

Varying-Coefficient Models and Functional Data Analyses for Dynamic Networks and Wearable Device Data

Jihui Lee

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2018

ABSTRACT

Varying-Coefficient Models and Functional Data Analyses for Dynamic Networks and Wearable Device Data

Jihui Lee

As more data are observed over time, investigating the variation across time has become a vital part of analyzing such data. In this dissertation, we discuss varying-coefficient models and functional data analysis methods for temporally heterogeneous data. More specifically, we examine two different types of temporal heterogeneity.

The first type of temporal heterogeneity stems from temporal evolution of relational pattern over time. Dynamic networks are commonly used when relational data are observed over time. Unlike static network analysis, dynamic network analysis emphasizes the importance of recognizing temporal evolution of relationship among observations. We propose and investigate a family of dynamic network models, known as varying-coefficient exponential random graph model (VCERGM), that characterize the evolution of network topology through smoothly varying parameters. The VCERGM directly provides an interpretable dynamic network model that enables the inference of temporal heterogeneity in dynamic networks.

Furthermore, we introduce a method that analyzes multilevel dynamic networks. If there exist multiple relational data observed at one time point, it is reasonable to additionally consider the variability among the repeated observations at each time point. The proposed method is an extension of stochastic blockmodels with a priori block membership and temporal random effects. It incorporates a variability among multiple relational structures at one time point and provides a richer representation of dependent engagement patterns at each time point. The method is also flexible in analyzing networks with time-varying networks. Its smooth parameters can be interpreted as evolving strength of engagement within and across blocks.

The second type of temporal heterogeneity is motivated by temporal shifts in continuously observed data. When multiple curves are obtained and there exists a common curvature shared by all the observed curves, understanding the common curvature may involve a preprocessing step of managing temporal shifts among curves. We explore the properties of continuous in-shoe sensor recordings to understand the source of variability in gait data. Our case study is based on measurements of three healthy subjects. The in-shoe sensor data we explore show both phase and amplitude variabilities; we separate these sources via curve registration. We examine the correlation of temporal shifts across sensors to evaluate the pattern of phase variability shared across sensors. We apply a series of functional data analysis approaches to the registered in-shoe sensor curves to examine their association with current gold-standard gait measurement, so called ground reaction force.

Contents

List of Figures	iv
List of Tables	viii
Acknowledgements	x
1 Introduction	1
1.1 Temporal Heterogeneity	1
1.1.1 Dynamic Networks	1
1.1.2 Wearable Device Data	4
1.2 Varying-Coefficient Models	6
1.3 Functional Data Analysis	7
2 Varying-Coefficient Exponential Random Graph Model	10
2.1 Introduction	10
2.2 Model	14
2.2.1 Exponential Random Graph Models	14
2.2.2 Varying-Coefficient Exponential Random Graph Models	17
2.2.3 Generalization to Higher Order Varying-Coefficient Exponential Random Graph Models	19
2.3 Estimation	20
2.3.1 Spline-Based Representation of Time-Varying Coefficients	20

2.3.2	Fast Estimation via Maximum Pseudo Likelihood	21
2.3.3	Penalized Logistic Regression	25
2.4	Testing for Heterogeneity	26
2.5	Simulation Study	29
2.5.1	Power Evaluation for Testing Heterogeneity	29
2.5.2	Estimation Performance	31
2.6	Applications	36
2.6.1	Political Voting Network	36
2.6.2	fMRI Dataset	38
2.7	Discussion	40
3	Multilevel Dynamic Stochastic Blockmodel	43
3.1	Introduction	43
3.1.1	Motivation	43
3.1.2	Literature Review	45
3.2	Method	46
3.2.1	Stochastic Blockmodel	47
3.2.2	Stochastic Blockmodel for Dynamic Networks	49
3.2.3	Multilevel Stochastic Blockmodel for Dynamic Networks	49
3.3	Estimation	51
3.3.1	Model Fitting	51
3.3.2	Computation	54
3.4	Application	55
3.5	Discussion	57
4	Curve Registration	59
4.1	Introduction	59
4.1.1	Motivation	59

4.1.2	Literature Review	63
4.2	Data	66
4.3	Methods	69
4.3.1	Curve Registration	69
4.3.2	Sources of Variability	70
4.3.3	Phase Variability Across Sensors	72
4.3.4	Association between In-Shoe Sensors and GRF	73
4.4	Results	74
4.4.1	Curve Registration	74
4.4.2	Sources of Variability	75
4.4.3	Phase Variability Across Sensors	76
4.4.4	Association between In-Shoe Sensors and GRF	78
4.5	Discussion	80
5	Conclusion	82
	Bibliography	87
	Appendix to Varying-Coefficient Exponential Random Graph Model	97
A.1	Stochastic Equivalence under the Difference Statistic Specification . .	97
A.2	Iterative Reweighted Least Squares (IRLS)	99
A.3	Additional Simulation Results	100
A.4	Estimation for Networks with Time-Varying Network Size	101
A.5	Estimation with Different Number of Basis Functions	101

List of Figures

1.1	U.S. Congress 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.	2
1.2	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 form an edge.	2
1.3	Gait measurements: Gait measurements observed from force plates placed on a treadmill (ground reaction force; left) and in-shoe sensor (heel sensor; right). Each panel shows 5 curves, corresponding to 5 stances of a healthy subject. Stances are color-coded.	5
2.1	Parameter estimates with 30 nodes: Estimated parameters for edges (top) and reciprocity (bottom). Black line is the true $\phi(t)$. 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.	34

2.2	Parameter estimates of political networks: Temporal heterogeneity is clearly presented for all three network statistics. Cross-sectional ERGMs (ERGM), ad hoc 2-step procedure (ERGM2) and VCERGM show similar estimates.	37
2.3	Parameter estimates of fMRI networks: Results of two randomly chosen individuals. For all three network statistics, one individual (first row; permutation test p -value = 0.117) displays slightly more fluctuations than the other individual (second row; permutation test p -value = 0.598). The ad hoc 2-step procedure and VCERGM show similar estimates.	39
3.1	Estimated probability of forming an edge between political parties from proposed model: Probability of forming an edge within Democrat (Red), within Republican (Blue), and across political parties (Black) in Congress 40 -113.	55
3.2	Estimated probability of forming an edge between political parties from cross-sectional SBM: Probability of forming an edge within Democrat (Red), within Republican (Blue), and across political parties (Black) in Congress 40 -113. The solid line represents the cross-sectional SBM estimates and the dashed line represents the cross-sectional SBM estimates after ad-hoc smoothing.	56
4.1	In-shoe sensor and GRF data: In-shoe sensor recordings and vertical GRF (VGRF) of a healthy subject's 5 stances. Data from in-shoe sensors at four locations are shown in the first four panels. The fifth panel shows the measurement of vertical GRF. Each curve represents a stance.	61

4.2	In-shoe sensor: (Left) NCPF sensor with wires embedded to measure the voltage response during impacts. (Right) Shoe equipped with NCPF foam sensors, microcontroller, and battery. The sensors were embedded under the insole at the heel, arch, ball, and toe.	67
4.3	Original measurements: Observed data for three runners are shown in rows; data from in-shoe sensors at various locations are shown in the first four columns and the direct measure of VGRF is shown in the fifth column. Curves are color-coded per subject and each curve represents a stance.	68
4.4	Curve after registration: Registered curves for three runners are shown in rows; curves from in-shoe sensors at various locations are shown in columns. Curves are color-coded per subject and each curve represents a stance.	74
4.5	Warping functions: Each panel represents warping function estimates of Subject 1 of four different in-shoe sensors. The clock time (x -axis) is the curve-specific time before registration and system time (y -axis) is the registered time which is common for every curve. Each curve represents a stance.	75
4.6	Functional principal components: First two FPCs of registered curves for three runners are shown in rows; the FPCs in four in-shoe sensors are shown in columns.	77
4.7	Function-on-function regression coefficient estimates: Heat map of function-on-function regression coefficient estimates with heel sensor as a predictor. VGRF is measured over t (y -axis) and heel sensor is measured over u (x -axis). Color gradation represents the magnitude of coefficient estimates.	79

4.8	Model fitting of function-on-function regression: (Top) Fitted values (Bottom) Residuals	80
A.1	Parameter estimates with 30 nodes: Estimated parameters for edges (top) and reciprocity (bottom). Black line is the true $\phi(t)$. The blue line indicates the average of 100 estimated curves and the shaded band illustrates the first and third quantiles.	102

List of Tables

2.1	Examples of network statistics in ERGM	15
2.2	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.	31
2.3	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.	35
2.4	Computing Time: Summary (Mean(SD)) of computing time (second) for dynamic networks with different number of time points K	35
4.1	Comparison of variability: Amplitude variability ($\times 10^5$) before and after registration.	76
4.2	Test statistic of pairwise comparison: For each subject, the upper diagonal elements are the mean correlation within stance across sensors.	77
A.1	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.	100

A.2	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.	101
A.3	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.	102

Acknowledgements

I would like to express my sincere gratitude to my advisors Dr. Jeff Goldsmith and Dr. Gen Li. I am so grateful to have two amazing advisors that have always encouraged me and supported me throughout my PhD program. They set a great example of a faculty, mentor, and researcher I hope to become one day.

Special thanks to my dissertation committee: Dr. Todd Ogden, Dr. Tian Zheng, and Dr. James D. Wilson. Their guidance and insightful feedback have greatly improved the quality of this dissertation. I feel absolutely privileged to have these supportive and inspiring mentors.

I cannot stress how lucky I feel to have awesome staffs in the Department of Biostatistics. They have helped me literally every step of my journey in the program. Also, I have so many good memories with friends and peers in the department and I will truly miss my time in the doctoral room.

Most importantly, I would like to thank my family for their endless love and support in every moment of my life. My parents, Jaewoo Lee and Youngmi Hong, have always believed in me and encouraged me to follow my dreams. My sister, Jiyeon Lee, has been always there for me as my best friend I thank her so much for that. I am also very grateful to William and Rosemary Diaz for supporting me in so many ways. I have been extremely fortunate to have them as my fabulous support team. My last yet biggest thanks to Anthony Diaz, who has been my greatest motivation and inspiration. May we continue to strive for our dreams and shine together.

Chapter 1

Introduction

1.1 Temporal Heterogeneity

The main focus of this dissertation is to understand the temporal heterogeneity in data observed over time. We begin with introducing two types of temporal heterogeneity we consider in this dissertation. The first type exists in dynamic networks, and the other type appears in observations from wearable device data.

1.1.1 Dynamic Networks

Networks are widely used to represent a relational structure of actors or individual units in a system. Nodes in a network are actors of the system, and two nodes form an edge if they have a relationship together. Aims of network analysis include visualizing the relational structure, detecting hidden communities within network, or directly modeling the system of interconnected nodes.

Dynamic networks are commonly used in applications where relational data are observed over time. When multiple networks are observed over time, it is of interest to examine the temporal dependencies between networks observed in time. For example, Figure 1.1 shows how the U.S. Congress co-voting networks fluctuate over time. In

each Congress, nodes represent Senators and an edge is formed between two nodes if they vote concurrently on at least 80% of the bills in the Congress. The nodes are color-coded by their political affiliation. Clusters of Republicans and Democrats are much more distinctive in Congress 113 compared to Congress 40 and Congress 70. The contribution of political parties in forming an edge has greatly changed in the course of 73 Congresses.

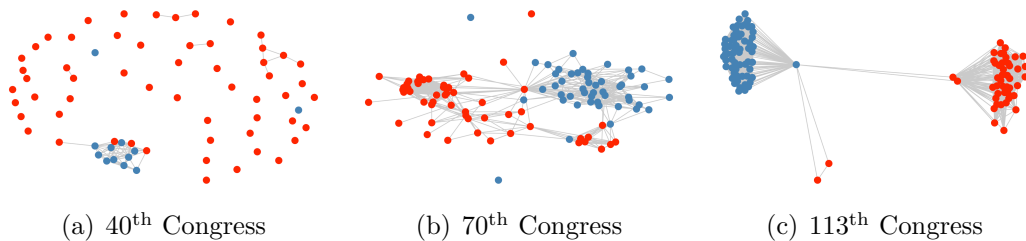


Figure 1.1: **U.S. Congress co-voting network:** Co-voting networks of U.S. Senators in Congress 40, 70, and 113. Red nodes represent Republican Senators and blue nodes represent Democratic Senators.

On the contrary, Figure 1.2 shows relatively stable structure of dynamic networks. The networks are defined based on brain activities obtained from the resting state functional magnetic resonance imaging (fMRI). Nodes in a network are brain regions and an edge between two nodes represents a high association between the two regions. Given that these images are obtained in a resting state, there is no major fluctuation over time except minor changes in local edges.

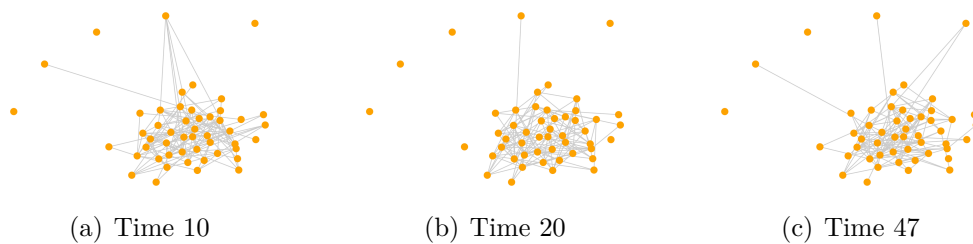


Figure 1.2: **Resting state fMRI network:** Resting state fMRI network at observed times 10, 20, and 47. Each node represents a brain region. The top 10% of partial correlation between regions form an edge.

These two examples above demonstrate how dynamic networks exhibit temporally fluctuating structure (either global or local). We refer to this feature as *temporal heterogeneity* in dynamic networks. Statistical models for dynamic network data should capture these temporal dependencies between networks observed across time as well as the structural dependencies among the nodes in each network.

In Chapter 2, we propose an exponential family of random graphs for dynamic networks that directly parametrizes the temporal heterogeneity in dynamic networks. The proposed method, called varying-coefficient exponential random graph model (VCERGM), characterizes the evolution of network topology using smoothly varying parameters. We establish how to fit the VCERGM through maximum pseudo-likelihood techniques, and provide a computationally tractable method for statistical inference of complex dynamic networks. We furthermore devise a bootstrap hypothesis testing framework for testing the temporal heterogeneity of an observed dynamic network sequence. We apply the VCERGM to the U.S. Congress co-voting networks (Figure 1.1) and a resting state fMRI networks (Figure 1.2) and show that our method provides relevant and interpretable patterns.

Chapter 3 proposes another method to model the temporal heterogeneity in dynamic networks. Specifically, it considers a case where there are repeatedly observed networks at one time point. Many studies have focused on creating a single network by taking a snapshot of relational structure or aggregating multiple observations or images in an interval of time. For example, the U.S. Congress co-voting networks in Figure 1.1 were created for each Congress by aggregating voting patterns observed from multiple bills. In this process, important dynamics across multiple bills within a Congress may be ignored. The proposed method in this chapter accommodates the repeatedly observed networks within a time frame and explains the variability among these multiple observations in a form of random effects. The proposed method is an extension of stochastic blockmodels, which assign the probability of connectivity

solely based on nodes' block membership. The overall trend of topological fluctuation is modeled as fixed effects via a varying-coefficient framework, while the random effects are added to incorporate the variability among repeatedly observed networks. In order to evaluate the performance of the proposed method, we analyze the U.S. Congress co-voting networks where a network is defined for each bill in a Congress.

1.1.2 Wearable Device Data

Recently, wearable technology has been widely implemented to record physical activities in observational or clinical studies. Records from a wearable device are collected very densely, resulting in a large number of observations in one time interval. The collection of observations measured in this time interval forms a curve and this curve can be viewed as a single unit of observation. The wearable device often cumulates the observations over a long period of time, and thus forms multiple curves. These curves in this context are also called functional data.

For example, Figure 1.3 shows gait measurements of a healthy subject. A curve represents a stance (defined as the period between the foot striking and lifting from the ground) and there are 5 curves in each panel. The curves in the left panel show ground reaction force (GRF), observed from force plates placed on a treadmill. The GRF is a force exerted by the ground on a body, and it provides valuable insights into biomechanics, locomotion, and the possible presence of pathology. The curves in the right panel are from in-shoe sensor (heel) corresponding to the 5 curves in the left panel. The in-shoe sensors recently have been pursued as a relatively inexpensive alternative to in-lab GRF measurement.

The curves in the left panel of Figure 1.3 show almost identical shape. It is expected because a healthy person is expected to show consistency in stances. However, data obtained by in-shoe sensors share a common curvature yet there exist unexpected drifting effects between curves. We refer to this time shifts as *temporal heterogeneity*

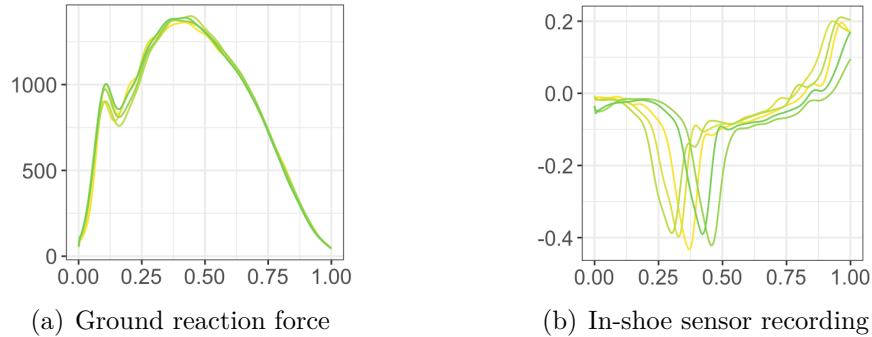


Figure 1.3: **Gait measurements:** Gait measurements observed from force plates placed on a treadmill (ground reaction force; left) and in-shoe sensor (heel sensor; right). Each panel shows 5 curves, corresponding to 5 stances of a healthy subject. Stances are color-coded.

in functional data.

Our goal to investigate the temporal heterogeneity in functional data mainly focuses on separating the source of variability between curves. We define two types of variabilities in functional data: amplitude variability and phase variability. For a case where there exist unexpected phase variability (i.e. temporal heterogeneity) between curves, removing the phase variability is an important step to appropriately analyze the amplitude variability. By realigning them, we remove the temporal heterogeneity in wearable device data and focus on analyzing the curvature shared by multiple curves.

As for the in-shoe sensor data presented in Figure 1.3, we conduct curve registration to remove the temporal heterogeneity. Essentially, we realign the originally observed curves to get rid of the unexpected time shifts. Using the realigned curves, we explore possible associations between in-shoe sensor recordings and GRF measurements to evaluate the in-shoe sensor recordings as a possible surrogate for in-lab GRF measurements. Functional data analysis methods are implemented to analyze the registered curves.

1.2 Varying-Coefficient Models

We identify the first type of temporal heterogeneity in temporally observed networks as time-varying dynamics in longitudinal data. Dynamic networks in Figure 1.1 and Figure 1.2 capture the evolution in relational structure over time. We view these networks as longitudinal data and analyze their temporal fluctuation using a varying-coefficient framework.

The varying-coefficient framework models the non-linear, dynamic pattern of covariates on a response variable (Hastie and Tibshirani 1993). It essentially is a linear additive model of predictors, but their coefficients are smoothly changing value of other predictors. Consider a response variable Y and two predictors X and R . In general, the varying-coefficient models have the form

$$Y = \beta_0 + X\beta_1(R). \quad (1.1)$$

The predictor R in model (1.1) is called an effect modifier, which governs the effects of X on Y ; depending on the value of R , the effect of X on Y varies. In a way, this model considers the interactive effects of X and R on Y .

In a longitudinal setting, data are collected over a period of time. The idea of dynamic pattern of covariate effects can be extended to analyze longitudinal data. The varying-coefficient models for longitudinal data parametrize the temporally fluctuating covariate effects on a response variable (Hoover et al. 1998; Zhang and Wang 2015). West, Harrison, and Migon 1985 referred to this model as dynamic generalized linear regression.

Let $Y(t)$ and $X(t)$ denote the response and predictor at time t , respectively. Define a smooth coefficient $\beta(t)$ that explains the effects of predictor on the response at time

t . We write the varying-coefficient model for longitudinal data as follows:

$$Y(t) = \beta_0(t) + \beta(t) X(t) + \varepsilon(t). \quad (1.2)$$

Multiple predictors can be considered in this framework, and a general idea of additive models is applied to include more than one predictor in model (1.2). The coefficient $\beta(t)$ flexibly models the change of effect parameters over time.

The varying-coefficient framework provides flexibility and interpretability in longitudinal data analysis. The multiplicative form $\beta(t) X(t)$ can be readily interpreted to explain the fluctuation in $Y(t)$. A detailed review of varying-coefficient models and their applications is provided in Fan and Zhang 2008.

We view the temporally observed networks as longitudinal data and analyze the temporal heterogeneity in these dynamic networks using the varying-coefficient framework. In Chapter 2 and Chapter 3, we apply the varying-coefficient framework to analyze the topological evolution of dynamic networks. The parameter in the proposed models is written as a smooth function of time, which is equivalent to $\beta(t)$ in model (1.2). We use the smoothness-based approaches to model the evolution of a network over time via varying-coefficient models. For estimating the coefficient in the varying-coefficient models, we employ basis splines (b-splines) (De Boor et al. 1978; Eilers and Marx 1996) to reduce the dimensionality of estimation.

1.3 Functional Data Analysis

While the first type of temporal heterogeneity we consider exists in longitudinal data, the second type of temporal heterogeneity is presented among repeatedly measured curves. The curves in Figure 1.3 are repeated samples of a smooth curve that contains full trajectories. We view these curves as functional data and apply functional data analysis approaches to manage the temporal heterogeneity among these curves.

Recently, more data have been collected in a complex, high-dimensional structure in various fields; e.g. wearable device data, electroencephalography, and screening images. Often, these data have curves or functions as unit of observations, and we refer to them as functional data. Key features of functional data include dense measurements, smooth curvatures, and repeated observations.

Suppose $Y_i(t)$ denote the i -th curve; a function of time $t \in [0, T]$ for some $T > 0$ and the sample data contain n curves. In practice, a curve consists of a collection of densely observed measurements. That is, we observe Y_{ij} , which denotes the j -th observation of the i -th curve observed at time t_j , $j = 1, \dots, m$.

Unlike classical statistical methods, where one observation consists of a single scalar value, functional data analysis (FDA) examines the trajectory (curvature) of smooth curves. Statistical methods such as regression models, clustering, or principal component analysis are extended to analyze functional data. Because of discrete observations with noise, important steps such as smoothing and interpolation are required when implementing those FDA methods. Main goals of FDA include visualizing the data to show characteristics of these curves, examining a predictive relationship, and exploring the variability among the curves. For a detailed introduction of FDA methods, see (Ramsay 2006).

In Chapter 4, we have gait measurements using in-shoe sensors. There exist multiple curves, representing multiple stances, and we view these curves as functional data. In recognizing the time shifts among curves, we explore the source of variabilities among these observed curves. Specifically, we conduct curve registration to separate the temporal heterogeneity from the observed curves. The estimated phase variability is further explored to understand the underlying pattern of temporal heterogeneity between in-shoe sensors. We fit a function-on-function regression to investigate the relationship between in-shoe sensor recordings and current gold-standard GRF measurements.

Let $x(t)$, $t \in \mathcal{T}$, be a predictor (in-shoe sensor recordings) and $y(t)$, $t \in \mathcal{T}$, denote a response (GRF measurements), respectively. The function-on-function regression we write the relationship between $x(t)$ and $y(t)$ as follows:

$$y(t) = \beta_0(t) + \int_{\mathcal{T}} x(u) \beta(u, t) du + \varepsilon(t). \quad (1.3)$$

The bivariate coefficient $\beta(u, t)$ in model (1.3) is smooth over both u and t and relates the predictor measured over u to the response measured over t . The bivariate smoothness of $\beta(u, t)$ allows the effect of predictor functions to vary over the domain of the response.

Chapter 2

Varying-Coefficient Exponential Random Graph Model

2.1 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 Figure 1.1 and resting state brain connectivity networks in Figure 1.2. 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 network. In parametric models, heterogeneity may result in significant changes in parameters that characterize the observed network. Consider the U.S. Senate co-voting network shown in Figure 1.1. 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 Figure 1.2 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 Sections 2.6.1 and 2.6.2.

In this chapter, 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 network 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 dynamic network models that have been investigated. We briefly describe these here. 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, Fu, and Xing 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. In fact, we show that in a wide range of situations the TERGM degenerates to a collection of independent and identically distributed ERGMs (see Appendix).

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 flexibly models the formation and dissolution of networks by separately parameterizing prevalence and duration of fluctuations. However, the STERGM is essentially a special case of first-order TERGM that it still cannot capture the temporal heterogeneity.

Another method for dynamic network modeling 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 parametrize the fluctuation in network topology. Sarkar and Moore 2005 and Sewell and Chen 2015 generalized the latent space model developed by Hoff, Raftery, and Handcock 2002 to dynamic networks. The dynamics of network structure is modeled through random effects in a latent space. It focuses on the transition between two time points and provides limited description on overall network.

Compared to time-invariant models described above, an alternative actor-based model was introduced in Hoff 2015, where dynamic networks are modeled using multi-linear 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. In the meantime, Kolar et al. 2010 emphasizes on 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 and Wang 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 *ibid.* 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.

2.2 Model

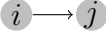
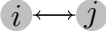
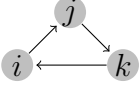
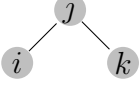
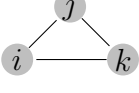
We begin by describing the exponential family of random graph models (ERGMs) and their temporal extension, the TERGM, since our proposed model is closely related to these specifications. We then introduce our proposed model: the VCERGM.

2.2.1 Exponential Random Graph Models

Suppose that $X \in \{0, 1\}^{n \times n}$ is 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. The ERGM is a probability distribution that characterizes the likelihood of X via a function of network statistics $\mathbf{h} : \{0, 1\}^{n \times n} \rightarrow \mathbb{R}^p$ that describe the topological structure of X . Table 2.1 provides examples of network statistics for a binary graph X with n nodes where x_{ij} represents whether there is an edge from node i to node j . For example, a reciprocity statistic quantifies the tendency of nodes in a directed network to exhibit mutual ties.

Given \mathbf{h} , the ERGM models X as a binary random matrix generated from the

Table 2.1: **Examples of network statistics in ERGM**

Type	Network Statistic	Definition
Directed	Edge density 	$\sum_{i \neq j} x_{ij}$
	Reciprocity 	$\sum_{i < j} x_{ij} x_{ji}$
	Cyclic triad 	$\sum_{i < j < k} x_{ij} x_{jk} x_{ki}$
Undirected	Two-star 	$\sum_{i < j < k} x_{ij} x_{jk}$
	Triangle 	$\sum_{i < j < k} x_{ij} x_{jk} x_{ki}$

following probability mass function

$$\mathbb{P}(X = x \mid \phi) = \frac{\exp\{\phi^T \mathbf{h}(x)\}}{\sum_{z \in \{0,1\}^{n \times n}} \exp\{\phi^T \mathbf{h}(z)\}}, \quad (2.1)$$

where $\phi \in \mathbb{R}^p$ 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 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, Kitts, and Morris 2009; Simpson, Hayasaka, and Laurienti 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; Robins et al. 2007.

The TERGM is an important and popular statistical model for inference of dynamic networks (Desmarais and Cranmer 2012; Hanneke, Fu, and Xing 2010), and can be described as follows. Consider a dynamic network $\mathbf{X} = \{X_1, X_2, \dots, X_T\}$ that is observed at T discrete and non-overlapping time periods, where each graph $X_t \in \{0, 1\}^{n \times n}$ from \mathbf{X} is unweighted, and observed for the set of vertices $[n] = \{1, \dots, n\}$. The TERGM is a generative model for \mathbf{X} that characterizes the conditional probability of X_t given $\mathbf{X}_t^- = \{X_s : s = 1, \dots, t-1\}$ 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.

$$\mathbb{P}(X_t = x_t \mid \mathbf{X}_t^- = \mathbf{x}_t^-) = \mathbb{P}(X_t = x_t \mid X_{t-1} = x_{t-1}) \quad (2.2)$$

Under (2.2), 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 $\mathbf{g} : \{0, 1\}^{n \times n} \times \{0, 1\}^{n \times n} \rightarrow \mathbb{R}^p$. 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 $t = 2, \dots, T$ as

$$\mathbb{P}(X_t = x_t \mid \mathbf{X}_t^-; \boldsymbol{\phi}) = \mathbb{P}(X_t = x_t \mid X_{t-1} = x_{t-1}; \boldsymbol{\phi}) = \frac{\exp\{\boldsymbol{\phi}^T \mathbf{g}(x_t, x_{t-1})\}}{\sum_{z \in \{0, 1\}^{n \times n}} \exp\{\boldsymbol{\phi}^T \mathbf{g}(z, x_{t-1})\}}, \quad (2.3)$$

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

$$\mathbb{P}(\mathbf{X} = \mathbf{x} \mid \boldsymbol{\phi}) = \mathbb{P}(X_1 = x_1 \mid \boldsymbol{\phi}) \prod_{t=2}^T \mathbb{P}(X_t = x_t \mid X_{t-1} = x_{t-1}, \boldsymbol{\phi}). \quad (2.4)$$

We note that in general if one is able to specify appropriate transition statistics, then the TERGM in (2.3) and (2.4) is readily generalized to higher-order Markov dependency; however, for discussion of our equivalence statement in Appendix, we consider the first-order TERGM only.

2.2.2 Varying-Coefficient Exponential Random Graph

Models

Let $\mathbf{X} = \{X_t : 0 \leq t \leq T\}$ be a stochastic sequence of temporally ordered networks observed continuously up to some time $T > 0$. At each time point t , $X_t \in \{0, 1\}^{n_t \times n_t}$ represents an unweighted, directed or undirected network with n_t nodes. 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 $\mathbf{h}(x_t, n_t) : \{0, 1\}^{n_t \times n_t} \rightarrow \mathbb{R}^p$ 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 $\boldsymbol{\phi}(t) = (\phi_1(t), \dots, \phi_p(t))^T \in \mathbb{R}^p$, the marginal likelihood of X_t at time t has an ERGM representation given by

$$\mathbb{P}(X_t = x_t \mid \boldsymbol{\phi}(t)) = \frac{\exp\{\boldsymbol{\phi}(t)^T \mathbf{h}(x_t, n_t)\}}{\sum_{z \in \{0, 1\}^{n_t \times n_t}} \exp\{\boldsymbol{\phi}(t)^T \mathbf{h}(z, n_t)\}}, \quad x_t \in \{0, 1\}^{n_t \times n_t}. \quad (2.5)$$

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.

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 $\phi(t)$.

The coefficients $\phi(t)$ in model (2.5) 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 (2.5). When a dynamic network evolves gradually over time, it is reasonable to believe the coefficients will also change gradually. In such a case, $\phi(t)$ 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 $\phi(t)$ 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 2.4, we introduce a formal hypothesis testing procedure to test the temporal heterogeneity of the coefficients.

2.2.3 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 (2.5) investigates the dynamics of coefficients for marginal network statistics. 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 (2.5) 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

$$\mathbb{P}(X_t = x_t | \phi_1(t), x_{t-1}) = \frac{\exp\{\phi_1^T(t) \mathbf{h}_1(x_t, x_{t-1})\}}{\sum_{z \in \{0,1\}^{n_t \times n_t}} \exp\{\phi_1(t)^T \mathbf{h}_1(z, x_{t-1})\}}, \quad x_t \in \{0, 1\}^{n_t \times n_t}. \quad (2.6)$$

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, Fu, and Xing 2010. In model (2.6), $\phi_1(t) = \{\phi_{1k}(t), k = 1, \dots, p\}$ 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 (2.6) 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 $\mathbf{h}_q(x_t, x_{t-1}, \dots, x_{t-q})$.

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 $\phi_q(t) \equiv \phi_q$. 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.

2.3 Estimation

2.3.1 Spline-Based Representation of Time-Varying Coefficients

Without any constraint, the collection of coefficients $\{\phi(t) : 0 \leq t \leq T\}$ contain an infinite number of parameters, making inference on (2.5) 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 (Konidakis, Osentoski, and Thomas 2011), and wavelets (Daubechies 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 $B_1(t), \dots, B_q(t)$, $0 \leq t \leq T$, 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 ℓ th basis function on $\phi_k(t)$. Let $\Phi =$

$\{\Phi_{k\ell}; k = 1, \dots, p, \ell = 1, \dots, q\}$ denote the $p \times q$ basis coefficient matrix and let $\mathbf{B}(t) = (B_1(t), \dots, B_q(t))^T$ be the length q vector of basis functions. We can represent the coefficients $\phi(t)$ as

$$\phi(t) = \Phi \mathbf{B}(t). \quad (2.7)$$

The set of q basis functions represents the smoothness of $\phi(t)$, and the coefficient matrix Φ determines the shape and trajectory of the fluctuations through time. Under the basis representation in (2.7), the distribution of \mathbf{X}_t in (2.5) is fully specified by the pq parameters in the coefficient matrix Φ .

2.3.2 Fast Estimation via Maximum Pseudo Likelihood

For an observed dynamic sequence of unweighted graphs $\mathbf{x} = \{x_s \in \{0, 1\}^{n_s \times n_s} : s = t_1, \dots, t_K, t_j < t_{j+1} \in [0, T]\}$, our goal is to estimate the coefficients $\{\phi(t) : 0 \leq t \leq T\}$ given the sequence \mathbf{x} . Let $\mathbf{B}_s = \{B_{s,\ell}; \ell = 1, \dots, q\}$ be a vector of length q of which elements are the basis functions evaluated at time s . By applying the basis representation in (2.7), we denote $\phi_s = \Phi \mathbf{B}_s$ as the smooth function $\phi(\cdot)$ evaluated at time s . Therefore, this estimation reduces to the task of estimating the $p \times q$ coefficient matrix Φ . A major obstacle in obtaining the maximum likelihood estimators of the parameters in Model (2.5), 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 Φ 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, Gile, and Handcock 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.

Without loss of generality, we assume that the numbers of nodes in different networks are the same over time, i.e., $n_s \equiv n$ for all s . In the case where n_s varies over time, one can simply use the standardized network statistics described in Section 2.2.2. For each observed time point $s = t_1, \dots, t_K$, 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 $s = t_1, \dots, t_K$, assume the conditional independence between edges. The marginal pseudo-likelihood function of Φ given x_s at time s is defined as

$$\text{PL}(\Phi|x_s) = \prod_{i,j \in [n]} \mathbb{P}(X_{ij}^s = x_{ij}^s | \mathbf{X}_{-(ij)}^s = \mathbf{x}_{-(ij)}^s). \quad (2.8)$$

Subsequently, the marginally independent composite pseudo likelihood of model (2.5) is

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

The MPLE $\hat{\Phi}$ is obtained by maximizing $\text{PL}(\Phi|\mathbf{x})$. 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 $\phi(t)$, the composite pseudo-likelihood function can be written as a product of marginal pseudo-likelihood functions at the observed time points t_1, \dots, t_k .

Let $x_{s,ij}^+$ denote the realization of x_s with x_{ij}^s set to 0 and let $x_{s,ij}^-$ be the realization of x_s with $x_{ij}^s = 0$. Define $\Delta_{ij}^s = \mathbf{h}(x_{s,ij}^+) - \mathbf{h}(x_{s,ij}^-)$ 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 $s = t_1, \dots, t_K$, the following relationship holds for all $i, j \in [n]$:

$$\begin{aligned} \text{logit} \left\{ \mathbb{P}(X_{ij}^s = 1 | \mathbf{X}_{-(ij)}^s = \mathbf{x}_{-(ij)}^s) \right\} &= \log \left\{ \frac{\mathbb{P}(X_{ij}^s = 1 | \mathbf{X}_{-(ij)}^s = \mathbf{x}_{-(ij)}^s)}{\mathbb{P}(X_{ij}^s = 0 | \mathbf{X}_{-(ij)}^s = \mathbf{x}_{-(ij)}^s)} \right\} \\ &= \log \left[\exp \left\{ \boldsymbol{\phi}(t)^T (\mathbf{h}(x_{s,ij}^+) - \mathbf{h}(x_{s,ij}^-)) \right\} \right] \\ &= \boldsymbol{\phi}_s^T \boldsymbol{\Delta}_{ij}^s \end{aligned} \quad (2.9)$$

Let $Y_{ij}^s = \text{logit} \left\{ \mathbb{P}(X_{ij}^s = 1 | \mathbf{X}_{-(ij)}^s = \mathbf{x}_{-(ij)}^s) \right\}$ and let $\mathbf{Y}_s = (Y_{11}^s, Y_{12}^s, \dots, Y_{nn}^s)^T$. Similarly, define $\boldsymbol{\Delta}_s = (\boldsymbol{\Delta}_{11}^s, \boldsymbol{\Delta}_{12}^s, \dots, \boldsymbol{\Delta}_{nn}^s)$ as the $p \times \binom{n}{2}$ matrix whose r th row contains the change in the r th network statistic when each edge changes from 0 to 1. Let $\text{vec}(X)$ be the operator that stacks the columns of X into a column vector, and let \otimes represent the Kronecker product operator. Combining (2.7) and (2.9) yields

$$\mathbf{Y}_s = \boldsymbol{\Delta}_s^T \boldsymbol{\Phi} \mathbf{B}_s = (\mathbf{B}_s \otimes \boldsymbol{\Delta}_s)^T \text{vec}(\boldsymbol{\Phi}), \quad s = t_1, \dots, t_K. \quad (2.10)$$

Let $\mathbf{Y} = (\mathbf{Y}_{t_1}, \dots, \mathbf{Y}_{t_K})^T$, 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 (2.10) connects to a logistic regression where \mathbf{H} represents a design matrix with its coefficient $\text{vec}(\boldsymbol{\Phi})$. In Strauss and Ikeda 1990, it was shown that maximizing the pseudo-likelihood $\text{PL}(\boldsymbol{\Phi} | x_s)$ in (2.8) is equivalent to finding the maximum likelihood estimator (MLE) of $\boldsymbol{\Phi}$ in the logistic regression model given in (2.9) with independent entries X_{ij}^s . Dependency among nodes in a network can be indirectly modeled by conditioning on the rest of 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 $\text{PL}(\Phi|\mathbf{x})$ is equivalent to calculating the MLE of Φ in the logistic regression model $\mathbf{Y} = \mathbf{H}^T \text{vec}(\Phi)$ treating $\{X_{ij}^s : i, j \in [n], s = t_1, \dots, t_K\}$ 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 be 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 (2.5). For a collection of basis functions $B_1(t), \dots, B_q(t)$, $0 \leq t \leq T$, 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 $\phi_{1k}(t)$ as a $\phi_{1k}(t) = \Phi_1 \mathbf{B}(t)$, with a $p \times q$ matrix of basis coefficients Φ_1 . The composite pseudo-likelihood functions for VCERGM(1) can be expressed as

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

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 (2.8). Let $\Delta_{ij}^s = h(x_{s,ij}^+, x_{s-1}) - h(x_{s,ij}^-, x_{s-1})$ denote the vector of differences in the transition statistics when $x_{s,ij}$ changes from 0 to 1. Then the relationship described in (2.9) can be

similarly applied to VCERGM(1). A logistic regression model like (2.10) is used for parameter estimation and thus the remainder of the estimation steps described above remain the same.

2.3.3 Penalized Logistic Regression

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

$$\mathcal{P}(\Phi_{(k)}) = \int \{D^2\phi_k(u)\}^2 du = \Phi_{(k)}^T \Omega \Phi_{(k)}$$

where a smoothness matrix Ω in this case can be specified as

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

For networks observed at discrete time points t_1, \dots, t_K , the (i, j) th element of Ω is

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

For more examples of possible penalties, see the Chapter 5 in Ramsay 2006. As the same collection of basis functions are used to express $\phi_k(t)$, $k = 1, \dots, p$, via basis representation, we impose the same Ω on all $\phi_k(t)$. Consequently, we add the penalty term

$$\mathcal{P}_\Omega(\Phi) = \sum_{k=1}^p \Phi_{(k)}^T \Omega \Phi_{(k)} = \text{vec}(\Phi)^T (\Omega \otimes \mathbf{I}_p) \text{vec}(\Phi)$$

to the logistic log pseudo likelihood function. Similar with \mathbf{Y} , let $\tilde{\mathbf{x}}$ denote a vec-

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

$$\tilde{\mathbf{x}}^T \mathbf{H}^T \text{vec}(\Phi) - \mathbf{1}^T \log[1 + \exp\{\mathbf{H}^T \text{vec}(\Phi)\}] - \lambda \mathcal{P}_\Omega(\Phi). \quad (2.11)$$

To fit (2.11), we implement an iteratively reweighted least squares (IRLS) algorithm. A detailed description of this procedure is available in Appendix.

2.4 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 a likelihood ratio test (LRT), which we will now describe.

We begin with the preconceived notion that \mathbf{x} is homogeneous, namely that the coefficients $\phi(t)$ under model (2.5) are fixed as constants over time. This serves as the null model, under which the VCERGM is equivalent to fitting independent and identically distributed ERGMs. Let $\phi_1^0, \dots, \phi_p^0$ be the estimates under the time-invariant model fitting. With fixed constants $\phi_1^0, \dots, \phi_p^0$, the null hypothesis corresponding to a homogeneous sequence of graphs can be written as

$$H_0 : \phi_1(t) = \phi_1^0, \dots, \phi_p(t) = \phi_p^0. \quad (2.12)$$

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 $\ell = 1, \dots, q$. In other words, the null hypothesis in (2.12) can be expressed more succinctly as

$$H_0 : \Phi = \Phi^0 = (\phi_1^0, \dots, \phi_p^0)^T \times \mathbf{1}_q^T,$$

where $\mathbf{1}_q$ is length q vector of 1's. Such simplification is applicable for spline basis functions and we have implemented 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, Fan, and Li 2000; Fan, Zhang, and Zhang 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 2.3, the joint pseudo likelihood consists of the distribution of X_{ij}^s given the rest of the data $\mathbf{X}_{-(ij)}^s$ for all $i, j \in [n], s = t_1, \dots, t_K$. Furthermore, maximizing the pseudo likelihood simplifies the estimation process as fitting a logistic regression. Namely, with observed networks $\mathbf{x} = \{x_s : s = t_1, \dots, t_K\}$ with n nodes, the pLRT compares the pseudo log likelihood function below under the null and alternative hypotheses:

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

Let $\hat{\Phi}_{H_0}$ and $\hat{\Phi}_{H_1}$ be the estimates of Φ under the null and alternative hypotheses.

The estimate $\hat{\Phi}_{H1}$ can be calculated by fitting the VCERGM specified in (2.5) and $\hat{\Phi}_{H0}$ is the estimate from the VERGM with a restriction of constant basis coefficients. Accordingly, let $\log \text{PL}(\hat{\Phi}_{H0}|\mathbf{x})$ and $\log \text{PL}(\hat{\Phi}_{H1}|\mathbf{x})$ denote the pseudo log likelihood functions under the null and alternative, respectively. Then, the test statistic is

$$\begin{aligned} T &= 2\{\log \text{PL}(\hat{\Phi}_{H1}|\mathbf{x}) - \log \text{PL}(\hat{\Phi}_{H0}|\mathbf{x})\} \\ &= 2 \sum_{s=t_1}^{t_K} \sum_{i,j \in [n]} \left[x_{ij}^s B_s^T (\hat{\Phi}_{H1} - \hat{\Phi}_{H0})^T \Delta_{ij}^s + \log \left\{ \frac{1 + \exp(B_s^T \hat{\Phi}_{H0}^T \Delta_{ij}^s)}{1 + \exp(B_s^T \hat{\Phi}_{H1}^T \Delta_{ij}^s)} \right\} \right]. \end{aligned} \quad (2.13)$$

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, Fan, and Li 2000; Fan and Zhang 2008; Huang, Wu, and Zhou 2002). It is preferable for moderate network size. Analogous to the work in De Brabanter et al. 2006; McLachlan 1987; Tekle, Gudicha, and Vermunt 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 (2.13) calculated based on B bootstrap samples successfully represent the null distribution of T .

1. Create B bootstrap samples. For each bootstrap, indexed by $b = 1, \dots, B$, $\mathbf{x}^{*(b)} = \{x_s^{*(b)} : s = t_1, \dots, t_K\}$ is a sample from $\mathbb{P}(\mathbf{X}|\hat{\Phi}_{H0})$.
2. For each bootstrap sample $\mathbf{x}^{*(b)}$, estimate Φ under the null and alternative hypotheses and denote them as $\hat{\Phi}_{H0}^{*(b)}$ and $\hat{\Phi}_{H1}^{*(b)}$, respectively.
3. Calculate the test statistic for each bootstrap sample as

$$T^{*(b)} = 2\{\log \text{PL}(\hat{\Phi}_{H1}^{*(b)}|\mathbf{x}^{*(b)}) - \log \text{PL}(\hat{\Phi}_{H0}^{*(b)}|\mathbf{x}^{*(b)})\}, \quad b = 1, \dots, B.$$

4. The critical value C_α is determined as the $(1-\alpha)$ th quantile of $(T^{*(1)}, \dots, T^{*(B)})$.

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

$$p\text{-value} = \frac{\sum_{b=1}^B I(T < T^{*(b)})}{B}.$$

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 (2.12) 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.

2.5 Simulation Study

The goal of our simulation study is two-fold: (i) to evaluate the power of the hypothesis testing procedure described in Section 2.4, and (ii) to assess the goodness of fit of the VCERGM on dynamic networks with various magnitudes of temporal heterogeneity. In Section 2.5.1, we evaluate the sensitivity of the hypothesis test in (2.12) for detecting temporal heterogeneity in a sequence of networks with fluctuating parameters using both the bootstrap and permutation procedures. Section 2.5.2 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.

2.5.1 Power Evaluation for Testing Heterogeneity

We first investigate the power of the hypothesis test for heterogeneity that we introduce in Section 2.4. 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 $\mathbf{x} = \{\mathbf{x}_1, \dots, \mathbf{x}_{100}\}$, where each sequence $\mathbf{x}_w = \{x_{w,1}, \dots, x_{w,K}\}$, contains K networks with 30 nodes observed at equally-spaced times t_1, \dots, 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 cases that we reject the null hypothesis at a $\alpha = 0.05$ level out of the 100 simulated dynamic network sequences. Table 2.2 reports these proportions when using the bootstrap procedure as well as permutation test. For 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 a 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 2.2, 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

use of 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).

Table 2.2: **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

2.5.2 Estimation Performance

We now evaluate the performance of VCERGM to accurately estimate fluctuating parameters $\phi(t)$, $t \in [0, T]$. We consider four different settings for $\phi(t)$: (i) sinusoidal curve $\phi(t) = a \sin\{(t + b)/c\} + d$ of varying amplitude a ; (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 (see Table 2.1). We simulate 100 dynamic sequences of directed graphs $\{\mathbf{x}_1, \dots, \mathbf{x}_{100}\}$ where each sequence $\mathbf{x}_w = \{x_{w,1}, \dots, x_{w,50}\}$ 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.

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 (2.1) 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)

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

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, Heath, and Wahba 1979).

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, \dots, t_K , namely

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

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 2.3 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 A.1 and A.2 in Appendix.

Without missing network, Figure 2.1 shows that cross-sectional ERGMs are more likely to introduce unexpected spikes or increased variability in estimating true $\phi(t)$, compared to VCERGM. Overall, the VCERGM presents smaller deviation from true $\phi(t)$ with smaller variability compared to cross-sectional ERGMs and ad hoc 2-step procedure. In the first three settings, the VCERGM performs better than cross-sectional ERGMs and ad hoc 2-step procedure with smaller IAE. In case of non-smooth functions, with true $\phi(t)$ not smooth but wiggly, the ad hoc 2-step procedure shows better performance than the VCERGM with respect to IAE. Both Table 2.3 and Figure 2.1 indicate that the VCERGM potentially misses random wiggles, which causes greater bias on average compared to cross-sectional ERGMs. Regardless, overall trend that the true non-smooth $\phi(t)$ presents is fairly well estimated by the VCERGM and the variability of the estimates is still smaller for the VCERGM compared to cross-sectional ERGMs. Overall, the performance of ad hoc 2-step procedure and VCERGM is comparable, but the VCERGM is more principled in terms of incorporating time-varying coefficients in the modeling step. For all four settings, 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 Table 2.4.

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

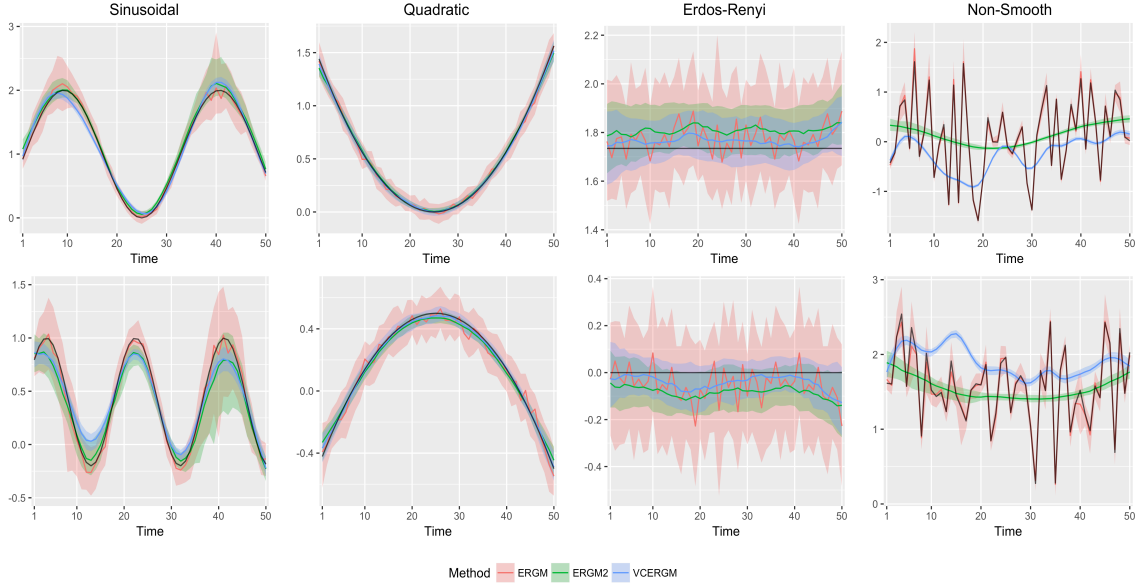


Figure 2.1: **Parameter estimates with 30 nodes:** Estimated parameters for edges (top) and reciprocity (bottom). Black line is the true $\phi(t)$. Red (ERGM) is for cross-sectional ERGMs, green (ERGM2) is for ad hoc 2-step procedure, and blue (VCERGM) is for VCERGM. For each method, solid line indicates the average of 100 estimated curves and the shaded band illustrates the first and third quantiles.

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 2.3, 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 record the computing time for VCERGM and cross-sectional ERGMs. Table 2.4 summarizes the computing times of 500 simulated dynamic network sequences and displays how computing time changes as the number of time points K changes. The maximum pseudo-likelihood approach is used for both ERGM and VCERGM estimation. According to Table 2.4, the VCERGM takes significantly less time than cross-sectional ERGMs to complete the parameter estimation. Even if both VCERGM and cross-sectional ERGMs show linear increase in computing time,

Table 2.3: **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)

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 2.4: **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	0.79 (0.07)	1.51 (0.08)	2.11 (0.10)	2.99 (0.07)	3.74 (0.11)
VCERGM	0.36 (0.04)	0.64 (0.04)	0.87 (0.06)	1.25 (0.11)	1.54 (0.06)

The VCERGM can analyze networks with time-varying network size. We conduct a simulation to evaluate the performance of VCERGM in such scenario. The results are presented in Appendix.

2.6 Applications

In this section we apply the VCERGM to two case studies which portray differing amounts of temporal heterogeneity. First, we analyze how the co-voting patterns among U.S. Senators have changed through time. In this example, we analyze the effects of political affiliation (Republican or Democrat) on the likelihood of the voting networks. In our second example, we investigate the structural changes of the resting-state functional magnetic resonance imaging (rfMRI) records of healthy individuals. For each case study, we first test for temporal heterogeneity of any statistic included in the model. We fit the VCERGM on the two data sets and also fit cross-sectional ERGMs and ad hoc 2-step procedure for comparison. In the first example, we witness a clear evidence of temporal heterogeneity that the importance of political affiliation on the likelihood of the voting networks significantly changes over time. On the contrary, relatively stable rfMRI networks show little fluctuations through time and our method is shown to accommodate it successfully as well.

2.6.1 Political Voting Network

We begin by analyzing the dynamic network that describes the co-voting patterns among U.S. Senators from 1867 (Congress 40) to 2015 (Congress 113). Three of the voting networks are shown in Figure 1.1. This network was first investigated in Moody and Mucha 2013 and has been subsequently analyzed in Wilson, Stevens, and Woodall 2016. The network is based off of 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 concurrently (both yay or both nay) on at least 80% of the bills to which they were both present.

As shown in Figure 1.1, there exist great fluctuations in network structure through time. Previous analyses in Moody and Mucha 2013; Wilson, Stevens, and Woodall 2016 have identified significant changes in the community structure of the network over time, and that this community structure is closely associated with political affiliation of the Senators. To account for these fluctuations, we incorporate a node-match statistic for political affiliation which counts the number of edges shared between Senators who have the same political affiliation. Furthermore, we include a statistic which models popularity (two-star) and clustering (triangle) of the Senator networks. See Table 1 for more details of these two statistics. We note that it is of separate interest to perform model selection for the VCERGM; however, we pursue this in future work. Here, we compare and contrast the results of a simple model when using several viable dynamic network models.



Figure 2.2: **Parameter estimates of political networks:** Temporal heterogeneity is clearly presented for all three network statistics. Cross-sectional ERGMs (ERGM), ad hoc 2-step procedure (ERGM2) and VCERGM show similar estimates.

The estimated parameters from i) cross-sectional ERGMs (ERGM), ii) ad hoc 2-step procedure (ERGM2) and iii) VCERGM are presented in Figure 2.2. Computing time for cross-sectional ERGMs and VCERGM are about 6 and 20 seconds, respectively. Notably, all three network statistics exhibit temporal heterogeneity. The permutation test p -value for testing heterogeneity is $< .001$. Consistent with our 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. The political nodematch parameter estimate reveals an important trend in the political network. We see that the coefficient value for this term generally increases over time, indicating the increasing importance of political affiliation on the co-voting habits of the Senators throughout U.S. history. More specifically, such increasing pattern in the coefficient reflects that the number of ties formed with the same political affiliation positively influences the likelihood. In particular since Congress 95, this coefficient has significantly increased. This finding matches the current theory of “political polarization” described in Moody and Mucha 2013. Figure 2.2 also suggests that the triangle coefficient has decreased in magnitude suggesting that clustering has become less and less important over time. In other words, the clustering has become not as influential as it was in the past in terms of forming ties.

2.6.2 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 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.

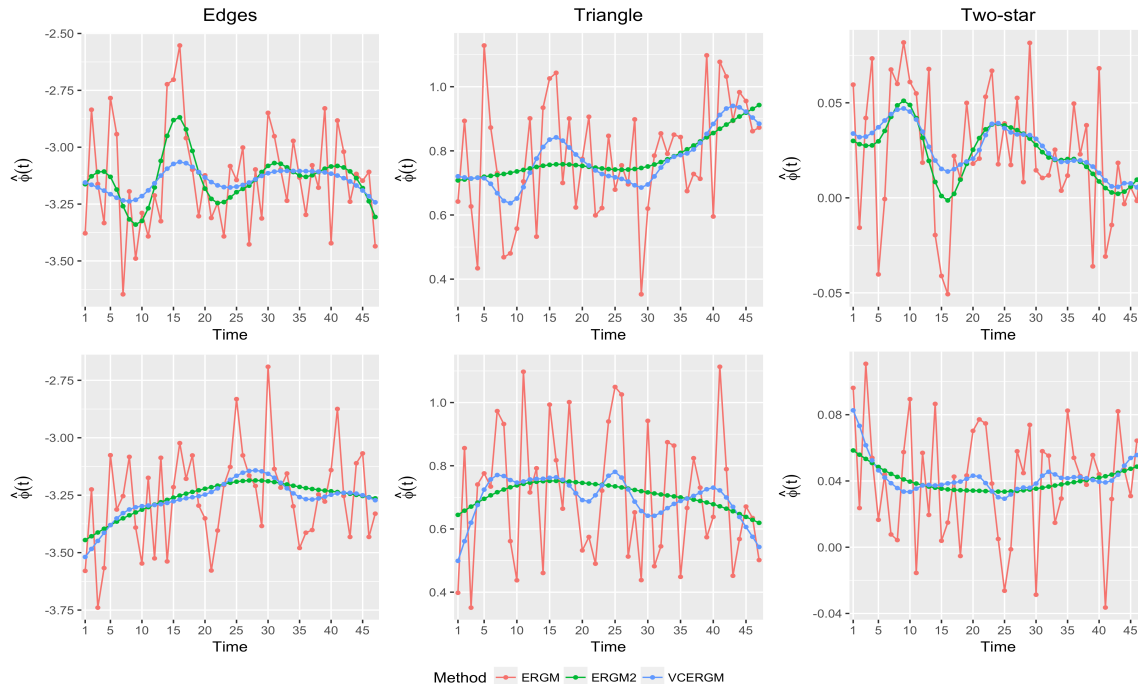


Figure 2.3: **Parameter estimates of fMRI networks:** Results of two randomly chosen individuals. For all three network statistics, one individual (first row; permutation test p -value = 0.117) displays slightly more fluctuations than the other individual (second row; permutation test p -value = 0.598). The ad hoc 2-step procedure and VCERGM show similar estimates.

Simpson, Hayasaka, and Laurienti 2011; Simpson, Moussa, and Laurienti 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 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.

Figure 2.3 shows the results of two individuals from this study. Computing time for cross-sectional ERGMs and VCERGM are about 1 second. 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.

2.7 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. 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, Goodreau, and Handcock 2008). 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.

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 MCMCMLE approach. There have been several studies exploring the relationship between the pseudo-likelihood and the likelihood of the ERGM, including Strauss and Ikeda 1990 and Desmarais and Cranmer 2012, where the efficacy of MPLE was empirically compared to MCMCMLE. More re-

cently, Schmid and Desmarais 2017 compared the performance of MCMCMLE 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 (Bindu and Thilagam 2016; Woodall et al. 2017). 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 identification of a change in network parameter - however, in future research we plan to pursue this idea further.

Chapter 3

Multilevel Dynamic Stochastic Blockmodel

3.1 Introduction

3.1.1 Motivation

Network topology changes over time by nature. Both local and global features of a network significantly changes through time. As relational data are collected over time, analyzing dynamic networks has become fundamental in demonstrating how the form of connection and relationship changes over time (Goldenberg et al. 2010). More studies emphasize the topological evolution of dynamic networks in various fields: examples include email communication among the employees of ENRON (Shetty and Adibi 2005), internet topology (Edwards et al. 2012) and biological systems of yeast response under oxidative stress (Gopalacharyulu et al. 2009). Temporal heterogeneity present in these examples can result in a dramatic change of the underlying process that best describes a network.

For example, the voting pattern of the U.S. Congress has been widely explored to understand the partisanship between Republican and Democrat. Andris et al. 2015

visualized and cross-sectionally examined the polarization between Democrat and Republican party members in the US House of Representatives from 1949 to 2012. Wilson, Stevens, and Woodall 2016 detected change points in relational structure and interpreted the topological changes to reflect the cohesion or polarization in US voting networks from Congress 40 to Congress 113. Lee, Li, and Wilson 2017 estimated a smooth parameter that demonstrates the overall evolution in topological structure and illustrated the increasing importance of political affiliation for constructing relational pattern among US senators.

When temporally observed relational data are used for dynamic network analysis, they are often processed to form networks. The real-time tracking of relational data has become more accessible, and many studies preprocess the temporal data to define dynamic networks. A common approach for preprocessing is to aggregate relational patterns observed in a time interval. For instance, many studies that analyzed the U.S. voting networks including the forementioned papers created a single network for each Congress that represents the overall relational structure for that Congress. The network is an aggregated form of voting patterns from multiple bills addressed in the certain Congress. It may result in a great loss of information; for example, there may be a variability in voting patterns across bills, and aggregating multiple bills ignores this variability and provides a limited picture of what happened in the Congress.

In this project, we propose a method for dynamic networks that does not aggregate the multiple relational patterns at each time point. Instead, the proposed method views them as repeated measurements and models them as random effects at each time point. We extend the idea of stochastic blockmodels for dynamic networks and parametrize the probability of engagement within and across blocks as a smooth function of time. We focus on the case in which the block membership is known to answer the effects of certain covariates, e.g. political affiliation for U.S. co-voting networks, in building an interaction across the different categories of the covariates.

3.1.2 Literature Review

The stochastic blockmodel (SBM) is a generative model for investigating community structure in random graphs; it identifies a subgroup structure of nodes (actors) in a network and assigns a probability of within-block and across-block ties (Holland, Laskey, and Leinhardt 1983; Wang and Wong 1987). Nodes in the same block are stochastically equivalent in that the likelihood of any given pattern of ties with a certain node in one block is the same as any other nodes in the same block. The block membership is often known, but in general, it has to be estimated along with the magnitude of connectivity across blocks.

The SBM has been explored in a variety of directions. For example, the mixed membership stochastic block model was proposed to allow an overlapping membership in multiple communities (Airoldi et al. 2008). More recently, a framework of SBM that incorporates edge- or node- level covariates was introduced to parametrize the covariate effects on forming ties between nodes (Leger 2016; Sweet 2015). The R package ‘`blockmodels`’ implemented an estimation process of SBM parameters using a variational expectation-maximization (EM) algorithm (Leger 2016; Mariadassou, Robin, and Vacher 2010).

When analyzing dynamic networks, it is of interest to examine the structural changes over time. A wide range of literature considers to detect change-points (anomalies) of communities of networks (Bhamidi, Jin, and Nobel 2015; Chen, Hendrix, and Samatova 2012; Greene, Doyle, and Cunningham 2010; Peel and Clauset 2015; Porter and White 2012; Sparks and Wilson 2016; Wilson, Stevens, and Woodall 2016). Types of anomaly are defined based on change in size and shape of communities through time. For example, Bhamidi, Jin, and Nobel 2015 formulates a non-parametric estimator for the change point based on observations of the network. Yang et al. 2011 introduces a model that captures the evolution of communities by explicitly modeling the transition of community membership for individual nodes in the

network. Meanwhile, Peel and Clauset 2014 employs the generalized hierarchical random graph model for distributional specification and detects network change-points by conducting Bayesian hypothesis testing. A hypothesis test for change-point detection compares two models, one with change and the other without change, in order to validate the existence of a change-point at the specific time point.

Considerable work has been done on the identification of structural changes in time-varying networks by directly modeling the temporal changes in dynamic networks (Pensky and Zhang 2017; Xu and Hero III 2013; Xu and Hero 2014). More specifically, the idea of state-space model (SSM) has been widely applied to model and detect the change-points in time-evolving networks. The SSM considers the linear transition between states of adjacent time points. For example, Matias and Miele 2017; Xu and Hero 2014 propose dynamic stochastic blockmodels with varying community memberships and connectivity parameters. Fu, Song, and Xing 2009; Ho, Song, and Xing 2011; Xing, Fu, and Song 2010 propose to model time-evolving networks with mixed membership stochastic blockmodel. Wilson, Stevens, and Woodall 2016 proposes the dynamic version of the degree corrected stochastic blockmodel.

The remainder of this paper is organized as follows. In Section 3.2, we introduce our model that incorporates the random effects at each time point in modeling time-varying group behavior via dynamic extension of stochastic blockmodels. We introduce the estimation process in detail in Section 3.3. We apply our proposed method to U.S. co-voting networks in Section 3.4. We conclude with discussion of our method and open areas for future research in Section 4.5.

3.2 Method

We start with describing a general idea of stochastic blockmodels for a static network in Section 3.2.1 and discuss the dynamic extension of stochastic blockmodel

for temporally observed networks in Section 3.2.2. In Section 3.2.3, our proposed model demonstrates the temporal extension of stochastic blockmodels with varying-coefficient approaches and random effects for multilevel dynamic networks.

3.2.1 Stochastic Blockmodel

The stochastic blockmodel (SBM) is a generative model to identify the community structure and quantify the connectivity between communities in a unweighted, symmetric network (Holland, Laskey, and Leinhardt 1983). Suppose $Y \in \{0, 1\}^{n \times n}$ is a undirected network with n nodes. Its (i, j) -th element, denoted by Y_{ij} , indicates an edge between node i and node j , $i, j = 1, \dots, n$: $Y_{ij} = 1$ if there is an edge, 0 if not. The edges in the network are assumed to be independent.

The SBM defines the stochastically equivalent likelihood of a relational structure solely based on the block membership of any pair of nodes in the network. Let K be the number of blocks and define a vector of length K , denoted by z_i , that indicates a block membership of node i . If node i belongs to the k -th block, $k = 1, \dots, K$, the k -th element of vector z_i is 1 and $\|z_i\| = 1$, $i = 1, \dots, K$. Let P be a symmetric $K \times K$ matrix that represents a probability of forming an edge across blocks. The (r, s) -th element of matrix P indicates the strength of connection between block r and block s , $r, s = 1, \dots, K$. Suppose that node i belongs to the k -th block and node j belongs to the k' -th block. The SBM specifies the probability of forming an edge between node i and node j as $P(Y_{ij} = 1) = P_{k,k'}$. Often, we parametrize the strength of connectivity across blocks via logistic regression model. That is, we write $\phi_{k,k'}$ as the logit of probability $P_{k,k'}$ and write $\text{logit}\{P(Y_{ij} = 1|\Phi)\} = \phi_{k,k'}$. In other words, similar with the probability matrix P , we define a $K \times K$ matrix $\Phi = \{\phi_{r,s} : r, s = 1, \dots, K\}$ whose elements are the logit of probabilities in matrix P of forming an edge across blocks. The (r, s) -th element of Φ , denoted by $\phi_{r,s}$, indicates the logit of connectivity between block r and s .

$$\Phi = \begin{pmatrix} \phi_{1,1} & \dots & \phi_{1,s} & \dots & \phi_{1,K} \\ \vdots & & \vdots & & \vdots \\ \phi_{r,1} & \dots & \phi_{r,s} & \dots & \phi_{r,K} \\ \vdots & & \vdots & & \vdots \\ \phi_{K,1} & \dots & \phi_{K,s} & \dots & \phi_{K,K} \end{pmatrix} \quad (3.1)$$

By using the matrix Φ in (3.1) and vectors of block membership for node i and node j , denoted by z_i and z_j , respectively, the SBM models the probability of forming an edge between the two nodes as follows:

$$\text{logit}\{P(Y_{ij} = 1|\Phi)\} = z_i' \Phi z_j \quad (3.2)$$

The model parameters Φ govern the edge probabilities in the network. More specifically, the nodes in the same block are stochastically equivalent in that the likelihood of any given pattern of ties with a certain node in one block is the same as any other nodes in the same block. For example, any edge $Y_{i',j'}$ is identically distributed with $Y_{i,j}$ if node i and node i' belong to the same block and node j and node j' belong to the same block, respectively. The number of blocks K is often pre-specified due to a priori block membership, but in general, it has to be estimated along with model parameters Φ .

The assumption of stochastic equivalence among nodes in the same block allows to directly interpret the parameters of SBM with respect to network structure. For example, suppose $P_{r,s} = p$ for all $r, s = 1, \dots, K$ for some known p . Then the SBM reduces to Erdős-Rényi random graph model with the probability of an edge in a pair. When the blocks are assortative, the nodes are more likely to form an edge within blocks and therefore the diagonal elements of matrices P and Φ are greater than off-diagonal elements. Reversely, dis-assortative block structure implies that the nodes in different blocks are likely to form a connection.

3.2.2 Stochastic Blockmodel for Dynamic Networks

Consider a network Y^t observed at time $t \in [0, T]$ for some $T > 0$ with n_t nodes. An edge between node i and node j in the network Y^t , $i, j = 1, \dots, n_t$, is denoted as the (i, j) -th element of matrix Y^t . There have been a variety of studies that propose to analyze dynamic networks via SBM. Most of them focus on detecting change-points, without directly modeling the temporal fluctuation in dynamic networks.

Xu and Hero III 2013; Xu and Hero 2014 propose a state-space model for dynamic networks that extends the static SBM to directly model the temporal fluctuation. Suppose that the block labels are pre-specified and the number of blocks K does not change across time. Define a $K \times K$ matrix P^t that represents the probability of forming an edge across K blocks. Its (r, s) -th element, denoted by $P_{r,s}^t$, indicates the probability of connectivity between block r and block s , $r, s = 1, \dots, K$. Xu and Hero III 2013 introduces a zero mean Gaussian noise matrix Z^t with variance $(\sigma_{r,s}^t)^2$ that explains the variability in connectivity over time. As a result, the probability of forming an edge across K groups at time t is modeled as $P^t + Z^t$. The probability P^t indicates the state of a dynamic system at time t and Z^t is viewed as noise in generating the sequence of observations.

A state-space model considers the state evolution within consecutive time points. A simple approach for the state evolution is random walk. This model considers the temporal fluctuation in connectivity over time, yet does not account for continuous, smooth evolution.

3.2.3 Multilevel Stochastic Blockmodel for Dynamic Networks

Assume there exist l_t networks observed at time $t \in [0, T]$ for some $T > 0$. A symmetric matrix $Y_l^t \in \{0, 1\}^{n_t \times n_t}$ denotes the l -th observed unweighted network at

time t , $l = 1, \dots, l_t$ with n_t nodes. Define the (i, j) -th element of matrix Y_l^t , denoted by $Y_{l,ij}^t$, as the connectivity between node i and node j , $i, j = 1, \dots, n_t$, in the l -th network at time t ; $Y_{l,ij}^t = 1$ if there exists an edge, 0 otherwise. For simplicity, we set the number of nodes is the same for all observed networks at time t (i.e. n_t), but the number of nodes does not have to be the same across l_t networks.

Suppose $Y_{l,ij}^t \sim \text{Bernoulli}(p_{l,ij}^t)$ where $p_{l,ij}^t = P(Y_{l,ij}^t = 1)$. We extend the idea of stochastic blockmodels (SBM) to multilevel dynamic networks, preserving the assumption that the probability of forming an edge between node i and node j – i.e. $p_{l,ij}^t$ for the l -th network at time t – solely depends on the block membership of node i , and j . Assume that the *a priori* block membership is known and let K be the number of pre-specified blocks. Define a smooth function $\phi_{r,s}(t)$ that represents the log odds of probability of forming an edge between block r and block s , $r, s = 1, \dots, K$ at time t . Higher value of $\phi_{r,s}(t)$ indicates the stronger connectivity between block r and block s . Any pair of nodes – one from block r and the other from block s – is modeled to have the same log odds of probability $\phi_{r,s}(t)$ of forming an edge. Note that for symmetric networks $\phi_{r,s}(t) = \phi_{s,r}(t)$. We then construct a $K \times K$ symmetric matrix $\Phi(t)$ with $\phi_{r,s}(t)$ as its (r, s) -th element.

A priori block membership of node i at time t is represented by a vector of length K , z_i^t : if the i -th node at time t belongs to the k -th block, the k -th element of the vector z_i^t is 1. We fix the number of blocks K to stay the same across time as our primary goal of the proposed model is to keep track of evolution of tightness in engagement within and across blocks.

Furthermore, we introduce random effects at time t to illustrate the variability in forming an edge block r and block s in l_t networks. The random effects between block r and block s at time t is denoted by $R_{r,s}(t)$, and we assume $R_{r,s}(t) \sim N(0, \sigma_{r,s}^2 I)$, $r = 1, \dots, K$, $s = r + 1, \dots, K$. The variance $\sigma_{r,s}^2$ penalizes the variation of the random effects deviated from 0 in relational structure between block r and block s .

Similar with the log odds, we construct a $K \times K$ symmetric matrix $\mathbf{R}(t)$ for random effects. For a pair of node i and node j , the random effects $R_{r,s}(t)$ can be expressed as a vector-matrix multiplication $z_i^{t'} \mathbf{R}(t) z_j^t$.

At time t , we model the probability of forming an edge between node i and j at time t as

$$\log\left(\frac{p_{l,ij}^t}{1 - p_{l,ij}^t}\right) = z_i^{t'} \{\mathbf{\Phi}(t) + \mathbf{R}(t)\} z_j^t \quad (3.3)$$

The proposed model assumes that the connectivity between blocks remains the same across l_t bills at time t . That is, the expected log odds of forming an edge between block s and r at time t is $\phi_{r,s}(t)$ across l_t bills. The $\mathbf{\Phi}(t)$ denotes the fixed effects of block membership shared by all edges from l_t networks with n_t nodes at time t . The fixed effects $\mathbf{\Phi}(t)$ can be interpreted as the smooth, overall trend of forming an edge between blocks over time. The random effects $\mathbf{R}(t)$ explain the deviation of connectivity within and between blocks away from the overall trend.

3.3 Estimation

In Section 3.3.1, we describe the overall estimation procedure. We clarify the parameters in the model and form a logistic regression model with random effects. Section 3.3.2 introduces how we achieve the computational efficiency by converting the binary response per network to binomial response across repeated networks.

3.3.1 Model Fitting

In practice, the networks are observed at discrete time points $t = 1, 2, \dots, T$. Define $\mathbf{Y}^t = \{Y_l^t : l = 1, \dots, l_t\}$ as a collection of networks observed at time t . Let $\mathbf{Y} = \{\mathbf{Y}^1, \dots, \mathbf{Y}^T\}$ be a collection of networks observed at T different time points. For a given number of blocks K , there exist $\frac{K(K+1)}{2}$ fixed and random effects. We now consider a vector of length $\frac{K(K+1)}{2}$ for fixed effects $\text{upper}(\mathbf{\Phi}(t)) = \{\phi_{r,s}(t) : r =$

$1, \dots, K, s = r+1, \dots, K\}$; it is a collection of upper triangular elements of $\phi(t)$ and is the parameter of our proposed model. Similarly, we define another vector of the same length for random effects $\text{upper}(\mathbf{R}(t)) = \{R_{r,s}(t) : r = 1, \dots, K, s = r + 1, \dots, K\}$.

Without any constraint, the coefficients $\phi_{r,s}(t)$, $0 \leq t \leq T$, contain an infinite number of parameters. We employ basis splines (b-splines) as a way to reduce the dimensionality of estimation (De Boor et al. 1978; Eilers and Marx 1996). Specify a vector of basis functions $\mathbf{B}(t) = (B_1(t), \dots, B_q(t))$, $0 \leq t \leq T$, of length q . Define a $\frac{K(K+1)}{2} \times q$ matrix Φ of basis coefficients; each row represents the basis coefficients for $\phi_{r,s}(t)$. We then approximate the upper($\Phi(t)$) by a linear combination of these functions as $\text{upper}(\Phi(t)) = \Phi\mathbf{B}(t)$. The set of q basis functions represents the smoothness of $\phi_{r,s}(t)$, and the coefficient matrix Φ determines the shape and trajectory of the fluctuations through time.

Let Φ^t be a $K \times K$ symmetric matrix of which (r, s) -th element $\phi_{r,s}^t$ is the $\phi_{r,s}(\cdot)$ evaluated at time t and let $\text{upper}(\Phi^t)$ denote a vector of length $\frac{K(K+1)}{2}$ with upper triangular elements of Φ^t . Accordingly, let $\mathbf{B}^t = (B_1^t, \dots, B_q^t)$ be a length q vector of basis functions evaluated at time t . Therefore, we write $\text{upper}(\Phi^t) = \Phi\mathbf{B}^t$ at time t . For each block connection between block r and block s , the random effects $\mathbf{R}_{r,s} = \{R_{r,s}^1, \dots, R_{r,s}^T\}$ follows a multivariate normal distribution.

In order to build design matrix for the logistic regression, we concatenate the edge information of all pairs of nodes across T discrete time points. That is, let $y_{l,ij}^t$ denote a binary response of forming an edge between node i and node j in the l -th bill at time t and build $\tilde{\mathbf{Y}}^t$ by stacking all $y_{l,ij}^t$'s for all $l = 1, \dots, l_t$, and $i, j = 1, \dots, n_t$. By stacking all $\tilde{\mathbf{Y}}^t$ at time points $t = 1, \dots, T$, we create a response variable $\tilde{\mathbf{Y}}$. Correspondingly, define the probability of forming an edge $p_{l,ij}^t = P(Y_{l,ij}^t = 1)$ for all $t = 1, \dots, T$, $l = 1, \dots, l_t$, and $i, j = 1, \dots, n_t$ and form \mathbf{p} .

Define b_{ij}^t is a vector of length $\frac{K(K+1)}{2}$ that indicates the block membership of $y_{l,ij}^t$. For example, when $K = 1$, the vector b_{ij}^t is a vector of length 3; $b_{ij}^t = (1, 0, 0)$ if both

node i and node j belong to block 1. $b_{ij}^t = (0, 1, 0)$ if node i belongs to block 1 and node j belongs to block 2. $b_{ij}^t = (0, 0, 1)$ if both node i and node j belong to block 2. Accordingly, $\text{upper}(\mathbf{R}^t)$ is a vector of length $\frac{K(K+1)}{2}$ that represents the random effects of each block at time t . We form a matrix \mathbf{R} that consists of T columns that indicates the $\frac{K(K+1)}{2}$ random effects at time t , $t = 1, \dots, T$.

$$\log\left(\frac{p_{l,ij}^t}{1 - p_{l,ij}^t}\right) = b_{ij}^{t'} \boldsymbol{\Phi} \mathbf{B}^t + b_{ij}^{t'} \mathbf{R} h^t = (\mathbf{B}^t \otimes b_{ij}^t)' \text{vec}(\boldsymbol{\Phi}) + b_{ij}^{t'} \text{upper}(\mathbf{R}^t)$$

where $\text{vec}(A)$ indicates a vectorized form of a matrix A . Let Q_t be the total number of pairs of edges across l_t bills at time t and set $Q = \sum_{t=1}^T Q_t$. Let \mathbf{b}^t represents the group membership of Q_t pairs of edges at time t ; it is a $\frac{K(K+1)}{2} \times Q_t$ matrix. The design matrices for fixed effects (\mathbf{X}) and random effects (\mathbf{Z}) are $Q \times \frac{K(K+1)}{2}$ matrices defined as

$$\mathbf{X} = \begin{pmatrix} \mathbf{B}^1 \otimes \mathbf{b}^1 \\ \vdots \\ \mathbf{B}^T \otimes \mathbf{b}^T \end{pmatrix}', \quad \mathbf{Z} = \begin{pmatrix} \mathbf{b}^1 \\ \vdots \\ \mathbf{b}^T \end{pmatrix}'.$$

Define \mathbf{g} as a component-wise link function $g(p) = \log \frac{p}{1-p}$. Thus, we write the model in (3.3) as

$$\mathbf{g}(\mathbf{p}) = \mathbf{X} \text{vec}(\boldsymbol{\Phi}) + \mathbf{Z} \text{vec}(\mathbf{R}).$$

We implement the penalized iteratively reweighted least squares (PIRLS) algorithm, which combines the step of determining conditional modes and the idea of IRLS for generalized linear regression. For the IRLS part where we update fixed effects parameters $\boldsymbol{\Phi}$, the log likelihood of $\boldsymbol{\Phi}$ given \mathbf{R} is given as

$$\ell(\boldsymbol{\Phi}|\mathbf{R}) = \tilde{\mathbf{Y}}'(\mathbf{X} \text{vec}(\boldsymbol{\Phi}) + \mathbf{Z} \text{vec}(\mathbf{R})) - \mathbf{1}' \log\{1 + \exp(\mathbf{X} \text{vec}(\boldsymbol{\Phi}) + \mathbf{Z} \text{vec}(\mathbf{R}))\}$$

where $\mathbf{1}$ is a vector of 1's with length Q .

To obtain smooth estimates of the time-varying coefficients ϕ^t , we further consider a roughness penalty on the coefficients of the basis functions (see Eilers and Marx 1996; Hastie and Tibshirani 1993; Hoover et al. 1998, for example). We use the integrated squared second derivative, $\Phi^T \Omega \Phi$, with pre-specified Ω . As a result, we update the log-likelihood of Φ given R with penalty as

$$\ell^*(\Phi|R) = \tilde{\mathbf{Y}}'(\mathbf{X}\text{vec}(\Phi) + \mathbf{Z}\text{vec}(\mathbf{R})) - \mathbf{1}' \log\{1 + \exp(\mathbf{X}\text{vec}(\Phi) + \mathbf{Z}\text{vec}(\mathbf{R}))\} - \lambda \Phi' \Omega \Phi$$

where $\mathbf{1}$ is a vector of 1's with length Q and λ is a tuning parameter. The generalized cross validation (GCV) is used to choose the tuning parameter λ (Golub, Heath, and Wahba 1979). Same as logistic regression model with random effects, we implement an iteratively reweighted least squares (IRLS) algorithm for parameter estimation.

3.3.2 Computation

At time t , there exist l_t networks with n_t nodes. Each network contains $\frac{n_t(n_t+1)}{2}$ unique edges, and therefore there are $l_t \times \frac{n_t(n_t+1)}{2}$ binary edges at each time t . Across T time points, the response vector T is of length $\sum_{t=1}^T l_t \times \frac{n_t(n_t+1)}{2}$. The response vector quickly gets long as either the number of nodes or the number of time points becomes bigger. Given the assumption that the number of nodes n_t stays the same across l_t networks at time t , we are able to convert binary response $y_{l,ij}^t \sim \text{Bernoulli}(p_{l,ij}^t)$, $l = 1, \dots, l_t$, into a binomial response by summing up edges across l_t networks. In other words, we can write $\sum_{l=1}^{l_t} y_{l,ij}^t \sim \text{Binomial}(l_t, p_{ij}^t)$. Note that the probability of forming an edge remains the same across l_t networks at time t . The probability p_{ij}^t can be expressed as in model (3.3). This conversion greatly reduces computational complexity.

3.4 Application

We now analyze the dynamic network that describes the co-voting patterns among U.S. Senators from 1867 (Congress 40) to 2015 (Congress 113). This network was first investigated in Moody and Mucha 2013 and has been subsequently analyzed in Wilson, Stevens, and Woodall 2016 and Lee, Li, and Wilson 2017. The network is based off of the roll call voting data from <http://voteview.com>, which contains the voting decision of each Senator (“Yay” or “Nay”) for every bill brought to Congress. Previous analyses in Moody and Mucha 2013; Wilson, Stevens, and Woodall 2016 have identified significant changes in the community structure of the network over time, and that this community structure is closely associated with political affiliation of the Senators. Therefore, a priori block membership in this application studies is political affiliation of the Senators and we model the fluctuating pattern of connectivity between and within political parties over time.

For the l -th bill in Congress t , we can create a network $Y_l^t \in \{0, 1\}^{n_t \times n_t}$, $l = 1, \dots, t$. The binary edge $y_{i,j}^t$ is formed if Senator i and Senator j vote concurrently in the l -th bill in Congress t . The overall temporal trend of connectivity across political parties is modeled as $\Phi(t)$ and we incorporate the random effects of time across bills as $\mathbf{R}(t)$ in (3.3).

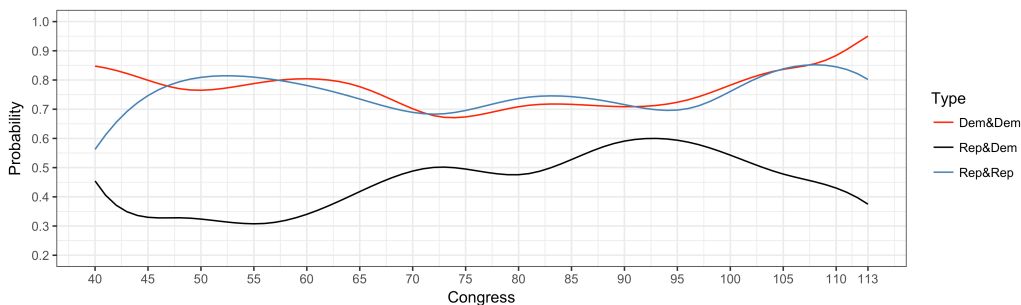


Figure 3.1: **Estimated probability of forming an edge between political parties from proposed model:** Probability of forming an edge within Democrat (Red), within Republican (Blue), and across political parties (Black) in Congress 40 -113.

Figure 3.1 above presents the estimated probability of forming an edge between

Senators depending on their political affiliation in Congress 40 - 113. The probability of connectivity across political parties are presented in different colors. The connectivity within the same political party remains stronger than the across-party connectivity. That is, Senators in the same political affiliation are more likely to form an edge in the network. Between the 40th to the 70th Congress, the connectivity within Democrat (red) and within Republican (blue) stays relatively flat, but increases after Congress 95. Inversely, the estimated connectivity between Democrat and Republican drops after Congress 95. This pattern indicates the polarization between political parties.

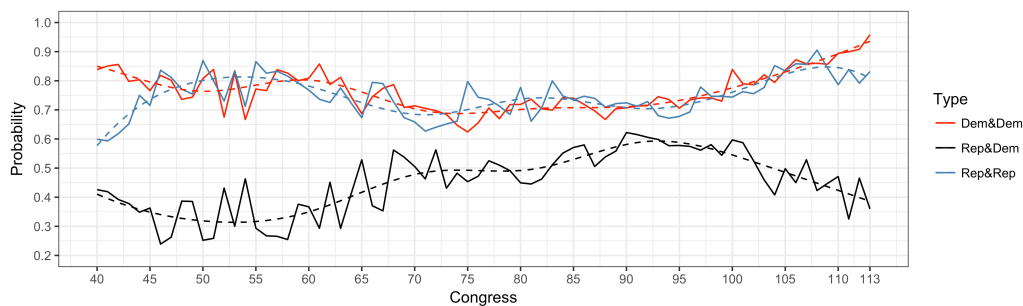


Figure 3.2: **Estimated probability of forming an edge between political parties from cross-sectional SBM:** Probability of forming an edge within Democrat (Red), within Republican (Blue), and across political parties (Black) in Congress 40 -113. The solid line represents the cross-sectional SBM estimates and the dashed line represents the cross-sectional SBM estimates after ad-hoc smoothing.

As a comparison, we fit the cross-sectional SBM and adapt an ad-hoc smoothing procedure. In Figure 3.2, the solid lines are for estimated probability of forming an edge from cross-sectional SBM and the dashed lines are after ad-hoc smoothing procedure. The estimates from cross-sectional SBM are spiky, but are in general congruous with those from our proposed model. The estimates with ad-hoc smoothing are similar with the estimated probability from our proposed model, implying that our proposed model consistently captures the overall, smooth trend across time.

3.5 Discussion

In this project, we propose a method for analyzing multilevel dynamic networks. When there are repeated networks observed at one time point, our proposed method combines the idea of stochastic blockmodel (SBM) and random effects model to analyze the evolution of relational structure. Particularly, the proposed method is valuable in that it considers the multiple networks as repeatedly observed networks and models the variability across the multiple networks via random effects. Often, aggregating a collection of relational patterns is implemented to define a single network at one time point. This process potentially introduces a loss of information; variabilities between multiple networks at one time point. The proposed method incorporates the variabilities in a form of random effects and considers them as a deviation from overall trend (i.e. fixed effects). We impose the smoothness on fixed effects and interpret the fixed effects as a smooth evolution of relational structure over time.

The parameter of our proposed model represents the smooth, overall trend of topological evolution across pre-specified blocks in dynamic networks. The random effects in the model explain the variability between repeated networks observed at the same time points. Similar with other versions of SBM, our model parameters can be estimated via fitting a logistic regression model. We propose a strategy to achieve the computational efficiency for multilevel structure of dynamic networks. We applied our proposed method to the US Congress co-voting networks to examine how the connectivity between two political parties has evolved over time from the Congress 40 to the Congress 113. Consistent with other studies that have analyzed the networks, our proposed model demonstrated the polarization between the two political parties since Congress 95. In addition to the overall trend, our model facilitates to examine the variability across bills across Congresses.

There are several directions we would like to pursue to investigate our proposed

model. First of all, we set our proposed model to be with pre-specified block membership and model how the strength of connectivity between blocks changes over time. Because our primary goal is to investigate the evolving strength of connectivity across blocks, a priori block membership is a reasonable assumption. In many cases, however, group labels of nodes are not known and estimating both the number of groups and group membership of nodes are important tasks. In such scenario, it would be important to carefully consider the identifiability of the model. Secondly, there have been studies that consider the mixed membership in stochastic blockmodels. It would be interesting to modify our model to incorporate the mixed membership in our proposed model. Lastly, Sweet 2015 discussed the covariate effects in stochastic blockmodels. Introducing the fluctuating contribution of covariates on forming an edge would be interesting to develop our model in the future; this approach allows us to further personalize the pattern of connection depending on each node's attributes. These directions remain as future work.

Chapter 4

Curve Registration

4.1 Introduction

4.1.1 Motivation

In order for humans to walk or run, the ground must exert a force on the bottom of the shoe, or foot, that results in acceleration of the human center of mass. This force is called the ground reaction force (GRF). The resultant GRF vector is often resolved into three orthogonal components: a vertical component and two horizontal components often called anterior-posterior (fore-aft) and medial-lateral (side-to-side). Various characteristics of the vertical component of walking and running GRF are measured, because they are associated with common musculoskeletal impairments and disease. For example, characteristics of the vertical GRF during walking and running are associated with knee joint health, including knee osteoarthritis onset and progression (Hyldahl et al. 2016; Seeley et al. 2017; Teng et al. 2017). Vertical GRF is also measured in order to discern progression of a musculoskeletal disease, like knee osteoarthritis, or the effectiveness of a clinical intervention designed to slow disease progression (Pietrosimone et al. 2017).

Currently, the accurate measurement of walking and running GRF requires a

subject to walk across a force platform that is either embedded in a laboratory floor or moveable surface (e.g., a force-sensing treadmill). Commercial force platforms and force-sensing treadmills are expensive; further, such instruments are restricted to laboratory environments and require extensive human resources (i.e., expertise) to manage. These challenges prohibit some researchers and most clinicians from making accurate measures of walking and running GRF. Additionally, real-world (i.e., out of the laboratory) measures of GRF are currently difficult or impossible to obtain.

These issues have motivated a development of mobile force sensors that can be used to measure GRF outside of a traditional motional analysis laboratory. Novel piezo-responsive foam sensors placed in athletic shoes have been recently developed to accurately estimate walking 3D GRF outside of the laboratory (Rosquist et al. 2017). The electromechanical behavior of these sensors has been validated for use in various large-strain applications (Bilodeau et al. 2015; Johnson et al. 2011). The strain-induced voltage is measured by attaching a conductive material embedded in the foam to a voltage sensing system, which correlates to the force of impact (Merrell et al. 2013). By embedding the foam into a shoe sole, it has been shown that the voltage response generated during gait accurately correlates to 3D GRF (Rosquist et al. 2017).

There are several potential advantages of using in-shoe sensor as a surrogate of GRF measurement. While the GRF is only recordable in controlled indoor laboratories, in-shoe sensors can be simply incorporated into the sole of a pair of shoes and deployed anywhere to analyze one's gait in various settings. Furthermore, Seliktar, Yekutieli, and Bar 1979 pointed out a chance of distortion in one's gait pattern when asked to walk on the force plates. Distortions that interfere with detecting a true gait pattern are less likely to happen when using the in-shoe sensors placed on a regular pair of shoes. Thus it is important to explore the properties of in-shoe sensor data to understand if in-shoe recordings can be a viable alternative of GRF measurement.

Figure 4.1 shows recorded values from both in-shoe sensors and gold-standard GRF measurement obtained via a force-sensing treadmill, during the ground contact phase of running (i.e., between heel strike and toe-off, called a “stance”) for five consecutive stances in a healthy individual. Recording for both in-shoe sensor and vertical GRF (VGRF) measurements are shown over time for each stance. In the first four panels, the x -axis corresponds to the standardized time frame and the y -axis corresponds to in-shoe sensor readings. The sensor is measuring a triboelectric effect as the embedded nano fillers rub against the base polymer in the foam. This effect may be amplified or concentrated by the voids in the foam which can function as a short duration capacitor that stores voltage and then discharges it. Larger forces produce larger displacements in the foam, which correspond to higher magnitude electrical response; in turn, the responses create more negative values in the sensor readings. The sensor has its lowest (most negative) values when reacting to the largest forces. The sensor has its lowest (most negative) values when reacting to the largest forces.

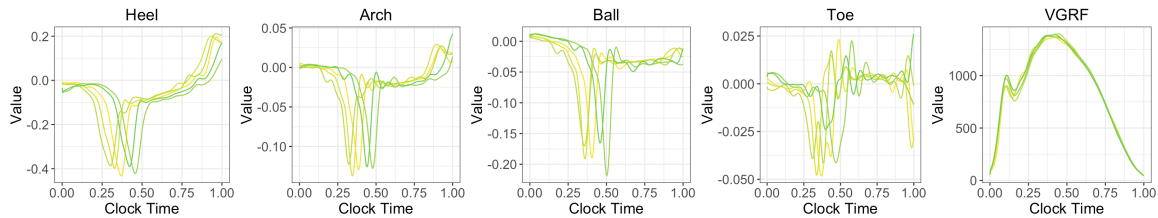


Figure 4.1: **In-shoe sensor and GRF data:** In-shoe sensor recordings and vertical GRF (VGRF) of a healthy subject’s 5 stances. Data from in-shoe sensors at four locations are shown in the first four panels. The fifth panel shows the measurement of vertical GRF. Each curve represents a stance.

The values observed from the sensors reflect the subject’s gait pattern. According to Figure 4.1, for example, each stance in in-shoe sensor data contains a large dip, after which the value increases and flattens until the completion of the stance. Recordings within a sensor share a common structure that is misaligned across stances: the exact timing of major features depends on stance. Furthermore, the magnitude of the common feature differs between stances. These observations relate to *phase* and

amplitude variability, respectively; phase variability relates to shifts in time, while amplitude variability relates to the change in the magnitude of measurements.

The VGRF measurement shown in the fifth panel of Figure 1 is more stable than in-shoe sensor recordings, and does not show significant time shifts across stances. Indeed, it has been documented that healthy runners are very consistent in their stances (Benedetti, Merlo, and Leardini 2013; Karamanidis, Arampatzis, and Brüggemann 2004). Thus, the phase variability in in-shoe sensor data across stances in Figure 4.1 is not expected, and the misalignment arises in the recording process rather than reflecting actual variability across stances. Removing phase variability from the in-shoe sensor data is a necessary step if these measurements are going to be used as surrogate measures of VGRF. Additionally, analyzing the sensor data without proper understanding of phase variability may lead us to draw misleading conclusions regarding the amplitude variability of common stance features (Sadeghi et al. 2000). Our goal, then, is to explore the elimination of phase variability of in-shoe sensor data without altering the values taken by the curves, so that differences in amplitude variability can be evaluated as alternative measures of the VGRF in studies of walking pattern and gait.

Temporal realignment of curves is referred to as *curve registration* in the functional data analysis literature. Specifically, curve registration shifts, stretches, and compresses the observations in time so that major features are aligned across curves. In this process, clock (originally observed) time is converted into the system (common across curves) time via *time warping functions*. Let t^* and t denote clock and system time, respectively. A warping function $h : [0, 1] \rightarrow [0, 1]$ represents the functional relationship between clock time and system time through $t^* = h(t)$. The warping functions are monotone increasing with $h(0) = 0$ and $h(1) = 1$. The principal challenge in registration, then, is the estimation of warping functions. Curve registration is often necessary before applying additional statistical methods to smooth

curves, and warping functions themselves can contain useful information about observed curves. In this manuscript, we use curve registration to understand phase and amplitude variability arising in in-shoe sensor data.

We investigate the variability present in in-shoe sensor data recorded for three representative healthy subjects, with more than 300 stances per subject. We emphasize the importance of understanding the source of variability and the utility of adequately addressing phase variability. Based on the hypothesis that time shifts across sensors within the same stance may be similar, we examine the correlation of warping functions using a permutation test. We further examine amplitude variability of in-shoe sensor recordings after registration, particularly with relation to GRF, by employing function-on-function regression models. All analyses are conducted separately for each subject due to unique running patterns.

4.1.2 Literature Review

In studies of gait, curve registration is valuable but underutilized in reducing intra-subject variability. In studies of joint mechanical power data, Sadeghi et al. 2000, 2003 emphasized the importance of curve registration as a preprocessing step. Both studies obtained joint mechanical power data for the right lower limb of healthy subjects using a 3D video-based system. Sadeghi et al. 2000 applied a straightforward registration method to align salient features of the observed curves before comparing key power bursts; Sadeghi et al. 2003 implemented more flexible registration method. These studies found that realigning the observed curves reduces the temporal variability induced by external sources and instrumental issue and facilitates the focus on sources of variability that reflect meaningful biomechanical features. More recently, curve registration was used in gait studies on healthy subjects (Crane et al. 2010; Helwig et al. 2011) and for comparing healthy individuals to stroke patients (Thies et al. 2009).

In functional data analysis, curve registration was introduced to identify a shared structural pattern in a sample of curves and to understand individual realizations of the shared pattern (Kneip et al. 2000; Kneip and Gasser 1992; Ramsay and Li 1998). There have been a variety of approaches to and applications of curve registration. For an in-depth history of curve registration, see Marron et al. 2014; Marron et al. 2015.

A simple approach to curve registration, referred to as landmark registration, locates important features of the observed curves by hand or using an automated process and realigns them using piecewise linear warping functions (Gasser and Kneip 1995; Kneip and Gasser 1992); landmark registration was implemented in Sadeghi et al. 2000. Although landmark registration is simple and straightforward to implement, it can be difficult and time consuming to determine landmark locations, and the performance of this approach may be poor in the area away from the landmarks.

More flexible methods for curve registration have been introduced. In principle, they estimate nonlinear, monotone warping functions that map the system (warped) time into observed clock time. Silverman 1995 proposes a method that does not require landmarks by considering uniform shifts in time to realign the observed curves. Ramsay 1998 and Ramsay and Li 1998 propose an iterative algorithm with two steps: first, estimate the cross-sectional mean of the registered curves using current warping function estimates, and second, update warping function estimates to minimize distance to the shared mean. This approach was used by Sadeghi et al. 2003. Building on this framework, functional principal component analysis (FPCA) has been widely used to model the common structure shared by registered curves (Earls and Hooker 2015a,b; Kneip and Ramsay 2008).

Srivastava et al. 2011a suggests a metric-based framework for registering elastic curves. The method, like other iterative algorithms for curve registration, alternates between two steps until convergence. In this approach, the Fisher-Rao Riemannian

metric and square-root velocity function (SRVF) are used to quantify the distance between two curves. The Fisher-Rao metric is a widely used tool to compare the shape of curves (Peter and Rangarajan 2006) and the SRVF maps the Fisher-Rao metric to Euclidean space (Srivastava et al. 2011b), which enables the comparison of distances in \mathbb{L}^2 space. In particular, Peter and Rangarajan 2006 represents landmark shapes as mixtures of Gaussians and use the parametric Fisher Rao metric to compare these density representations. On the other hand, Srivastava et al. 2011b use the extension of the nonparametric Fisher Rao metric directly on the space of functions and use its parameterization invariance to separate phase and amplitude variability. The SRVF then maps this metric to \mathbb{L}^2 , which greatly simplifies computation in this case. In terms of shape analysis, i.e., analyzing shapes of Euclidean curves with dimension higher than 1, the SRVF maps a specific case of an elastic metric to \mathbb{L}^2 . This method is extended to generative models in Tucker, Wu, and Srivastava 2013 and the analysis of shape of elastic curves in Euclidean space in Srivastava et al. 2011b; it has also been applied to proteomics data and spike train data (Tucker, Wu, and Srivastava 2014; Wu and Srivastava 2014).

The remainder of this paper is organized as follows. Section 4.2 introduces the gait data and describes the questions of interest. Section 4.3 illustrates the use of curve registration, constructs a permutation test framework for evaluating correlation within stance across sensors, describes a metric that quantifies the amplitude variability before and after curve registration, and introduces functional regression methods to model the association between in-shoe sensor recordings and VGRF curves. We present the results of curve registration, permutation testing, quantification of amplitude variability, and regression model fitting in Section 4.4. We conclude with a discussion of our findings and open areas for future research in Section 4.5.

4.2 Data

Data were collected in a biomechanics laboratory at Brigham Young University. Subjects were instrumented with a Cosmed K4b2 portable metabolic analyzer (Cosmed K4b2, Cosmed, Rome, Italy) and standardized athletic shoes instrumented with the nanocomposite piezo-responsive foam (NCPF) sensors, accompanying electrical components, and an accelerometer attached to the dorsal aspect of the shoe. Next, to be able to account for any NCPF sensor drift, subjects completed a 15-minute warm-up run at 2.68 m/s. After this warm-up, subjects completed five different trials in a randomized order. Each trial consisted of four minutes of walking or running, at one of the following speeds: 1.34, 2.23, 2.68, 3.13, or 3.58 m/s. A 1-min walk (1.34 m/s) was completed before and after every trial, to be able to characterize for any drift in the NCPF sensors. Voltages, recorded via the NCPF sensors and microcontroller (1000 Hz), energy expenditure, recorded via the Cosmed (breath by breath), and accelerations, measured via the shoe accelerometer (16 Hz), were measured throughout the entire collection period (the warm-up period and all five trials), which lasted approximately 50 minutes.

The data set analyzed here consists of measurements obtained from three healthy female subjects who wear size 8 shoes. The mass and height of the subjects are (52, 60, 60) kg and (161, 165, 169) cm, respectively. Each subject was required to 1) be between the ages of 18 and 30; 2) have no history of lower-extremity injury within the past 6 months; 3) have no history of lower-extremity surgery in their lifetime; 4) be able to walk and run without pain; and 5) be able to comfortably run consistently for at least 5 continuous km. Each subject ran at 6mph (2.68 m/s) for 4 minutes and completed 362, 321, and 308 stances, respectively. Consistent with much of the scientific literature in this area, we define “stance” as a period from heel-strike (point at which vertical GRF rises above 50 N) to toe-off (point at which VGRF drops below 50 N). From each stance, information such as VGRF and in-shoe sensor measurements



Figure 4.2: **In-shoe sensor:** (Left) NCPF sensor with wires embedded to measure the voltage response during impacts. (Right) Shoe equipped with NCPF foam sensors, microcontroller, and battery. The sensors were embedded under the insole at the heel, arch, ball, and toe.

at four locations (heel, arch, ball, and toe) has been collected. Completing a stance takes different amount of time; once a stance is extracted, the curves are interpolated to a common domain with 200 discrete time points. The time frame is then scaled to the unit interval $[0, 1]$.

The resulting data are shown in Figure 4.3. For each in-shoe sensor and subject, a common feature exists across stances but is misaligned in time. In addition to the observed phase variability, there is amplitude variability across individuals and stances. The distinctive patterns in the in-shoe sensor and VGRF data across subjects are due to different running styles. We classify the subject in purple as a mid-foot striker; this runner is more likely to land near the middle of the foot, rather than the heel, leading to low amplitude variability in the heel sensor and high amplitude variability in the arch and ball sensors. This running pattern also results in the absence of an impact transient, or initial local maximum, in the VGRF. In contrast, the remaining subjects are heel strikers who show greater amplitude variability in the heel sensor, lower amplitude variability in the arch, ball, and toe sensors, and have,

to differing degrees, the impact transient in the VGRF.

These distinctive running patterns may have long term implications for biomechanics and the health of joints involved in running (Rice, Jamison, and Davis 2016). The initial peak or impact transient is an important element of GRF data, as is the slope, or load rate, of the impact transient. A steeper impact transient load rate is thought to be correlated with certain musculoskeletal injury (Daoud et al. 2012), and the absence of this peak for the subject in purple might suggest that she is at lower risk for certain types of injuries (Rice, Jamison, and Davis 2016).

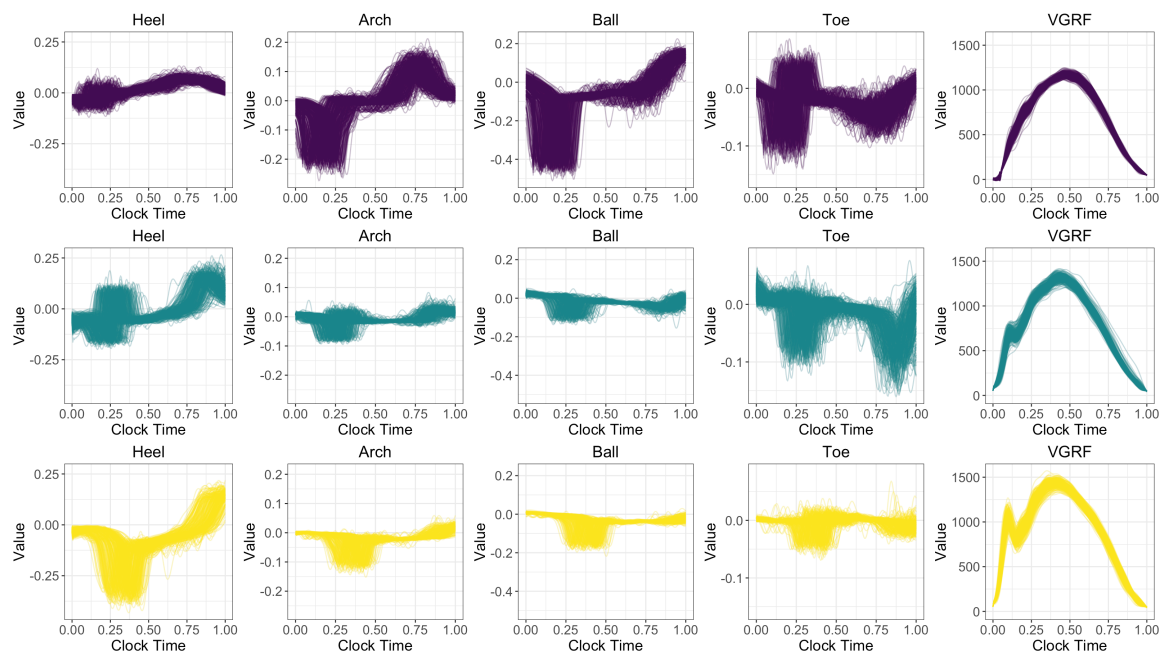


Figure 4.3: **Original measurements:** Observed data for three runners are shown in rows; data from in-shoe sensors at various locations are shown in the first four columns and the direct measure of VGRF is shown in the fifth column. Curves are color-coded per subject and each curve represents a stance.

As noted in Section 4.1.1, time shifts in in-shoe sensor data are not expected for healthy subjects. The stance-level force measurements are expected to be relatively consistent, and shifts away from a common structure shared by every curve are an issue of measurement rather than true phase variability; this is emphasized by the consistency of concurrently-measured VGRFs during data collection. There

are several possible reasons for the presence of phase variability in in-shoe sensor data, including the need for a warming-up duration before sensors output consistent voltages corresponding to consistent impacts and the possibility of misalignment in the sensor recording mechanisms (e.g. inconsistent identification of heel strike). It is also unclear whether the phase variation is similar across sensors within the same stance; that is, whether recordings made by the difference sensors on the same stance are time-shifted in a similar way. For a new and developing technology such as NCPF sensors, preprocessing the observed data is valuable for understanding how the technology can be used and improved.

4.3 Methods

Unlike classic statistical methods, where one observation consists of a single scalar value, functional data analysis considers the basic unit of observation to be a smooth curve. The gait data illustrated in Figure 4.3 are an example of functional data, since stances recorded by each sensor within each subject are the observations of interest. We introduce notation to denote curves for each subject and conduct curve registration to each sensor and subject separately. Let $y_i^s(t^*)$, $t^* \in [0, 1]$, denote the observed curve for the i -th stance for sensor s , where $i = 1, 2, \dots, I$, and $s = 1$ (Heel), 2 (Arch), 3 (Ball), 4 (Toe). Here, the clock time t^* is curve specific; the goal of curve registration is to estimate warping functions $t^* = h_i^s(t)$ that map the shared system time t to curve-specific clock time t^* .

4.3.1 Curve Registration

We implement the curve registration proposed by Srivastava et al. 2011a. Like other approaches, this method is based on an iterative algorithm that alternates two steps until convergence. In the first step, the mean of registered curves using current

warping functions is estimated; this mean is referred to as a template. In the second step, the warping functions are updated to minimize the distance between the curves and the current template using the square-root velocity function (SRVF) to calculate the distance between two curves. The SRVF represents the Fisher-Rao Riemannian metric, which is a widely used tool to compare the shape of curves, in \mathbb{L}^2 and alleviates the computational complexity of the algorithm.

More concretely, for an absolutely continuous function $y_i^s(t^*)$, $t \in [0, 1]$ with its derivative $y_i^{s'}(t^*)$, define the SRVF $q_i^s : [0, 1] \rightarrow \mathbb{R}$ as

$$q_i^s(t^*) = \text{sign}(y_i^{s'}(t^*))\sqrt{|y_i^{s'}(t^*)|}.$$

In the first iteration, the template $\mu(t^*)$ is taken to be the observed $q_i^s(t^*)$ that is closest to the sample mean $\frac{1}{I} \sum_{j=1}^I q_j^s(t^*)$. Let $q_{i,h}^s(t^*) = q(h^{-1}(t^*))\sqrt{(h^{-1})'(t^*)}$ denote the SRVF of registered curve $y_i^s(h^{-1}(t^*))$ with warping function $t^* = h(t)$. For any curves $y_l^s(t^*)$ and $y_m^s(t^*)$, $i, i' \in \{1, 2, \dots, I\}$ and warping function $h(t)$, the distance between SRVFs of registered curves is the same as that between SRVFs of unregistered curves. That is, $\|q_{i,h}^s(t^*) - q_{i',h}^s(t^*)\| = \|q_i^s(t^*) - q_{i'}^s(t^*)\|$. Then, in the second step of the iterative algorithm, for each observed curve $y_i^s(t^*)$, warping function estimate is updated as $\text{argmin}_h \|\mu(t^*) - q_{i,h}^s(t^*)\|$. That is, the warping function estimate minimizes the distance of SRVF between the template and registered curves. In subsequent iterations, the template is updated as a mean of SRVFs q_{i,h_i}^s with current warping function estimates. The algorithm iterates the two steps of calibrating the template $\mu(t^*)$ and updating the warping functions until convergence.

4.3.2 Sources of Variability

We conduct exploratory analyses on registered curves to evaluate the utility of curve registration. Following Tucker, Wu, and Srivastava 2013 and Kneip and Ramsay 2008,

we define the amplitude variability of observed curves $\{y_i^s(t^*), i = 1, 2, \dots, I, t^* \in [0, 1]\}$ for sensor s as

$$V^s = \frac{1}{I-1} \int_0^1 \sum_{i=1}^I \left(y_i^s(t^*) - \frac{1}{I} \sum_{j=1}^I y_j^s(t^*) \right)^2 dt^*.$$

Similarly, we define the amplitude variability after curve registration with time transformation $t^* = h_i^s(t)$ and denote the resulting quantity as V_h^s . This definition quantifies the amplitude variability as the mean integrated sum of squared differences between curves in a sample and their mean. We expect that the amplitude variability before curve registration, V^s , may contain variability attributable to misalignment in time. Therefore, we compare the amplitude variability before and after curve registration to describe the utility of curve registration.

Additionally, we conduct functional principal component analysis (FPCA) to understand the patterns observed in amplitude variability after curve registration (Goldsmith, Greven, and Crainiceanu 2013; Yao, Müller, and Wang 2005). FPCA allows a parsimonious representation of registered curves by decomposing the observed functions into a mean, scores, and shared functional principal components:

$$y_i^s(t) = \mu^s(t) + \sum_{k=1}^K c_{ik}^s \phi_k^s(t) + \epsilon_i^s(t) \quad (4.1)$$

The representation in (4.1) is based on the Karhunen-Loève representation of the $y_i^s(t)$ in which $\mu(t)$ is the population mean, the ϕ_k^s are population level basis functions obtained through an eigendecomposition of the covariance $\text{Cov}(y_i^s(t), y_i^s(t'))$ with corresponding eigenvalues λ_k^s such that $\lambda_1^s \geq \lambda_2^s \geq \dots$, the subject-specific scores $c_{ik}^s \sim (0, \lambda_k^s)$ are uncorrelated random variables, and the $\epsilon_i^s(t)$ are white noise errors.

4.3.3 Phase Variability Across Sensors

We are interested in exploring the similarity of phase shifts across sensors within stances. Similarity in time shifts across sensors within stance may reasonably be expected given the biomechanical process underlying these data; dissimilarity would suggest that phase shifts are not consistent across sensors in the same stance.

To assess similarity, we use a permutation test based on a functional analog of Pearson's correlation that considers the relationship between two warping functions after subtracting the identity function (i.e. $h(t) = t$). Namely, we define the correlation between two warping functions from sensor s and s' within the i -th stance as

$$\text{Corr}_i = \frac{\int_0^1 \{h_i^s(t) - t\} \{h_i^{s'}(t) - t\} dt - \int_0^1 \{h_i^s(t) - t\} dt \int_0^1 \{h_i^{s'}(t) - t\} dt}{\sqrt{[\int_0^1 \{h_i^s(t) - t\}^2 dt - [\int_0^1 \{h_i^s(t) - t\} dt]^2][\int_0^1 \{h_i^{s'}(t) - t\}^2 dt - [\int_0^1 \{h_i^{s'}(t) - t\} dt]^2]}}.$$

Terms in the numerator and denominator are based on the expectations and variances that appear in the usual definition of Pearson's correlation. Let $\hat{h}_i^s(t)$ denote the estimated warping function of the i -th stance for sensor s , where $i = 1, 2, \dots, I$ and $s = 1, 2, 3, 4$. The test statistic used to evaluate the similarity across sensors is the mean of correlations within stance Corr_i ; that is, we compute the correlation between two sensors within each stance, and then average across stances.

To determine statistical significance, we randomly permute the stance label of second sensor, recompute within-stance correlations, and average across stances. The test statistic for k -th permuted sample, denoted as T_k , is defined as a mean of correlation in the permuted sample. We repeat the permutation process K times to obtain a null distribution for our test statistic. The p -value for permutation test is defined to be

$$p\text{-value} = 2 \times \frac{\sum_{k=1}^K I\{T < T_k\}}{K}.$$

4.3.4 Association between In-Shoe Sensors and GRF

We are interested in the relationship between VGRF curves and the aligned in-shoe sensor recordings because any meaningful relationship between VGRF curves and in-shoe sensor recordings may support the utilization of inexpensive in-shoe sensor measurement as a surrogate for GRF measurement.

To understand possible associations, we apply function-on-function regression models using VGRF as an outcome and in-shoe sensor recording as a predictor. Let $\text{GRF}_i(t)$ be the measure of VGRF from the i -th stance and let $y_i^s(u)$, $u \in [0, 1]$ denote the registered curve of the i -th stance for sensor s . We fit a model with VGRF as a response and the sensor s recordings as a predictor:

$$\text{GRF}_i(t) = \beta_0(t) + \int_0^1 y_i^s(u) \beta(u, t) du + \varepsilon_i(t). \quad (4.2)$$

We fit separate models for each subject sensor predictor using the tensor-product spline approach described in Scheipl, Staicu, and Greven 2015 and implemented in the `pffr` function in the `refund` package in R (Goldsmith et al. 2016). The bivariate coefficient surface $\beta(u, t)$ is smooth over both u and t and relates the predictor measured over u to the response measured t , respectively. In this model, fitted values at time t are obtained fixing t and viewing $\beta(u, t)$ as a univariate coefficient function over u , which is multiplied by the predictor and integrated over u . The bivariate smoothness of $\beta(u, t)$ allows the effect of predictor functions to vary over the domain of the response.

4.4 Results

4.4.1 Curve Registration

Curve registration is applied separately to each individual and each measurement. Figure 4.4 shows curves after registration, and can be directly compared to Figure 4.3. Curve registration achieves a reduction in phase variability; curves are better aligned and amplitude variability is more easily understood. For example, the spikes in the early phase of the heel sensor for each subject were largely obscured before registration. In the middle of Subject 1’s arch and Subject 2’s toe sensors, there is some variability after registration.

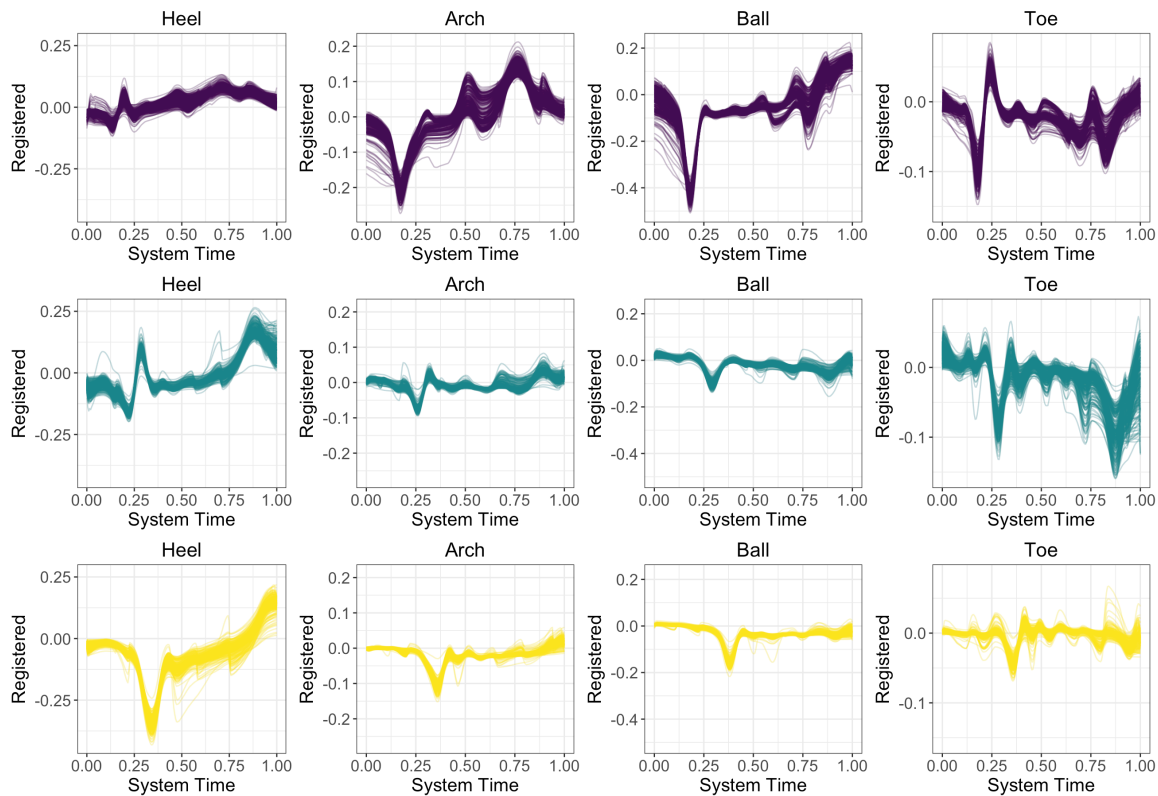


Figure 4.4: **Curve after registration:** Registered curves for three runners are shown in rows; curves from in-shoe sensors at various locations are shown in columns. Curves are color-coded per subject and each curve represents a stance.

Inverse warping functions for such realignments of Subject 1 are presented in Fig-

ure 4.5, with observed clock times on the x -axis and aligned system times on the y -axis. Roughly speaking, inverse warping functions above the identity line shift an early peak in the observed time and map it to a later system time, while inverse warping functions below the identity warping shift later peaks to an earlier system time. In the arch sensor, the range of system time when clock time is 0.5 is wide compared to other sensors, which is reasonable given the wide time shifts across stances in the originally observed curves. Visual inspection of inverse warping functions suggests that adjacent sensors, such as the heel and arch or the ball and toe, are similar for this subject, which may imply some association in time shifts for these sensors. The similarity in warping functions across sensors within a subject motivates the pairwise comparison of warping functions across sensors presented in Section 4.4.3.

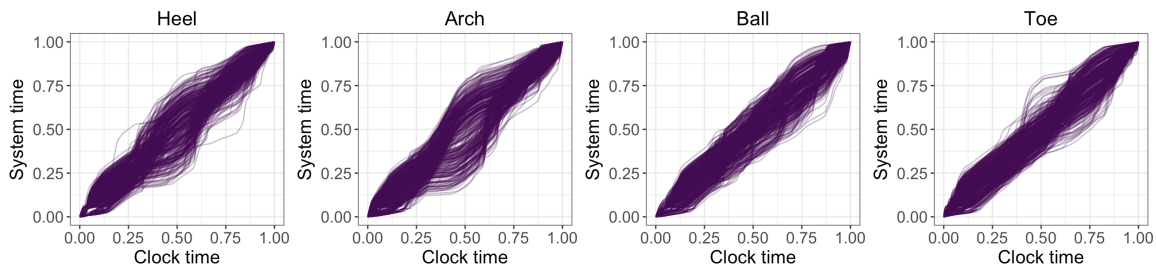


Figure 4.5: **Warping functions:** Each panel represents warping function estimates of Subject 1 of four different in-shoe sensors. The clock time (x -axis) is the curve-specific time before registration and system time (y -axis) is the registered time which is common for every curve. Each curve represents a stance.

4.4.2 Sources of Variability

Table 4.1 presents the amplitude variability, defined in Section 4.3.2, before and after curve registration. As expected, all sensors across subjects have smaller amplitude variability after curve registration. Many sensors have large reductions in amplitude variability, often by as much as 80 – 90 %; for example, the heel sensor of Subject 2 and the heel and arch sensors of Subject 3. The substantial reduction in amplitude

variability indicates that the registration has a substantial effect in understanding the sources of variability in sensor data.

Table 4.1: **Comparison of variability:** Amplitude variability ($\times 10^5$) before and after registration.

	Heel		Arch		Ball		Toe	
	Before	After	Before	After	Before	After	Before	After
Subject 1	38.50	20.70	183.84	49.39	519.77	78.77	48.43	9.01
Subject 2	196.76	28.76	24.44	5.09	45.62	9.22	51.95	19.80
Subject 3	340.98	39.93	30.94	3.48	57.22	4.83	9.09	2.65

After curve registration, we can better understand the patterns that underly amplitude variability. We conduct FPCA to identify the dominant direction of variation in the registered curves. The first two FPCs for all subjects and sensors are plotted in Figure 4.6. Between 24 and 50% of curve variability is explained by the first FPC and the first two FPCs together explain between 39 and 69%. The FPCs for each subject, even those coming from the same sensor, are very different; this supports the uniqueness of running patterns from subject to subject. It is also noteworthy that for only a few subjects, major patterns of variation coincide with the location of major peaks, while in most cases the FPCs are relatively flat in the areas where major features are observed. For example, the peak in the heel sensor does not show up in either the first or second FPC, suggesting much of the variability in the sensor is not related to the magnitude of the largest force but instead lies elsewhere in the sensor recordings.

4.4.3 Phase Variability Across Sensors

Using the permutation test framework described in Section 4.3.3 with 1000 permuted datasets, correlations between warping functions within stance across sensors are significant for all subjects and all sensor pairs; the p-value for each pairwise comparison is less than 0.001. These results indicate that the phase variability in-shoe sensors are

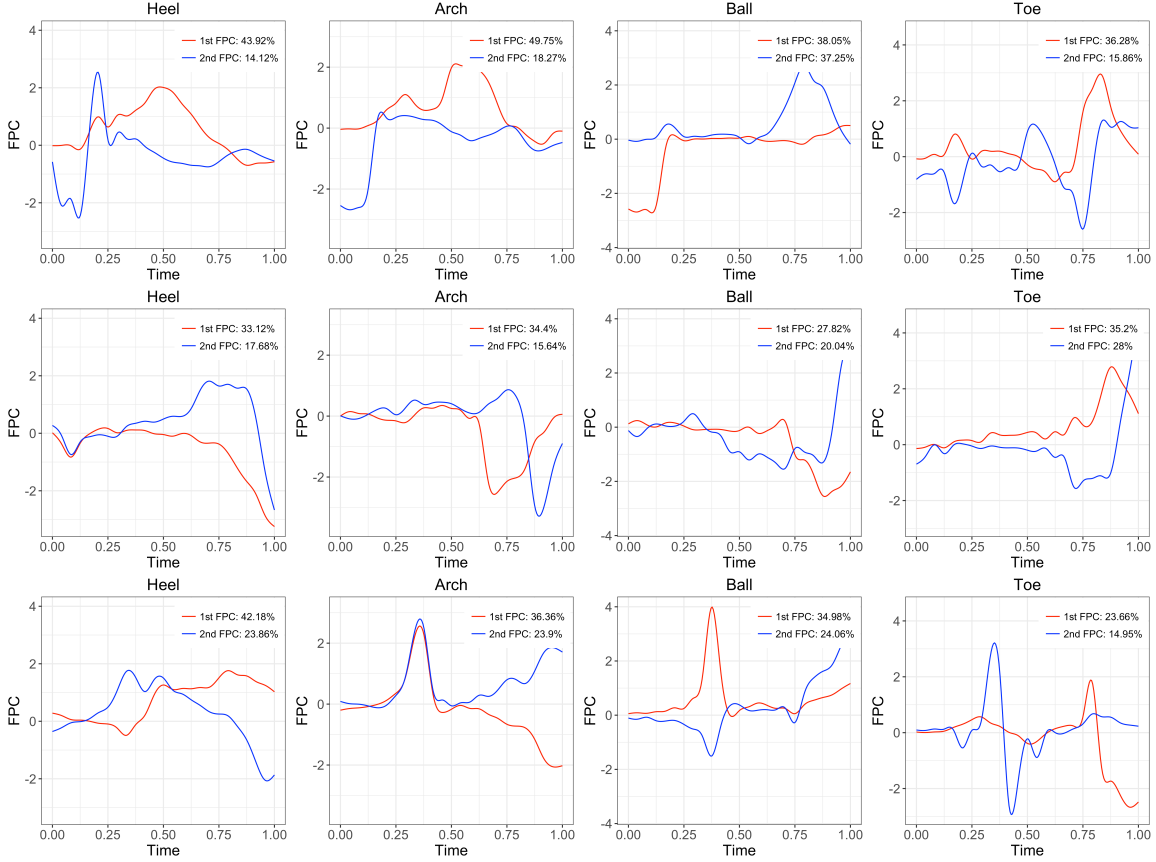


Figure 4.6: **Functional principal components:** First two FPCs of registered curves for three runners are shown in rows; the FPCs in four in-shoe sensors are shown in columns.

more similar within a stance than across stances, perhaps suggesting that the process underlying phase variation depends on the stance itself.

Table 4.2: **Test statistic of pairwise comparison:** For each subject, the upper diagonal elements are the mean correlation within stance across sensors.

	Subject 1			Subject 2			Subject 3		
	Arch	Ball	Toe	Arch	Ball	Toe	Arch	Ball	Toe
Heel	0.617	0.240	0.094	0.268	0.264	0.411	0.378	0.280	0.238
Arch		0.315	0.125		0.262	0.258		0.337	0.300
Ball			0.501			0.239			0.276

Test statistics for each pairwise comparison of all subjects are presented in Table 4.2. The magnitude of the correlation varies across sensors. and, in many cases, the correlation is relatively small. Together with the finding of strong statistical signif-

icance, the low observed correlations suggest that although some phase variability in sensors is similar due to processes underlying the stance, a large proportion of phase variability is dissimilar across sensors in the same stance. Stated differently, although the warping functions obtained through the registration process are similar for different sensors in the same stance, substantial dissimilarities in warping functions remain.

Recall that Subject 1 is a mid-foot striker and the other subjects are heel strikers. The pairwise correlation values in Table 4.2 suggest different correlation patterns depending on the running style. In case of mid-foot striker, the heel sensor shows the biggest correlation with arch (0.617) while the smallest correlation appears between heel and toe (0.094). More generally, adjacent sensors have greater strength of correlation within stance. Heel strikers in contrast have roughly uniform correlation across sensors; adjacency does not matter. It is possible that the mid-foot striker has a smoother transition of forces in different parts of the foot within a stance compared to the heel strikers, which may help explain why mid-foot strikers have been hypothesized to have lower risks of certain musculoskeletal injuries. Analyses of additional runners with varying gait patterns will help clarify this hypothesis.

4.4.4 Association between In-Shoe Sensors and GRF

In this section we use the VGRF curves as a response and fit function-on-function regression model; we present results using the heel sensor as an example for the approach and interpretation of the results. We mainly focused on the relationship of heel sensor with VGRF curve; in many gait studies, the value of VGRF curve is interpreted with its relation to heel movement.

Figure 4.7 shows estimated coefficient surfaces of the function-on-function regression model in (4.2) for 3 subjects. The coefficient surface is interpreted by integrating the product of the predictor and surface at a time t for the $\text{GRF}(t)$. For example,

fixing $t = 0.1$, which is roughly the location of the initial peak, the coefficient $\beta(u, t)$ is the contrast between sensor values in the middle of the u domain and values at the ends of the domain. In particular, for all three subjects, stances with high starting values may have higher initial peaks in the VGRF.

The estimated coefficients are distinctively different across subjects, implying the unique relationship between in-shoe sensors and VGRF curves for each subject. Heel strikers, Subjects 2 and 3, seem to have relatively similar estimated coefficients compared to Subject 1, the mid-foot striker. This finding further suggests to include more subjects in this study, both heel strikers and mid-foot strikers, and examine the relationship between heel sensor data and VGRF curve depending on their gait pattern.

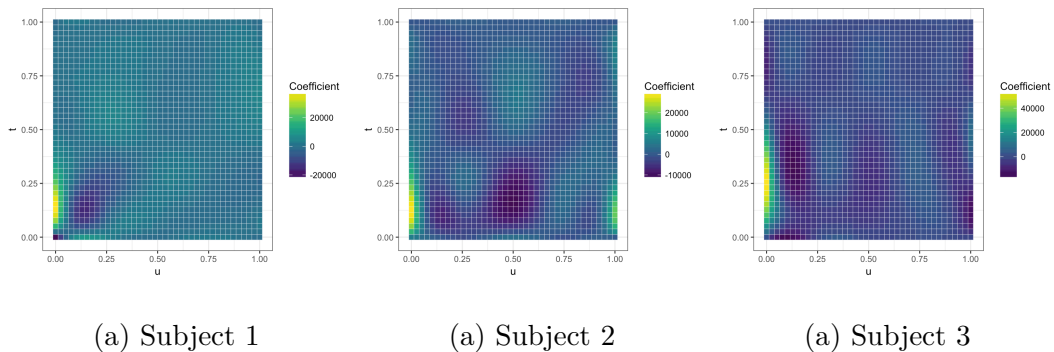


Figure 4.7: **Function-on-function regression coefficient estimates:** Heat map of function-on-function regression coefficient estimates with heel sensor as a predictor. VGRF is measured over t (y -axis) and heel sensor is measured over u (x -axis). Color gradation represents the magnitude of coefficient estimates.

The fitted values from this regression, presented in the top row of Figure 4.8, are fairly similar to the observed VGRF measurements (shown in the second row and fifth column in Figure 4.3); the function-on-function regression model captures the major features of observed curves, including the overall shape and impact transient. This suggests that in-shoe sensor data may be indeed useful for predicting important features or even the whole of VGRF curve. Residuals, shown in the top row of Figure

4.8, indicate that variability in the VGRF is not wholly captured by the function-on-function regression model, especially near the impact transient. Model performance may be improved by including additional predictors, such as data from remaining sensors, or using alternative modeling strategies.

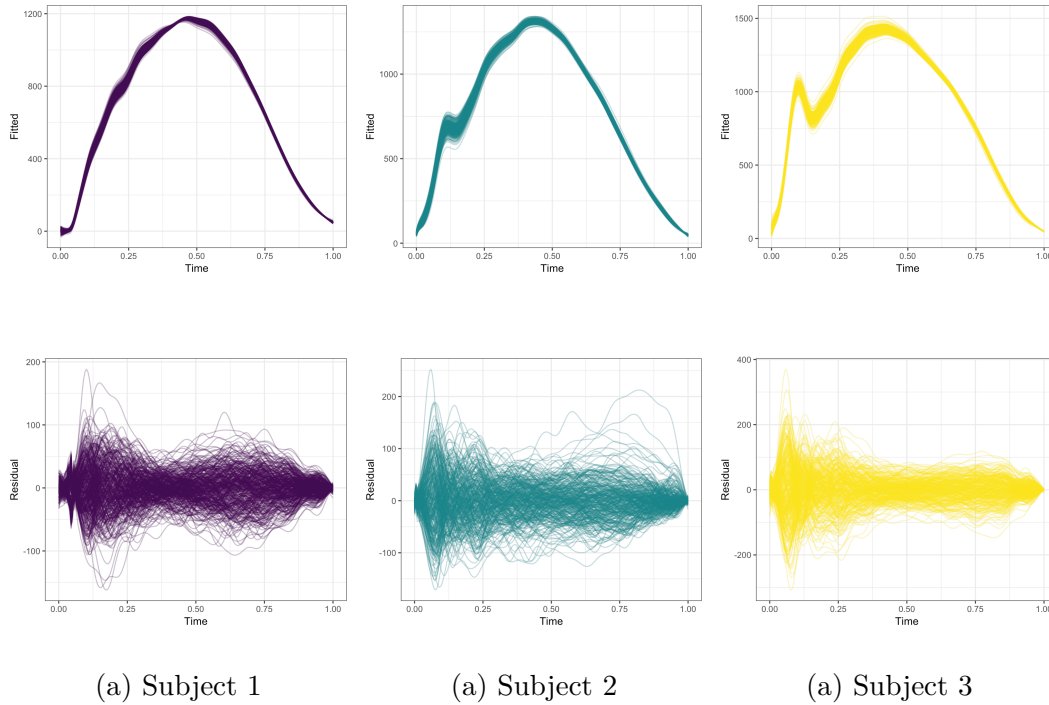


Figure 4.8: **Model fitting of function-on-function regression:** (Top) Fitted values (Bottom) Residuals

4.5 Discussion

In this project, we explored in-shoe sensor data observed from three healthy subjects obtained during an experiment evaluating gait. We examined both phase and amplitude variability in the observed data and illustrated the importance of aligning the observed curves via curve registration. Because the observed phase variability is not expected in stances of healthy individuals, the processing of data using registration methods is an important step in obtaining reliable data from in-shoe sensors; the

results of registration also shed light on the properties of the in-shoe sensors and may help to refine the development of this new technology.

In our analyses, we examined the utility of curve registration by comparing the amplitude variability before and after curve registration, which identified a reduction in amplitude variability after curve registration. We further investigated the similarity in estimated warping functions to understand the sources of phase variability across sensors within stances. Our permutation test results indicate that within each stance, time shifts are related across sensors, but that much of the phase variability across stances is dissimilar within the same sensor. This correlation may support the development of a hierarchical approach to understand the shared and sensor-specific phase variation within a stance.

We used function-on-function regression to begin to evaluate in-shoe sensor data as surrogate of VGRF measurements. Our results provide some evidence for a relationship between key features of in-shoe sensor and VGRF curves. The results are encouraging but it remains unclear, based on our limited case study, whether in-shoe sensor data be used as a replacement for in-lab measurements, and the expansion of this approach to include many subjects and stances is an important future direction.

Although we have focused on healthy subjects, it is also of interest to use in-shoe data to diagnose pathologies. Both phase and amplitude variability can be an evidence of pathological conditions such as movement disorders (Parkinson’s disease and Huntington’s disease; Hausdorff et al. 1998), age effects (Owings and Grabiner 2004), and psychological disorders (major depressive disorder and bipolar disorder; Hausdorff et al. 2004). Including individuals with pathological disorders and investigating their time shifts will emphasize the clinical importance of curve registration and enhance the quality of gait analysis after preprocessing.

Chapter 5

Conclusion

In this dissertation, we emphasized the prevalence of temporal heterogeneity in temporally observed data. Specifically, we focused on two types of temporal heterogeneity; one is the temporal evolution of relational structure in dynamic networks, and the other is from unexpected time shifts among curves observed from a wearable device. For both types of temporal heterogeneity, we pointed out that acknowledging the presence of temporal heterogeneity is an important step of appropriately analyzing the temporal data.

As for dynamic networks, we were interested in modeling the topological fluctuation over time via varying-coefficient models. We proposed two models that directly investigate the temporal heterogeneity in dynamic networks. The first method proposed in Chapter 2 applies the idea of varying-coefficients model to exponential random graph model (ERGM) for dynamic network analysis. The exponential random graph model efficiently summarizes the formation of a network into a combination of network statistics: either network-specific topological features (endogenous) or node-/edge-specific attributes (exogenous). The proposed model, called varying-coefficient exponential random graph model (VCERGM), directly parametrizes the smoothly evolving contribution of these network statistics on the marginal likelihood of a net-

work at each time point. Unlike other competing methods for analyzing temporal networks using ERGMs, the VCERGM is capable of flexibly explaining temporally heterogeneous relational structure. Its model parameter, a smooth function of time, provides meaningful interpretation of temporal heterogeneity. Our simulation studies showed the performance of VCERGM in terms of its efficient and robust estimation. Furthermore, our application studies on both U.S. Congress co-voting networks and resting-state fMRI networks suggested that our proposed method can be utilized to analyze both temporally fluctuating and relatively stable networks over time.

Even though we mainly described the VCERGM with no lag in this dissertation, the VCERGM can be extended to model the temporal dependencies among consecutive networks. Similar with temporal ERGM, we can replace the marginal network statistics into change statistics by considering the difference in topological features in consecutive networks. Although we did not discuss it in this dissertation, the VCERGM can be further explored with respect to its model selection procedure. As multiple networks are observed through time, we can consider a cross-validation approach to conduct model selection in VCERGM. In any type of ERGM models, the endogenous network statistics are intrinsically correlated, and little studies have done regarding variable selection. Its predictive performance for missing edges and goodness-of-fit are also the important factors to be explored in the future.

For dynamic networks with multiple observations at one time point, we proposed a method to incorporate the variability among repeated networks at the same time point. The proposed method combines a stochastic blockmodel (SBM), which determines the structure of forming a connection between two nodes solely depending on their block membership, and the concept of random effects for repeatedly observed networks. The multilevel SBM with a priori block membership presented in Chapter 3 models how the strength of connectivity between blocks changes over time. Fixed effects in the model are smooth functions of time and they are the parameters

explaining the temporal heterogeneity in connectivity between blocks in dynamic networks. While the fixed effects explain overall trend of connectivity evolving over time, random effects are accountable for illustrating the deviation of repeatedly observed networks at one time point from the overall trend.

The multilevel SBM was applied to analyze the U.S. Congress co-voting networks. Both VCERGM and multilevel SBM analyzed these Congressional networks and demonstrated the polarization between political parties in U.S. Congress co-voting patterns over time. In order to apply the VCERGM, we aggregated multiple bills in one Congress and formed a single network per Congress to summarize the co-voting patterns in the Congress. However, multilevel SBM does not require this preprocessing step and analyzes the bill-based networks by incorporating random effects to the model. Therefore, the multilevel SBM additionally provides information regarding the variability in connectivity between and within political parties across bills at each time point.

The multilevel SBM in this dissertation assumes that the node's block membership is pre-specified. Extending the proposed method to analyze dynamic networks with unknown block membership remains a future work. It would require an extra step of determining the number of blocks as well as estimating the block membership of each node at each time point with carefully stated assumptions for model identifiability. If some nodal attributes are known (e.g. political affiliation in U.S. Congress networks), we may consider a hierarchical structure to calibrate the block membership using these attributes. In this dissertation, we only considered random effects on time points. As a future work, we may consider random effects on edges to explain the variability in connectivity specific to each pair of nodes in a network.

One of the main research questions for dynamic network analysis is change-point detection. The smooth parameter in both VCERGM and multilevel SBM can be examined to detect change-points in temporally evolving networks. The parameter

itself can be interpreted to understand the topological fluctuation in dynamic networks, but a formal test procedure to determine statistically significant change-points would be additionally helpful to scrutinize the temporal heterogeneity. Meanwhile, nodal or edge-specific attributes may greatly affect a relational structure in a network. When analyzing the U.S. Congress networks, both VCERGM and multilevel SBM considered political party of Senators to understand their polarization over time, but extending this idea to incorporating more attributes (e.g. age, seniority) in both VCERGM and multilevel SBM remains an important future task.

The other type of temporal heterogeneity arises when multiple curves that share a common structure are presented with unexpected time shifts. Analyzing the observed curves with respect to their common curvature may be inappropriate without removing their time shifts. In Chapter 4, we examined the continuously observed gait measurements using in-shoe sensors to emphasize the importance of removing temporal heterogeneity between curves. We defined that a completion of a step (i.e. from heel-strike to toe-off) forms a stance, and multiple stances were collected from a subject. We viewed these observations as functional data and applied functional data analysis approaches to analyze the data. First, we detected unexpected time shifts across stances and conducted curve registration to remove the temporal variability among curves. After realigning the originally observed curves, we further applied methods in functional data analysis to understand the pattern of amplitude variability among curves. We also investigated the estimated phase variability from curve registration to understand the intrinsic pattern of time shifts between different sensors within a shoe. Our study was a small case study with only three subjects, but the results provided promising insights, especially in terms of relating in-shoe sensor's measurements to ground reaction force, the current gold-standard gait measurement. The results suggested further investigation of in-shoe sensor recordings to examine their utility in gait studies.

Wearable devices have become more accessible and more data are collected using these wearable devices. These wearable devices collect the measurements very densely, and thus it is reasonable to view these observations as functional objects and apply functional data analysis approaches to analyze them. However, as seen in gait measurements using in-shoe sensors, these observations often involve instrumental time shifts. Removing these time shifts via curve registration is an important preprocessing step before appropriately analyzing the data. Many previous gait studies that focused on analyzing a key feature of a stance (e.g. impact transient), instead of examining a whole sequence of the stance. Functional data analysis methods applied in Chapter 4 investigated the whole curve of a stance and helped us to understand one's running pattern more thoroughly. In general, wearable device data can be treated as functional data and examined by using functional data analysis methods.

In summary, this dissertation proposed varying-coefficient framework and functional data analysis methods to recognize the temporal heterogeneity and analyze the densely observed data over time. The proposed methods on dynamic network analysis allow us to directly model the temporal heterogeneity in evolving relational pattern using the varying-coefficient framework. A series of analysis on gait measurements using functional data analysis approaches suggests the importance of implementing proper preprocessing step of removing unexpected time shifts in functional data. Carefully considering temporal fluctuation in repeatedly observed data over time not only gives us meaningful interpretation but also allows us to properly analyze the variability across observations.

Bibliography

- Airoldi, Edoardo M et al. (2008). “Mixed membership stochastic blockmodels.” In: *Journal of Machine Learning Research* 9.Sep, pp. 1981–2014.
- Andris, Clio et al. (2015). “The rise of partisanship and super-cooperators in the US House of Representatives.” In: *PloS one* 10.4, e0123507.
- Benedetti, MG, A Merlo, and A Leardini (2013). “Inter-laboratory consistency of gait analysis measurements.” In: *Gait & posture* 38.4, pp. 934–939.
- Bhamidi, Shankar, Jimmy Jin, and Andrew Nobel (2015). “Change point detection in Network models: Preferential attachment and long range dependence.” In: *arXiv preprint arXiv:1508.02043*.
- Bilodeau, R Adam et al. (2015). “Evolution of nano-junctions in piezoresistive nanos-trand composites.” In: *Composites Part B: Engineering* 72, pp. 45–52.
- Bindu, PV and P Santhi Thilagam (2016). “Mining social networks for anomalies: Methods and challenges.” In: *Journal of Network and Computer Applications* 68, pp. 213–229.
- Brockwell, Peter J and Richard A Davis (2013). *Time series: theory and methods*. Springer Science & Business Media.
- Cai, Zongwu, Jianqing Fan, and Runze Li (2000). “Efficient estimation and inferences for varying-coefficient models.” In: *Journal of the American Statistical Association* 95.451, pp. 888–902.
- Chen, Zhengzhang, William Hendrix, and Nagiza F Samatova (2012). “Community-based anomaly detection in evolutionary networks.” In: *Journal of Intelligent Information Systems* 39.1, pp. 59–85.
- Crane, Elizabeth A et al. (2010). “Effect of registration on cyclical kinematic data.” In: *Journal of biomechanics* 43.12, pp. 2444–2447.

- Cranmer, Skyler J. and A. Desmarais Bruce (2011). “Inferential Network Analysis with Exponential Random Graph Models.” In: *Political Analysis* 19.1, pp. 66–86.
- Daoud, Adam I et al. (2012). “Foot strike and injury rates in endurance runners: a retrospective study.” In: *Med Sci Sports Exerc* 44.7, pp. 1325–1334.
- Daubechies, Ingrid et al. (1992). *Ten lectures on wavelets*. Vol. 61. SIAM.
- De Boor, Carl et al. (1978). *A practical guide to splines*. Vol. 27. Springer-Verlag New York.
- De Brabanter, Joseph et al. (2006). “Generalized likelihood ratio statistics based on bootstrap techniques for autoregressive models.” In: *IFAC Proceedings Volumes* 39.1, pp. 790–795.
- Desmarais, Bruce A and Skyler J Cranmer (2012). “Statistical inference for valued-edge networks: the generalized exponential random graph model.” In: *PloS one* 7.1, e30136.
- Earls, Cecilia and Giles Hooker (2015a). “Adapted Variational Bayes for Functional Data Registration, Smoothing, and Prediction.” In: *arXiv preprint arXiv:1502.00552*.
- (2015b). “Combining functional data registration and factor analysis.” In: *arXiv preprint arXiv:1502.00587*.
- Edwards, Benjamin et al. (2012). “Internet topology over time.” In: *arXiv preprint arXiv:1202.3993*.
- Eilers, Paul HC and Brian D Marx (1996). “Flexible smoothing with B-splines and penalties.” In: *Statistical Science*, pp. 89–102.
- Fan, Jianqing, Chunming Zhang, and Jian Zhang (2001). “Generalized likelihood ratio statistics and Wilks phenomenon.” In: *Annals of Statistics*, pp. 153–193.
- Fan, Jianqing and Wenyang Zhang (2000). “Simultaneous Confidence Bands and Hypothesis Testing in Varying-coefficient Models.” In: *Scandinavian Journal of Statistics* 27.4, pp. 715–731.
- (2008). “Statistical methods with varying coefficient models.” In: *Statistics and its Interface* 1.1, p. 179.
- Fellows, Ian and Mark S Handcock (2012). “Exponential-family Random Network Models.” In: *arXiv preprint arXiv:1208.0121*.

- Fienberg, Stephen E (2012). “A brief history of statistical models for network analysis and open challenges.” In: *Journal of Computational and Graphical Statistics* 21.4, pp. 825–839.
- Fu, Wenjie, Le Song, and Eric P Xing (2009). “Dynamic mixed membership block-model for evolving networks.” In: *Proceedings of the 26th annual international conference on machine learning*. ACM, pp. 329–336.
- Gasser, Theo and Alois Kneip (1995). “Searching for structure in curve samples.” In: *Journal of the American Statistical Association* 90.432, pp. 1179–1188.
- Goldenberg, Anna et al. (2010). “A survey of statistical network models.” In: *Foundations and Trends® in Machine Learning* 2.2, pp. 129–233.
- Goldsmith, Jeff, Sonja Greven, and CIPRIAN Crainiceanu (2013). “Corrected confidence bands for functional data using principal components.” In: *Biometrics* 69.1, pp. 41–51.
- Goldsmith, Jeff et al. (2016). *refund: Regression with Functional Data*. R package version 0.1-16. URL: <https://CRAN.R-project.org/package=refund>.
- Golub, Gene H, Michael Heath, and Grace Wahba (1979). “Generalized cross-validation as a method for choosing a good ridge parameter.” In: *Technometrics* 21.2, pp. 215–223.
- Goodreau, Steven M, James A Kitts, and Martina Morris (2009). “Birds of a feather, or friend of a friend? Using exponential random graph models to investigate adolescent social networks.” In: *Demography* 46.1, pp. 103–125.
- Gopalacharyulu, Peddinti V et al. (2009). “Dynamic network topology changes in functional modules predict responses to oxidative stress in yeast.” In: *Molecular BioSystems* 5.3, pp. 276–287.
- Greene, Derek, Donal Doyle, and Padraig Cunningham (2010). “Tracking the evolution of communities in dynamic social networks.” In: *Advances in social networks analysis and mining (ASONAM), 2010 international conference on*. IEEE, pp. 176–183.
- Guo, Fan et al. (2007). “Recovering temporally rewiring networks: A model-based approach.” In: *Proceedings of the 24th international conference on Machine learning*. ACM, pp. 321–328.
- Hanneke, Steve, Wenjie Fu, Eric P Xing, et al. (2010). “Discrete temporal models of social networks.” In: *Electronic Journal of Statistics* 4, pp. 585–605.

- Hastie, Trevor and Robert Tibshirani (1993). “Varying-coefficient models.” In: *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 757–796.
- Hausdorff, Jeffrey M et al. (1998). “Gait variability and basal ganglia disorders: Stride-to-stride variations of gait cycle timing in Parkinson’s disease and Huntington’s disease.” In: *Movement disorders* 13.3, pp. 428–437.
- Hausdorff, Jeffrey M et al. (2004). “Gait unsteadiness and fall risk in two affective disorders: a preliminary study.” In: *BMC psychiatry* 4.1, p. 39.
- Helwig, Nathaniel E et al. (2011). “Methods to temporally align gait cycle data.” In: *Journal of biomechanics* 44.3, pp. 561–566.
- Ho, Qirong, Le Song, and Eric P Xing (2011). “Evolving cluster mixed-membership blockmodel for time-varying networks.” In:
- Hoff, Peter D, Adrian E Raftery, and Mark S Handcock (2002). “Latent space approaches to social network analysis.” In: *Journal of the American Statistical Association* 97.460, pp. 1090–1098.
- Hoff, Peter D et al. (2015). “Multilinear tensor regression for longitudinal relational data.” In: *The Annals of Applied Statistics* 9.3, pp. 1169–1193.
- Holland, P.W. and S. Leinhardt (1981). “An exponential family of probability distributions for directed graphs.” In: *Journal of the American Statistical Association* 76.373, pp. 33–50.
- Holland, Paul W, Kathryn Blackmond Laskey, and Samuel Leinhardt (1983). “Stochastic blockmodels: First steps.” In: *Social networks* 5.2, pp. 109–137.
- Hoover, Donald R et al. (1998). “Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data.” In: *Biometrika* 85.4, pp. 809–822.
- Huang, Jianhua Z, Colin O Wu, and Lan Zhou (2002). “Varying-coefficient models and basis function approximations for the analysis of repeated measurements.” In: *Biometrika* 89.1, pp. 111–128.
- Hunter, D.R. and M.S. Handcock (2006). “Inference in curved exponential family models for networks.” In: *Journal of Computational and Graphical Statistics* 15.3, pp. 565–583.
- Hunter, David R, Steven M Goodreau, and Mark S Handcock (2008). “Goodness of fit of social network models.” In: *Journal of the American Statistical Association* 103.481, pp. 248–258.

- Hyldahl, Robert D et al. (2016). “Running decreases knee intra-articular cytokine and cartilage oligomeric matrix concentrations: a pilot study.” In: *European journal of applied physiology* 116.11-12, pp. 2305–2314.
- Johnson, Oliver K et al. (2011). “Optimization of nickel nanocomposite for large strain sensing applications.” In: *Sensors and Actuators A: Physical* 166.1, pp. 40–47.
- Karamanidis, Kiros, Adamantios Arampatzis, and Gert-Peter Brüggemann (2004). “Reproducibility of electromyography and ground reaction force during various running techniques.” In: *Gait & posture* 19.2, pp. 115–123.
- Kneip, A et al. (2000). “Curve registration by local regression.” In: *Canadian Journal of Statistics* 28.1, pp. 19–29.
- Kneip, Alois and Theo Gasser (1992). “Statistical tools to analyze data representing a sample of curves.” In: *The Annals of Statistics*, pp. 1266–1305.
- Kneip, Alois and James O Ramsay (2008). “Combining registration and fitting for functional models.” In: *Journal of the American Statistical Association* 103.483, pp. 1155–1165.
- Kolar, Mladen et al. (2010). “Estimating time-varying networks.” In: *The Annals of Applied Statistics*, pp. 94–123.
- Konidaris, George, Sarah Osentoski, and Philip Thomas (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, Pavel N (2012). “Exponential-family random graph models for valued networks.” In: *Electronic Journal of Statistics* 6, p. 1100.
- Krivitsky, Pavel N and Mark S Handcock (2014). “A separable model for dynamic networks.” In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 76.1, pp. 29–46.
- Lee, Jihui, Gen Li, and James D Wilson (2017). “Varying-coefficient models for dynamic networks.” In: *arXiv preprint arXiv:1702.03632*.
- Leger, Jean-Benoist (2016). “Blockmodels: A R-package for estimating in Latent Block Model and Stochastic Block Model, with various probability functions, with or without covariates.” In: *arXiv preprint arXiv:1602.07587*.

- Mariadassou, Mahendra, Stéphane Robin, and Corinne Vacher (2010). “Uncovering latent structure in valued graphs: a variational approach.” In: *The Annals of Applied Statistics*, pp. 715–742.
- Marron, JS et al. (2014). “Statistics of time warpings and phase variations.” In: *Electronic Journal of Statistics* 8.2, pp. 1697–1702.
- Marron, James Stephen et al. (2015). “Functional data analysis of amplitude and phase variation.” In: *Statistical Science* 30.4, pp. 468–484.
- Matias, Catherine and Vincent Miele (2017). “Statistical clustering of temporal networks through a dynamic stochastic block model.” In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 79.4, pp. 1119–1141.
- McLachlan, Geoffrey J (1987). “On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture.” In: *Applied statistics*, pp. 318–324.
- Merrell, A Jake et al. (2013). “Applications of nano-composite piezoelectric foam sensors.” In: *ASME 2013 Conf. on Smart Materials, Adaptive Structures and Intelligent Systems*.
- Moody, James and Peter J Mucha (2013). “Portrait of political party polarization.” In: *Network Science* 1.01, pp. 119–121.
- Owings, Tammy M and Mark D Grabiner (2004). “Variability of step kinematics in young and older adults.” In: *Gait & posture* 20.1, pp. 26–29.
- Peel, Leto and Aaron Clauset (2014). “Detecting change points in the large-scale structure of evolving networks.” In: *arXiv preprint arXiv:1403.0989*.
- (2015). “Detecting Change Points in the Large-Scale Structure of Evolving Networks.” In: *AAAI*, pp. 2914–2920.
- Pensky, Marianna and Teng Zhang (2017). “Spectral clustering in the dynamic stochastic block model.” In: *arXiv preprint arXiv:1705.01204*.
- Peter, Adrian and Anand Rangarajan (2006). “Shape analysis using the Fisher-Rao Riemannian metric: Unifying shape representation and deformation.” In: *Biomedical Imaging: Nano to Macro, 2006. 3rd IEEE International Symposium on*. IEEE, pp. 1164–1167.
- Pietrosimone, Brian et al. (2017). “Biochemical markers of cartilage metabolism are associated with walking biomechanics 6-months following anterior cruciate ligament reconstruction.” In: *Journal of Orthopedic Research*.

- Porter, Michael D and Gentry White (2012). “Self-exciting hurdle models for terrorist activity.” In: *The Annals of Applied Statistics* 6.1, pp. 106–124.
- Qu, Annie and Runze Li (2006). “Quadratic Inference Functions for Varying-Coefficient Models with Longitudinal Data.” In: *Biometrics* 62.2, pp. 379–391.
- Ramsay, JO and Xiaochun Li (1998). “Curve registration.” In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 60.2, pp. 351–363.
- Ramsay, James O (1998). “Estimating smooth monotone functions.” In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 60.2, pp. 365–375.
- (2006). *Functional data analysis*. Wiley Online Library.
- Rice, Hannah, Steve Jamison, and Irene Davis (2016). “Footwear Matters: Influence of Footwear and Foot Strike on Load Rates During Running.” In: 48.12, pp. 2462–2468.
- Robins, Garry et al. (2007). “An introduction to exponential random graph (p^*) models for social networks.” In: *Social Networks* 29.2, pp. 173–191.
- Rosquist, Parker G et al. (2017). “Estimation of 3D Ground Reaction Force Using Nanocomposite Piezo-Responsive Foam Sensors During Walking.” In: *Annals of Biomedical Engineering*, pp. 1–13.
- Sadeghi, Heydar et al. (2000). “Reduction of gait data variability using curve registration.” In: *Gait & posture* 12.3, pp. 257–264.
- Sadeghi, Heydar et al. (2003). “Continuous curve registration as an intertrial gait variability reduction technique.” In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 11.1, pp. 24–30.
- Sarkar, Purnamrita and Andrew W Moore (2005). “Dynamic social network analysis using latent space models.” In: *ACM SIGKDD Explorations Newsletter* 7.2, pp. 31–40.
- Scheipl, Fabian, Ana-Maria Staicu, and Sonja Greven (2015). “Functional additive mixed models.” In: *Journal of Computational and Graphical Statistics* 24.2, pp. 477–501.
- Schmid, Christian S and Bruce A Desmarais (2017). “Exponential Random Graph Models with Big Networks: Maximum Pseudolikelihood Estimation and the Parametric Bootstrap.” In: *arXiv preprint arXiv:1708.02598*.

- Seeley, Matthew K et al. (2017). “Walking mechanics for patellofemoral pain subjects with similar self-reported pain levels can differ based upon neuromuscular activation.” In: *Gait & posture* 53, pp. 48–54.
- Seliktar, R, M Yekutieli, and A Bar (1979). “Gait consistency test based on the impulse-momentum theorem.” In: *Prosthetics and Orthotics International* 3.2, pp. 91–98.
- Sewell, Daniel K and Yuguo Chen (2015). “Latent space models for dynamic networks.” In: *Journal of the American Statistical Association* 110.512, pp. 1646–1657.
- Shetty, Jitesh and Jafar Adibi (2005). “Discovering important nodes through graph entropy the case of enron email database.” In: *Proceedings of the 3rd international workshop on Link discovery*. ACM, pp. 74–81.
- Silverman, Bernard W (1995). “Incorporating parametric effects into functional principal components analysis.” In: *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 673–689.
- Simpson, Sean L, Satoru Hayasaka, and Paul J Laurienti (2011). “Exponential random graph modeling for complex brain networks.” In: *PloS one* 6.5, e20039.
- Simpson, Sean L, Malaak N Moussa, and Paul J Laurienti (2012). “An exponential random graph modeling approach to creating group-based representative whole-brain connectivity networks.” In: *Neuroimage* 60.2, pp. 1117–1126.
- Snijders, Tom AB (2001). “The statistical evaluation of social network dynamics.” In: *Sociological methodology* 31.1, pp. 361–395.
- Sparks, Ross and James D Wilson (2016). “Monitoring communication outbreaks among an unknown team of actors in dynamic networks.” In: *arXiv preprint arXiv:1606.09308*.
- Srivastava, Anuj et al. (2011a). “Registration of functional data using fisher-rao metric.” In: *arXiv preprint arXiv:1103.3817*.
- Srivastava, Anuj et al. (2011b). “Shape analysis of elastic curves in euclidean spaces.” In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33.7, pp. 1415–1428.
- Staicu, Ana-Maria et al. (2014). “Likelihood ratio tests for dependent data with applications to longitudinal and functional data analysis.” In: *Scandinavian Journal of Statistics* 41.4, pp. 932–949.

- Strauss, D. and M. Ikeda (1990). “Pseudolikelihood estimation for social networks.” In: *Journal of the American Statistical Association* 85.409, pp. 204–212.
- Sweet, Tracy M (2015). “Incorporating covariates into stochastic blockmodels.” In: *Journal of Educational and Behavioral Statistics* 40.6, pp. 635–664.
- Székely, Eszter et al. (2016). “Childhood peer network characteristics: Genetic influences and links with early mental health trajectories.” In: *Journal of Child Psychology and Psychiatry and Allied Disciplines* 57.6, pp. 687–694.
- Tekle, Fetene B, Dereje W Gudicha, and Jeroen K Vermunt (2016). “Power analysis for the bootstrap likelihood ratio test for the number of classes in latent class models.” In: *Advances in Data Analysis and Classification*, pp. 1–16.
- Teng, Hsiang-Ling et al. (2017). “Gait Characteristics Associated With a Greater Increase in Medial Knee Cartilage T1 ρ and T2 Relaxation Times in Patients Undergoing Anterior Cruciate Ligament Reconstruction.” In: *American Journal of Sports Medicine*, p. 0363546517723007.
- Thies, Sibylle B et al. (2009). “Movement variability in stroke patients and controls performing two upper limb functional tasks: a new assessment methodology.” In: *Journal of neuroengineering and rehabilitation* 6.1, p. 2.
- Tucker, J Derek, Wei Wu, and Anuj Srivastava (2013). “Generative models for functional data using phase and amplitude separation.” In: *Computational Statistics & Data Analysis* 61, pp. 50–66.
- Tucker, J Derek, Wei Wu, Anuj Srivastava, et al. (2014). “Analysis of proteomics data: Phase amplitude separation using an extended Fisher-Rao metric.” In: *Electronic Journal of Statistics* 8.2, pp. 1724–1733.
- Van Duijn, Marijtje AJ, Krista J Gile, and Mark S Handcock (2009). “A framework for the comparison of maximum pseudo-likelihood and maximum likelihood estimation of exponential family random graph models.” In: *Social Networks* 31.1, pp. 52–62.
- Van Essen, David C et al. (2012). “The Human Connectome Project: a data acquisition perspective.” In: *Neuroimage* 62.4, pp. 2222–2231.
- Wang, Yuchung J and George Y Wong (1987). “Stochastic blockmodels for directed graphs.” In: *Journal of the American Statistical Association* 82.397, pp. 8–19.
- Wasserman, Stanley and Philippa Pattison (1996). “Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p^* .” In: *Psychometrika* 61.3, pp. 401–425.

- West, Mike, P Jeff Harrison, and Helio S Migon (1985). “Dynamic generalized linear models and Bayesian forecasting.” In: *Journal of the American Statistical Association* 80.389, pp. 73–83.
- Wilson, James D, Nathaniel T Stevens, and William H Woodall (2016). “Modeling and estimating change in temporal networks via a dynamic degree corrected stochastic block model.” In: *arXiv preprint arXiv:1605.04049*.
- Wilson, James D et al. (2017). “Stochastic weighted graphs: Flexible model specification and simulation.” In: *Social Networks* 49, pp. 37–47.
- Woodall, William H et al. (2017). “An overview and perspective on social network monitoring.” In: *IISE Transactions* 49.3, pp. 354–365.
- Wu, Wei, Anuj Srivastava, et al. (2014). “Analysis of spike train data: Alignment and comparisons using the extended Fisher-Rao metric.” In: *Electronic Journal of Statistics* 8.2, pp. 1776–1785.
- Xing, Eric P, Wenjie Fu, and Le Song (2010). “A state-space mixed membership blockmodel for dynamic network tomography.” In: *The Annals of Applied Statistics*, pp. 535–566.
- Xu, Kevin S and Alfred O Hero III (2013). “Dynamic Stochastic Blockmodels: Statistical Models for Time-Evolving Networks.” In: *SBP*. Springer, pp. 201–210.
- Xu, Kevin S and Alfred O Hero (2014). “Dynamic stochastic blockmodels for time-evolving social networks.” In: *IEEE Journal of Selected Topics in Signal Processing* 8.4, pp. 552–562.
- Yang, Tianbao et al. (2011). “Detecting communities and their evolutions in dynamic social networks: a Bayesian approach.” In: *Machine learning* 82.2, pp. 157–189.
- Yao, Fang, Hans-Georg Müller, and Jane-Ling Wang (2005). “Functional data analysis for sparse longitudinal data.” In: *Journal of the American Statistical Association* 100.470, pp. 577–590.
- Zhang, Xiaoke, Jane-Ling Wang, et al. (2015). “Varying-coefficient additive models for functional data.” In: *Biometrika* 102.1, pp. 15–32.

Appendix to Varying-Coefficient Exponential Random Graph Model

A.1 Stochastic Equivalence under the Difference Statistic Specification

Comparing the first-order TERGM with model (2.1), 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 $\mathbf{Z} = \{Z_1, \dots, Z_T\}$ 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

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

where $\mathbf{h} : \{0, 1\}^{n \times n} \rightarrow \mathbb{R}^p$ 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 $X_t \in \mathbf{X}$ as an independent realization from the *same* exponential family probability mass function. This is made precise in the next proposition.

Proposition 1 *Let \mathcal{X} denote the family of unweighted dynamic graph sequences on n vertices with $T \geq 1$ discrete observations. Suppose that $\mathbf{X} = \{X_1, \dots, X_T\} \in \mathcal{X}$ is generated under the TERGM in (2.3), where for $t = 2, \dots, T$*

$$X_t \mid \mathbf{X}_t^- \sim \mathbb{P}(X_t = x_t \mid X_{t-1} = x_{t-1}; \phi) = \frac{\exp\{\phi^T \mathbf{g}(x_t, x_{t-1})\}}{\sum_{z \in \{0,1\}^{n \times n}} \exp\{\phi^T \mathbf{g}(z, x_{t-1})\}}.$$

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 \mathbf{X}_t^- and can be generated as an independent realization of an ERGM with the following probability mass function

$$X_t \mid \mathbf{X}_t^- \sim \mathbb{P}(X_t = x \mid \phi) = \frac{\exp\{\phi^T \mathbf{h}(x)\}}{\sum_{z \in \{0,1\}^{n \times n}} \exp\{\phi^T \mathbf{h}(z)\}}$$

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.

A.2 Iterative Reweighted Least Squares (IRLS)

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

$$\tilde{\mathbf{x}}^T \mathbf{H}^T \text{vec}(\Phi) - \mathbf{1}^T \log[1 + \exp\{\mathbf{H}^T \text{vec}(\Phi)\}] - \lambda \mathcal{P}_\Omega(\Phi). \quad (\text{A.2})$$

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

$$\text{vec}(\Phi^{(u+1)}) = (\mathbf{H}^T \mathbf{W}^{(u)} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{W}^{(u)} \left\{ \mathbf{H} \text{vec}(\Phi^{(u)}) + (\tilde{\mathbf{x}} - \boldsymbol{\mu}^{(u)}) \cdot g'(\boldsymbol{\mu}^{(u)}) \right\}, \quad (\text{A.3})$$

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

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

With the penalty term $\mathcal{P}(\Phi)$, we only need to replace $\mathbf{H}^T \mathbf{W}^{(u)} \mathbf{H}$ by $\mathbf{H}^T \mathbf{W}^{(u)} \mathbf{H} + \lambda(\Omega \otimes \mathbf{I}_p)$ in (A.3). The generalized cross validation (GCV) is used to choose the tuning parameter λ (Golub, Heath, and Wahba 1979). Namely, the λ is a minimizer of $G(\lambda)$, which is defined as

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

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

A.3 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 A.1: **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 A.2: **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)

A.4 Estimation for Networks with Time-Varying Network Size

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).

A.5 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 $\phi(t)$ from temporal

Table A.3: **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)

networks with $K = 50$. There is no significant difference in estimation performance when the number of basis functions is greater than 10.

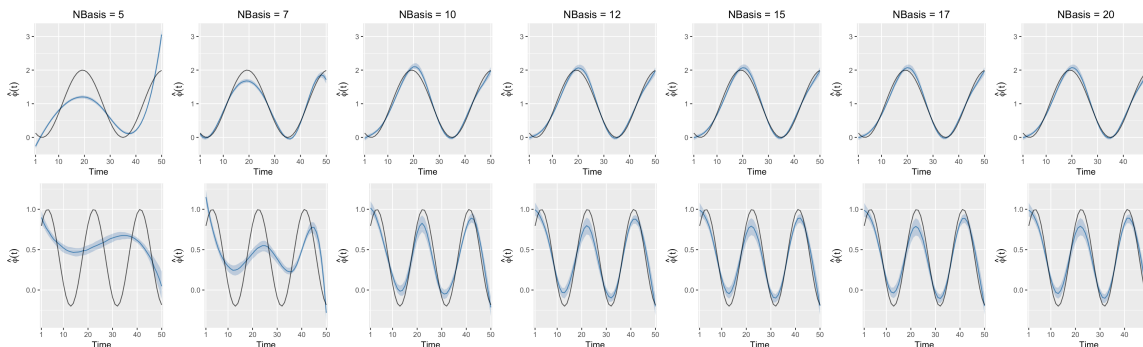


Figure A.1: **Parameter estimates with 30 nodes:** Estimated parameters for edges (top) and reciprocity (bottom). Black line is the true $\phi(t)$. The blue line indicates the average of 100 estimated curves and the shaded band illustrates the first and third quantiles.