Y.R., Kantarcioglu, M., Lyubchich, V., Thuraisingham, B., 2019. Graphboot: Quantifying uncertainty in node feature learning on large networks. IEEE Trans. Knowl. Data Eng.

Almquist, Z.W., Butts, C.T., 2014. Logistic network regression for scalable analysis of networks with joint edge/vertex dynamics. Sociol. Methodol. 44 (1), 273–321.

Bhamidi, S., Chakraborty, S., Cranmer, S., Desmarais, B., 2018. Weighted exponential random graph models: Scope and large network limits. J. Stat. Phys. 173 (3–4), 704–735.

Bindu, P., Thilagam, P., 2016. Mining social networks for anomalies: Methods and challenges. J. Netw. Comput. Appl. 68, 213–229.

Brockwell, P.J., Davis, R.A., 2013. Time Series: Theory and Methods. Springer Science & Business Media.

Cai, Z., Fan, J., Li, R., 2000. Efficient estimation and inferences for varying-coefficient models. J. Amer. Statist. Assoc. 95 (451), 888–902.

Chen, S., Onnela, J.-P., 2019. A bootstrap method for goodness of fit and model selection with a single observed network. Sci. Rep. 9 (1), 1–12.

Cranmer, S.J., Desmarais, A., 2011. Inferential network analysis with exponential random graph models. Polit. Anal. 19 (1), 66–86.

Daubechies, I., et al., 1992. Ten Lectures on Wavelets, Vol. 61. SIAM.

De Boor, C., De Boor, C., Mathématicien, E.-U., De Boor, C., De Boor, C., 1978. A Practical Guide to Splines, Vol. 27. Springer-Verlag New York.

De Brabanter, J., Pelckmans, K., Suykens, J., De Moor, B., 2006. Generalized likelihood ratio statistics based on bootstrap techniques for autoregressive models. IFAC Proc. Vol. 39 (1), 790–795.

Desmarais, B.A., Cranmer, S.J., 2012. Statistical inference for valued-edge networks: the generalized exponential random graph model. PLoS One 7 (1), e30136.

Eilers, P.H., Marx, B.D., 1996. Flexible smoothing with B-splines and penalties. Statist. Sci. 89–102.

Fan, J., Zhang, W., 2000. Simultaneous confidence bands and hypothesis testing in varying-coefficient models. Scand. J. Stat. 27 (4), 715–731.

Fan, J., Zhang, W., 2008. Statistical methods with varying coefficient models. Stat. Interface 1 (1), 179.

Fan, J., Zhang, C., Zhang, J., 2001. Generalized likelihood ratio statistics and wilks phenomenon. Ann. Statist. 153–193.

Fellows, I., Handcock, M.S., 2012. Exponential-family random network models. arXiv preprint arXiv:1208.0121.

Fienberg, S.E., 2012. A brief history of statistical models for network analysis and open challenges. J. Comput. Graph. Statist. 21 (4), 825–839.

Goldenberg, A., Zheng, A.X., Fienberg, S.E., Airoldi, E.M., 2010. A survey of statistical network models. Found. Trends® Mach. Learn. 2 (2), 129–233.

Golub, G.H., Heath, M., Wahba, G., 1979. Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics 21 (2), 215–223.

Goodreau, S.M., Kitts, J.A., Morris, M., 2009. Birds of a feather, or friend of a friend? Using exponential random graph models to investigate adolescent social networks. Demography 46 (1), 103–125.

Guo, F., Hanneke, S., Fu, W., Xing, E.P., 2007. Recovering temporally rewiring networks: A model-based approach. In: Proceedings of the 24th International Conference on Machine Learning. ACM, pp. 321–328.
Handcock, M.S., Robins, G., Snijders, T., Moody, J., Besag, J., 2003. Assessing degeneracy in statistical models of social networks. J. Amer. Statist. Assoc. 76, 33–50.
Hanneke, S., Fu, W., Xing, E.P., et al., 2010. Discrete temporal models of social networks. Electron. J. Stat. 4, 585–605.
Hastie, T., Tibshirani, R., 1993. Varying-coefficient models. J. R. Stat. Soc. Ser. B Stat. Methodol. 757–796.
Hoff, P.D., Raftery, A.E., Handcock, M.S., 2002. Latent space approaches to social network analysis. J. Amer. Statist. Assoc. 97 (460), 1090–1098.
Hoff, P.D., et al., 2015. Multilinear tensor regression for longitudinal relational data. Ann. Appl. Stat. 9 (3), 1169–1193.
Holland, P., Leinhardt, S., 1981. An exponential family of probability distributions for directed graphs. J. Amer. Statist. Assoc. 76 (373), 33–50.
Hoover, D.R., Rice, J.A., Wu, C.O., Yang, L.-P., 1998. Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. Biometrika 85 (4), 809–822.
Huang, J.Z., Wu, C.O., Zhou, L., 2002. Varying-coefficient models and basis function approximations for the analysis of repeated measurements. Biometrika 89 (1), 111–128.
Hunter, D.R., Goodreau, S.M., Handcock, M.S., 2008b. Goodness of fit of social network models. J. Amer. Statist. Assoc. 103 (481), 248–258.
Hunter, D., Handcock, M., 2006. Inference in curved exponential family models for networks. J. Comput. Graph. Statist. 15 (3), 565–583.
Hunter, D.R., Handcock, M.S., Butts, C.T., Goodreau, S.M., Morris, M., 2008a. ergm: A package to fit, simulate and diagnose exponential-family models for networks. J. Stat. Softw. 24 (3), 1–29.
Kolar, M., Song, L., Ahmed, A., Xing, E.P., 2010. Estimating time-varying networks. Ann. Appl. Stat. 94–123.
Konidaris, G., Osentoski, S., Thomas, P., 2011. Value function approximation in reinforcement learning using the fourier basis. In: Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence. AAAI Press, pp. 380–385.
Krivitsky, P.N., 2012. Exponential-family random graph models for valued networks. Electron. J. Stat. 6, 1100.
Krivitsky, P.N., Handcock, M.S., 2014. A separable model for dynamic networks. J. R. Stat. Soc. Ser. B Stat. Methodol. 76 (1), 29–46.
Leifeld, P., Cranmer, S.J., Desmarais, B.A., 2018. Temporal exponential random graph models with btergm: Estimation and bootstrap confidence intervals. J. Stat. Softw. 83 (6).
McLachlan, G.J., 1987. On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. Appl. Stat. 318–324.
Moody, J., Mucha, P.J., 2013. Portrait of political party polarization. Netw. Sci. 1 (01), 119–121.
Qu, A., Li, R., 2006. Quadratic inference functions for varying-coefficient models with longitudinal data. Biometrics 62 (2), 379–391.
Ramsay, J.O., 2006. Functional Data Analysis. Wiley Online Library.
Rinaldo, A., Fienberg, S., Zhou, Y., 2009. On the geometry of discrete exponential families with application to exponential random graph models. Electron. J. Stat. 3, 446–484.
Robins, G., Pattison, P., Kalish, Y., Lusher, D., 2007. An introduction to exponential random graph (p^*) models for social networks. Social Networks 29 (2), 173–191.
Sarkar, P., Moore, A.W., 2005. Dynamic social network analysis using latent space models. ACM SIGKDD Explor. Newsl. 7 (2), 31–40.
Schmid, C.S., Desmarais, B.A., 2017. Exponential random graph models with big networks: Maximum pseudolikelihood estimation and the parametric bootstrap. In: 2017 IEEE International Conference on Big Data. Big Data, IEEE, pp. 116–121.
Schweinberger, M., 2011. Instability, sensitivity, and degeneracy of discrete exponential families. J. Amer. Statist. Assoc. 106 (496), 1361–1370.
Sewell, D.K., Chen, Y., 2015. Latent space models for dynamic networks. J. Amer. Statist. Assoc. 110 (512), 1646–1657.
Simpson, S.L., Hayasaka, S., Laurienti, P.J., 2011. Exponential random graph modeling for complex brain networks. PLoS One 6 (5), e20039.
Simpson, S.L., Moussa, M.N., Laurienti, P.J., 2012. An exponential random graph modeling approach to creating group-based representative whole-brain connectivity networks. Neuroimage 60 (2), 1117–1126.
Snijders, T.A., 2001. The statistical evaluation of social network dynamics. Sociol. Methodol. 31 (1), 361–395.
Snijders, T.A., Borgatti, S.P., 1999. Non-parametric standard errors and tests for network statistics. Connections 22 (2), 161–170.
Snijders, T., Pattison, P., Robins, G., Handcock, M., 2006. New specifications for exponential random graph models. Sociol. Methodol. 36 (1), 99–153.
Staicu, A.-M., Li, Y., Crainiceanu, C.M., Ruppert, D., 2014. Likelihood ratio tests for dependent data with applications to longitudinal and functional data analysis. Scand. J. Stat. 41 (4), 932–949.
Stillman, P.E., Wilson, J.D., Denny, M.J., Desmarais, B.A., Bhamidi, S., Cranmer, S.J., Lu, Z.-L., 2017. Statistical modeling of the default mode brain network reveals a segregated highway structure. Sci. Rep. 7 (1), 11694.
Stillman, P.E., Wilson, J.D., Denny, M.J., Desmarais, B.A., Cranmer, S.J., Lu, Z.-L., 2019. A consistent organizational structure across multiple functional subnetworks of the human brain. NeuroImage.
Strauss, D., Ikeda, M., 1990. Pseudolikelihood estimation for social networks. J. Amer. Statist. Assoc. 85 (409), 204–212.
Székely, E., Pappa, M., Wilson, J.D., Bhamidi, S., Jaddoe, V., Verhulst, F., Tiemeier, H., Shaw, P., 2016. Childhood peer network characteristics: Genetic influences and links with early mental health trajectories. J. Child Psychol. Psychiatry 57 (6), 687–694.
Tekle, F.B., Gudicha, D.W., Vermunt, J.K., 2016. Power analysis for the bootstrap likelihood ratio test for the number of classes in latent class models. Adv. Data Anal. Classif. 1–16.
Van Duijn, M.A., Gile, K.J., Handcock, M.S., 2009. A framework for the comparison of maximum pseudo-likelihood and maximum likelihood estimation of exponential family random graph models. Social Networks 31 (1), 52–62.
Van Essen, D.C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T., Bucholz, R., Chang, A., Chen, L., Corbetta, M., Curtiss, S.W., et al., 2012. The Human Connectome Project: a data acquisition perspective. Neuroimage 62 (4), 2222–2231.
Wasserman, S., Pattison, P., 1996. Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p^*. Psychometrika 61 (3), 401–425.
Wilson, J.D., Denny, M.J., Bhamidi, S., Cranmer, S.J., Desmarais, B.A., 2017. Stochastic weighted graphs: Flexible model specification and simulation. Social Networks 49, 37–47.
Wilson, J.D., Stevens, N.T., Woodall, W.H., 2019. Modeling and detecting change in temporal networks via the degree corrected stochastic block model. Qual. Reliab. Eng. Int. 35 (5), 1363–1378.
Woodall, W.H., Zhao, M.J., Paynabar, K., Sparks, R., Wilson, J.D., 2017. An overview and perspective on social network monitoring. IISE Trans. 49 (3), 354–365.
Wyatt, D., Choudhury, T., Bilmes, J., 2010. Discovering long range properties of social networks with multi-valued time-inhomogeneous models. In: Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, pp. 630–636.
Zhang, X., Wang, J.-L., et al., 2015. Varying-coefficient additive models for functional data. Biometrika 102 (1), 15–32.