Latent position network models with applications in time series analysis

Hardeep Kaur

19200482

The thesis is submitted to University College Dublin in fulfilment of the requirements for the degree of
Doctor of Philosophy

School of Mathematics and Statistics

Head of School: Prof. Nial Friel

Primary Supervisor: Dr. Riccardo Rastelli

Co-Supervisor: Prof. Nial Friel

Doctoral Studies Panel: Dr Michael Fop
Dr Michelle Carey
Prof. Nial Friel

November 2024

Table of contents

Abstract

Declaration

Sponsor

Collaborations

Acknowledgements

Publications

1 Introduction **1**

Introduction **1**

1.1 Theoretical background of time series models 1
1.2 Overview of network models 3
1.3 Latent position network models 4
1.4 Network embedded count time series analysis: foundational framework and research motivation 5
1.5 Foundational methods for modeling time series using network approaches . . 6
1.6 Expanding the horizon: time series analysis with dynamic network structures 9
1.7 Overview of Stan and inferential procedures 11
1.8 Other inferential procedures 13
1.9 Motivational datasets 14
1.10 Thesis overview and chapter summaries 15

2 Latent Position Network Models **24**

2.1 Introduction 24
2.2 Notation 25
2.3 Latent Position Models 26

2.3.1 The original distance model and projection model . . . 26
2.3.2 Latent distance models with clustering . . . 27
2.3.3 Latent distance models with node-specific random effects . . . 28
2.4 Inference for the LPM . . . 29
2.4.1 Estimation . . . 29
2.4.2 Interpretation of the posterior samples . . . 30
2.4.3 Other inferential approaches . . . 32
2.5 An illustrative example for the distance LPM . . . 33
2.6 State of the art . . . 37
2.6.1 LPMs for multiview networks . . . 37
2.6.2 LPMs for dynamic networks . . . 38
2.6.3 LPMs for weighted networks . . . 39
2.6.4 LPMs for rank data . . . 40
2.6.5 Theory of LPMs . . . 40
2.6.6 Extensions of the projection LPM . . . 41
2.6.7 Extensions of the LPCM . . . 41
2.7 Software . . . 42
2.8 Open research questions . . . 43
2.9 Acknowledgements . . . 44

3 A latent space model for multivariate count data time series analysis 51
3.1 Introduction . . . 52
3.1.1 Literature review . . . 53
3.2 Time Series Latent Position Model (TSLPM) . . . 57
3.2.1 Poisson log-linear models . . . 57
3.2.2 Latent position models . . . 58
3.2.3 The TSLPM . . . 60
3.3 Inferential procedure . . . 65
3.3.1 Estimation . . . 65
3.3.2 Procrustes analysis of nodal positions . . . 67
3.4 Simulation study . . . 69
3.4.1 Estimation performance using LBFGS . . . 70
3.4.2 Forecasting . . . 73
3.5 Chicago crime data . . . 78
3.5.1 Methodology . . . 78
3.5.2 TSLPM for the crime dataset . . . 80
3.5.3 Model selection . . . 82
3.5.4 Crime data 5 clusters . . . 83

3.5.5 Chicago's eleven-regional network data . . . 87
3.6 Conclusion and discussion . . . 92
3.7 Acknowledgement . . . 94
3.8 Appendix . . . 95
3.8.1 Supplementary tables . . . 95
3.8.2 TSLPM(p) . . . 95
3.8.3 Mean count for the time series . . . 96
3.8.4 Goodness of fit . . . 96

4 A dynamic latent space time series model to assess the spread of mumps in England 107
4.1 Abstract . . . 107
4.2 Introduction . . . 108
4.2.1 Related Work . . . 110
4.3 Dynamic Time series Latent Position model (DTSLPM) . . . 113
4.3.1 Data . . . 113
4.3.2 Model Specification . . . 113
4.3.3 Stability of network time series process . . . 116
4.3.4 Prior specification . . . 116
4.4 Model Fitting . . . 117
4.4.1 Estimation . . . 117
4.4.2 Post-processing of posterior samples . . . 117
4.5 Simulation study . . . 118
4.5.1 Static time series distance model . . . 118
4.5.2 Large scale simulation study using optimization . . . 119
4.5.3 Dynamic time series distance model . . . 122
4.6 Analysis of mumps cases . . . 130
4.6.1 Static model for the mumps cases . . . 132
4.6.2 Results for dynamic time series distance model . . . 134
4.6.3 Risk of contagion . . . 140
4.7 Goodness of model fit . . . 141
4.8 Conclusions . . . 143

5 Conclusion 148

Conclusion 148

Abstract

In this dissertation, we develop a novel framework for modeling multivariate count time series with a network structure, focusing on the use of latent position models (LPM) to uncover hidden relationships among time series. Traditional approaches to time series analysis within network frameworks have primarily concentrated on continuous data, leaving a gap in the literature for count data, particularly when the network structure evolves over time. This work addresses this gap by proposing statistical models that reveal completely unknown latent network structures underlying count time series data, thus enabling a deeper understanding of the complex connection patterns between time series. The proposed models provide a comprehensive framework that captures both static and dynamic interdependencies, offering theoretical advancements as well as practical tools for visualizing and interpreting complex relationships in time series data.

The core objective of this research is to develop methodologies for estimating embedded network structures within time series data informed by latent factors. To illustrate the effectiveness of these approaches, we apply them to two datasets: burglary counts in Chicago and mumps cases in England, contributing to fields such as epidemiology and criminology. The findings demonstrate the models' ability to reveal relationships between burglary occurrences across regions in Chicago and the fixed or temporal patterns in the spread of mumps infection across regions in England, uncovering connections that are not readily observable using conventional time series analysis techniques.

We demonstrate network inference from the proposed modeling frameworks using optimization methods, such as the L-BFGS algorithm, and sampling methods such as the Hamiltonian Monte Carlo method. These approaches efficiently estimate model parameters, such as the latent positions of nodes within the network. A collection of simulation studies are presented to evaluate the performance of the proposed models across various settings. These studies demonstrate the models' capacity to produce consistent parameter estimates and to uncover underlying network structures in both static and dynamic contexts. The simulation results further support the model's applicability to real-world data, illustrating how they can uncover hidden interdependencies in network-embedded count time series data.

Statement of Original Work

"This thesis is submitted to the University College Dublin in support of my application for the degree of Doctor of Philosophy. I confirm that this work is entirely my own and was completed while I was registered as a candidate for the degree mentioned on the Title Page. I also confirm that I have not received any other degree based on the research presented in this thesis."

Hardeep Kaur
November 2024

Sponsor

This publication has emanated from research supported in part by a grant from the Insight Centre for Data Analytics which is supported by Science Foundation Ireland under grant number $12/RC/2289\_P2$ and supported by the School of Mathematics and Statistics, University College Dublin, Ireland.

Collaborations

Riccardo Rastelli	As my principal supervisor, Dr. Riccardo Rastelli provided supervision throughout the projects and assisted with the editing of Chapter 2, Chapter 3, and Chapter 4.
Nial Friel	Prof. Nial Friel collaborated on the work in Chapter 2 by reviewing and editing the final version of the manuscript.
Adrian E. Raftery	Prof. Adrian E. Raftery contributed to the work in Chapter 2 by reviewing and editing the final version of the manuscript.

Acknowledgements

My PhD path has profoundly impacted me, influencing my intellectual and personal development in several ways. Over the past four years, I have seen incredible growth in myself, from gaining a deeper understanding of complex ideas to becoming more confident in presenting my work. This journey has taught me invaluable skills like organization, time management, and resilience, along with many others. Nevertheless, these represent but a fraction of the knowledge I have acquired over my journey. I am incredibly thankful for this opportunity to grow and learn in ways that far exceeded my expectations. None of these achievements or transformations would have been possible without the support and guidance of some wonderful people, to whom I would like to express my heartfelt gratitude.

The first person who, without a doubt, deserves the greatest credit for making this PhD journey a successful and fulfilling experience is my supervisor, Dr. Riccardo Rastelli. Riccardo, I am immensely grateful to you for accepting me as a PhD student and believing in my potential from the very beginning. Your mentorship, guidance, and invaluable feedback in our weekly meetings have been instrumental in every step of this journey.

I am especially thankful for your unwavering commitment and diligent efforts, which rendered this dissertation achievable. Your help in grasping complex statistical concepts, as well as the one-on-one board sessions whenever I encountered challenges, greatly enhanced my understanding. I also appreciate how you improved my coding skills by introducing me to good coding practices, which have become essential to my work. Our many discussions on effective presentation skills - emphasizing the importance of “telling a story” through slides rather than memorizing content - have also profoundly impacted my approach to public speaking.

Additionally, your insights on crafting reports and plots to academic standards, along with your meticulous feedback on refining my writing to be clear, precise, and well-structured, have transformed my approach to research. These experiences have brought about significant changes in my thought process and working style, which I will carry forward into future projects. For all these lessons and more, I am truly grateful.

I am extremely grateful to Prof. Nial Friel for his unwavering support throughout my PhD journey. Your guidance and feedback on the projects, coupled with your administrative support and understanding, have been invaluable during various phases of this academic path. I

would also like to extend my heartfelt thanks to my Research Studies Panel members, Dr. Michael Fop and Dr. Michelle Carey, for their constructive feedback and insightful suggestions. Your input has been instrumental in shaping this thesis and significantly enhancing the quality of the work. I truly appreciate the time and effort you invested in guiding me, and your contributions have played a vital role in bringing this dissertation to fruition. I would also like to express my gratitude to Dr. Adrian E. Raftery for collaborating with me, providing constant feedback, and taking the time to review the final work on our collaborative project. Your insights and commitment have greatly contributed to the quality of this research.

I gratefully acknowledge the support of the Insight Centre for Data Analytics, Science Foundation Ireland, and University College Dublin throughout my PhD. Special thanks to the administrative staff at the Insight Centre for organizing valuable development sessions for PhD students and to the staff of the School of Mathematics and Statistics at UCD for their consistent assistance.

I would like to thank a very special person, my husband, Gurpreet. You were the one who encouraged me to pursue a PhD, and I am so grateful for your belief in me. This journey has been a roller coaster, and I couldn't have done it without you by my side. Thank you for being my constant source of support and strength. Your warm hugs during tough times, your reminders that I am hardworking and capable, and your constant encouragement helped me through moments of doubt. You lifted me up during difficult times, encouraged me to give my best and then let go, and made the joyful moments even brighter with your warmth and laughter. Thank you for reminding me to take care of my health, even if I didn't always follow through. This journey wouldn't have been nearly as enjoyable or meaningful without you by my side.

I would like to extend my heartfelt thanks to my wonderful family. To my parents, who never stopped cheering me on and constantly prayed for my success - your love and support mean everything to me. To my little sister, thank you for reminding me to enjoy life alongside my studies and encouraging me to maintain a balanced routine. To my brother, thank you for your cheerful support and encouragement throughout this journey. I am also deeply grateful to my parents-in-law, who offered their blessings and motivated me to earn the Dr. title, just as my husband did. Your collective encouragement and belief in me have been invaluable.

I would also like to thank my Ireland family—Prabhleen, Akhilesh, and Ankur. Thank you for making this journey fun-filled and joyful, celebrating every small moment as if it were a grand occasion. From get-togethers and camping trips to trekking adventures, long drives, and more, you made my time here unforgettable. Thank you for being such a special part of this journey.

Publications

The material in Chapter 2 has been peer-reviewed and published as a book chapter in *The Sage Handbook of Social Network Analysis, 2e*. The material in Chapter 3 has been submitted to the *Journal of the Royal Statistical Society, Series A (Statistics in Society)* and is currently under review. The material in Chapter 4 has been submitted to the *Journal of the Royal Statistical Society, Series C (Applied Statistics)* and is currently under review.

Peer-reviewed book chapter:

- Kaur, H., Rastelli, R., and Friel, N. (2023). “Latent Position Network Models”. In: *The Sage Handbook of Social Network Analysis*, edited by J. McLevey, P. Carrington and J. Scott, SAGE Publications. Chap. 36, pp. 526–541. ISBN: 9781529614671. https://doi.org/10.4135/9781529614695.n36

Submitted article (under review):

- Kaur, H., Rastelli, R. (2024). "A latent space model for multivariate count data time series analysis". Pre-print currently available on arXiv.

Submitted article (under review):

- Kaur, H., Rastelli, R. (2024). "A dynamic latent space time series model to assess the spread of mumps in England". Pre-print currently available on arXiv.

Chapter 1

Introduction

1.1 Theoretical background of time series models

In recent years, the collection and analysis of time series data sets has grown substantially. Time series data often exhibit similarities between past and present values, leading to autocorrelation within the data. Researchers have made notable advancements in time series analysis, developing various modeling frameworks. One widely used method for modeling univariate time series is the AutoRegressive (AR) model. An AR model is a simple model that linearly regresses the current series based on one or more past values of the same series. An AR(p) model takes the form:

$$y_t = \alpha + \gamma_1 y_{t-1} + \gamma_2 y_{t-2} + \cdots + \gamma_p y_{t-p} + \epsilon_t$$

where, y_t is a time series, ϵ_t is a white noise and γ_i denotes the strength of the dependence of future values on past values. This model is commonly used for forecasting a single variable of interest (Box et al., 2015).

However, many real-world scenarios involve multiple interacting variables, making multivariate time series analysis crucial for capturing the relationships among different entities and generating accurate forecasts which univariate models like AR(p) do not account for. To address this limitation, the Vector AutoRegressive (VAR) model (Luetkepohl, 2005) has been widely adopted for multivariate time series involving continuous data as it allows for the modeling of interactions between multiple time series. A VAR(1) model is defined as:

$$y_{it} = \alpha + \sum_{j=1}^{N} \gamma_{ij} y_{j(t-1)} + \epsilon_{it}$$

where, ϵ_{it} white noise processes that may be contemporaneously correlated. The coefficients γ_{ij}s are the (i, j)th entries of a Γ matrix, termed as coefficient matrix with autoregressive

coefficient, in the vector form of the VAR model. The coefficient γ_{ii} captures the influence of the previous value of variable y_i on itself, while the coefficient γ_{ij} captures the influence of the previous value of variable y_j on y_i. For reliable forecasting, it is essential that the time series be stationary, meaning that its statistical properties, like mean and variance, remain constant over time. Stationary series are easier to model because they have a consistent structure, which allows us to make predictions based on past patterns. In contrast, non-stationary series can be challenging to model effectively, as their behavior may vary unpredictably over time, reducing the reliability of forecasts.

In a univariate time series, a specific type of non-stationarity occurs when the absolute value of the autoregressive coefficient, γ_i exceeds 1. This situation, often called explosive non-stationarity, causes the series to grow exponentially over time. This exponential growth makes the series unpredictable and can lead to inaccurate or even nonsensical forecasts. In multivariate cases, stationarity requires that the absolute values of the eigenvalues of the coefficient matrix Γ remain less than 1. However, as the number of series increases, the VAR faces significant limitations, as it becomes increasingly challenging to avoid over-parametrization, making the model computationally demanding.

While VAR models are effective for continuous data, many interesting empirical questions involve the analysis of count data, where multivariate time series modeling is less developed. According to the findings of Fokianos et al. (2020) and Fokianos (2024), there are currently two widely recognized methodologies for inferential procedures in count time series models: parameter-driven and observation-driven processes. Our focus is specifically on observation-driven models due to their appeal in treating current parameters as deterministic functions of lagged dependent variables, along with contemporaneous and lagged exogenous variables. This feature makes them particularly useful for capturing time-dependent dynamics in count data. Within the field of observation-driven models, linear and log-linear models are common and straightforward generalizations of Poisson autoregressive models for count data. A Poisson log-linear model of order 1 introduced by Fokianos et al. (2020), takes the form:

$$\mathbf{y}_{it} \sim \text{Poisson}(\boldsymbol{\lambda}_{it}) \quad \log(\boldsymbol{\lambda}_{it}) = \alpha + \sum_{j=1}^{N} \gamma_{ij} \log(y_{j(t-1)} + 1)$$

The Poisson linear model, a variant of the above model takes a similar form but excludes the log function. Heinen and Erick (2007), Andreassen (2013), and Lee et al. (2018) primarily focused on the study of linear models, while Fokianos and Tjøstheim (2011) discussed the notable advantages of log-linear models.

The selection of an appropriate time series model depends on several key factors, including the type of data, the specific application, and the properties required from the model. For instance, continuous data often call for models like VAR or Vector AutoRegressive Moving Average (VARMA), which are suited to capturing linear dependencies, while count data benefit

from models like the Poisson or log-linear model, which can handle discrete, non-negative values. Additionally, the choice of model should align with the underlying structure and goals of the analysis—whether it's forecasting or uncovering structural dependencies. Thus, a careful evaluation of these aspects is essential to ensure that the chosen model effectively captures the dynamics of the data.

While various time series models exist for continuous and count data, our research specifically addresses the gap in the literature surrounding count time series models, particularly those with network structures. In Chapters 3 and 4 of this thesis, we delve into the modeling of count time series data. One of the primary motivations for this focus is the relative scarcity of research on count time series models, which presents a valuable opportunity to contribute to this evolving area. Building on the proven utility of log-linear models in existing research, we aim to extend this framework. Specifically, our objective is to adapt the log-linear model to capture both the temporal dynamics of count data and the underlying network structures, enabling a deeper analysis of network-based interactions within time series data.

1.2 Overview of network models

The study of statistical networks, encompassing areas such as computer, biological, and social networks, is a multidisciplinary field drawing from social sciences, biology, finance, physics, and other disciplines. This area has recently gained prominence due to its wide-ranging applications. In particular, social networks are used to study the relationships and interactions between actors or entities, referred to as nodes, with pairwise interactions between them represented by edges. The formulation of statistical network models is critical for quantifying network structures, creating interpretable representations of the underlying connections, understanding their formation and growth, and predicting the behavior of networked systems.

One of the foundational models in this area is the random graph model (Erdös and Rényi, 1959), where edges are randomly assigned between nodes. This model provides a basic framework for understanding network connectivity but assumes independent edges and equal probabilities of connectivity between node pairs, making it unsuitable for modeling many real-world networks. An extension of this is the configuration model (Newman, 2018), which better replicates real-world networks with heterogeneous degree distributions. However, it still assigns edges between nodes randomly, failing to capture the structured and non-random relationships seen in most networks.

More advanced models, such as the preferential attachment model (Newman, 2018), simulate real-world phenomena where certain nodes attract more connections over time, as observed in social and biological networks. Despite its utility, not all real-world networks exhibit the stark hub-and-spoke structure predicted by this model, where a small number of nodes have extremely high degrees while most have very few connections.

The Exponential Random Graph Model (ERGM) addresses some of these limitations by allowing dependencies between network edges. Instead of assuming independence, ERGMs (Robins et al., 2007) model the probability of a network based on observed structures like triangles or k-stars, making them particularly effective for capturing clustering and transitivity. It is by far the most studied network model, especially within the statistics literature. However, they can face challenges such as model degeneracy in complex networks. Another important model is the Stochastic Block Model (SBM), which uncovers community structure by partitioning nodes into groups or blocks (Holland et al., 1983; Holland and Leinhardt, 1981; Wang and Wong, 1987). Nodes within the same group tend to exhibit similar behaviors, and interactions are modeled probabilistically based on group memberships, making SBMs ideal for community detection. Finally, the Latent Position Model (LPM) proposed by Hoff et al. (2002), posits that each node occupies a position in an unobserved latent space, with the probability of an edge between two nodes depending on their distance in this space. LPMs are particularly effective in capturing transitivity and reciprocity in networks, offering a geometric interpretation of relationships.

These models are essential for a wide range of applications, offering valuable insights into social behaviors, biological interactions, disease transmission, and financial systems. Their ability to capture the complexity of real-world networks makes them indispensable in both theoretical research and applied studies. The choice of model, however, is guided by the specific research problem at hand, as different models are tailored to address varying network structures and dynamic phenomena. This flexibility allows researchers to select the most appropriate model to best represent the underlying relationships in their data.

For our research, the LPM is particularly well-suited for analyzing network connections, as it provides a clear geometric representation of relationships within a latent space. This feature is crucial for uncovering hidden structures and patterns in network time series data, especially when dealing with dynamic interactions. In the following section, we will explore this model in greater detail and examine its significance to our proposed framework.

1.3 Latent position network models

Hoff et al. (2002) introduced the LPMs, also known as the Latent Space Model (LSM), to the study of social networks. This development led to a remarkable surge in research on the topic, establishing LPMs as a widely adopted and influential statistical tool for network analysis. The LSM was introduced as a way to model the underlying relationships in network data by positioning each node in a d-dimensional Euclidean space. The likelihood of an edge existing between two nodes is modeled as a function of the distance between their respective positions, with closer nodes having a higher probability of forming an edge. This approach

allows for a simple, easily interpretable representation of network structures, particularly in capturing complex characteristics such as reciprocity and transitivity in the data.

Two primary types of latent space models are the distance model and the projection model. The distance model is widely used due to its simplicity, where the probability of an edge between two nodes decreases as the Euclidean distance between them increases. Given z_i denotes the latent position of node i occupied in the Euclidean space, the log odds of edge ϵ_{ij} connecting node i and j is given as:

$$\zeta_{ij} = \text{logodds}(\epsilon_{ij} = 1|\mathbf{z}_i, \mathbf{z}_j, \alpha) = \alpha - |\mathbf{z}_i - \mathbf{z}_j|$$

where α is a real intercept and |.| denotes the Euclidean norm. This model is particularly suitable for undirected graphs or networks with strong reciprocity. In contrast, the projection model is designed to handle asymmetric graphs more effectively. In the projection model, the log odds of an edge between two nodes are based on the angle they form in a latent space and are given as:

$$\zeta_{ij} = \text{logodds}(\epsilon_{ij} = 1|\mathbf{z}_i, \mathbf{z}_j, \alpha) = \alpha + \mathbf{z}_i^\top \mathbf{z}_j$$

Here, a narrow angle between nodes implies a higher likelihood of an edge, while wider angles suggest lower probabilities of connection. This model is particularly suited for directed networks where relationships between nodes may not be reciprocal. Both of these models offer unique strengths in uncovering latent patterns in networks, making them powerful tools for statistical network analysis

In Chapter 2 of this dissertation, we provide a more in-depth exploration of the LPM. The chapter offers a concise overview of the traditional LPM and its underlying modeling framework, while also highlighting key extensions that have been developed for the model. These extensions demonstrate the flexibility and broad applicability of LPM in various contexts. Furthermore, in Chapter 3 and Chapter 4, we demonstrated the efficacy of this model in count time series analysis.

1.4 Network embedded count time series analysis: foundational framework and research motivation

Over the last few years, networks have gained considerable popularity as a framework for inferring and analyzing cross-sectional dependencies among multiple variables in a multivariate time series system, denoted as, $\mathcal{X} = (\mathbf{y}_t, \mathcal{G})$, where $\mathbf{y}_t$ represents a multivariate

time series with an underlying network $\mathcal{G} = (\mathcal{K}, \mathcal{E})$. Here, $\mathcal{G}$ is a network with N nodes, where each time series $\mathbf{y}_{it}$ is associated with node $i \in \mathcal{K}$ and $\mathcal{E} \subset \mathcal{K} \times \mathcal{K}$ represents the set of edges.

The time series data $\mathbf{y}_{it}$ data observed at each node can either be continuous or represent counts, depending on the application. The edges in the network $\mathcal{E}$ can take various values, such as binary, discrete, or continuous values, representing different types of relationships between the nodes. Furthermore, the network $\mathcal{G}$ itself can be either undirected, where ($\mathcal{E}_{ij} = \mathcal{E}_{ji}$) or directed, where ($\mathcal{E}_{ij} \neq \mathcal{E}_{ji}$) allowing for a wide range of possible network dynamics to be captured. Additionally, based on whether the edge connection $\mathcal{E}_{ij}$ is independent or dependent on time, the network can be classified as either static or dynamic, with static networks having fixed relationships over time and dynamic networks allowing the relationships to evolve.

In this thesis, we use the network analysis technique to better understand and interpret the complex cross-sectional dependencies within a multivariate time series system. By representing such a system as a graph, the structure of the graph can be leveraged to study the underlying properties and interactions within the system.

Building on this foundation, our research aims to address the specific challenge of formulating statistical models for multivariate time series data studied through the lens of a network approach. The central goal of this work is to obtain a model-based summary of the data by uncovering the latent network structure, thereby providing clear, network-driven visualizations of complex interactions among the time series and revealing hidden behavioral patterns that may not be immediately apparent through traditional analysis methods. In doing so, our approach promotes parsimony and enhances interpretability, offering a more streamlined view for understanding of the underlying dynamics driving the data.

1.5 Foundational methods for modeling time series using network approaches

The study of multivariate time series models using network approaches has gained significant attention over recent years resulting in the development of models that couple time series with network structures. One of the early contributions to this field is by Knight et al. (2016), who introduced the Network AutoRegressive (Integrated) Moving Average (NARIMA) models. These models are designed for multivariate continuous time series that are coupled with a fixed network structure. A significant advancement in this area was made by Zhu et al. (2017), who proposed the Network Vector AutoRegressive (NAR) model. This model incorporates network structure into the traditional VAR framework, allowing node values to depend on those of their neighbors. Their work has inspired a range of extensions and adaptations by other researchers, some of which are discussed in Knight et al. (2020), Zhu et al. (2019b), Chen et al. (2023b), Zhu et al. (2019a), and Zhu and Pan (2020). However, it is important

to note that this body of literature typically assumes continuous time series data and a fixed network structure across time.

Despite the significant advancements in multivariate time series modeling with network structures, research on studying multivariate count time series models with embedded network frameworks remains limited. In this relatively unexplored domain, Armillotta and Fokianos (2024) introduced the Poisson Network Autoregressive (PNAR) model, specifically designed for count data in non-random network structures.

$$\mathbf{y}_{it} \sim \text{Poisson}(\boldsymbol{\lambda}_{it})$$

$$\log(\boldsymbol{\lambda}_{it}) = \alpha + \beta_1 \left(n_i^{-1} \sum_{\substack{j=1 \\ j \neq i}}^{N} \gamma_{ij} \log(y_{j(t-1)} + 1) \right) + \beta \log(y_{i(t-1)} + 1)$$

where, $n_i = \sum_{j \neq i} \gamma_{ij}$ i.e. the total number of nodes with which i has an edge in the known network structure. This model represents an extension of the NAR framework to handle count data, thus providing a foundation for further investigation into multivariate count time series with network dependencies. A few other notable contributions in this area include Armillotta et al. (2022), Liu and Nason (2023), and Yin et al. (2024), which propose innovative methodologies for network-based count time series modeling. The above studies generally assume that the underlying network structure is known and uses the connected variables to describe the dynamic behavior of the data, as well as provide better forecasting results. However, in many practical situations, the complete network structure is unknown. Our work in Chapter 3 aims to address this problem by advancing the study of count time series with fixed network structures which is completely unknown, contributing to a more comprehensive understanding of this growing field. Specifically, we are motivated by the dataset on the counts of burglaries in the census block groups of Chicago, which consists of monthly time series data. The connection between burglaries occurring in census blocks in Chicago can be studied through a network framework, with nodes corresponding to census blocks or regions and monthly time series of burglaries associated with region i. However, the regional connections and their strengths are unknown.

There is extensive literature on time series models with unknown network structures. This body of work primarily focuses on fitting Vector Autoregressive (VAR) models to infer inter-dependencies between variables. By estimating the non-zero entries of the VAR coefficient matrix, interconnections (edges) between the components of vector time series in a Granger causality network are recovered. Various approaches for inferring networks of causal relationships have been discussed in the literature. For example, Bolstad et al. (2011) addresses the problem of inferring sparse causal networks using the Group Lasso (gLasso) procedure, while Basu et al. (2015) introduces a group lasso regression regularization framework to

learn network structures from temporal panel data. Barigozzi and Brownlees (2019) inferred networks through the use of the long-run partial correlation matrix between multiple time series, and Owens et al. (2023) developed the fnets R package to estimate and visualize networks, focusing on strong spatial and temporal correlations in time series data through a factor-adjusted vector autoregressive model. However, these studies predominantly focus on continuous time series data, revealing a clear gap in the literature for count time series models with unknown network structures.

We address this gap by proposing a novel methodology, termed as Time Series Latent Position Model (TSLPM) that combines the multivariate log-linear count time series model with the Latent Position Model (LPM), a well-established network model. Building on the discussion of modeling count time series from Armillotta and Fokianos (2024), we employ a log-linear time series model to capture the dynamics of the multivariate count time series $\mathbf{y}_t$, where the underlying network structure is entirely unknown.

The Latent Position Model (LPM) provides a novel and intuitive framework for uncovering the network structure within multivariate count time series data. Another class of models for capturing interdependencies is the spatial time series models, which share conceptual similarities with LPM and focus on capturing spatial dependencies rather than purely networked relationships. These approaches extend the analysis of interdependencies by incorporating spatial structures, making them a powerful tool for studying spatio-temporal dynamics.

Among these, the endemic-epidemic (EE) framework, introduced by Held et al. (2005) and extended by Held and Paul (2012), Paul et al. (2008), Bracher and Held (2022), and Meyer and Held (2017), has been widely applied for multivariate spatio-temporal count predictions in surveillance data.

Recent advancements in this domain have further expanded the range of spatial count time series models. For example, Jahn et al. (2023) proposes spatio-temporal INGARCH models capable of handling both bounded and unbounded spatio-temporal counts while accommodating negative parameters and autocorrelation. Similarly, Lee et al. (2023) utilizes a multivariate integer-valued GARCH model to analyze weekly syphilis case counts across multiple U.S. states, capturing key spatial and temporal dependencies.

Incorporating spatial-temporal patterns into multivariate integer-valued GARCH models, Chen et al. (2023a) effectively addresses characteristics such as over-dispersion, spatial dependency, and zero inflation in infectious disease data. Building on these developments, Chen and Chen (2024) introduce spatial hurdle and zero-inflated generalized Poisson (ZIGP) INGARCH models, offering robust tools for analyzing complex time-series counts, such as dengue fever cases.

However, the primary objective of our work is to uncover the hidden interdependency structure within the time series data—essentially identifying the underlying network and using this structure for a model-based summary of the data. To achieve this, we adopt a new

perspective that has not yet been explored in the literature. Specifically, we employ the Latent Position Model (LPM), a well-established network model, framework to estimate the VARs coefficient matrix which captures the cross-connections among the series. Leveraging LPM's ability to determine network connections, each time series is associated as a node and positioned in a latent space. The relationships between nodes can then be both visualized intuitively and quantified using LPM's model formulation.

This methodology is discussed in detail in Chapter 3, where we employ the projection latent position model as the networking approach to map the time series within the latent space. We then use a Bayesian estimation procedure to estimate these positions based on some latent information, thereby uncovering the underlying dependency structure. This approach provides deeper insights into the complex interactions within the time series data.

Chapter 3 extends the application of TSLPM to a real-world dataset by analyzing burglary incidents in Chicago, USA. This dataset consists of counts of burglaries recorded across 552 census block groups over a period of 72 months. By focusing on aggregated count time series data from clusters of five and eleven census block groups, TSLPM uncovers patterns of connections in burglary activity within these regions. Leveraging the latent projection network model, the methodology provides straightforward and interpretable visualizations of these connections, based on latent features, which reveal geographical patterns in burglary incidence across clustered regions. This analysis offers valuable insights into the spatial dynamics of crime.

1.6 Expanding the horizon: time series analysis with dynamic network structures

With the development of network science, network models with non-random connections have proven inadequate for capturing the dynamic processes that occur in multivariate time series data. In real-world networks, interactions between entities evolve rapidly, and models resulting in fixed network structures fail to account for these changing relationships. Dynamic or temporal networks offer a more robust framework where the network evolves with time. Dynamic network captures the changing interactions between a set of nodes, represented through snapshots of networks at different time points. This helps in understanding evolving associations, allowing researchers to gauge the real underlying relationships over time. There is extensive literature on dynamic network models due to their properties in closely capturing real-world phenomena. LPM has been extended in various ways to accommodate dynamic networks, including the work of Sarkar and Moore (2005), who incorporated time dynamics into the model parameters. Sewell and Chen (2015) further extended it by introducing a new modeling framework network data is modeled via temporal trajectories in a latent Euclidean space to study the time-varying network connections. A few other contributions to dynamic

extensions of LPM are discussed in Sewell and Chen (2016), Friel et al. (2016), Durante and Dunson (2016), and Rastelli and Corneli (2023).

The work in Chapter 4 extends the modeling framework discussed in Chapter 3 by introducing a dynamic approach to capture time-varying patterns in the spread of mumps infection across nine regions of England. This dynamic extension is motivated by a need to understand how connections between regions evolve over time in response to changing infection patterns. Recently, there has been a growing interest in using dynamic network models to examine dependency structures in multivariate time series. Kang et al. (2022) utilized dynamic networks to model non-stationary vector autoregressive processes, while Krampe (2019) proposed a model treating dynamic networks with a fixed number of nodes but using a doubly stochastic matrix to represent the changing network structure. Additionally, Knight et al. (2020) introduced the Generalized Network Autoregressive (GNAR) model, which incorporates time-dependent weights and neighborhood structures, allowing the model to adapt to network changes over time. The aforementioned modeling frameworks primarily focus on continuous time series data. In contrast, Castro et al. (2012) fill a gap by introducing a framework for count time series with an embedded network structure. Their approach combines spatial and temporal dependencies, applying a spatial structure to latent continuous variables while accommodating both time-stationary and time-varying temporal correlations. Despite these advances, the literature on dynamic embeddings remains sparse, even though they are likely to be more realistic for many practical applications compared to static models.

In our approach, detailed in Chapter 4, we build on the earlier modeling framework to capture time-varying cross-connections among the time series variables using a dynamic network modeling approach. Similarly to the work in Chapter 3, we employ LPM to provide a structured representation of interactions between time series by locating them within a latent space. In this extended framework, the latent positions of the time series are treated as model parameters, similar to our previous framework, but evolve over time, following the approach of Sarkar and Moore (2005). In this chapter, we focus on the distance model, which uncovers the strength of connections between the nodes based on the Euclidean distance between them. This approach allows for a straightforward assessment of how connection levels change over time, as captured by changes in Euclidean distances. Additionally, LPM provides a latent space visualization of the temporal trajectories of the time series, offering a clear and interpretable view of these evolving relationships.

One key application of this extended framework is our analysis of mumps cases in England, which serves as a dataset motivating our research. This dataset records the quarterly number of mumps cases across nine regions of England, spanning 2008(Q1) to 2020(Q2) for a total of 50 quarters. Recognizing the complex, interrelated factors influencing mumps transmission, we adapt our framework to examine both static and dynamic network models with a focus on understanding the spread of mumps across England. We employ the LPM to structure

interactions but with a specific emphasis on a distance-based approach to uncover connection strengths among regions.

The results of this analysis reveal both geographic and indirect transmission pathways, which shift over time. The northernmost regions consistently show the highest incidence of mumps cases, while London emerges as the central node within the network, both influencing and being influenced by other regions. Furthermore, our analysis demonstrates a gradual weakening of connections over time, suggesting a decline in mumps transmission across regions as time progresses.

1.7 Overview of Stan and inferential procedures

Stan is an open-source probabilistic programming language developed by a team of statisticians, computer scientists, and mathematicians, and was officially released in the year 2012. Its primary applications encompass statistical modeling, inference algorithms for model fitting, predictive analytics, high-performance statistical computation, and posterior analysis tools for the evaluation of results. Furthermore, it offers optimization tools designed for finding maximum likelihood estimates, along with tools for variational inference. It functions as an interface to various programming environments, including R (through rstan), Python (via pystan), and CmdStan, which is a command-line interface. It is designed to handle a wide variety of models accurately and efficiently, with a focus on robust performance for real-world data, see https://mc-stan.org/docs/stan-users-guide/ for detailed discussion. It was specifically developed to provide a more flexible and powerful framework for fitting complex probabilistic models compared to other existing tools, particularly in the context of Bayesian inference.

Stan implements the Hamiltonian Monte Carlo (HMC) and No-U-Turn Sampler (NUTS) algorithms, both of which offer greater efficiency and scalability for high-dimensional, complex models compared to traditional Markov Chain Monte Carlo (MCMC) methods (Gelman et al., 2013). A notable limitation of methods like the Gibbs sampler and the Metropolis algorithm is their reliance on random walk behavior, which can lead to inefficient exploration of the target distribution. In particular, the random walk characteristic causes these algorithms to spend considerable time “zigging and zagging” through the distribution, slowing down convergence. Although techniques such as reparameterization and optimized jumping rules can mitigate this issue to some extent, random walk inefficiencies persist in high-dimensional settings and complex models. To address the inefficiencies caused by the local random walk behavior inherent in the Metropolis algorithm, Hamiltonian Monte Carlo (HMC) incorporates principles from physics to enable more efficient exploration of the target distribution. Specifically, HMC introduces an auxiliary momentum variable m_j for each parameter θ_j in the target space. These momentum variables are generally multivariate normally distributed with mean $\mathbf{0}$ and covariance matrix Σ, termed as “mass matrix”.

The parameters $\boldsymbol{\theta}$ and the momentum variables $\mathbf{m}$ are sampled from a joint distribution:

$$p(\boldsymbol{\theta}, \mathbf{m}) \propto \exp(\mathcal{L}(\boldsymbol{\theta}) - \frac{1}{2}\mathbf{m}^\top \mathbf{m})$$

where $\mathcal{L}(\boldsymbol{\theta})$ represents the log-probability for $\boldsymbol{\theta}$. In this formulation, the proposed updates for $\boldsymbol{\theta}$ are influenced by the momentum $\mathbf{m}$ and and the gradient of $\mathcal{L}(\boldsymbol{\theta})$, allowing the sampler to make informed, larger moves in the parameter space.

HMC updates the pair $(\boldsymbol{\theta}, \mathbf{m})$ simultaneously over time t by discretely simulating Hamiltonian dynamics. This update involves L "leapfrog steps", each scaled by a factor ϵ. The L leapfrog steps proceed as follows:

$$\begin{aligned}
\mathbf{m}_{t+\frac{\epsilon}{2}} &= \mathbf{m}_t + \frac{\epsilon}{2}\nabla_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta}_t) \\
\boldsymbol{\theta}_{t+\epsilon} &= \boldsymbol{\theta}_t + \epsilon\, \mathbf{m}_{t+\frac{\epsilon}{2}} \\
\mathbf{m}_{t+\epsilon} &= \mathbf{m}_{t+\frac{\epsilon}{2}} + \frac{\epsilon}{2}\nabla_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta}_{t+\epsilon})
\end{aligned}$$

After completing L such steps, a proposed state $(\hat{\boldsymbol{\theta}}, \hat{\mathbf{m}})$ is obtained. This proposed state is then subjected to a Metropolis acceptance step by comparing it with the current state $(\boldsymbol{\theta}_t, \mathbf{m}_t)$. The acceptance probability is computed as:

$$\mathsf{r} = \frac{p(\hat{\boldsymbol{\theta}}, \hat{\mathbf{m}})}{p(\boldsymbol{\theta}_t, \mathbf{m}_t)} \tag{1.1}$$

The new state $\boldsymbol{\theta}_{t+1}$ is then set according to:

$$\boldsymbol{\theta}_{t+1} = \begin{cases} \hat{\boldsymbol{\theta}} & \text{with probability } \min(\mathsf{r}, 1) \\ \boldsymbol{\theta}_t & \text{otherwise} \end{cases}$$

This accept-reject mechanism ensures that the Markov chain remains reversible and preserves the target distribution as its stationary distribution.

As with other MCMC algorithms, this process is iterated until convergence criteria are met. Convergence diagnostics, such as the $\hat{R}$ being near 1 and the effective sample size being large enough for all quantities of interest, are used to confirm that the samples are representative of the posterior distribution.

To address more complex problems, an extension of HMC called the No-U-Turn Sampler (NUTS) was developed. NUTS adaptively determines the number of leapfrog steps per iteration, rather than using a fixed count. It also adjusts the mass matrix Σ and step size ϵ during a warm-up phase, after which these parameters remain fixed for posterior sampling. The full NUTS algorithm is more sophisticated, as it explores the parameter space by expanding

trajectories both backward and forward along the Hamiltonian path, all while ensuring that detailed balance is maintained. This adaptive approach helps to avoid inefficient “U-turns” in the trajectory, making it particularly well-suited for high-dimensional problems where a fixed step count could result in suboptimal exploration of the posterior distribution.

In Chapter 3 and Chapter 4, we implemented HMC and NUTS for the Bayesian inference of proposed models. However, for the large-scale simulation studies, where the number of parameters grows with the number of time series and data size, HMC can become computationally intensive. For this reason, we opted for numerical optimization methods as a practical alternative for the simulation studies performed in both chapters.

Stan provides a variety of numerical methods to solve optimization problems, including three optimizers: a Newton optimizer, and two quasi-Newton algorithms, BFGS and L-BFGS (Nocedal and Wright, 2006). These methods find the modes of the joint posterior distribution, which can serve as parameter estimates or approximations to the Bayesian posterior. We employed the L-BFGS algorithm, which is particularly suitable for large-scale optimization problems due to its efficient use of limited memory, bypassing the need to compute the full Hessian matrix. The algorithm iteratively updates the parameter estimates by combining past gradients and curvature differences until convergence is achieved. Although L-BFGS does not guarantee convergence to a global minimum, it demonstrates favorable convergence properties with minimal computational demands.

1.8 Other inferential procedures

Apart from the inferential methods discussed in Section 1.7, alternative approaches for parameter estimation can also be explored. One such approach involves optimization using PyTorch (Paszke et al., 2019), which provides a library of optimizers capable of updating parameters based on computed gradients. We implemented this approach in the initial phase of this PhD research to explore its feasibility. However, applying this method presents several challenges.

A key issue is the selection of an appropriate learning rate. The optimal learning rate may vary across model parameters, making it difficult to identify a suitable value for models with numerous parameters to estimate. Another significant challenge encountered when using PyTorch optimizers is ensuring the convergence of model parameters. Monitoring the convergence of each parameter proved difficult, which rendered this approach unsuitable for the proposed models.

Another viable inferential procedure that has not yet been explored but could be implemented is the use of Markov Chain Monte Carlo (MCMC) methods. MCMC can be employed to obtain samples from the predictive distribution, providing a robust framework for inference. In related works by Armillotta and Fokianos (2024) and Guo and Zhu (2024), the

quasi-likelihood method has been utilized for parameter estimation. However, their models assume that the network structure is known, which is not the case in our setting. Despite this difference, such approaches could be adapted, tested, and potentially extended to suit the requirements of our proposed models.

1.9 Motivational datasets

This thesis is motivated by the analysis of three datasets comprising count-valued multivariate time series. The first and second datasets record the number of burglaries across 552 census block groups (regions) in Chicago, aggregated into 5 and 11 broader regions over a 72-month period. The third dataset captures quarterly mumps cases reported across nine counties in England from 2008 (Q1) to 2020 (Q2).

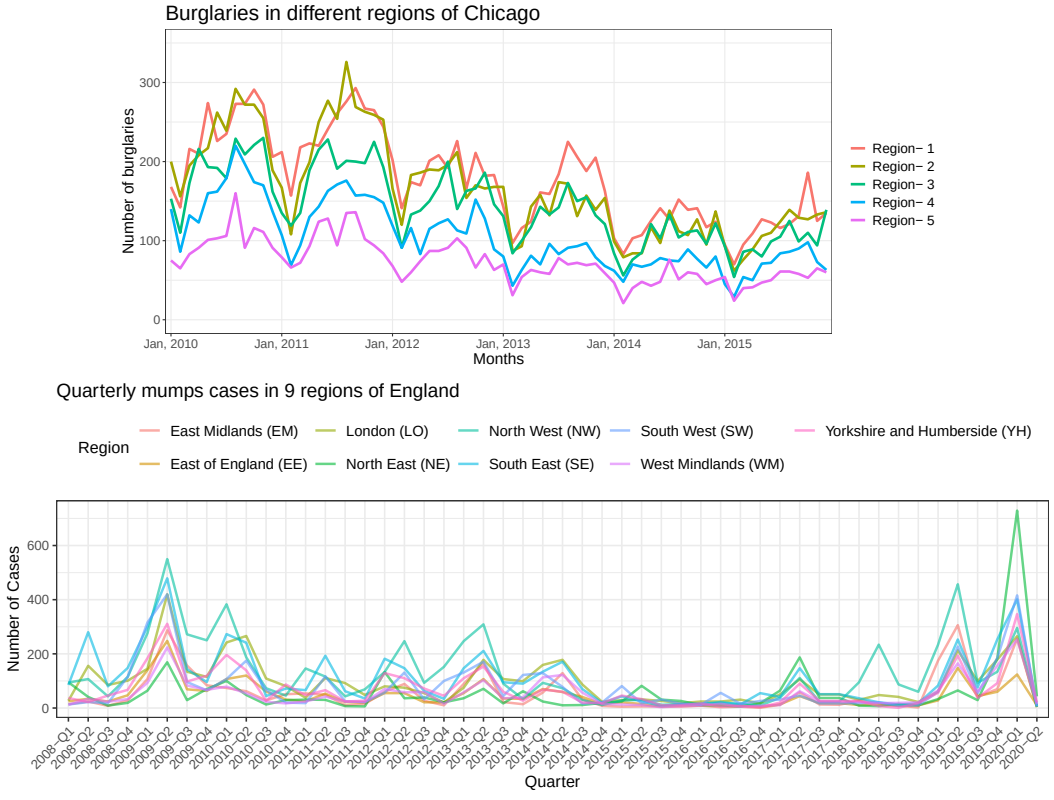

Fig. 1.1 Top: Monthly number of cases of burglaries in the south side of Chicago (2010-2015) aggregated over 5 regions. Bottom: Quarterly number of mumps cases in 9 regions of England.

Each dataset consists of time-varying counts of burglary or mumps cases across different regions in Chicago and England, respectively. This work focuses on the study of multivariate count time series within these datasets, where a key aspect of our framework is that each time

series is represented as a node within a network. By leveraging network structures, we aim to uncover latent dependencies among time series, thereby revealing important underlying characteristics within the data.

Each dataset here consists of a time-varying count of burglaries/mumps cases in different regions of Chicago/England. Our work is centered around studying multivariate count time series in the aforementioned datasets, with the key feature in our setting, that each time series is observed as a node of a network. Here, we aim to study multivariate count time series data by uncovering the network structure. This study will reveal some important underlying features in the time series data.

Consider a system of N time series to be studied through a lens of networking approach, where each time series is now observed as a node of an N sized network. The adjacency matrix defines the structure of a network by describing the pairwise connections between the nodes. In this thesis, the network adjacency matrix corresponds to the matrix with autoregressive coefficients in a time series model. We use network models to estimate the pairwise connections described by the adjacency matrix of the network. We assume the underlying network that we tend to identify is undirected and weighted connections in both Chapters. Our approach will allow us to quantify the connections in the burglary occurrences across regions of Chicago beyond their geographical location for the Chicago burglary dataset is discussed in detail in Chapter 3. Furthermore, in the mumps cases dataset, we are motivated to identify time-varying patterns in the transmission of mumps infection between the regions of England and study the spillover effect caused by outbreaks between different geographical areas, potentially leading to local or global epidemics. The network-based time series modeling framework will explain the complex transmission patterns in the mumps data based on latent factors. Moreover, the expansion and contraction of the latent space will measure any possibility of spillovers and contagion, as discussed in 4.

1.10 Thesis overview and chapter summaries

The primary goal of this dissertation is to address the gap in the literature on multivariate count time series with network structures. Traditional modeling approaches for count time series with fixed network structures are relatively scarce. Most current models focus on analyzing interdependencies between time series under the assumption that the underlying network structure is known. However, the literature on models that account for evolving network structures in count time series is even more limited. This dissertation aims to fill these gaps by developing and implementing new statistical models tailored specifically for the multivarite count time series model, with a focus on identifying their underlying network structures, and whether they remain constant or change over time. In doing so, it aims to improve our understanding of pairwise relationships by providing a model-derived overview of

the time series data and revealing the intricate network structure. A central component of this approach is the Latent Position Model, which uncovers the hidden network structure and enables clear and interpretable visual representation.

Chapter 2 of this thesis presents the foundational theory on Latent Position Models (LPMs) for modeling networks where the relationships between nodes can be inferred based on their positions in a latent space. This model is crucial for uncovering hidden structures in network data and has a broad range of applications across different fields. This chapter provides an overview of recent literature on LPMs and their various extensions, highlighting their versatility.

Chapter 3 extends the LPM framework by incorporating it into a multivariate time series model. Specifically, we focus on count time series data, where the LPM is used to model the interdependencies between time series. This integration facilitates the visualization and quantification of pairwise relationships among time series, offering a more interpretable framework for uncovering underlying connections, making it well-suited for both prediction and deeper analysis. Additionally, we address issues related to model identifiability, stationarity, and parameter interpretation, demonstrating efficient inference within a Bayesian framework. The methodology is then applied to analyze burglary count data across Chicago's census blocks.

Chapter 4 addresses the challenge of modeling time series data with a dynamic network structure, where both the network structure and the time series evolve together. This chapter presents a progression in network-based time series analysis from static network structures to dynamic ones, offering more advanced techniques to uncover relationships. We propose a novel modeling approach that integrates a latent distance model with a multivariate log-linear count time series model. Both static and dynamic variants of the model are developed, with our results indicating distinct advantages and specific applications for each. Additionally, we demonstrate efficient Bayesian inference of the model and apply this methodology to mumps case data in England, utilizing both the static and dynamic models. The model provides a robust framework for analyzing mumps transmission across the network.

Chapter 5 offers the concluding remarks of this thesis, examining the major contributions and implications of this research in the context of network-integrated time series analysis, shedding light on how it advances our understanding of the complex interdependencies among time series. Additionally, the broader impact of this work on shaping future research in network-embedded time series modeling is discussed. The chapter also identifies areas for further investigation, outlining potential directions for continued exploration and development in this evolving field.

References

Andreassen, C. (2013). *Models and inference for correlated count data*. PhD thesis, Department of Mathematics, Aarhus University.

Armillotta, M. and Fokianos, K. (2024). Count network autoregression. *Journal of Time Series Analysis*, 45(4):584–612.

Armillotta, M., Fokianos, K., and Krikidis, I. (2022). *Generalized Linear Models Network Autoregression*, pages 112–125. Springer International Publishing.

Barigozzi, M. and Brownlees, C. (2019). Nets: Network estimation for time series. *Journal of Applied Econometrics*, 34(3):347–364.

Basu, S., Shojaie, A., and Michailidis, G. (2015). Network Granger Causality with Inherent Grouping Structure. *Journal of Machine Learning Research*, 16(1):417–453.

Bolstad, A., Van Veen, B., and Nowak, R. (2011). Causal Network Inference Via Group Sparse Regularization. *IEEE Transactions on Signal Processing*, 59(6):2628–2641.

Box, G., Jenkins, G., Reinsel, G., and Ljung, G. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Bracher, J. and Held, L. (2022). Endemic-epidemic models with discrete-time serial interval distributions for infectious disease prediction. *International Journal of Forecasting*, 38(3):1221–1233.

Castro, M., Paleti, R., and Bhat, C. (2012). A latent variable representation of count data models to accommodate spatial and temporal dependence: Application to predicting crash frequency at intersections. *Transportation Research Part B: Methodological*, 46(1):253–272.

Chen, C. and Chen, C. (2024). Spatial-temporal hurdle model vs. spatial zero-inflated garch model: analysis of weekly dengue fever cases. *Stochastic Environmental Research and Risk Assessment*, 38:2119–2134.

Chen, C., CS., C., and Hsiung, M. (2023a). Bayesian modeling of spatial integer-valued time series. *Computational Statistics Data Analysis*, 188:107827.

Chen, E., Fan, J., and Zhu, X. (2023b). Community network auto-regression for high-dimensional time series. *Journal of Econometrics*, 235(2):1239–1256.

Durante, D. and Dunson, D. (2016). Locally adaptive dynamic networks. *Annals of Applied Statistics*, 10(4):2203–2232.

Erdös, P. and Rényi, A. (1959). On random graphs I. *Publicationes Mathematicae Debrecen*, 6:290–297.

Fokianos, K. (2024). Multivariate count time series modelling. *Econometrics and Statistics*, 31:100–116.

Fokianos, K., Støve, B., Tjøstheim, D., and Doukhan, P. (2020). Multivariate count autoregression. *Bernoulli*, 26(1):471 – 499.

Fokianos, K. and Tjøstheim, D. (2011). Log-linear poisson autoregression. *Journal of Multivariate Analysis*, 102(3):563–578.

Friel, N., Rastelli, R., Wyse, J., and Raftery, A. (2016). Interlocking directorates in irish companies using a latent space model for bipartite networks. *Proceedings of the National Academy of Sciences of the United States of America*, 113(24):6629–6634.

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., and Rubin, D. (2013). *Bayesian Data Analysis*. Chapman and Hall/CRC.

Guo, X. and Zhu, F. (2024). Softplus negative binomial network autoregression. *Stat*, 13(1):e638.

Heinen, A. and Erick, R. (2007). Multivariate autoregressive modeling of time series count data using copulas. *Journal of Empirical Finance*, 14(4):564–583.

Held, L., Hòhle, M., and Hofmann, M. (2005). A statistical framework for the analysis of multivariate infectious disease surveillance counts. *Statistical Modelling*, 5(3):187–199.

Held, L. and Paul, M. (2012). Modeling seasonality in space-time infectious disease surveillance data. *Biomedical Journal*, 54(6):824–843.

Hoff, P., Raftery, A., and Handcock, M. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460):1090–1098.

Holland, P., Laskey, K., and Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social Networks*, 5:109–137.

Holland, P. and Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association*, 76(373):33–50.

Jahn, M., Weiß, C., and Kim, H. (2023). Approximately linear ingarch models for spatio-temporal counts. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 72(2):476–497.

Kang, X., Ganguly, A., and Kolaczyk, E. (2022). Dynamic networks with multi-scale temporal structure. *Sankhya A: The Indian Journal of Statistics*, 84(1):218—260.

Knight, M., Leeming, K., Nason, G., and Nunes, M. (2020). Generalized Network Autoregressive Processes and the GNAR Package. *Journal of Statistical Software*, 96(5):1–36.

Knight, M., Nunes, M., and Nason, G. (2016). Modelling, detrending and decorrelation of network time series. *arXiv: 1603.03221*.

Krampe, J. (2019). Time series modeling on dynamic networks. *Electronic Journal of Statistics*, 13(2):4945 – 4976.

Lee, S., Kim, D., and Kim, B. (2023). Modeling and inference for multivariate time series of counts based on the ingarch scheme. *Computational Statistics Data Analysis*, 177:107579.

Lee, Y., Lee, S., and Tjøstheim, D. (2018). Asymptotic normality and parameter change test for bivariate poisson ingarch models. *Test*, 27:52–69.

Liu, H. and Nason, G. (2023). New methods for network count time series. *arXiv: 2312.01944*.

Luetkepohl, H. (2005). *The New Introduction to Multiple Time Series Analysis*. Springer Berlin, Heidelberg.

Meyer, S. and Held, L. (2017). Incorporating social contact data in spatio-temporal models for infectious disease spread. *Biostatistics*, 18(2):338 – 351.

Newman, M. (2018). *Networks*. Oxford university press.

Nocedal, J. and Wright, S. (2006). *Numerical Optimization*. Springer, New York, NY, USA, 2e edition.

Owens, D., Cho, H., and Barigozzi, M. (2023). fnets: An R package for network estimation and forecasting via factor-adjusted var modelling. *The R Journal*, 15(3):214–239.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc.

Paul, M., Held, L., and Toschke, A. (2008). Multivariate modelling of infectious disease surveillance data. *Statistics in Medicine*, 27(29):6250 – 6267.

Rastelli, R. and Corneli, M. (2023). Continuous latent position models for instantaneous interactions. *Network Science*, 11(4):560–588.

Robins, G., Pattison, P., Kalish, Y., and Lusher, D. (2007). An introduction to exponential random graph (p *) models for social networks. *Social Networks*, 29:173–191.

Sarkar, P. and Moore, A. (2005). Dynamic social network analysis using latent space models. In *Proceedings of the 18th International Conference on Neural Information Processing Systems*, NIPS'05, pages 1145–1152. MIT Press.

Sewell, D. and Chen, Y. (2015). Analysis of the formation of the structure of social networks by using latent space models for ranked dynamic networks. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, 64(4):611–633.

Sewell, D. and Chen, Y. (2016). Latent space models for dynamic networks with weighted edges. *Social Networks*, 44:105–116.

Wang, Y. and Wong, G. (1987). Stochastic blockmodels for directed graphs. *Journal of the American Statistical Association*, 82(397):8–19.

Yin, H., Safikhani, A., and Michailidis, G. (2024). A functional coefficients network autoregressive model. *arXiv: 2402.07373*.

Zhu, X., Chang, X., Li, R., and Wang, H. (2019a). Portal nodes screening for large scale social networks. *Journal of Econometrics*, 209(2):145–157.

Zhu, X. and Pan, R. (2020). Grouped network vector autoregression. *Statistica Sinica*, 30(3):1437–1462.

Zhu, X., Pan, R., Li, G., Liu, Y., and Wang, H. (2017). Network vector autoregression. *The Annals of Statistics*, 45(3):1096 – 1123.

Zhu, X., Wang, W., Wang, H., and Härdle, W. (2019b). Network quantile autoregression. *Journal of Econometrics*, 212(1):345–358.

Acknowledgement of Authorship Form

This form must be completed for each paper to be included and placed in front of that paper in the thesis. The purpose of this is to ensure that shared authorship is appropriately explained in a thesis.

1. Examination Candidate Details
Examination Candidate Name: **Hardeep Kaur**
Examination Candidate UCD Student Number: **19200482**
Research Degree for which thesis is being submitted: **Doctor of Philosophy**
Title of Research thesis: **Latent position network models with applications in time series analysis**
2. Details of the Paper
Title of Paper: **Latent Position Network Models**
Current Status of the Paper: ☐ Drafted ☐ Submitted. ☐ Under review ☐ Under revision ☐ Accepted for publication ☐✓ Published
Name of Journal (if known): **SAGE Publication**
Digital Object Identifier (if published): **https://doi.org/10.4135/9781529614695.n36**
AUTHOR CONTRIBUTIONS
Part A: Contribution of Examination Candidate: *The overall percentage contribution must be specified, as well as a full description the Examination Candidate's intellectual and practical contribution to the research presented in the paper. Each section of the paper should be listed separately i.e. Introduction, Methods, Figure 1, Table 2 etc.*
Percentage Contribution of Examination Candidate to this paper: **80** %
Description of Examination Candidate's contribution for each section of paper: (Max 500 words): **As the primary author of this paper, I took on the majority of the work involved including an extensive review of relevant literature. I conducted the formal analysis, ensuring rigor and accuracy in interpreting the data and results. Additionally, I was responsible for writing the paper, and organizing and presenting our findings clearly and cohesively. I also contributed to the initial coding development for one of the sections, providing a foundational structure for the analyses presented.**
Part B: **Contribution(s) of Co-Authors to this paper**: *Outline briefly the contribution of all co-authors to the paper.* *Note:* *1. The combined contributions should be 100%.* *2. The signature of a co-author is not required, if their contribution is < 10% of the paper.*

Co-Author 1	
Name of Author No.1	**Dr. Riccardo Rastelli**
Contribution of Co- Author No.1	% Contribution: **10%** *Brief outline of contribution:* (Max 200 words) **Dr. Riccardo Rastelli supervised the work, offering guidance and reviewing progress through weekly meetings. He also contributed to the code development in one section by reviewing, correcting, and making additional enhancements.**
Signature of Co-Author No.1	
Co-Author 2	
Name of Co-Author No. 2	**Prof. Nial Friel**
Contribution of Co-Author No.2	% Contribution: **5%** *Brief outline of contribution:* (Max 200 words) **Prof. Nial Friel collaborated on the work by attending monthly meetings, offering valuable suggestions, and reviewing and editing the final manuscript.**
Signature of Co- Author 2	
Co-Author 3	
Name of Co-Author 3	**Prof. Adrian E. Raftery**
Contribution of Author 3	% Contribution: **5%** *Brief outline of contribution:* (Max 200 words) **Prof. Adrian E. Raftery contributed to the work by attending monthly meetings, offering valuable suggestions, and reviewing and editing the final manuscript.**
Signature of Author 3	
Add more Co-Author details as required.	

Submission Date for Thesis: 12th November, 2024 *Date thesis is submitted to the University for examination*
Signature of Examination Candidate

Note for Notation: Chapter 2 of this thesis introduces the popular network model and its extensions. In this chapter, $Y = (y_{ij})$ as the adjacency matrix of a network, consistent with the literature on network models. However, in Chapters 1, 3 and 4, we use $Y = (y_{it})$ to represent a time series, following standard notation in time series analysis. To determine cross-dependencies between time series, we assume an underlying network structure. Specifically, we use network models to uncover such dependencies by treating each time series as a node in the network. In this setting, $Y = (y_{it})$ denotes the observed time series associated with each node.

Chapter 2

Latent Position Network Models

Abstract

In this chapter, we present a review of latent position models for networks. We review the recent literature in this area and illustrate the basic aspects and properties of this modeling framework. Through several illustrative examples we highlight how the latent position model is able to capture important features of observed networks. We emphasize how the canonical design of this model has made it popular thanks to its ability to provide interpretable visualizations of complex network interactions. We outline the main extensions that have been introduced to this model, illustrating its flexibility and applicability.

Keywords: Social Network Analysis; Latent Position Models; Bayesian inference; Network Visualization.

2.1 Introduction

Statistical network analysis has recently emerged as a prominent area of research, with applications in many fields including social sciences, biology, finance and physics. Social networks are used to study actors and the pairwise interactions between them. The formulation of statistical models for such network data plays an important role in describing the network's global topology and in providing interpretable representations of the data. Some of the pioneering work in network analysis dates back to Erdös and Rényi (1959), who introduced the famous Erdős-Rényi model with two connected variants for the generation of random graphs or the evolution of a random network. An initial version of what are called today latent position models was developed soon afterwards by Gilbert (1961), who introduced the so-called spatially embedded random networks. The key aspect of these models is that they

define a generative framework for the edges of a graph based on the positions of the nodes in a Euclidean space.

Hoff et al. (2002) adapted similar concepts and ideas to the analysis of social networks, by introducing a new framework called the Latent Position Model (LPM) or Latent Space Model (LSM). This work renewed interest in spatially embedded models as a tool for modelling the complex networks arising in many applied fields. As a result, the literature on the topic has increased rapidly, and the LPM has become a widely used statistical model for network analysis. The model has been used in the analysis of corporate governance (Friel et al., 2016), financial risk (Tafakori et al., 2022), trophic food webs (Chiu and Westveld, 2011, 2014), protein sequence data (Ding et al., 2019), interbank networks (Linardi et al., 2020), trade networks (Ward et al., 2013), social influence (McFowland III and Shalizi, 2021; Sweet and Adhikari, 2020), music contests (D'Angelo et al., 2019), migration flows (Xiao et al., 2022), anomaly detection (Lee et al., 2021), political networks (Ng et al., 2021), conflict networks (Westveld and Hoff, 2011), flows of controlled substances (Berlusconi et al., 2017), among others. It has also been used in a variety of other research areas, including neuroscience (Durante et al., 2017; Wilson et al., 2020), item response theory (Jin and Jeon, 2019), mediation analysis (Liu et al., 2021), education research (Sweet et al., 2013), and epidemiology (Chu et al., 2021).

The literature on LPMs has also been reviewed in previous articles on network analysis, including Rastelli et al. (2016), Salter-Townshend et al. (2012), Matias and Robin (2014), Raftery (2017), Smith et al. (2019), Kim et al. (2018), and Sosa and Buitrago (2021). This paper provides a review of LPMs and of the recent literature that originated from the seminal paper of Hoff et al. (2002). We include an explanation of the modelling framework in Section 2.3, and an application to an original dataset in Section 2.5. This is followed by a detailed literature review in Section 2.6.

2.2 Notation

Throughout this paper, we consider data on the relations between a set of nodes $\{1, \ldots, n\}$. The relations are described by the set of edges $\mathcal{E}$. We denote by $Y = (y_{ij})_{1 \leq i < j \leq n}$ the $n \times n$ adjacency matrix of the observed undirected network, where y_{ij} is the edge value between the nodes i and j. These edge values are usually binary, discrete or continuous. A collection of dyad-specific covariates may be available. We denote these by $X = (x_{ijk})$, where (i, j) is an ordered pair of nodes and $k = 1, \ldots, p$ indexes the covariates. We denote by $\boldsymbol{\theta}$ the collection of parameters that do not refer to edges or nodes; we call these the global parameters. We denote by Z the unobserved latent variables, i.e. the latent positions associated with the nodes. Here z_i represents the unknown position of the corresponding node in a d-dimensional latent space.

2.3 Latent Position Models

2.3.1 The original distance model and projection model

Latent position models (also called latent space models) were introduced by Hoff et al. (2002), for undirected binary networks. This introduced a new family of latent variable models for network data, with the goal of providing a visualization of a relational dataset through a latent geometric social space. Crucially, the authors introduced an inferential framework to estimate this model for networks of moderate size (of around a few hundred nodes).

The social space introduced by the authors consists of a set of n nodes, each having an unknown position z_i in a d-dimensional latent space, typically $\mathbb{R}^d$. In a Bayesian framework, the positions are assumed to be independent and identically distributed according to a multivariate Gaussian distribution $\mathcal{MND}(\mathbf{0}, \Sigma)$. One fundamental property of this model is that, conditionally on the latent positions Z, the relational ties $y_{ij},\ \forall\ i, j = 1, \ldots, n$ are independent. In particular, the probability of the presence of a generic edge y_{ij} is modeled as some function of the latent positions z_i and z_j of the two nodes involved, as shown in Eq 3.3.

$$\mathbb{P}(Y|Z, X, \boldsymbol{\theta}) = \prod_{i<j} \mathbb{P}(y_{ij}|z_i, z_j, x_{ij}, \boldsymbol{\theta}), \tag{2.1}$$

where the product is taken over all pairs (i, j), where $i < j$, for $i, j = 1, \ldots, n$.

Thanks to their geometric framework, these models are naturally able to represent common network features such as reciprocity, transitivity, and homophily. In other words, nodes with similar characteristics (i.e. positions) tend to possess higher probability of forming a tie. The similarity of latent characteristics may be defined in different ways: the two approaches proposed by Hoff et al. (2002) are called the distance model and the projection model.

Distance model

The latent distance model proposed by Hoff et al. (2002) assumes that each node has an unobserved latent position z_i in an Euclidean latent space, typically in the plane $\mathbb{R}^2$. Then, the closer two nodes' latent positions are to each other, the higher the probability that there is a connection between them. Conversely, the probability of connection decreases as the distance between the nodes increases. The Euclidean distance is most commonly used, but any other distance function may be considered. This framework permits easy visualization and interpretation of the latent social space. The formulation of Hoff et al. (2002) defines the log odds of a tie between nodes i and j as:

$$\eta_{ij} = \text{logodds}(y_{ij} = 1|z_i, z_j, x_{ij}, \alpha, \beta) = \alpha + \beta^\top x_{ij} - |z_i - z_j|, \tag{2.2}$$

where α and β are real-valued and $|\cdot|$ indicates the Euclidean norm.

This parameterization is for a logistic regression model with $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta})$ where α is an intercept term and $\boldsymbol{\beta}$ is a vector of coefficients for covariate effects. The distance model is particularly suited for networks with undirected or directed relations exhibiting strong reciprocity.

Projection model

Hoff et al. (2002) postulated the projection model as an alternative to the distance model. This proposes that two nodes have a higher probability of connecting if their respective latent positions are in the same direction $({z_i}^\top z_j > 0)$, or conversely, they are less likely to connect if they point in opposite directions $({z_i}^\top z_j < 0)$. In other words, two nodes are likely to form a tie if the angle (with respect to the center of the space) between them is small, and less likely to form a tie if the angle between them is obtuse. The projection model is defined as follows:

$$\eta_{ij} = \text{logodds}(y_{ij} = 1 | z_i, z_j, x_{ij}, \alpha, \beta) = \alpha + \beta^\top x_{ij} + \frac{|z_i^\top z_j|}{|z_j|} \tag{2.3}$$

The latent effects depend on $\frac{|z_i^\top z_j|}{|z_j|}$ which is the signed magnitude of the projection of z_i in the direction of z_j. This quantity can be interpreted as the extent of shared characteristics among nodes i and j, multiplied by the activity level of i.

Similarly to the distance model, the projection model provides a latent view of the social space. The latent space, in this case, is more easily thought of in terms of polar coordinates, where the position of a node is interpreted as a direction and a magnitude, indicating the social orientation and sociality effect, respectively. While the distance models naturally yields only symmetric edge probabilities, the projection model can also represent asymmetric edge probabilities, and may thus be more suitable for directed networks. However, it should also be noted that the distance model typically provides a more clear and intuitive representation of the latent space, and arguably more flexibility in representing some network topologies, such as community structures.

2.3.2 Latent distance models with clustering

An extension of the latent space model was proposed by Handcock et al. (2007), who introduced the Latent Position Cluster Model (LPCM). Their innovation was to model the clustering of highly connected nodes in the network via a latent mixture mode. This model has been extended by others, including Krivitsky et al. (2009), Krivitsky and Handcock (2008), Salter-Townshend and Murphy (2013), Ryan et al. (2017), Gormley and Murphy (2010), Sewell and Chen (2017), and Aliverti and Durante (2019). This model integrates the latent

space distance models with model-based clustering of the nodes, hence connecting to the literature on stochastic blockmodels (Snijders and Nowicki, 1997; Wang and Wong, 1987).

In the paper, the latent distance model is given as

$$\text{logodds}(y_{ij} = 1|z_i, z_j, x_{ij}, \beta) = \beta^\top x_{ij} - \nu|z_i - z_j|. \tag{2.4}$$

The prior structure is modified by assuming that the latent positions $z_i \in \mathbb{R}^d$ arise from a finite mixture of multivariate normal distributions with G components:

$$z_i \sim \sum_{g=1}^{G} \lambda_g MND_d(\mu_g, \sigma_g^2 I_d),$$

where λ_g is the probability that a node belongs to the g-th group, and $\sum_{g=1}^{G} \lambda_g = 1$. The proposed model captures network features such as transitivity and homophily, but is also capable of representing clustering in a more explicit and natural way.

This work introduces a model-based criterion to select the best number of clusters G and the number of latent dimensions d jointly. The criterion is inspired by the BIC (Bayesian Information Criterion, Fraley and Raftery (1998)) and it combines two principled approximations: one determined by d, the dimension of the latent space and one determined by G, the number of latent clusters or mixture components. While this addresses a gap in the literature by providing a principled approach for model choice, it also emphasizes that the method itself (in particular, the BIC) may not be ideal for the selection of the number of dimensions d since the underlying asymptotic approximation result has not been shown to hold in this case.

2.3.3 Latent distance models with node-specific random effects

Krivitsky et al. (2009) combined the approaches of Handcock et al. (2007), which involves model based clustering of the latent space positions, and Hoff (2005), which uses actor-specific random effects, into a new model called the Latent Position Cluster Random Effect model (LPCMRE). The proposed model is

$$\eta_{ij} = \text{logodds}(y_{ij} = 1|z_i, z_j, x_{ij}, \beta) = \sum_{k=1}^{p} \beta_k x_{ijk} - |z_i - z_j| + \delta_i + \gamma_j, \tag{2.5}$$

where, in addition to the specification of Handcock et al. (2007), δ_i, γ_j are introduced as actor-specific sender and receiver effects. The purpose of these parameters is to represent different levels of sociality of the nodes, so that a wider variety of degree distributions can be represented.

2.4 Inference for the LPM

2.4.1 Estimation

Hoff et al. (2002) considered two main approaches to inference for the LPM. The first is based on maximum likelihood maximization and it consists of a two-step procedure. In the first step, we compute the maximum likelihood estimator of the pairwise distances between all nodes. The log-likelihood is a convex function of the distances between the actors, so it can be easily maximized using numerical procedures to obtain an estimator. In the second step, the optimal distances are used to derive the actual positions of the nodes in the latent space. The authors proposed performing this step using multidimensional scaling. The two-step procedure does not guarantee that the global maximum of the likelihood is obtained, but empirically it has given good results with low computational demands. This is appealing as the set of inferred positions may be used as a starting point for potentially better, but more computationally intensive procedures.

The second approach proposed by Hoff et al. (2002) is a Bayesian approach based on Markov chain Monte Carlo (MCMC). Prior distributions are specified for the model parameters, and then MCMC is used to obtain approximate samples of the parameters from the posterior distribution. For the distance model, the prior distributions are typically specified as:

$$z_i \overset{\text{iid}}{\sim} MVN(0, \Sigma),$$
$$\theta_k \overset{\text{iid}}{\sim} N(0, \sigma^2),$$

for some covariance matrix Σ and variance parameter σ^2.

The prior distribution is combined with the likelihood function, namely

$$\mathbb{P}(Y|Z,\theta) = \prod_{i \neq j} \mathbb{P}(y_{ij} = 1|Z,\theta)^{y_{ij}} \mathbb{P}(y_{ij} = 0|Z,\theta)^{1-y_{ij}} = \prod_{i \neq j} \frac{\exp(y_{ij}\eta_{ij})}{1 + \exp(\eta_{ij})}, \tag{2.6}$$

to obtain the posterior distribution of the model:

$$\pi(Z,\theta|Y) \propto \mathcal{L}_Y(Z,\theta)\pi(\theta)\pi(Z).$$

The MCMC approach is a standard Metropolis-within-Gibbs algorithm with random walk proposals, where each parameter of the model is sampled in turn from its full conditional distribution. The full conditional distributions for the parameters are not in standard form and are given by:

$$\pi(z_i|Z_{-i},\theta,Y) \propto \pi(z_i) \prod_{i \neq j} \mathbb{P}(y_{ij} = 1|Z,\theta)^{y_{ij}} \mathbb{P}(y_{ij} = 0|Z,\theta)^{1-y_{ij}}, \tag{2.7}$$

$$\pi(\theta_k|\theta_{-k}, Z, Y) \propto \pi(\theta_k)\mathcal{L}_Y(Z, \theta), \tag{2.8}$$

where the negative subcripts indicate the collection of parameters with the exception of the one negated.

One challenge with LPMs is the high computational complexity, as highlighted by Eq. 2.7 and Eq. 2.8. These updates make the number of calculations that are required for the procedure grow with the square of the number of nodes. This quadratic cost is not scalable and it can limit the applicability of the original model.

The MCMC sampling proceeds as follows:

1. Choose an initial guess for all the model parameters Z and θ.
2. For every node $i = 1, \ldots, n$:
   (a) Sample a vector z_i^* from a multivariate Gaussian proposal $q(z_i^* \rightarrow z_i)$ which is centered in the current position of this node.
   (b) Calculate the ratio between the full-conditionals $r_Z = \frac{\pi(z_i^*|Z_{-i},\boldsymbol{\theta})}{\pi(z_i|Z_{-i},\boldsymbol{\theta})}$.
   (c) The new proposed values are accepted with probability $\min(1, r_Z)$, otherwise the current values are retained in the sample.
3. For every node $k = 1, \ldots, K$:
   (a) Sample a new parameter θ_k^* from a Gaussian proposal $q(\theta_k^* \rightarrow \theta_k)$ which is centered in the current value of this parameter.
   (b) Calculate the ratio between the full-conditionals $r_\theta = \frac{\pi(\theta_k^*|\boldsymbol{\theta}_{-k},Z)}{\pi(\theta_k|\boldsymbol{\theta}_{-k},Z)}$.
   (c) The new proposed value is accepted with probability $\min(1, r_\theta)$, otherwise the current value is retained in the sample.

This algorithm generates a Markov chain whose stationary distribution is the posterior distribution sought.

2.4.2 Interpretation of the posterior samples

The likelihood defined in Hoff et al. (2002) depends on the latent positions only through the pairwise distances between the nodes. Hence, the model parameters are non-identifiable with respect to any distance-preserving transformations of the latent positions. These transformations include rotations, reflections and translations of the latent positions. In principle, if we focus only on one single configuration of the parameters drawn from the posterior distribution, the non-identifiability issues may be negligible, because a translation, reflection, or rotation of the points would not affect our interpretations of the latent space.

However, in a Bayesian setting, we obtain a posterior sample $\{Z^{(1)}, Z^{(2)}, \dots\}$, and we are interested in summarising that sample to obtain estimators such as the posterior mean. In this case, applying a transformation (e.g. a translation) to any of the configurations of the sample would directly affect the value of the posterior summaries. Crucially, we have no way to determine whether the latent space was rotated, translated or reflected during the sampling procedure. For this reason, the posterior sample is not identifiable, and we cannot draw meaningful summaries from it.

As a solution to this problem, Hoff et al. (2002) considered Procrustes matching. This approach is based on a comparison of one configuration of points with another via their respective coordinate matrices (we denote these matrices by A and B in this section). In Procrustes matching, a rigid transformation is applied to A to make it as close as possible to the reference B. The matrix A is rotated, reflected and translated to achieve the best match to the reference matrix B, where the best match is defined as the one that minimises the sum of squared distances between corresponding points. This is given by:

$$R^2 = \sum_{i=1}^{n} \sum_{k=1}^{d} (b_{ik} - a_{ik})^2.$$

We can set up an optimization problem by specifying the following transformation of the matrix A:

$$\boldsymbol{a}_i^\top = \mathbf{O}^\top \boldsymbol{a}_i + \mathbf{t},$$

where $\boldsymbol{a}_i$ is a vector of coordinates of point i in the matrix A, and $\mathbf{O}$ is an orthogonal matrix that induces a rotation and/or reflection and $\mathbf{t}$ is a translation vector. Then, the optimal values of $\mathbf{O}$, and $\boldsymbol{t}$ can be calculated by minimizing the sum of squared distances between points, namely

$$R^2 = \sum_{i=1}^{n} (\boldsymbol{b}_i - \mathbf{O}^\top \boldsymbol{a}_i + \mathbf{t})^\top (\boldsymbol{b}_i - \mathbf{O}^\top \boldsymbol{a}_i + \mathbf{t}).$$

The optimal values obtained are then applied to the original A configuration leading to the minimum R^2 value (called the Procrustes sum of squares). The computation for the projection model is slightly different from that for the distance model. Since the projection model is not invariant under translation, the optimization problem involving transformation of matrix A reduces to:

$$\boldsymbol{a}_i^\top = \mathbf{O}^\top \boldsymbol{a}_i.$$

Thus the modified sum of squared distances is

$$R^2 = \sum_{i=1}^{n} (\boldsymbol{b}_i - \mathbf{O}^\top \boldsymbol{a}_i)^\top (\boldsymbol{b}_i - \mathbf{O}^\top \boldsymbol{a}_i).$$

The problem of identifiability has been further studied by Shortreed et al. (2006) to extend the work of Hoff et al. (2002). They considered three point estimators for the positions of the nodes, namely, the maximum likelihood estimator, the posterior mode estimator, and the posterior mean estimator (defined after Procrustes matching). They argued that these three estimators may provide inaccurate point estimates of the nodes, since they rely exclusively on the estimated distances to derive the positions. They introduce an original fourth procedure which aims at minimizing the Kullback-Leibler divergence between the distribution implied by the fitted positions, and the one implied by the parameters minimizing the expected posterior loss. This fourth approach is most commonly used (Handcock et al., 2007; Krivitsky et al., 2009) and is implemented in the `latentnet` `R` package.

2.4.3 Other inferential approaches

The computational complexity of the log-likelihood function in Eq. 2.6 is on the order of the square of the number of nodes. This is a problem when inferring LPMs, because likelihood-based procedures require a quadratic computing cost, which does not scale well with the size of the data. The full Bayesian approach, on other hand, based on MCMC is particularly slow and becomes impractical for networks of more than a few hundred nodes. On the other hand, the MLE approach tends to be faster, yet it still requires a quadratic cost and thus it too does not scale well.

A number of papers have specifically addressed this research impasse, by proposing various strategies that can speed up the inferential procedures and thus make the algorithms scale better with large datasets. The variational Bayesian inference procedure proposed by Salter-Townshend and Murphy (2013) proposes to replace the posterior distribution of the model parameters in the Latent position cluster model with a variational posterior. This leads to an approximation that has been widely used in a variety of different settings with latent variables (Blei et al., 2017; Jordan et al., 1999). The advantage of this approximation is that it allows one to characterize the posterior distribution of interest without having to resort to sampling methods like MCMC, which can be much more computationally intensive. However, the approximation error induced is difficult to quantify, and it has been shown that this error can be relevant when assessing the uncertainty around the point estimates provided.

Another approach is considered by Raftery et al. (2012), who introduce a likelihood approximation using case-control sampling. This takes advantage of the fact that the positions of the nodes are primarily regulated by the edges between nodes, rather than by the missing edges. This is especially true in sparse networks, where the few edges that are present give us information on which nodes should be located close to each other. By contrast, the missing edges tend to not provide as much information, so can be thinned out using a case-control sampling strategy. As a consequence, only a fraction of the missing edges are considered in the likelihood calculation, and their contributions are reweighted to make them represent all

the missing edges. This approximation yields an unbiased estimate of the likelihood function, while scaling up the procedure to networks of several thousand nodes.

Ryan et al. (2017) extend the distance LPCM by including conjugate priors in the model specification. The key advantage of this approach is that a number of model parameters can be integrated out from the posterior distribution thanks to the conjugacy property. This leads to a marginal posterior which can be characterized through a MCMC sampler. The advantage of this approach is that it can speed up the sampling procedure since many of the model parameters have been integrated out from the posterior. The sampler can explore different models with a different number of latent clusters in one run. This is a key advantage because it allows one to perform model choice on the number of groups without having to fit each of the models individually.

These methodological ideas were introduced for binary graphs, but they have been extended and adopted in a variety of different settings with dynamic and multiview networks. More recently, some other methodological ideas have been considered (Aliverti and Russo, 2021; Liu and Chen, 2022; Rastelli et al., 2024; Spencer et al., 2022) signalling that this research problem remains one of active interest.

2.5 An illustrative example for the distance LPM

We now describe a motivating example on an original coauthorship network. The data (collected from Scopus on 17 January, 2022) consists of all the bibliographic entries which cite Hoff et al. (2002). A total of 929 articles are part of this dataset, and 1,758 unique authors are involved. From these, we constructed a coauthorship network for all the authors of those papers, where two authors are linked if they coauthored at least one paper. We removed all the authors that published only one paper in total, to focus on the most active authors and to reduce the computational burden. This led to a final undirected binary network of 279 nodes.

We then fit a distance LPM on these data using the `R` package `latentnet`. We compared a number of models using the BIC criterion, as implemented in `latentnet`. The resulting latent space is shown in Figure 2.1, and a more detailed representation of the results through an interactive plot is publicly available on GitHub[1].

The results highlight a strong presence of communities, which reflect the presence of various research groups. Some of the authors tend to have overlapping positions, suggesting strong similarity of research collaborators. Others are positioned in between communities, highlighting that they may have connections with more than one research group or with a more diverse set of collaborators. In Figure 2.2 we show the posterior sample for the intercept parameter, and the distribution of the log-odds for all edges.

[1] URL for GitHub: https://github.com/riccardorastelli/CoauthorshipsLPM

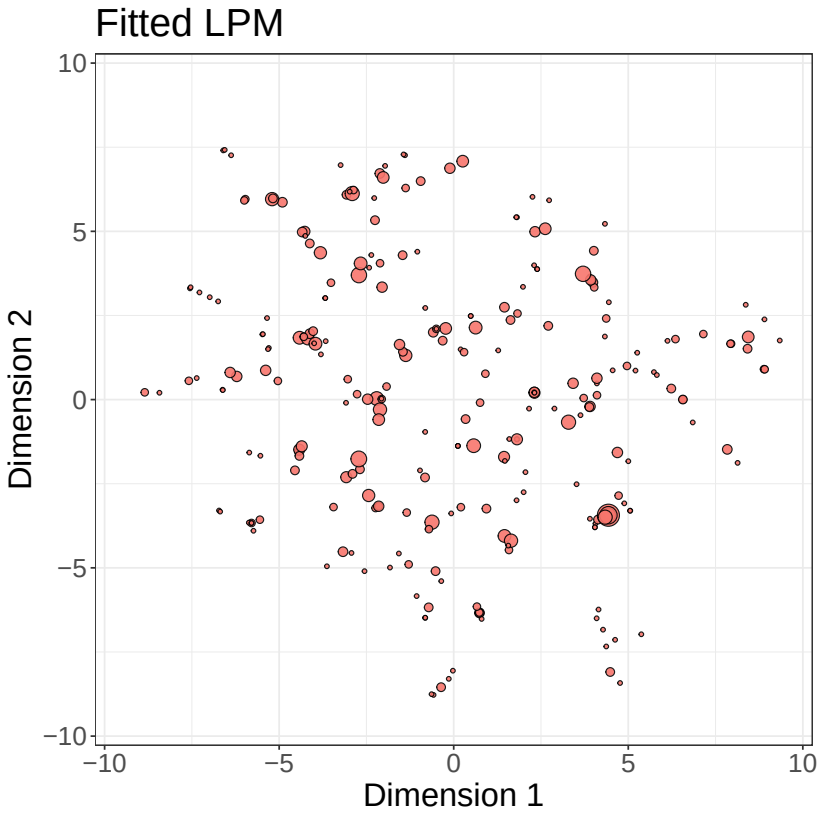

Fig. 2.1 Distance LPM fitted on the coauthorship data. Each node represents a different author. The size of a node represent the number of publications in which the author participated.

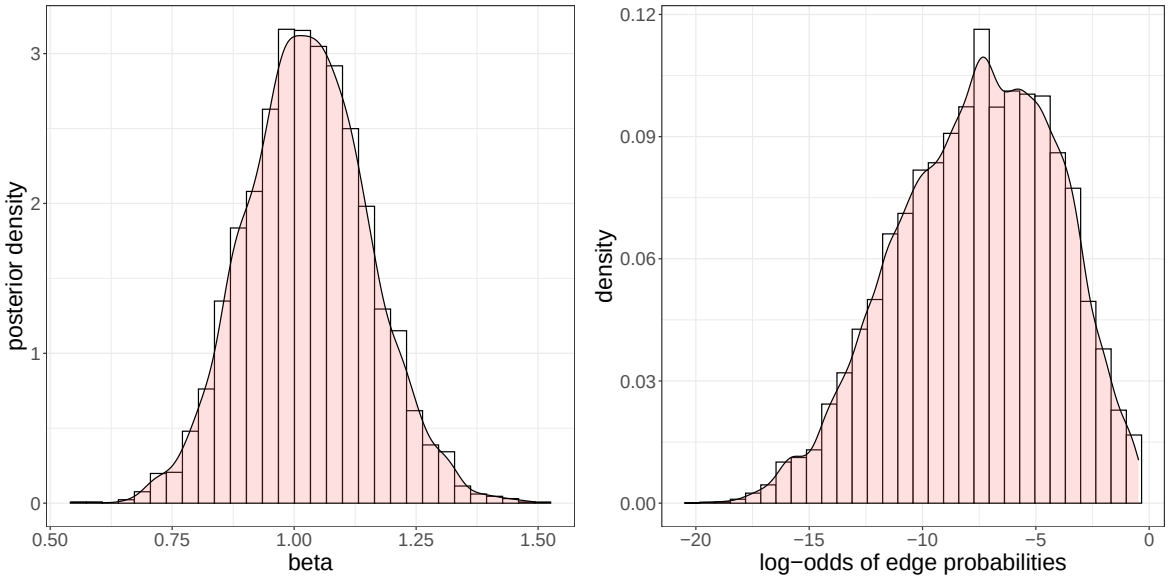

Fig. 2.2 The left panel shows the approximate posterior sample for the intercept parameter, called β. The right panel shows instead the η parameters representing the log-odds for the edges appearing. The distribution represents these values across all dyads.

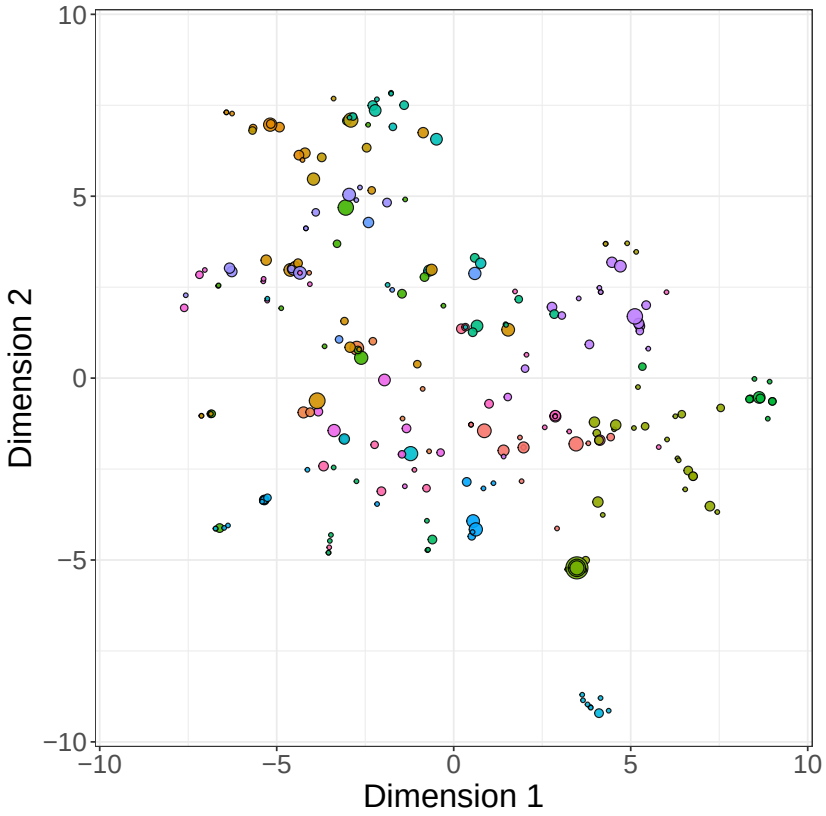

Fig. 2.3 Distance LPCM fitted on the coauthorship data. The size of the nodes represent the total number of publications to which the authors participated. The colors of the nodes represent their partitioning.

These plots show that the latent space plays an important role in determining the presence of edges. The left panel shows that the intercept parameter concentrates around the value 1, signalling a strong effect of the latent space. This is confirmed on the right panel of the same figure, where we can see that the log-odds exhibit large variability, implying edge probabilities that range from 0.00 to 0.38. This means that the model is able to represent the heterogeneity in the data by leveraging the geometric framework induced by the LPM.

In the fitted LPM of Figure 2.1 we can explore the presence of communities by studying how "clustered" the latent space is. However, the LPCM allows one to include this information directly in the modelling, and thus obtain a partitioning of the nodes into groups. For this reason, we also fitted a LPCM with up to 30 groups, and selected the best model using the BIC criterion advocated by Handcock et al. (2007). The results (27 groups) are shown in Figure 2.3. In this case we notice strong agreement between the partitioning of the nodes and the clustering of points in the latent space. The research groups are fairly well captured and they convey an interesting model-based view of the data.

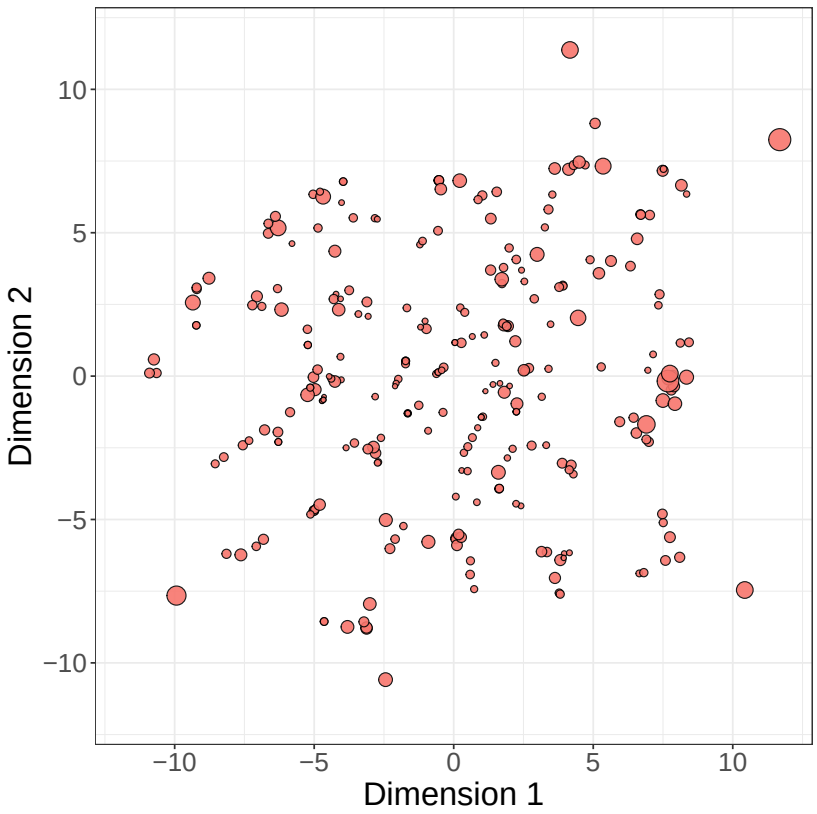

Fig. 2.4 Distance LPCMRE fitted on the coauthorship data. The size of the nodes represent their "sociality" random effect parameter.

In addition, we also provide a third example using the model of Krivitsky et al. (2009). Also in this case we fit the model for different numbers of groups and we select the best fit using BIC. The results for the best model (shown in Figure 2.4) are obtained when only one group is chosen. This is reasonable since the random effects used in the model can explain a substantial part of the variability in the data, at the expense of the latent space. Nonetheless, the visualization that the model can offer still includes all the information regarding the model parameters since the sociality of nodes is shown through their size.

In this case, we observe that the sociality is similar for most of the nodes that are positioned in the center of the latent space. These nodes are well characterized by their latent positions, and the model does not require any random effect to “correct” their connectivity patterns. Some nodes with a relatively central position also benefit from a high sociality parameter. This represents the fact that some authors will collaborate with diverse research groups and may require the creation of long-distance edges. The sociality parameter addresses this situation, helping to bridge that gap and to create connections regardless of the pairwise distance. However, some of the largest sociality values are observed for nodes in the fringes of the latent space. These authors do not have many connections; for this reason it is difficult

to position them centrally or close to other nodes. As a consequence, they are pushed to the edge of the latent space. Since they do have connections to more central nodes, a large sociality parameter can provide a compensation for their loose positioning thus well justifying their observed connections.

2.6 State of the art

The work of Hoff et al. (2002) has led to a strand of literature that focuses on LPMs and their extensions. In this section, we review some methodological contributions that extend the work of Hoff et al. (2002) in different directions. We organize our exposition by categorizing the papers into the type of network data they refer to. However many of the papers are generalizable to multiple contexts and provide novelty in a number of different ways.

2.6.1 LPMs for multiview networks

Multiview or multiplex networks are observed when there are multiple relationships among the same set of nodes, hence describing a multirelational network. An example of this framework may be a social network where we observe various types of ties between actors, such as professional ties, family ties or friendship ties. Also, temporal networks are closely connected to multiview networks as they may be seen as a special case where a time dependency is created to connect contiguous network views.

LPMs have been extended to the multiview framework in a number of papers. Gollini and Murphy (2016) introduced the latent space joint model which displays the different relations on the same nodes to be encapsulated in the same latent space. This also introduces a scalable variational approach to speed up the estimation of the latent space model. Another novelty is that they use the squared Euclidean distance instead of the absolute distance in their model.

Instead, Durante et al. (2017) develop an extension of the latent projection model of Hoff et al. (2002) using a novel non-parametric Bayesian approach to model the latent space and the clustering of the nodes. They propose an application to a population of brain connectivity networks, whereby nodes refer to anatomical brain regions and edges refer to the structural interconnections between them. They consider data collected from a number of different samples/individuals, which is where the multiview aspect of their analysis comes from. The application of their model can be seen in a subsequent paper (Durante and Dunson, 2018) in which the model is used to characterize and test for differences in brain connectivity networks. A second paper in a similar direction is Aliverti and Durante (2019).

Salter-Townshend and McCormick (2017) propose a new model for multiview networks which focuses on the interrelation of dyads across different network views. They use a

multivariate Bernoulli likelihood whose parameters are determined by multiple latent spaces and by the distances between nodes in different views. The framework permits a model-based assessment of the correlation between the multiple edges, providing a measure of association between the various network views. The price to pay for this additional flexibility however, is increased computational requirements and a more challenging model-choice task.

D'Angelo et al. (2019) and D'Angelo et al. (2020) introduce a Euclidean distance LPM for multiview networks to model the voting patterns in the Eurovision contest. The authors specify a common latent space to represent the positions of the nodes, but they let the other parameters be different across the various network views. This is a key aspect of the model since it directly allows one to quantify the effect that the latent space can have on the generation of the data. Similarly to Eq. 2.4, a coefficient for the latent distances approaching zero would mean that the latent distances do not represent a strong effect, and thus the data can be well represented through a homogeneous model. The authors use this aspect of the LPM to determine the effect of the latent space as a way to measure the voting bias of the countries.

Sewell (2019) introduces a new method to analyze cognitive social structures based on the LPM. These data refer to the perception that each node has of the network, so it is related to multiview networks in that the observed data correspond to a collection of networks. He derives an LPM approach which combines the advantages of the latent space visualization, with additional model parameters which can measure the nodes' biases in perceiving the network.

2.6.2 LPMs for dynamic networks

Dynamic or temporal networks refer to networks that evolve over time. Most typically, dynamic networks describe the interactions between the same set of nodes, and are observed through network snapshots, i.e. at some specific points in time. For this reason, they can be formally represented as a collection of graphs $Y^{(1)}, \ldots, Y^{(K)}$. The LPM of Hoff et al. (2002) has been extended to dynamic networks in a variety of ways. Sarkar and Moore (2005) extend it to include time-dynamics in the model parameters. They propose an application to a dataset of friendship relationships and study how these change over time. The proposed model embeds the nodes in the latent space in such a way that the positions of nodes are regularized at each time point to ensure that their positions do not change by much.

Sewell and Chen (2015b) introduced a new model to extend that of Sarkar and Moore (2005). In their framework, the longitudinal network data is modeled via temporal trajectories in a latent Euclidean space. One key novel aspect of this work is that the formulated model introduces a social reach parameter which resembles the nodal random effects of Krivitsky et al. (2009). These parameters aim at representing different levels of activity of the nodes and their sociability, regardless of the latent position.

Friel et al. (2016). define a model for bipartite dynamic networks, and apply their method to a network of companies and their directors in Ireland. The model aims at providing a model-based assessment of how appointments of directors to multiple boards may have been associated with financial instability during the 2008 crisis. The authors embed the nodes of bipartite networks into a single latent space, similarly to Gormley and Murphy (2008). They assume that the nodes can move along trajectories, hence capturing the time dependencies in the data. In addition, they note that the data present a strong persistence of edges and non-edges. This refers to dyads that tend not to change their status over time. In order to address this aspect of the data, which is common to many dynamic network datasets, the authors also introduce a regime-switching framework which selects a different intercept parameter based on the previous network realization. The model captures the heterogeneity shown by the data very well through latent positions and persistence features by modeling the structure of the intercept parameters.

Durante and Dunson (2016) further extend the LPM by allowing the positions of nodes to evolve continuously over time through a nested Gaussian process, which results in time-varying smoothness. The authors incorporate locally adaptive dynamics to study face-to-face dynamic contact network, highlighting the flexibility of their approach in capturing data observed at unequally spaced intervals and structural changes in the nodes trajectories.

In a similar direction, Rastelli and Corneli (2023) introduce a fully continuous latent position model, which extends the application of LPMs to time-stamped interaction data. They consider and infer the continuous trajectories of the nodes in a two-dimensional latent space, using only their frequency of pairwise connections.

2.6.3 LPMs for weighted networks

Weighted networks are characterized by edges that carry additional numerical information. Typically, the edges of these networks carry a real value, for example representing the intensity or the length of the interaction. Although binary networks are easier to study, weighted networks arise often in practice. Hoff (2005, 2021) proposed a general framework that includes various types of latent variable models as special cases, including the distance LPM and the projection LPM. The network variables are considered as the response variables of a regression model, and are characterized by a combination of nodal effects, dyadic effects and other types of network-derived information. Westveld and Hoff (2011) extended this framework to model a network of international trade and conflict.

Sewell and Chen (2016) introduce a dynamic weighted network framework. They define the temporal dependencies through a Markovian property, and consider two models for the weights: a Poisson model for count data and a Tobit model for nonnegative continuous data. They also consider nodal random effects similarly to Sewell and Chen (2015b), to define a possibly different social reach for each node.

2.6.4 LPMs for rank data

Ranked network data can be used to represent rankings of items or entities by the nodes. A bipartite network can be used to represent how a set of nodes rank various items, or the other nodes themselves. Gormley and Murphy (2008) combined the LPM of Hoff et al. (2002) with the Plackett-Luce model (Plackett, 1975) to create a new type of LPM which can be used to study bipartite rank networks, and applied their method to the voters' preferences in the 2002 Irish elections.

Sewell and Chen (2015a) further extended the approach of Gormley and Murphy (2008) to account for temporal changes in the networks as well as a social reach effect on the nodes. Their LPM is dynamic in that the nodes move along trajectories, and their positions are used to characterize the various network snapshots at different times. This permits an analysis of the evolution of the latent space, and thanks to the node-specific parameters, it allows for an assessment of the popularity and stability of individual actors. In both papers, inference is carried out using Bayesian framework and Procrustes matching. As concerns scalability, the methods proposed are particularly demanding since they require an even higher complexity than the original LPM. This points to an important research question for future work in this direction.

2.6.5 Theory of LPMs

The papers we consider in this section are general in that they are not restricted to a particular network type. We first highlight a number of theoretical papers that have studied the LPM framework or some closely connected variants. Rastelli et al. (2016) provide a detailed analysis of the LPM framework by considering some theoretical and empirical aspects that arise from the model structure. They address how the LPM can capture some relevant network features of interest such as clustering, heavy-tailed degree distributions, small world behaviuor, and assortative mixing. They also propose a variant of Eq. 3.4 in which the logistic expression is replaced by a Gaussian kernel. This choice permits a detailed analysis of the theoretical properties of the model, creating a connection to the physics literature on the topic (Newman, 2018).

Smith et al. (2019) focus on the role played by the geometry of the latent space in an LPM. They create a general framework which includes a variety of latent variable models, and they study how switching to a hyperbolic geometry can lead to more flexible network models which can better represent the observed data. In particular, they show that a negative curvature in the space permits more complexity without the need to change any other aspect of the models, hence providing a more parsimonious modeling choice.

Caron and Fox (2017) consider a general latent variable framework, and argue that commonly used frameworks such as the LPM may not be adequate for large networks since

they cannot capture sparsity in asymptotic settings. They then propose a new framework based on exchangeable random measures which is capable of maintaining ideal features like network sparsity along with exchangeability properties.

2.6.6 Extensions of the projection LPM

The distance LPM and the projection LPM introduced by Hoff et al. (2002) generated two quite separate sets of literature. On the one hand, the distance LPM has been widely applied thanks to its clear representations and easy interpretability. On the other hand, the projection model has been intensely studied due to its tractability and also its similarity to other types of latent factor models. In the projection LPM literature, the work of Hoff (2005) has initiated a research avenue (Hoff, 2007, 2011, 2021) which has evolved into a flexible and widely applicable framework for network analysis.

Nickel (2008) and Young and Scheinerman (2007) introduce a general latent variable multiplicative structure which is closely connected to the projection model, and call this the Random Dot Product graph model. In its original formulation, the model does not include the logit link for the multiplicative effects. However, this becomes an advantage because it makes the model quite tractable, so a number of network properties can be derived analytically. A number of recent papers extend the random dot product graph and use this framework in network applications (Athreya et al., 2021; Ng and Murphy, 2019; Passino et al., 2021; Xie and Xu, 2021; Zhang et al., 2020).

2.6.7 Extensions of the LPCM

The distance LPCM of Handcock et al. (2007) has been an especially influential since it inherits the advantages of the basic LPM of Hoff et al. (2002) while gaining more flexibility and interpretability through the clustering structure. This has led to a rich literature about this model and its applications.

Fosdick et al. (2019) introduce a new latent variable network model inspired by the LPCM. They argue that the complexity of networks may be better captured when different statistical models are combined at different level of resolution. This means that the models may be either stacked into different hierarchical layers, or they may model different aspects of the data. They illustrate their approach by proposing a new model, the latent space stochastic blockmodel, which combines the stochastic blockmodel of Wang and Wong (1987) with the LPM of Hoff et al. (2002). Nodes are clustered into groups following a stochastic blockmodel structure, while the probabilities of connections for nodes belonging to the same group are determined by a group-specific LPM. This effectively creates different resolutions in the model where the inner level, represented by the LPM, can break down and capture the within-group

heterogeneous connectivity patterns. The result is a framework that can scale better than that of Handcock et al. (2007) while maintaining the flexibility and interpretability of the LPMs.

Another approach which combines LPMs and clustering is that of Sewell (2020). He considers a clustering framework on the edges rather than on the nodes. This reflects the idea that edges are determined within a particular context they refer to, for example because the nodes interacted within that context. He then introduces a latent space to formalize the presence of these contexts that drive the formation of the edges. As a result, the nodes involved in a particular edge are chosen based on their relative positions with respect to the context of the edge. The methodology can provide latent space representations (as well as clustering) of edges and nodes, hence creating a strong connection with the LPCM. However, differently from the LPCM, the computational complexity of the approach proposed scales with the number of edges in the graph, so it may be particularly suitable for sparse networks.

A dynamic extension of the LPCM is introduced by Sewell and Chen (2017). In this work, the authors consider a dynamic LPM where both the latent positions of the nodes, and their cluster allocations, can change over time. The the position of a node is influenced by its previous position and its previous cluster membership. The clustering variables are assumed to have a Markovian structure and depend on their previous value. The authors consider an estimation strategy based on variational inference (for large networks), and another strategy based on MCMC (for smaller networks due to higher computational costs). Also, they perform model choice using BIC to select the number of clusters, and DIC to select the number of latent dimensions.

Gormley and Murphy (2010) extend the framework of Handcock et al. (2007) to include node covariates. While the LPCM can accommodate edge covariates directly into the edge probabilities through the logit link, the inclusion of a node's information is not as straightforward. Gormley and Murphy (2010) address this gap by proposing a mixture of experts' models as a prior for the latent positions, whereby the nodes' covariates can be used to characterize the group allocation probabilities associated to the nodes. This defines a framework in which covariates can be introduced via the likelihood, and also into the structure of the model, providing more flexibility and possibly a different interpretation of the data.

2.7 Software

An essential aspect of latent position modeling is the implementation of the methods, which include MCMC, variational methods and other procedures. MCMC estimation of a range of latent position network models is carried out by the `R` package `latentnet` (Krivitsky and Handcock, 2008; Krivitsky et al., 2024). This implements the methodologies outlined in Handcock et al. (2007); Hoff et al. (2002); Krivitsky et al. (2009); Shortreed et al. (2006). This package has been used to produce the examples in this paper.

2.8 Open research questions

LPMs have had great success in the last two decades thanks to their ability to capture some common network properties, such as transitivity, homophily and clustering. Research on these models is active, and several methodological questions remain open.

One key issue concerns computational efficiency and scalability. While this is a relevant topic throughout network science, it is especially important in the context of LPMs. As mentioned earlier, the original LPM has a quadratic cost which arises from the likelihood calculation. The model requires information from each dyad of the network, and this dyadic information is possibly always different, generating the quadratic cost. While a number of ideas have been proposed, the current approaches can only scale to thousands of nodes, whereas observed datasets can be much larger. This gap highlights the need to define models and methods that can take computational efficiency directly into account, while providing a solid structure for statistical analyses.

Another aspect of the research on LPMs refers to its flexibility in capturing features of interest. A closely related latent variable model, the stochastic blockmodel, can capture both assortative and disassortative behaviors of the nodes. The original LPM falls short in this regard, since the model can capture assortative mixing but it cannot capture disassortative mixing. While social networks tend to often show assortative mixing, other types of networks such as biological networks can exhibit strong disassortative patterns (Newman, 2018). For this reason, variations of the LPM have been recently considered to make the model more flexibile in this regard, for example by combining it with the stochastic blockmodel (Fosdick et al., 2019; Ng et al., 2021; Rastelli, 2018).

Another key open research question relates to model choice, and, in particular, to the selection of the dimensionality of the latent space. Most commonly, LPMs are fitted in two latent dimensions, due to easier interpretations and clearer visualizations. However, a number of papers have highlighted how the choice of this parameter, the number of latent dimensions, can be set up as one of model choice. A BIC criterion was introduced by Handcock et al. (2007) to select the number of clusters and the number of latent dimensions jointly. However, the authors note that the asymptotic properties of BIC may not be valid when selecting the number of dimensions. A Deviance Information Criterion (DIC, Spiegelhalter et al. (2002)) has been used by D'Angelo et al. (2019); Friel et al. (2016); Sewell and Chen (2017). However, more research on the model choice issue is needed.

2.9 Acknowledgements

This publication has emanated from research supported in part by a grant from the Insight Centre for Data Analytics which is supported by Science Foundation Ireland under Grant number $12/RC/2289\_P2$.

References

Aliverti, E. and Durante, D. (2019). Spatial modeling of brain connectivity data via latent distance models with nodes clustering. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 12(3):185–196.

Aliverti, E. and Russo, M. (2021). Stratified stochastic variational inference for high-dimensional network factor model. *Journal of Computational and Graphical Statistics*, 31(2):502–511.

Athreya, A., Tang, M., Park, Y., and Priebe, C. (2021). On estimation and inference in latent structure random graphs. *Statistical Science*, 36(1):68–88.

Berlusconi, G., Aziani, A., and Giommoni, L. (2017). The determinants of heroin flows in europe: A latent space approach. *Social Networks*, 51:104–117.

Blei, D., Kucukelbir, A., and McAuliffe, J. (2017). Variational inference: A review for statisticians. *Journal of the American statistical Association*, 112(518):859–877.

Caron, F. and Fox, E. (2017). Sparse graphs using exchangeable random measures. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 79(5):1295.

Chiu, G. and Westveld, A. (2011). A unifying approach for food webs, phylogeny, social networks, and statistics. *Proceedings of the National Academy of Sciences of the United States of America*, 108(38):15881–15886.

Chiu, G. and Westveld, A. (2014). A statistical social network model for consumption data in trophic food webs. *Statistical Methodology*, 17(C):139–160.

Chu, A., Chan, T., So, M., and Wong, W.-K. (2021). Dynamic network analysis of covid-19 with a latent pandemic space model. *International Journal of Environmental Research and Public Health*, 18(6):3195.

D'Angelo, S., Alfò, M., and Murphy, T. (2020). Modeling node heterogeneity in latent space models for multidimensional networks. *Statistica Neerlandica*, 74(3):324–341.

D'Angelo, S., Murphy, T., and Alfò, M. (2019). Latent space modelling of multidimensional networks with application to the exchange of votes in eurovision song contest. *Annals of Applied Statistics*, 13(2):900–930.

Ding, X., Zou, Z., and Brooks III, C. (2019). Deciphering protein evolution and fitness landscapes with latent space models. *Nature Communications*, 10(1).

Durante, D. and Dunson, D. (2016). Locally adaptive dynamic networks. *Annals of Applied Statistics*, 10(4):2203–2232.

Durante, D. and Dunson, D. (2018). Bayesian inference and testing of group differences in brain networks. *Bayesian Analysis*, 13(1):29–58.

Durante, D., Dunson, D., and Vogelstein, J. (2017). Rejoinder: Nonparametric bayes modeling of populations of networks. *Journal of the American Statistical Association*, 112(520):1547–1552.

Erdös, P. and Rényi, A. (1959). On random graphs I. *Publicationes Mathematicae Debrecen*, 6:290–297.

Fosdick, B., McCormick, T., Murphy, T., Ng, T., and Westling, T. (2019). Multiresolution network models. *Journal of Computational and Graphical Statistics*, 28(1):185–196.

Fraley, C. and Raftery, A. (1998). How many clusters? which clustering method? answers via model-based cluster analysis. *The Computer Journal*, 41(8):578–588.

Friel, N., Rastelli, R., Wyse, J., and Raftery, A. (2016). Interlocking directorates in irish companies using a latent space model for bipartite networks. *Proceedings of the National Academy of Sciences of the United States of America*, 113(24):6629–6634.

Gilbert, E. (1961). Random plane networks. *Journal of the Society for Industrial and Applied Mathematics*, 9(4):533–543.

Gollini, I. and Murphy, T. (2016). Joint modeling of multiple network views. *Journal of Computational and Graphical Statistics*, 25(1):246–265.

Gormley, I. and Murphy, T. (2008). *A latent space model for rank data*, pages 90–102.

Gormley, I. and Murphy, T. (2010). A mixture of experts latent position cluster model for social network data. *Statistical Methodology*, 7(3):385–405.

Handcock, M., Raftery, A., and Tantrum, J. (2007). Model-based clustering for social networks (with discussion). *Journal of the Royal Statistical Society. Series A: Statistics in Society*, 170(2):301–354.

Hoff, P. (2005). Bilinear mixed-effects models for dyadic data. *Journal of the American Statistical Association*, 100(469):286–295.

Hoff, P. (2007). Modeling homophily and stochastic equivalence in symmetric relational data. *NIPS'07: Proceedings of the 21st International Conference on Neural Information Processing Systems*, 20:657 – 664.

Hoff, P. (2011). Hierarchical multilinear models for multiway data. *Computational Statistics & Data Analysis*, 55(1):530–543.

Hoff, P. (2021). Additive and multiplicative effects network models. *Statistical Science*, 36(1):34–50.

Hoff, P., Raftery, A., and Handcock, M. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460):1090–1098.

Jin, I. and Jeon, M. (2019). A doubly latent space joint model for local item and person dependence in the analysis of item response data. *Psychometrika*, 84(1):236–260.

Jordan, M., Ghahramani, Z., Jaakkola, T., and Saul, L. (1999). An introduction to variational methods for graphical models. *Machine learning*, 37(2):183–233.

Kim, B., Lee, K., Xue, L., and Niu, X. (2018). A review of dynamic network models with latent variables. *Statistics Surveys*, 12:105–135.

Krivitsky, P. and Handcock, M. (2008). Fitting position latent cluster models for social networks with latentnet. *Journal of Statistical Software*, 24(5):1–23.

Krivitsky, P., Handcock, M., Raftery, A., and Hoff, P. (2009). Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Social networks*, 31(3):204–213.

Krivitsky, P., Handcock, M., Shortreed, S., Tantrum, J., Hoff, P., Wang, L., Li, K., Fisher, J., and Bates, J. (2024). *Package 'latentnet'*. The Statnet Project. R package version 2.11.0.

Lee, W., McCormick, T., Neil, J., Sodja, C., and Cui, Y. (2021). Anomaly detection in large scale networks with latent space models. *Technometrics*, 64(2):241–252.

Linardi, F., Diks, C., van der Leij, M., and Lazier, I. (2020). Dynamic interbank network analysis using latent space models. *Journal of Economic Dynamics and Control*, 112:103792.

Liu, H., Jin, I., Zhang, Z., and Yuan, Y. (2021). Social network mediation analysis: A latent space approach. *Psychometrika*, 86(1):272–298.

Liu, Y. and Chen, Y. (2022). Variational inference for latent space models for dynamic networks. *Statistica Sinica*, 32(4):2147–2170.

Matias, C. and Robin, S. (2014). Modeling heterogeneity in random graphs through latent space models: a selective review. *ESAIM: Proceedings and Surveys*, 47:55–74.

McFowland III, E. and Shalizi, C. (2021). Estimating causal peer influence in homophilous social networks by inferring latent locations. *Journal of the American Statistical Association*, 118(541):707–718.

Newman, M. (2018). *Networks*. Oxford university press.

Ng, T. and Murphy, T. (2019). Generalized random dot product graph. *Statistics & Probability Letters*, 148:143–149.

Ng, T., Murphy, T., Westling, T., McCormick, T., and Fosdick, B. (2021). Modeling the social media relationships of irish politicians using a generalized latent space stochastic blockmodel. *The Annals of Applied Statistics*, 15(4):1923–1944.

Nickel, C. (2008). *Random dot product graphs a model for social networks*. PhD thesis, Johns Hopkins University.

Passino, F., Bertiger, A., Neil, J., and Heard, N. (2021). Link prediction in dynamic networks using random dot product graphs. *Data Mining and Knowledge Discovery*, 35(5):2168–2199.

Plackett, R. (1975). The analysis of permutations. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 24(2):193–202.

Raftery, A. (2017). Comment: Extending the latent position model for networks. *Journal of the American Statistical Association*, 112(520):1531–1534.

Raftery, A., Niu, X., Hoff, P., and Yeung, K. (2012). Fast inference for the latent space network model using a case-control approximate likelihood. *Journal of Computational and Graphical Statistics*, 21(4):901–919.

Rastelli, R. (2018). The sparse latent position model for nonnegative weighted networks. *arXiv: 1808.09262*.

Rastelli, R. and Corneli, M. (2023). Continuous latent position models for instantaneous interactions. *Network Science*, 11(4):560–588.

Rastelli, R., Friel, N., and Raftery, A. (2016). Properties of latent variable network models. *Network Science*, 4(4):407–432.

Rastelli, R., Maire, F., and Friel, N. (2024). Computationally efficient inference for latent position network models. *Electronic Journal of Statistics*, 18(1):2531 – 2570.

Ryan, C., Wyse, J., and Friel, N. (2017). Bayesian model selection for the latent position cluster model for social networks. *Network Science*, 5(1):70–91.

Salter-Townshend, M. and McCormick, T. (2017). Latent space models for multiview network data. *Annals of Applied Statistics*, 11(3):1217–1244.

Salter-Townshend, M. and Murphy, T. (2013). Variational bayesian inference for the latent position cluster model for network data. *Computational Statistics and Data Analysis*, 57(1):661–671.

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

Sarkar, P. and Moore, A. (2005). Dynamic social network analysis using latent space models. In *Proceedings of the 18th International Conference on Neural Information Processing Systems*, NIPS'05, pages 1145–1152. MIT Press.

Sewell, D. (2019). Latent space models for network perception data. *Network Science*, 7(2):160–179.

Sewell, D. (2020). Model-based edge clustering. *Journal of Computational and Graphical Statistics*, 30(2):390–405.

Sewell, D. and Chen, Y. (2015a). Analysis of the formation of the structure of social networks by using latent space models for ranked dynamic networks. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, 64(4):611–633.

Sewell, D. and Chen, Y. (2015b). Latent space models for dynamic networks. *Journal of the American Statistical Association*, 110(512):1646–1657.

Sewell, D. and Chen, Y. (2016). Latent space models for dynamic networks with weighted edges. *Social Networks*, 44:105–116.

Sewell, D. and Chen, Y. (2017). Latent space approaches to community detection in dynamic networks. *Bayesian analysis*, 12(2):351–377.

Shortreed, S., Handcock, M., and Hoff, P. (2006). Positional estimation within a latent space model for networks. *Methodology*, 2(1):24–33.

Smith, A., Asta, D., and Calder, C. (2019). The geometry of continuous latent space models for network data. *Statistical Science*, 34(3):428–453.

Snijders, T. and Nowicki, K. (1997). Estimation and prediction for stochastic blockmodels for graphs with latent block structure. *Journal of Classification*, 14(1):75–100.

Sosa, J. and Buitrago, L. (2021). A review of latent space models for social networks. *Revista Colombiana de Estadistica*, 44(1):171–200.

Spencer, N., Junker, B., and Sweet, T. (2022). Faster MCMC for Gaussian latent position network models. *Network Science*, 10(1):20–45.

Spiegelhalter, D., Best, N., Carlin, B., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639.

Sweet, T. and Adhikari, S. (2020). A latent space network model for social influence. *Psychometrika*, 85(2):251–274.

Sweet, T., Thomas, A., and Junker, B. (2013). Hierarchical network models for education research: Hierarchical latent space models. *Journal of Educational and Behavioral Statistics*, 38(3):295–318.

Tafakori, L., Pourkhanali, A., and Rastelli, R. (2022). Measuring systemic risk and contagion in the european financial network. *Empirical economics*, 63(1):345–389.

Wang, Y. and Wong, G. (1987). Stochastic blockmodels for directed graphs. *Journal of the American Statistical Association*, 82(397):8–19.

Ward, M., Ahlquist, J., and Rozenas, A. (2013). Gravity's rainbow: A dynamic latent space model for the world trade network. *Network Science*, 1(1):95–118.

Westveld, A. and Hoff, P. (2011). A mixed effects model for longitudinal relational and network data, with applications to international trade and conflict. *The Annals of Applied Statistics*, 5(2A):843–872.

Wilson, J., Cranmer, S., and Lu, Z.-L. (2020). A hierarchical latent space network model for population studies of functional connectivity. *Computational Brain and Behavior*, 3(4):384–399.

Xiao, T., Oppenheimer, M., He, X., and Mastrorillo, M. (2022). Complex climate and network effects on internal migration in south africa revealed by a network model. *Population and Environment*, 43:289–318.

Xie, F. and Xu, Y. (2021). Efficient estimation for random dot product graphs via a one-step procedure. *Journal of the American Statistical Association*, 118(541):651–664.

Young, S. and Scheinerman, E. (2007). Random dot product graph models for social networks. In *International Workshop on Algorithms and Models for the Web-Graph*, pages 138–149. Springer.

Zhang, X., Xue, S., and Zhu, J. (2020). A flexible latent space model for multilayer networks. volume 119 of *Proceedings of Machine Learning Research*, pages 11288–11297. PMLR.

Chapter 3

A latent space model for multivariate count data time series analysis

Abstract

Motivated by a dataset of burglaries in Chicago, USA, we introduce a novel framework to analyze time series of count data combining common multivariate time series models with latent position network models. This novel methodology allows us to gain a new latent variable perspective on the crime dataset that we consider, allowing us to disentangle and explain the complex patterns exhibited by the data, while providing a natural time series framework that can be used to make future predictions. Our model is underpinned by two well known statistical approaches: a log-linear vector autoregressive model, which is prominent in the literature on multivariate count time series, and a latent projection model, which is a popular latent variable model for networks. The role of the projection model is to characterize the interaction parameters of the vector autoregressive model, thus uncovering the underlying network that is associated with the pairwise relationships between the time series. Estimation and inferential procedures are performed using an optimization algorithm and a Hamiltonian Monte Carlo procedure for efficient Bayesian inference. We also include a simulation study to illustrate the merits of our methodology in recovering consistent parameter estimates, and in making accurate future predictions for the time series.

Keywords: time series; vector autoregressive models; latent position models; statistical network analysis.

3.1 Introduction

Modeling of multivariate count time series have become prominent in various applied fields. Historically, such analyses were predominantly utilized in economics and finance literature. In recent years, and in particular over the last decade, multivariate time series analysis has been increasingly associated with network analyses. In such cases, these interrelated time series are visualized through networks, with each time series acting as a node and the connections between them represented as edges. This network representation allows for a deeper understanding of the complex interdependencies within the data. For example, it helps elucidate how the behavior or attributes of individual nodes may influence or be influenced by other nodes in the network at any given time. This approach not only enhances predictive accuracy but also provides richer insights into the dynamic interactions within the data, thereby improving decision-making across various domains such as epidemiology (Dahlhaus, 2000), finance (Billio et al., 2012; Diebold and Yılmaz, 2014; Härdle et al., 2016), biology (Shojaie and Michailidis, 2010), and social network analysis (Zhu and Pan, 2020).

The dataset that we consider in this paper describes the number of burglaries in a number of areas of south Chicago, USA (Clark and Dixon, 2021). The data exhibit complex patterns of interdependencies which are rooted in the geographical and social layout of the city (Mohler et al., 2011; Yuan et al., 2019). Trends in the number of crimes can be associated to events and other types of social dynamics, and their frequency can exhibit self-exciting or cross-exciting patterns, which may be resembling a contagious process (Johnson, 2008; Mohler et al., 2011). Disentangling and inferring these complex patterns of the data can be a very challenging task, since they are not directly observed and typically cannot be objectively measured.

In this paper, we integrate a type of Latent Position Model (LPM) to model the hidden network representing the relations between the time series. Our objective is to ultimately uncover the network's latent structure and then use this to give a model-based visualization of the time series data, thus obtaining a highly interpretable output that can be used for prediction or qualitative analyses. This extends the existing literature on multivariate count time-series network models by introducing a novel methodology, which can directly address the modeling challenges that are posed by complex datasets such as the crime dataset that we consider here.

More in detail, we modify a multivariate log-linear count time series model by incorporating a network modeling framework, namely the LPM. This integration is novel within the domain of multivariate count time series with network structures, particularly because the network structure is not observed and has to be inferred from the time series data. The rationale for integrating LPM into the multivariate time series context is to facilitate a straightforward and interpretable visualization, which aids in understanding pairwise relationships and unraveling

the complex network structure. Moreover, our approach not only provides a geometric interpretation of these relationships but also quantifies the strength of interdependencies between the time series. In our study, the connections between series are not merely binary; they manifest as positive, negative, or neutral, reflecting various types of interdependences. This nuanced view of connectivity introduces additional complexity into the model, opening up a relatively new research direction for the latent variable network models. Overall, our approach considerably enhances both theoretical insight and practical application in the study of network structures associated to multivariate count time series.

3.1.1 Literature review

Multivariate time series analysis comes into play in determining appropriate functions for a system of variables, generating reliable forecasts, and capturing the relationships and interactions among different entities. To model such data, we can consider each time series individually, using, for example, with an autoregressive model (Box et al., 2015). However, the consequence of such an approach would be a loss in capturing cross-sectional dependencies. The Vector AutoRegressive (VAR) model (Luetkepohl, 2005) is widely used for modeling multivariate time series data with continuous response variables. However, as the number of time series increases, the complexity of the model also increases, leading to computationally demanding estimation and over-parametrization. Moreover, when dealing with multivariate time series for continuous variables, models such as the Vector AutoRegressive Moving Average (VARMA) (Luetkepohl, 2005) or multivariate Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) (Francq and Zakoian, 2019) are also commonly used for analyses.

As regards the literature on multivariate time series analysis for count data, we note that the literature on this is significantly less advanced compared to the continuous case. Some recent contributions in this topic are, for example, in electrical system reliability (Ertekin et al., 2015); neuroscience (Brown et al., 2004; Hall and Willett, 2015); econometrics (Bermúdez and Karlis, 2011); in finance (Pedeli and Karlis, 2013b) and in environmental science (Livsey et al., 2018).

In the field of multivariate time series analysis, particularly for modeling systems with count data, various methodological approaches have been proposed. Cox et al. (1981) classified time series models into two main categories: parameter-driven models and observation-driven models. In parameter-driven models, parameters vary over time based on some dynamic unobserved processes. These models are also termed as state space models. Some of the applications of these models for univariate time series modeling are discussed in Zeger (1988), Harvey and Fernandes (1989), Durbin and Koopman (2000), and Frühwirth-Schnatter and Wagner (2006). In addition, Jørgensen et al. (1996), Jung et al. (2011), and Ravishanker et al. (2014) studied multivariate state space models. The estimation procedure of these

models is based on likelihood approaches or on Bayesian approaches. These models are simple but computationally demanding.

Lastly, we have a class of observation-driven models. This is a class where the current parameters are considered as deterministic functions of lagged dependent variables, along with contemporaneous and lagged exogenous variables. In this approach, the parameters change randomly over time, yet they are predictable for one step ahead using past information. While univariate observation-driven models are quite prominent and widely discussed in the statistical literature; see for example Davis et al. (2000); Davis and Liu (2012); Fokianos et al. (2009); Fokianos and Tjøstheim (2011), multivariate observation-driven count time series models represent a relatively novel and emerging area of research.

Within this growing field, multivariate INteger AutoRegressive (INAR) models stand out as one of the foundational contributions, first introduced by Franke and Rao (1993) and Latour (1997) introduced multivariate INAR models. The application of these models has been discussed in the papers by Pedeli and Karlis (2013a), Pedeli and Karlis (2013b), Scotto et al. (2014) and Darolles et al. (2019). Estimation methods for INAR models typically involve classical least squares and likelihood-based approaches. However, the application of likelihood theory, even in the simpler context of univariate INAR models, becomes particularly complex with higher-order models. Consequently, despite their suitability for modeling certain basic data structures, these models present significant challenges in estimation and prediction, particularly when dealing with models of a higher order.

Beyond INAR models, the literature on observation-driven count time series models has expanded to include linear and log-linear models, which serve as simple and popular generalizations of Poisson autoregressive models for count data. Heinen and Erick (2007), Andreassen (2013), and Lee et al. (2018) primarily focused on studying linear models. The properties of the linear model have been discussed in Streett (2000), and Ferland et al. (2006). Another relevant contribution is by Fokianos and Tjøstheim (2011), who talked about several advantageous characteristics of log-linear models. Finally, Fokianos et al. (2020) discussed both multivariate linear and log-linear models and used a copula-based construction for the joint distribution of the counts.

We now review the literature that connects time series models with network models. The main objective of these models is to estimate the pairwise relationship between the entities and predict future events or outcomes. A commonly studied type of multivariate time series associated with a network structure refers to situations where we observe time series data on the nodes of a network. In a recent work, Knight et al. (2016) introduced the Network AutoRegressive (Integrated) Moving Average (NARIMA) models, defining multivariate continuous time series that are coupled with a network structure. These models are designed to effectively handle dynamic variations in the network structure associated with multivariate time series. Recently, Zhu et al. (2017) introduced a Network vector AutoRegressive model

(NAR), a variant of the traditional VAR model that integrates a network structure. This model assumes that the value of a node depends not only on its own historical values but also on the average historical values of its neighboring nodes. The NAR model further incorporates node-specific characteristics or covariates, significantly simplifying the complexity by relying on a limited set of parameters. This simplification enhances manageability and interpretability. Inference is performed using least squares in two asymptotic regimes: (a) increasing time sample size ($T \to \infty$) with a fixed number of nodes N, and (b) both N and time frames T increasing. This methodology was further expanded by Knight et al. (2020) who put forward the Generalized Network Autoregressive model (GNAR) for continuous random variables. In this setting, they introduced the node-specific effects based on neighborhood sizes.

A different strand of literature is related to the work by Barigozzi and Brownlees (2019), who introduces a network methodology for analyzing large panels of time series data. The authors introduced a sparse VAR model by specifically focusing on two representations: a directed graph representing predictive Granger relations and an undirected graph representing contemporaneous partial correlations. This approach provides a parsimonious synthesis of the data and offers insights into the underlying structure of the panel. Zhou et al. (2020) introduce a network GARCH model that leverages information obtained from a well-defined network structure. Their approach reduces the number of unknown parameters and significantly simplifies the computational complexity involved in the model. Some further extensions of these models are discussed in Zhu et al. (2019b) that captures the changing behavior of quantiles over time. Chen et al. (2023) enhances the flexibility of the model by Zhu et al. (2017) by introducing the capability to accommodate diverse network effects across various network communities within the autoregressive model. The model proposed by Zhu et al. (2019a) takes into account the sparse structure of network effects in the Network Autoregressive Model (NAM), allowing for the possibility of heterogeneity among network nodes. Zhu and Pan (2020) proposed the grouped network vector autoregression (GNAR) model, which groups network nodes by shared characteristics, with each group defined by unique parameters to capture specific behaviors and interactions. Zhu et al. (2020) introduced a spatial autoregressive model tailored for large social networks. Additionally, Kang et al. (2022) developed a model that integrates multi-scale modeling with network-based neighborhood selection, aiming to capture temporal local structures and identify significant temporal network changes while maintaining sparsity in interactions.

A central reference for our work is Zhu et al. (2017) which combines multivariate time series with network structures and its extensions discussed so far considers continuous response data. The literature on the discrete response variable is sparse and has received relatively less attention in research. This gap is notable despite the applicability of such data across various domains, including social media, email communication, collaboration networks, disease spread modeling, financial systems, transportation, and ecological food

webs. Our work contributes to this sparse literature. It relates to the multivariate time series with underlying network structure considered in the studies mentioned earlier. The key difference is that, in our setting, the time series are observed on the nodes of the network and take the form of counting processes. Additionally, the weights on the edges quantify various types of interdependencies.

Theoretical and methodological advancements in modeling networks for multivariate count time series are currently active areas of ongoing research. Some more recent work by Armillotta and Fokianos (2024) has made a significant contribution by demonstrating the suitability of multivariate observation-driven count time series models for modeling time-varying network data. Their extension of the NAR model to count data led to the formulation of the Poisson Network Autoregression (PNAR) model. This model uses linear and log-linear approaches and incorporates copulas to capture interdependencies among count variables. It also allows for alternative distributions like the negative binomial. Their work extensively analyzes two asymptotic regimes discussed in Zhu et al. (2017) and advances a theoretical framework for asymptotic inference. The authors employ Quasi Maximum Likelihood Estimation (QMLE) for parameter estimation.

Bracher and Held (2022) also discusses an application of a model akin to the Network Autoregressive (NAR) model for multivariate count data. Armillotta et al. (2022) in their work, discuss a statistical framework that consolidates the findings from previous studies by Zhu et al. (2017) and Armillotta and Fokianos (2024). The authors present a comprehensive framework that accommodates the case of both continuous and count responses measured over time for each node of a known network, demonstrating its practical potential. For high-dimensional discrete data, Mark et al. (2017) constructs a mathematical model that includes a statistical learning guarantee for estimating parameters in complex multivariate point process models. Pandit et al. (2020) presents a p-lag multivariate discrete valued autoregressive model. The main challenge in this model is the presence of a large number of unknown parameters that need to be estimated. The analysis is done using a $l1$-regularized M-estimator when the loss functions are strongly convex.

With regards to the networks literature, our work pivots on LPMs, which were introduced and popularized as a statistical model for social networks by Hoff et al. (2002). The highly influential work of Hoff et al. (2002) lead to a significant increase in the literature on LPMs. However, despite this increase in research, we are not aware of any work integrating LPMs with multivariate time series, and, specifically, count time series data. This conceptual link represents one of our original contributions to both the networks and time series research literature. One paper in this area is by Ahelegbey et al. (2020), which presents a Bayesian hierarchical model combining a Covariance Graphical Model (CGM), and a VAR model. However, their work focused on financial time series, whose values are continuous and not counts. This is a major distinction from our paper, as their contribution adds to the literature

on multivariate continuous time series using specifically a sparse network structure. Another significant difference is that Ahelegbey et al. (2020) model the network using a CGM. In contrast, in our paper, we use a LPM framework both for modeling and inference purposes.

Another relevant contribution in this context is that of Tafakori et al. (2022), where a latent space approach is used to model default probabilities of financial institutions. In a similar framework, latent variable models have also been employed by Hledik and Rastelli (2023) to study systemic risk and financial stability. To our knowledge, no work has specifically applied LPM to the study of multivariate count time series data, highlighting a clear gap in the research literature.

The remainder of this paper is structured as follows: Section 3.2 describes the Poisson Log-linear model, LPM, and the proposed Time series Latent Position Model (TSLPM). Section 4.4 outlines the inferential process. Section 4.5 contains simulation studies exploring the performance of the TSLPM in terms of parameter estimation and prediction. Section 3.5 includes the application of the proposed model using the Chicago burglary dataset. Section 4.8 concludes the paper with a conclusion and discussion on the possible extensions.

3.2 Time Series Latent Position Model (TSLPM)

3.2.1 Poisson log-linear models

Let $\mathbf{y}_t = \{y_{it},\ i = 1, 2, \ldots, N\}$, for $t = 1, 2, \ldots, T$, denote a multivariate count time-series made of N time series and T time points. Assume a Poisson distribution on $\mathbf{y}_t$ and let $\boldsymbol{\lambda}_t = \{\lambda_{it},\ i = 1, \ldots, N\}$, for $t = 1, \ldots, T$, be the corresponding intensity process. Define $\mathscr{F}_t$ be the σ-field generated by $\mathbf{y}_{t-1}$. According to the model specification, we make the assumption that $\boldsymbol{\lambda}_t = E(\mathbf{y}_t|\mathscr{F}_{t-1})$.

Considerable emphasis has been placed on observation-driven models, particularly the Poisson log-linear model, for the estimation and inference of multivariate time series for count data (Davis et al., 2021; Doukhan et al., 2020; Fokianos, 2024). A Poisson log-linear model of order 1 introduced by Fokianos et al. (2020), is formalized as follows:

$$\begin{aligned} \mathbf{y}_t|\mathscr{F}_{t-1}^{\mathbf{Y},\boldsymbol{\lambda}} &\sim \text{Pois}(\boldsymbol{\lambda}_t) \\ \log(\boldsymbol{\lambda}_t) &= \boldsymbol{\alpha} + \Gamma \log(\mathbf{y}_{t-1} + \mathbf{1}_N) \end{aligned} \tag{3.1}$$

where $\log(\boldsymbol{\lambda}_t)$ is the log link applied on all elements of $\boldsymbol{\lambda}_t$ and $\mathbf{1}_N$ is a N-dimensional unit vector. The value of $\log(\boldsymbol{\lambda}_t)$ is determined by the lagged values of the actual time series via $\log(\mathbf{y}_{t-1} + \mathbf{1}_N)$ in Eq. 3.1. In the given model, $\boldsymbol{\alpha}$ is a vector of dimension N, with entries that

can either be fixed or vary across the series, and Γ is a matrix of size $N \times N$. These are the parameters that must be estimated.

By contrast, the linear model variant for multivariate Poisson VAR(1) discussed by Fokianos et al. (2020) is defined for each $i = 1, 2, \ldots, N$, as follows:

$$\begin{aligned} \mathbf{y}_t | \mathscr{F}_{t-1}^{\mathbf{Y},\boldsymbol{\lambda}} &\sim \text{Pois}(\boldsymbol{\lambda}_t) \\ \boldsymbol{\lambda}_t &= \boldsymbol{\alpha} + \Gamma \mathbf{y}_{t-1} \end{aligned} \tag{3.2}$$

In this case, the entries of $\boldsymbol{\alpha}$ and Γ are all assumed to be non-negative. This constraint is imposed on the model parameters to ensure that the intensity vector is always non-negative.

Despite its convenient formulation, the linear model has several limitations that result in the log-linear model typically being a more suitable modeling choice; see Fokianos and Tjøstheim (2011) for univariate and Fokianos (2024) for multivariate time series modeling and related discussions. The rationale behind favoring a log-linear autoregressive models over linear autoregressive models is that the former do not impose any constraints on the parameters regarding positivity, facilitating the analysis of datasets exhibiting both positive and negative correlations. Additionally, the log-linear model may be expanded through the inclusion of covariates thus making it an optimal choice for a more comprehensive statistical analysis of multivariate count time series. For a detailed discussion on these models, see Fokianos (2024).

3.2.2 Latent position models

LPMs (Hoff et al., 2002) have gained prominence for their ability to offer a readily interpretable visualization of a network via a latent social space. The extensive body of literature on LPMs has found widespread application across various disciplines, particularly within the domain of social network analysis. For network data, we often represent the relationships between nodes using an $N \times N$ symmetric adjacency matrix, which we can generically denote as $\mathbf{A}$. In this context, N refers to the number of observed nodes, and the entries a_{ij} in the matrix indicate the relationship between node i and node j. Typically, an adjacency matrix is a binary matrix of 1s and 0s which describe the presence or absence of an edge, respectively. In this section we review this binary framework, although we note that later in this paper we will be more interested in a weighted network framework, where a_{ij} can take any value in $\mathbb{R}$, again, to represent the type of relation that the two corresponding nodes have.

Let $\mathbf{Z}$ be a $N \times d$ matrix representing the coordinates of the N nodes in a d-dimensional social space. Using the notation $\mathbf{z}_i$, we refer to the i-th row of the matrix $\mathbf{Z}$, which indicates the d coordinates of the i-th node in the latent space. We highlight that inference aimed at estimating the number of latent dimensions d from the data is becoming a prominent aspect of the research on LPMs; see Handcock et al. (2007), Sewell and Chen (2017), Friel et al.

(2016), D'Angelo et al. (2020), and Gwee et al. (2023) for their contribution. However, in this paper, we only focus on the case $d = 2$. This choice facilitates a clear visualization of the network structure and it simplifies inference. This choice is indeed very common in the literature and the simplification allows us to focus on and develop other aspects of our methodology. We emphasize that the choice of the number latent dimensions is an important future direction of our work.

We adopt a modeling approach based on conditional independence, assuming that each edge is determined solely by the positions of the nodes, without direct influence from other edges in the network. Consequently, the probability of the presence of an edge existing between nodes i and j is expressed as a function that captures the relationship between the latent positions $\mathbf{z}_i$ and $\mathbf{z}_j$ associated with these nodes, given by Eq. 3.3.

$$\mathbb{P}(\boldsymbol{Y}|\boldsymbol{Z}, \alpha) = \prod_{i<j} \mathbb{P}(y_{ij}|\mathbf{z}_i, \mathbf{z}_j, \boldsymbol{\theta}) \tag{3.3}$$

here, the product notation is intended over all pairs of nodes (i, j) such that $i < j$, for $i, j = 1, 2, \ldots, N$, whereas $\boldsymbol{\theta}$ is the collection of parameters that do not refer to edges or nodes.

In these models, the nodes with similar characteristics (i.e. positions) tend to possess a higher probability of forming an edge between them. The definition of similarity in latent characteristics can be approached in various ways, and Hoff et al. (2002) has introduced two distinct approaches known as the distance model and the projection model. For a comprehensive review of both models, see (Kaur et al., 2023; Salter-Townshend et al., 2012).

The distance model by Hoff et al. (2002) posits that as the proximity between two nodes increases, the probability of a connection between them also increases. Conversely, as the distance between nodes grows larger, the probability of a connection decreases. This framework provides an easily interpretable visualization of geometric social space. The log odds that node i and j will form a tie are usually given as:

$$\zeta_{ij} = \text{logodds}(y_{ij} = 1|\mathbf{z}_i, \mathbf{z}_j, \alpha) = \alpha - |\mathbf{z}_i - \mathbf{z}_j| \tag{3.4}$$

where $\alpha \in \mathcal{R}$ is called the intercept and $|\cdot|$ represents the Euclidean norm. The Euclidean distance is frequently employed for this purpose, although alternative distance functions can be explored. For a few related contributions to the latent distance model, see Handcock et al. (2007) for clustering of highly connected nodes in the network via a latent mixture mode, Krivitsky and Handcock (2008) which uses actor-specific random effects along with model-based clustering of nodes, D'Angelo et al. (2019) and D'Angelo et al. (2020) for introducing a Euclidean distance LPM for multi-view networks to model. Further related contributions have appeared over the last years, including (Aliverti and Durante, 2019; Gollini and Murphy, 2016; Gormley and Murphy, 2008; Hoff, 2021; Rastelli and Corneli, 2023).

An alternative latent position model, also proposed by Hoff et al. (2002) is called the projection model. This model proposes that two nodes are highly likely to form a tie if their respective latent positions are pointing in the same direction ($\mathbf{z}_i^\top \mathbf{z}_j \gg 0$), or conversely, they are very unlikely to connect if they point in opposite directions ($\mathbf{z}_i^\top \mathbf{z}_j \ll 0$). In other words, the probability that the two nodes will connect is higher when the angle between them is small, and the probability is lower when the angle gets closer to 180 degrees. Here, $(.)^\top$ represents the transpose of a vector or matrix. The log odds for this model are defined as:

$$\zeta_{ij} = \text{logodds}(y_{ij} = 1|\mathbf{z}_i, \mathbf{z}_j, \alpha) = \alpha + \mathbf{z}_i^\top \mathbf{z}_j \tag{3.5}$$

Here, the magnitude of the dot product between the latent positions $\mathbf{z}_i^\top \mathbf{z}_j$ of node i and node j quantifies the degree of similarity or dissimilarity between the nodes. In the latent projection model, it is more straightforward to visualize the latent space using polar coordinates, where the positions of nodes can be understood as a combination of direction and magnitude (i.e. distance from the center of the space). This polar coordinate representation allows for a clearer understanding of the relationship between the nodes, with the direction indicating their orientation and the magnitude representing the strength of their latent characteristics. The statistical literature centered around this model includes Hoff (2005), Hoff (2007), Hoff (2011), Hoff (2021) ,Young and Scheinerman (2007), and Nickel (2008). Some applications of this model have been explored by Durante et al. (2017) and Durante and Dunson (2018), where the model is used to characterize and test for differences in brain connectivity networks.

Both the latent distance model and the latent projection model provide meaningful insights through geometric latent space representations. The latent distance models are popular due to their clarity and intuitive interpretations. However, latent projection models offer advantages over distance models, particularly in their ability to accommodate both positive and negative relationships between entities. Since our primary aim is to effectively visualize and quantify the pairwise relationships among time series, which may be involve either positive or negative interactions, we focus our attention on the latent projection model. Ultimately, this model can facilitate a model-based visualization of the data using a latent embedding of the series.

3.2.3 The TSLPM

Model definition

We introduce a hierarchical model consisting of a multivariate log-linear autoregressive time series model combined with latent projection network model. We call our model the Time Series Latent Position Model (TSLPM), of order 1, and define it as follows:

$$
\begin{aligned}
y_{it}|\mathscr{F}_{t-1} &\sim \text{Pois}(\lambda_{it}) \\
\log(\lambda_{it}) &= \alpha_i + \sum_{j=1}^{N} \gamma_{ij} \log(y_{j(t-1)} + 1) + \sum_{k=1}^{K} \delta_k x_{ik} \\
\gamma_{ij} &= \begin{cases} \beta_i & i = j \\ \mathbf{z}_i^\top \mathbf{z}_j & i \neq j \end{cases}
\end{aligned} \tag{3.6}
$$

for all $i = 1, \ldots, N$ and $t = 1, \ldots, T$. The model parameters are $\boldsymbol{\alpha} \in \{\alpha_1, \ldots, \alpha_N\} \in \mathbb{R}^N$, $\boldsymbol{\beta} = \{\beta_1, \ldots, \beta_N\} \in \mathbb{R}^N$, $\mathbf{Z} = \{\mathbf{z}_i \in \mathbb{R}^2 | i = 1, \ldots, N\}$ and $\boldsymbol{\delta} = \{\delta_1, \ldots, \delta_K\} \in \mathbb{R}^K$.

The model states that the Poisson log rate of the marginal conditional mean of any series i is determined by two main components. The first component is the log rate of the previous count of series i itself, associated to the autoregressive coefficient β_i. The second component consists of the log rates of the counts of all the other series, each weighted by a suitable coefficient γ_{ij}, for all js. The parameter α_is, which may be consistent across all series or vary among them, allows for the possibility of a non-zero mean process and significantly affects the long-term forecasts generated by the model. Another aspect of this model is the presence of covariates, indicated in Eq. 3.6 by the term x_{ik}, for K observed covariates. These additional variables are included in the log-linear model using the coefficients $\delta_1, \ldots, \delta_K$. We point out that, similarly to the $\boldsymbol{\alpha}$ parameter, the parameters δ_k may be equal across all series or differ among them. In this version of the model, we consider the same α and the same δs for all series.

The parameter γ_{ij} plays a significant role in informing the network structure that identifies the interrelations between the time series. In the TSLPM, the (i, j)th element of matrix Γ, represented by γ_{ij}, corresponds to the autoregressive coefficients in the Poisson log-linear model specified by the Eq. 3.1. We modify the structure of the log-linear model by incorporating the projection modeling approach on the elements of matrix Γ. Consequently, each non-diagonal entry of the matrix Γ is the dot product of the latent positions associated to two different series, series i, and series j, thus measuring the dependence of the series $\mathbf{y}_i$ and $\mathbf{y}_j$. As regards interpretation of these parameters, the relationships between series are significantly influenced by the angular proximity of their latent position vectors. Vectors that are closely aligned, demonstrated by smaller angles between them, are associated with a positive association between the two corresponding series. Conversely, vectors that form obtuse angles, are indicative of negative associations. Orthogonal vectors indicate the absence of a meaningful relationship, i.e. the two series in this case are conditionally independent given the model parameters.

Conditions to impose stationarity

In this section, we discuss on stationarity and ergodicity of the family of time series that we consider in this paper, building on the discussions presented by Doukhan et al. (2020) and Fokianos et al. (2020).

We define the ℓ^1-norm of a matrix Γ as $\|\Gamma\|_1 = \underset{1\leq j\leq N}{max} \sum_{i=1}^{N} |\gamma_{ij}|$, which corresponds to the maximum absolute column sum of the matrix. Furthermore, the spectral norm of a matrix Γ is defined as $\|\Gamma\|_2 = \sqrt{\rho(\Gamma^\top \Gamma)} = \sigma_{max}(\Gamma)$, where $\rho(.)$ denotes the spectral radius of a matrix and σ_{max} is the largest singular value of Γ. In the case where Γ is a symmetric matrix in $\mathcal{R}^{N\times N}$ (i.e., $\Gamma^\top = \Gamma$) then $\|\Gamma\|_2 = \rho(\Gamma) = \underset{1\leq i\leq N}{max} |\Omega_i|$, where, Ω_i denotes the ith eigenvalue of matrix Γ.

The TSLPM model is a variant of the Poisson VAR(1) model. It is perhaps worth pointing out the stability condition of the VAR model. The condition is as follows:

Proposition 1. *(Luetkepohl, 2005) Assume $\mathbf{y}_t = (y_{1t}, y_{2t}, ..., y_{Nt})^\top$ denotes an $N \times 1$ vector of time-series. The basic VAR of order 1 has the form:*

$$\mathbf{y}_t = \alpha + \Gamma \mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t, \quad t = 1, \ldots, T$$

where Γ is a $N \times N$ matrix with autoregressive coefficients and $\boldsymbol{\epsilon}_t$ is an $N \times 1$ zero mean white noise vector process.

The VAR(1) is stable if the roots of $det(\mathbf{I}_N - \Gamma z) = 0$ lie outside the complex unit circle (i.e., have a modulus greater than one), or, equivalently, if all the eigenvalues of the matrix Γ have absolute value less than one. Assuming the process has been initialized in the infinite past, a stable VAR(1) process is stationary and ergodic with time-invariant means, variances, and autocovariances.

Next, we discuss the conditions that are critical for achieving stationarity and ergodicity of our TSLPM, borrowing from the literature on related models. The main goal is to obtain stationarity and ergodicity of the joint process $(\mathbf{y}_t, \boldsymbol{\lambda}_t)$ due to the dependency structure between the count data $\mathbf{y}_t$ and the intensity process $\boldsymbol{\lambda}_t$.

Proposition 2. *(Doukhan et al., 2020; Fokianos et al., 2020) Consider the model of Eq.* (3.1)*. Suppose that $\|\Gamma\|_1 < 1$. Then there exists a unique causal solution {$(\mathbf{y}_t, \log(\boldsymbol{\lambda}_t)$} to the model of Eq.* (3.1) *which is stationary, ergodic and satisfies E$\|\log(\mathbf{y}_t + \mathbf{1}_N)\|_r^r < \infty$ and E$\|\log(\boldsymbol{\lambda}_t)\|_r^r < \infty$ and E$[\exp(r\|\log(\boldsymbol{\lambda}_t)\|_1)] < \infty$ for any $r \in \mathcal{N}$.*

The above proposition also holds true whenever $\|\Gamma\|_2 < 1$, thus this condition is sufficient to guarantee stationarity and ergodicity of the joint process $(\mathbf{y}_t, \boldsymbol{\lambda}_t)$. For proof and detailed discussion, see Fokianos et al. (2020). Leveraging the above results, we have obtained an alternative condition for stationarity and ergodicity, i.e. the condition $\underset{1\leq i\leq N}{max} |\Omega_i(\Gamma)| < 1$.

We now discuss further on conditions that are sufficient to achieve stationarity, extending Proposition 2. We rely here on the Gershgorin Circle Theorem (GCT) which has been a cornerstone for introducing bounds on eigenvalues of a square matrix.

Proposition 3. *(Varga, 2011) Given a real symmetric $N \times N$ matrix Γ with generic element γ_{ij}, define the Gershgorin disc for the i-th row as the closed disc $D(\gamma_{ii}, r_i)$ centered at γ_{ii} with radius r_i, which is given by:*

$$r_i = \sum_{i \neq j} |\gamma_{ij}|$$

where r_i is the sum of the absolute values of the non-diagonal entries in the i-th row. According to the Gershgorin Circle Theorem (GCT), every eigenvalue of matrix Γ lies within at least one of the Gershgorin discs $D(\gamma_{ii}, r_i)$. Consequently, the boundaries of the union of all discs establish the potential range of the eigenvalues. Thus, $\min_i(\gamma_{ii} - r_i)$ and $\max_i(\gamma_{ii} + r_i)$ establish the lower and upper bounds on the eigenvalues, respectively:

$$\min_i(\gamma_{ii} - r_i) < \Omega_i(\Gamma) < \max_i(\gamma_{ii} + r_i)$$

Hence, the Proposition 3 supplements these criteria by providing explicit bounds on the eigenvalues, derived from the Gershgorin disc analysis. This introduces a rigorous set of necessary conditions to guarantee stationarity of a TSLPM:

$$-1 < \min_i(\gamma_{ii} - r_i) < |\Omega_{max}(\Gamma)| < \max_i(\gamma_{ii} + r_i) < 1 \tag{3.7}$$

Satisfying the above inequality ensures the stationarity and ergodicity of the joint process $(\boldsymbol{y}_t, \boldsymbol{\lambda}_t)$ ensuring that the process remains well-behaved over time. Critically, we note that these conditions are sufficient, but not necessary for stationarity and ergodicity of the joint process.

Example. For instance, let us assume that we model a multivariate time series $\{Y_1, \dots Y_5\}$ using a TSLPM of order 1. Let us consider the following interaction matrix Γ to analyze and determine the stationarity and ergodicity of the associated multivariate time series process.

$$\Gamma = \begin{bmatrix} 0.12 & 0.02 & 0.08 & 0.08 & -0.14 \\ 0.02 & 0.20 & -0.02 & -0.02 & 0.10 \\ 0.08 & -0.02 & 0.07 & 0.06 & -0.13 \\ 0.08 & -0.02 & 0.06 & 0.06 & -0.12 \\ -0.14 & 0.10 & -0.13 & -0.12 & 0.27 \end{bmatrix}$$

By applying the Gershgorin Circle Theorem (GCT) as stated in Proposition 3, we obtain 5 discs given by $D(0.12, 0.32),\ D(0.2, 0.16),\ D(0.07, 0.29),\ D(0.06, 0.28)$ and $D(0.27, 0.49)$. These discs suggest that the eigenvalues of Γ are bounded as follows:

$$-0.22 < \Omega_i(\Gamma) < 0.76$$

Given the actual eigenvalues of Γ, which are $0.5,\ 0.21,\ 0.01,\ 0.003,\ -0.0009$, we confirm that the largest eigenvalue is 0.5. This observation is consistent with the established bounds and also satisfies the condition from Eq. 3.7.

Prior specification

We apply a Bayesian estimation framework to estimate the model parameters. Specifically, we incorporate prior information for the parameters α, $\boldsymbol{\beta}$, $\mathbf{Z}$, and $\boldsymbol{\delta}$ by assigning independent and noninformative priors to each of them. Each parameter, namely α, $\boldsymbol{\beta}$, $\mathbf{Z}$, and $\boldsymbol{\delta}$ is assigned an independent normal prior with a mean of zero and a standard deviation of 100. This choice of a high standard deviation is deliberate, aiming to minimize the influence of prior assumptions on the posterior distributions. By employing broad, noninformative priors, we ensure that the analysis remains less biased by prior beliefs, allowing the observed data to play a more significant role in shaping the posterior densities. Furthermore, the specifications allow α, $\boldsymbol{\beta}$, $\mathbf{Z}$, and $\boldsymbol{\delta}$ to take both positive and negative values, reflecting their potential impacts in the model.

Parameter interpretation

In this section we discuss more in detail how the TSLPM parameters can be interpreted. To facilitate this, we also express our TSLPM model using an alternative form which is equivalent to the form given in Eq. 3.6:

$$\lambda_{it} = \exp\left(\alpha_i + \sum_{k=1}^{K} \delta_k x_{ik}\right) (y_{i(t-1)} + 1)^{\beta_i} \prod_j (y_{j(t-1)} + 1)^{\gamma_{ij}} \tag{3.8}$$

The exponential function in the model translates the log-linear relationship into a multiplicative form for $\boldsymbol{\lambda}$. This approach provides a different perspective on the interpretation of parameters, focusing on proportional changes rather than additive changes. In other words, this emphasis is particularly meaningful for predicting the effect of individual parameters on the count data generated from TSLPM.

To begin with, the node specific intercept α and covariates effects $\boldsymbol{\delta}$ determine a baseline for the count data. The effect of α on the Poisson rate is common across all series and is multiplicative through an exponential function, so a positive α will tend to increase the

time series values, and a negative α will tend to decrease them. Similarly, the coefficients δ can add more nuances by differentiating the baselines for the series using the covariates information.

The series-specific autoregressive parameter β determines instead the magnitude of the multiplicative effect of the preceding period on current counts. A value equal to zero means that the preceding time series value has no effect on the current one. A higher β value gives a typical autoregressive structure with positive autocorrelations and sustained trends. By contrast, a negative value suggests an alternation of positive and negative autocorrelations and a more erratic behavior.

As regards γ_{ij}, each of these parameters determines the strength of a multiplicative effect given by $y_{j(t-1)} + 1$. Similarly to the β parameters, also the γ_{ij} are autoregressive terms that capture the interactions between the series. A value of zero indicates no associations between series i and j. A value greater than zero gives positive cross autocorrelations and sustained trends. A negative value suggests a more erratic behavior through negative cross autocorrelations.

3.3 Inferential procedure

3.3.1 Estimation

In this paper, we explore two main approaches for the inference of the TSLPM. The first approach involves optimizing the likelihood function using the Limited-memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) algorithm. This algorithm is particularly suited for large-scale optimization problems, where the computing cost increases at least with the square of the number of series. The LBFGS algorithm is efficient in this optimization task as it utilizes limited memory to store information from previous iterations, thereby circumventing the need for full Hessian matrix computation. The algorithm iteratively updates the current candidate solution by efficiently combining past gradients and curvature differences. The steps of the LBFGS algorithm as discussed by Nocedal and Wright (2006) are outlined in the Algorithm 1.

The algorithm is not guaranteed to converge to a global maximum of the objective function, however, it will converge to a local optimum and it generally shows good performance on our problem using very little computation demand.

The second approach which we have opted for is Hamiltonian Monte Carlo (HMC). HMC is a Markov chain Monte Carlo (MCMC) algorithm that efficiently navigates high-dimensional spaces by using derivatives of the density function to inform sampling decisions from the posterior distribution (Betancourt and Girolami, 2015; Neal, 2012). Unlike simpler sampling methods such as the random-walk Metropolis or Gibbs sampling (Duane et al., 1987; Neal,

Algorithm 1 The LBFGS algorithm

Require: 1. Choose an initial guess for all the model parameters $\alpha, \boldsymbol{\beta}$, and $\mathbf{Z}$. Set $r = 0$.

2. Compute the gradient of the objective function at the current point.

3. Update the search direction p_r by approximating the inverse Hessian matrix $\mathbf{H}_r^{-1}$ using information from the previous iterations.

4. Select a step size that sufficiently reduces the objective function.

5. Update the current point by taking the step determined in the line search.

6. Store the updated estimate of the inverse Hessian matrix by using the information from the current iteration and the previous iterations.

7. Check for convergence criteria. If the criteria are met, stop. Otherwise, increment r and repeat steps 2-6.

2012), HMC leverages the gradient information of the log-density function, facilitating a more effective exploration of the parameter space.

Hamiltonian dynamics are utilized in the Hamiltonian Monte Carlo (HMC) method to simulate two distinct components. The first component is a d-dimensional parameter space, denoted as $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots \theta_d)$. The second component comprises d-dimensional auxiliary momentum variables, $\mathbf{m} = (m_1, m_2, \ldots, m_d)$. These momentum variables typically follow a multivariate normal distribution. This distribution has a mean of zero and is characterized by a covariance matrix ($\boldsymbol{\Sigma}$). The purpose of this setup is to help rotate and scale the target distribution, as described by Betancourt and Stein (2011) in their work on the geometry of Hamiltonian Monte Carlo.

The joint distribution from which HMC samples can be expressed reads as follows:

$$\mathsf{p}(\boldsymbol{\theta}, \mathbf{m}) \propto \exp(\mathcal{L}(\boldsymbol{\theta}) - \frac{1}{2}\mathbf{m}^\top \mathbf{m})$$

where $\mathcal{L}(\boldsymbol{\theta})$ denotes the log-probability for $\boldsymbol{\theta}$. This augmented non-normalized model acts as a fictitious Hamiltonian system reflecting the total negative energy of the particle $\log(\mathsf{p}(\boldsymbol{\theta}, \mathbf{m}))$ for some particle's negative potential energy $\mathcal{L}(\boldsymbol{\theta})$ and the kinetic energy $\frac{1}{2}\mathbf{m}^\top \mathbf{m}$. The HMC algorithm procedures are detailed in Algorithm 2.

The algorithm (Hoffman et al., 2019) employs the leapfrog integrator to numerically simulate the Hamiltonian dynamics over time. This integrator preserves volume and is reversible, conserving the Hamiltonian to an approximation. Given the step size ϵ and time

point t, the leapfrog integrator updates the states as follows:

$$\begin{aligned}\mathbf{m}_{t+\frac{\epsilon}{2}} &= \mathbf{m}_t + \frac{\epsilon}{2}\nabla_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta}_t)\\ \boldsymbol{\theta}_{t+\epsilon} &= \boldsymbol{\theta}_t + \epsilon\, \mathbf{m}_{t+\frac{\epsilon}{2}}\\ \mathbf{m}_{t+\epsilon} &= \mathbf{m}_{t+\frac{\epsilon}{2}} + \frac{\epsilon}{2}\nabla_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta}_{t+\epsilon})\end{aligned} \tag{3.9}$$

Following L repetitions of these steps, the proposed state $(\hat{\boldsymbol{\theta}}, \hat{\mathbf{m}})$ is evaluated using a Metropolis acceptance step in Eq. 3.10 to decide on its inclusion in the Markov chain. If the proposal is not accepted, the previous parameter value is returned for the next draw and used to initialize the next iteration.

Algorithm 2 The HMC algorithm

Require: 1. Choose the initial state $s_{t-1} = (\boldsymbol{\theta}, \mathbf{m})$ from the parameter state of interest.

2. Sample the new momentum variable $\mathbf{m}^{\star} \sim N(\mathbf{0}, \mathbf{1})$.

3. Update position and momentum $(\boldsymbol{\theta}, \mathbf{m}^{\star})$ by applying L leapfrog steps starting from state $s^{*}_{t-1} = (\boldsymbol{\theta}, \mathbf{m}^{\star})$.

4. The new state $\hat{s_t} = (\hat{\boldsymbol{\theta}}, \hat{\mathbf{m}})$ is proposed.

5. The Metropolis algorithm is applied to calculate the acceptance probability of the proposal $(\hat{\boldsymbol{\theta}}, \hat{\mathbf{m}})$, which has been generated by transitioning from the state $s^{*}_{t-1} = (\boldsymbol{\theta}, \mathbf{m}^{\star})$ and is given as:

$$\mathsf{p}_{accept}(s^{*}_{t-1}) = min\left\{1, \frac{p(\hat{\boldsymbol{\theta}}, \hat{\mathbf{m}})}{p(\boldsymbol{\theta}, \mathbf{m}^{*})}\right\} \tag{3.10}$$

6. Accept the move from s^{*}_{t-1} to s_t with probability p_{accept}, otherwise the current state is preserved within the sample and used to initialize the next iteration.

The leapfrog integrator is accurate up to $\mathcal{O}(\epsilon^2)$. To maintain a high target acceptance rate (commonly set at 0.8), the algorithm adapts to use smaller step sizes ϵ. This can improve sampling efficiency (effective samples per iteration) at the cost of increased L (number of iterations) to keep the total distance traveled roughly constant. This increase is costly due to the higher number of gradient computations per iteration as indicated in the leapfrog steps in Eq. 3.9. Thus, the goal is to balance the acceptance rate to optimize both efficiency and computational expense.

3.3.2 Procrustes analysis of nodal positions

Parameter identification poses a very intricate challenge in the estimation procedure of latent variable models. In this study, we have adopted the projection model to estimate the positions

of nodes within the latent space. This method brings to light a critical issue: the model depends on the latent positions only through the pairwise distances between them. As a result, the parameters become non-identifiable when subjected to certain transformations of the point process. Differently from other types of latent variable network models (e.g. the distance model), these transformations are limited to rotations and reflections, since the projection model lacks translation invariance.

Focusing on a single parameter configuration, as it would be the case in an optimization setting, removes to some extent the issues related to non-identifiability, since these transformations would not alter our interpretations of the latent space. However, the situation becomes more complex in a Bayesian framework. Here, we generate posterior samples and attempt to summarize these by calculating estimators, such as the posterior mean. Applying transformations like rotations or reflections to any configuration of the sample will affect the posterior summaries. One way to explain this is by noting that it is impossible for us to know if the latent space was rotated or reflected during the sampling process. Thus, in a Bayesian framework, the posterior sample of the TSLPM lacks identifiability, creating an obstacle in providing meaningful insights or summaries.

To address this issue, we employ a popular methodology known as Procrustes matching, also referred to as Procrustes analysis or Procrustes superimposition. This technique seeks to find the best alignment or correspondence between two or more sets of points, typically represented as coordinate matrices in a multidimensional space, denoted here as $\mathbf{Z}$ and $\hat{\mathbf{Z}}$. In Procrustes matching, $\hat{\mathbf{Z}}$ undergoes a rigid transformation, consisting of rotations or reflections (pertinent to the projection model), to get as near to the reference $\mathbf{Z}$ as possible. The optimal match minimizes the sum of squared distances between corresponding points across the matrices, defined as:

$$R^2 = \sum_{i=1}^{N} \sum_{k=1}^{d} (z_{ik} - \hat{z}_{ik})^2$$

where $\hat{\mathbf{z}}_i$ and $\mathbf{z}_i$ are the coordinates of point i in configuration $\mathbf{Z}$ and $\hat{\mathbf{Z}}$, respectively.

Considering that the projection model lacks translation invariance, we formulate an optimization problem by introducing just rotations and reflections onto the matrix $\hat{\mathbf{Z}}$. For a more detailed discussion, see Hoff et al. (2002). The transformation applied to $\hat{\mathbf{Z}}$ is represented as:

$$\hat{\mathbf{z}}_i' = \mathbf{O}^\top \hat{\mathbf{z}}_i$$

where $\mathbf{O}$ is an orthogonal matrix that facilitates a rotation and/or reflection transformation. Therefore, the modified sum of squared distances between points is expressed as:

$$R^2 = \sum_{i=1}^{n} (\mathbf{z}_i - \mathbf{O}^\top \hat{\mathbf{z}}_i)^\top (\mathbf{z}_i - \mathbf{O}^\top \hat{\mathbf{z}}_i). \tag{3.11}$$

By minimizing the above equation, we can estimate the optimal **O** which can then be applied to the original $\hat{\mathbf{Z}}$ configuration so it is best matched with **Z**. The measure of the "match" between the two configurations is the minimized value of R^2 – known as the Procrustes sum of squares.

We adopt this procedure on all posterior samples that we obtain using HMC. As regards the reference configuration, we typically choose the maximum-a-posteriori solution, so that all other points configurations are rotated to match this one. As an exception, in the simulation studies that follow in the next section, we use the true positions as a reference whenever these are available. We emphasise again that this operation does not change the qualitative interpretation of each individual solution, in that the posterior value associated to any given sample remains unchanged before and after the Procrustes matching.

3.4 Simulation study

We implement a simulation study to assess the performance of the TSLPM model in estimating all the model parameters. For this goal, we run a simulation study on a variable number of time series with different sample sizes, so that a multitude of settings is examined.

Since our parameters' priors are noninformative, we devise a data-generating process that is capable of generating realistic time series data, as follows:

$$\alpha \sim \text{Uniform}(0, 3)$$
$$\boldsymbol{\beta} \sim \text{Uniform}(-1, 1)$$
$$\mathbf{Z} \overset{iid}{\sim} \text{Normal}(0, \sigma^2)$$

The intercept parameter α is generated from a uniform distribution on a narrow interval, specifically chosen to prevent the series from experiencing exponential growth. Similarly, the parameter $\boldsymbol{\beta}$ is simulated from a uniform distribution, with values ranging from -1 to 1. The rationale behind this choice is also to avoid degeneracy, and the specific choice of values is motivated by the results in Proposition-2 in Section 3.2.3 of the paper. The mean counts of the series generated with the chosen parameters are relatively small, as detailed in Section 3.8.3. It is important to note that no constraints are imposed on the parameters during the inferential procedure. As a result, there is a possibility that the inference obtained through both methods may lead to a non-stationary series.

In the proposed model, a two-dimensional latent space is employed to facilitate the visualization of network structure. The latent positions, denoted as $\mathbf{z}_1, \ldots, \mathbf{z}_N$, are generated using independent and identical distributions within this space. Each component z_{ik} of a latent position $\mathbf{z}_i$ follows a normal distribution with a mean of zero and variance $\sigma^2 > 0$.

Initially, the variance σ^2 for these normal distributions is set to a small value, specifically 0.01. This small value makes it very likely that the stability conditions set in Section 3.2.3 are satisfied. So, the initial latent space that we generate gives stable series, but also it exhibits an almost-independence of the series, due to the interaction terms being almost zero. Thus, before proceeding with the time series generation, we transform the latent space to make it more interesting, and closer to a borderline setting. We do this by iteratively adjusting the latent space, by expanding the latent positions by a small factor. This iterative adjustment is continued until the derived interaction matrix fail to meet the stability conditions outlined in Section 3.2.3. With this expansion procedure, we are still guaranteed to satisfy the stability conditions for the time series, while making the datasets more interesting since the latent space would actually play a stronger role. This contrasts with a setting where latent positions are all close to the center of the space, the latent space effect remains weak, and the generated series are independent.

We use the STAN programming language (Stan Development Team, 2024) to formulate our statistical models and to perform parameter estimation. For this purpose, we utilized two distinct approaches: an optimization procedure, notably the LBFGS algorithm and a Bayesian method, specifically Hamiltonian Monte Carlo (HMC). STAN implements a framework to use efficiently both procedures.

In the subsections below, we go through some synthetic data experiments that we conducted to explore the performance of the TSLPM in parameter estimation under varying network sizes and time series sample sizes, and in prediction under varying network sizes.

3.4.1 Estimation performance using LBFGS

We consider datasets with $N = \{10, 30, 50\}$ and time frames $T = \{50, 500, 1000\}$ and we simulate multivariate time series datasets with associated network structures using the TSLPM. In the simulation procedure, we generate 500 different realizations of a multivariate time series with different random seeds, for each choice of N and T. In this study, we employ the LBFGS optimization procedure to obtain point estimates of the parameters by maximizing the joint posterior from the model. We illustrate the estimation error for the parameters for α and $\boldsymbol{\beta}$ using violin plots in Figure 3.1, across the 500 generated datasets.

Similarly, we report a comparison between true and estimated values in Figure 3.2.

Based on the simulated results, we can draw the following conclusions. The violin plots in Figure 3.1 show some trends as N and T change. These plots illustrate a notable reduction in estimation errors as the sample points increase for both parameters α and $\boldsymbol{\beta}$. Additionally, with the expansion of network sizes, the range of estimation errors also demonstrates a decreasing pattern. Analogous conclusions can be drawn from the analysis of scatter plots in Figure 3.2. In datasets with limited sample sizes, the estimated values exhibit a relatively higher level of dispersion, whereas they become more concentrated and approach the true value as the

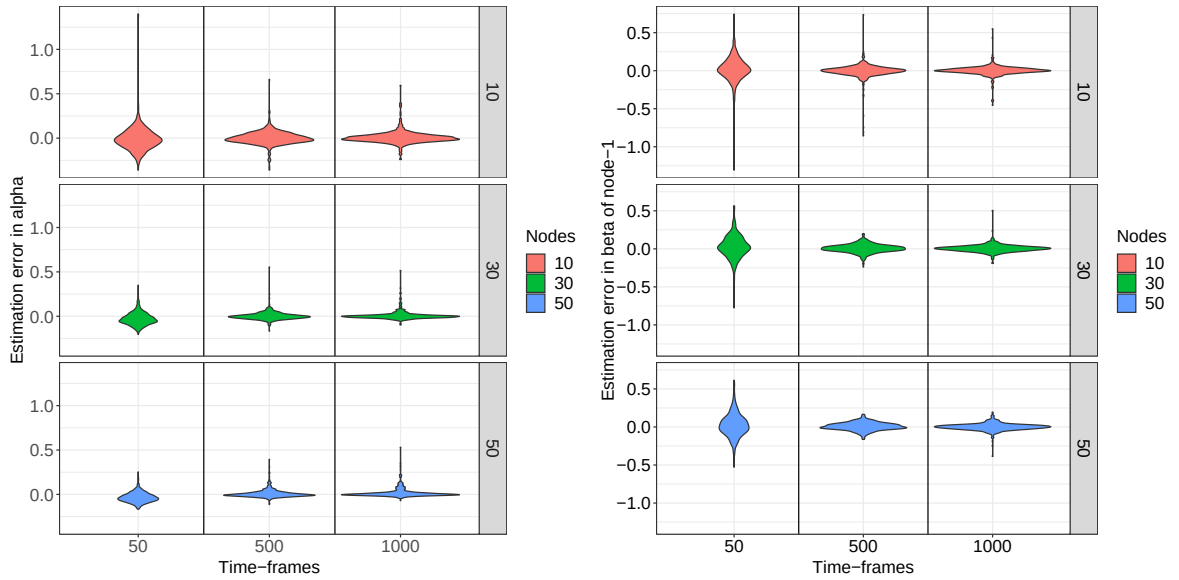

Fig. 3.1 Estimation errors across 500 different network datasets: Left panel shows the error for the parameter α, and the right panel shows the error for the parameter β, with node counts $N = \{10, 30, 50\}$ and time points $T = \{50, 500, 1000\}$.

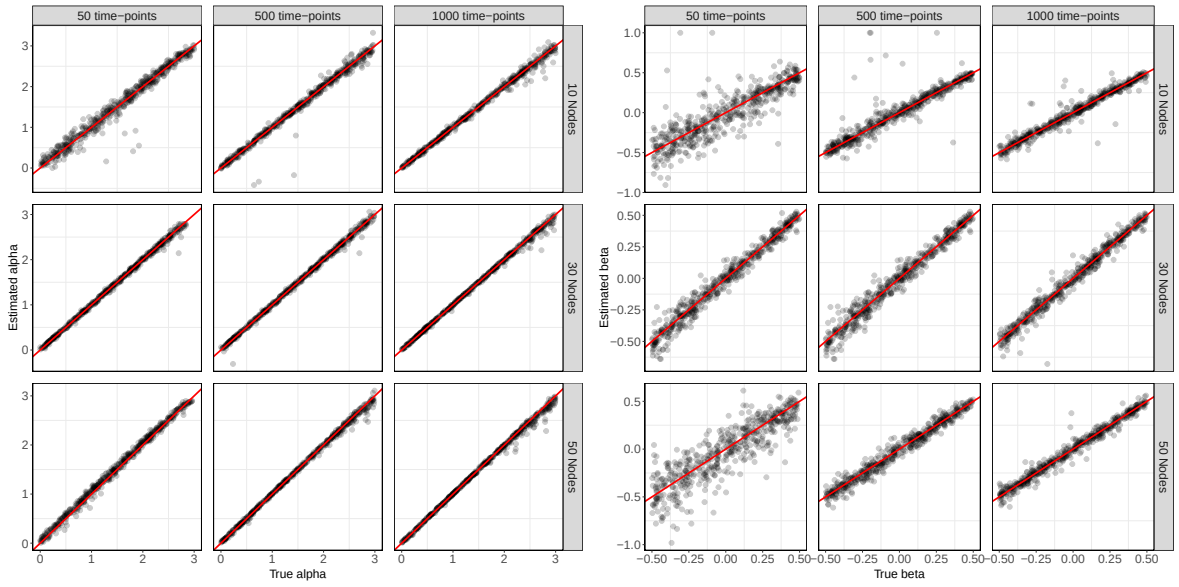

Fig. 3.2 True values (x-axis) versus estimated values (y-axis) for the parameters across 500 different network datasets with nodes $N = \{10, 30, 50\}$ and time points $T = \{50, 500, 1000\}$. Left: α, Right: β

sample size increases. This observation aligns with the expectation that larger sample sizes enhance the accuracy and precision of parameter estimation. These findings highlight the effectiveness of the LBFGS optimizer in achieving satisfactory parameter estimation results, even in the presence of just limited data.

Another noteworthy observation emerges from the examination of scatterplots comparing the true and estimated values of α in Figure 3.2. In datasets with larger sample sizes, a larger error is evident for higher values of α. The higher value of α leads to a systematic deviation in the estimated values towards higher magnitudes, which introduces challenges in accurately estimating the parameter α. This reaffirms the rationale behind the imposed bounds on α during the data generation process.

The plot in Figure 3.3 displays the average curve for the distribution of ratios of pairwise distances, similarly to the approach used by Sewell and Chen (2015, 2016).

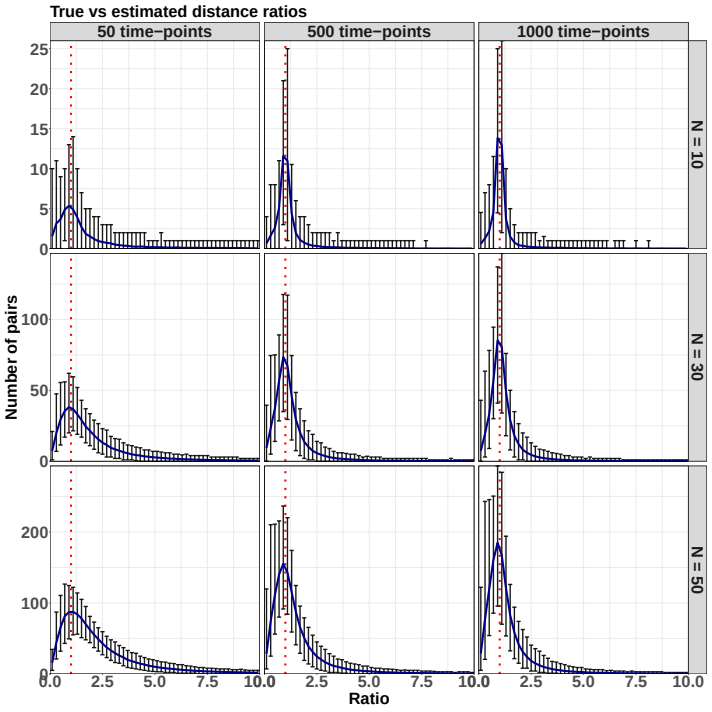

Fig. 3.3 Distribution of pairwise distance ratios, comparing estimated latent positions with true latent positions among 500 different network datasets with nodes $N = \{10, 30, 50\}$ and time points $T = \{50, 500, 1000\}$.

The distance ratios of N nodes are calculated by taking the ratio of the pairwise distances between the estimated latent positions and the pairwise distances between the true latent positions. So for each dataset we look at $\frac{\|\hat{\mathbf{z}_i} - \hat{\mathbf{z}_j}\|}{\|\mathbf{z}_i - \mathbf{z}_j\|}$, giving us the distribution curve consists of $\frac{N(N-1)}{2}$ such ratios and a total of 500 curves for 500 different realizations. A narrow resultant

plot centered at 1 conveys that the obtained latent space is quite similar to the true latent space.

The bars on the graph of Figure 3.3 represent the 95% quantile, indicating the number of pairs that fall within a specific ratio range. It is notable that these average distributions are consistently centered around 1, as indicated by a red dotted line in the plot. Moreover, as the sample points increase, these distributions exhibit a remarkable narrowing around the central value of 1. This trend holds across various network sizes, underlining the robustness of the estimation results for the latent space.

3.4.2 Forecasting

In this section, we employ forecasting as an evaluation tool to test the predictive performance of TSLPM, comparing it against two established models - the Vector AutoRegression (VAR) (Lütkepohl, 1999; Sim, 1980) and the Poisson Network AutoRegression (PNAR) (Armillotta and Fokianos, 2024), across various network sizes. Our comparative analysis involves out-of-sample testing at each node using distinct datasets. It is important to note that the VAR model does not account for network structure explicitly, whereas the PNAR model incorporates and leverages a *known* underlying patterns or mechanisms that govern node connections. We emphasize that our TSLPM is similar to the PNAR approach in that it uses a network structure to determine the interactions, however, differently from PNAR, our method assumes that the network is not observed.

Predictions for a small-sized network

We consider here an initial setting where we generate small-sized 5-node networks, with a time series observed on each node for 1000 observations each. A total of 300 datasets are generated with this format. For each simulated network, we use different seeds and parameter specifications, employing the same data-generating process as discussed in Section 4.5. We fit each dataset using our TSLPM, but also two similar other methods, i.e. PNAR (Armillotta and Fokianos, 2024) and VAR (Lütkepohl, 1999; Sim, 1980).

All the predictions that we make are intended as h steps ahead, meaning that once we reach the last available observation, we predict the following h values of each series, for some $h \geq 1$. We use two distinct techniques to make the h steps ahead forecast of the time series. The first method uses one-step forecasting steps, where predictions are generated one step at a time for future periods, each time using the real observed time series. This setup mimics a live prediction system where more data becomes available for each prediction. In contrast, the second approach, which we refer to as multi-step forecasting, entails generating forecasts for multiple future time steps simultaneously within a single computation cycle. Since in this simulation study we possess the true generated data, we proceed by splitting the data into

training set and test set. The first 80% of the observations are used as training, whereas the remaining 20% is used as test.

For all models and for all prediction tasks, we calculate the conditional means to obtain point estimates, i.e. $E(\mathbf{y}_{t+h|t}) = \hat{\mathbf{y}}_{t+h}$. Then, to measure the accuracy of our predictions, we consider the Root Mean Squared Error (RMSE). We evaluate the RMSE for $h = 1, \ldots, 5$, for all three models using both one-step (left panel) and multi-step forecasting algorithms (right panel) in Figure 3.4. The evaluation and comparison of each model's performance involve

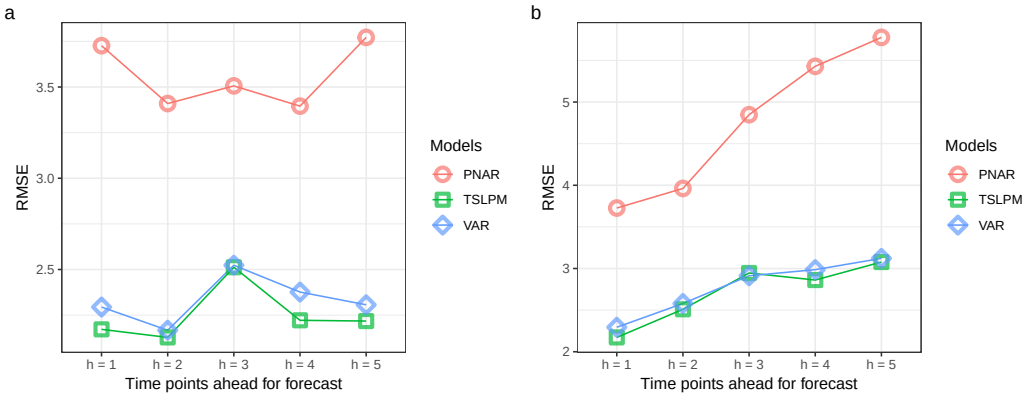

Fig. 3.4 Forecasting, case $N = 5$. RMSE of three models namely TSLPM, VAR, and PNAR in small-sized networks at each h-step ahead in the future using two different forecasting algorithms. a) One-step forecasting algorithm, and b) Multi-step forecasting algorithm.

estimating these errors at each future time point. In both the left and right plots of Figure 3.4, the mean of 300 different simulated time series observed on Node-1 of the networks is depicted. These plots consistently show that the PNAR models yield the highest RMSE at each future time point. This is quite unexpected since, in this application, PNAR uses the true underlying network structure as input. By contrast, slight variations are observed between the VAR and TSLPM models, with TSLPM demonstrating marginally better performance at certain time points in both one-step and multi-step forecasts. The similarity of results between TSLPM and VAR is not particularly suprising because, ultimately, our TSLPM is a variant of VAR for count data where we add more structure on the interaction terms.

In Figure 3.5 we plot instead the distribution of the RMSE as a boxplot for each model. In this case, the panels show only the 5-step ahead forecast that is either obtained by cumulating five 1-step ahead predictions or via a joint 5-steps ahead prediction. In no scenario the PNAR model outperforms the other two models, with similar distributions of errors for the VAR and TSLPM, echoing our findings from Figure 3.4.

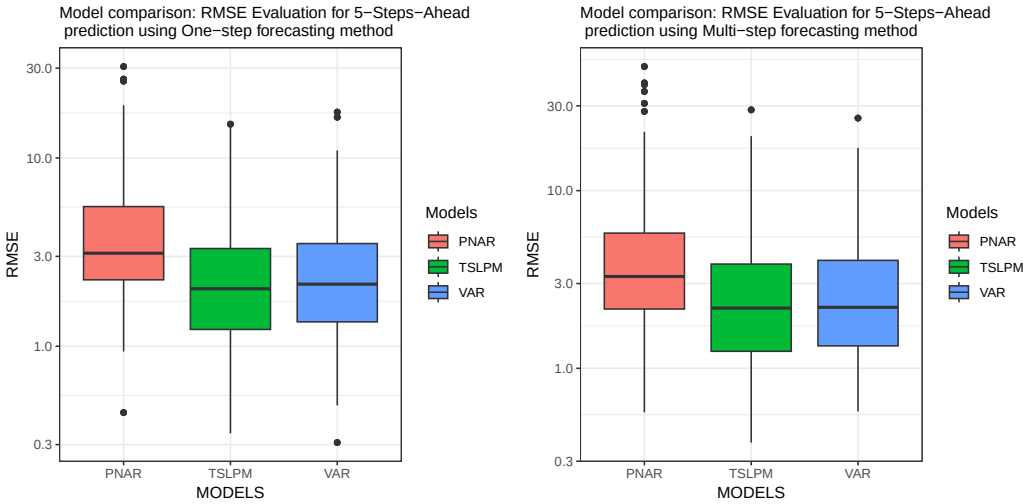

Fig. 3.5 Forecasting, case $N = 5$. RMSE evaluation in small-sized networks using TSLPM, VAR, and PNAR models across two future time points using two different forecasting algorithms. Left: 5-steps ahead prediction using a five 1-step ahead forecast (aggregated error). Right: 5-steps ahead prediction using a multi-step approach (aggregated error).

Moderately sized network

We consider another setting involving a moderately sized network comprising of 30 nodes, each with a time series observed over a length of $T = 1000$. The simulation regime used in generating the time series on the nodes of the networks, as well as the approach used to generate the forecast, remains unchanged from that used for small-sized networks. We replicate the same plots to compare the performance of the models. In Figure 3.6, on average, the RMSE is smaller in TSLPM compared to both the other models using a multi-step forecasting algorithm. The distributions of RMSE in Figure 3.7 convey similar findings

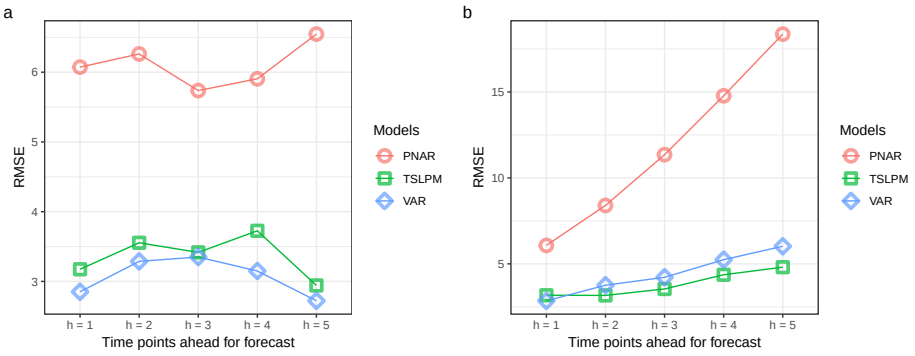

Fig. 3.6 Forecasting, case $N = 30$. RMSE of three models namely TSLPM, VAR, and PNAR in small-sized networks at each h-step ahead in the future using two different forecasting algorithms. a) One-step forecasting algorithm, and b) Multi-step forecasting algorithm.

regarding the models as observed in the small-sized networks. The PNAR model continues to perform poorly, while the distribution of errors appears similar for VAR and TSLPM.

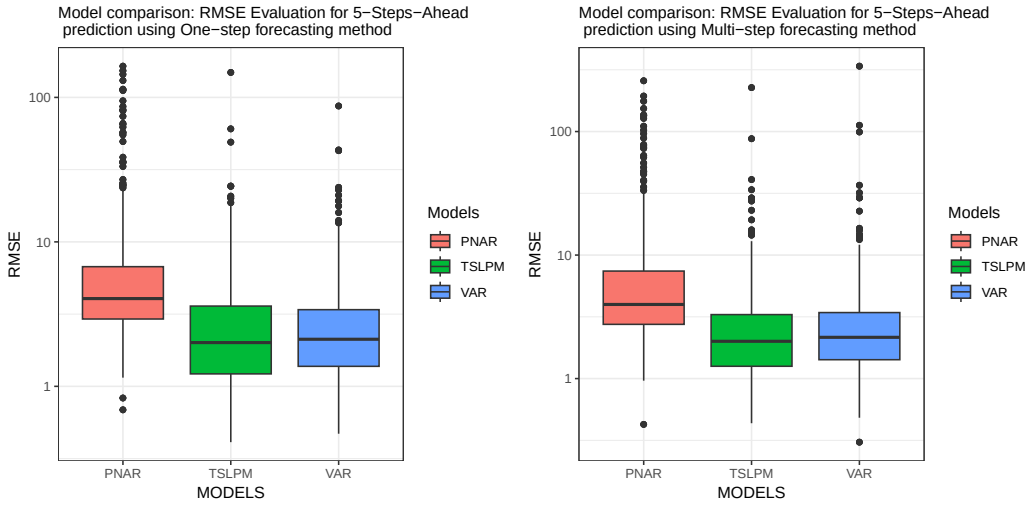

Fig. 3.7 Forecasting, case $N = 30$. RMSE evaluation in small-sized networks using TSLPM, VAR, and PNAR models across two future time points using two different forecasting algorithms. Left: 5-steps ahead prediction using a five 1-step ahead forecast (aggregated error). Right: 5-steps ahead prediction using a multi-step approach (aggregated error).

Large network

We now assess the predictive performance of the model to examine networks comprising of 50 nodes, each with a time series observed over a length of $T = 1000$. The simulation regime and forecasting approach employed remains consistent with those used for smaller networks. We reproduce the same type of plots to facilitate a comparative analysis of the model's performance. The Figures 3.8 and 3.9 depict whether the errors are at individual time points in the future or aggregated up to five steps ahead. The PNAR model consistently exhibits poorer performance, while the VAR and TSLPM demonstrate similar predictive accuracy.

The computational time for Bayesian estimation is efficient. More specifically, the execution time for one replication, based on the log-linear INGARCH model with ZIGP distribution, is approximately four minutes of CPU time on an Intel Core i7 desktop computer when performed using FORTRAN.

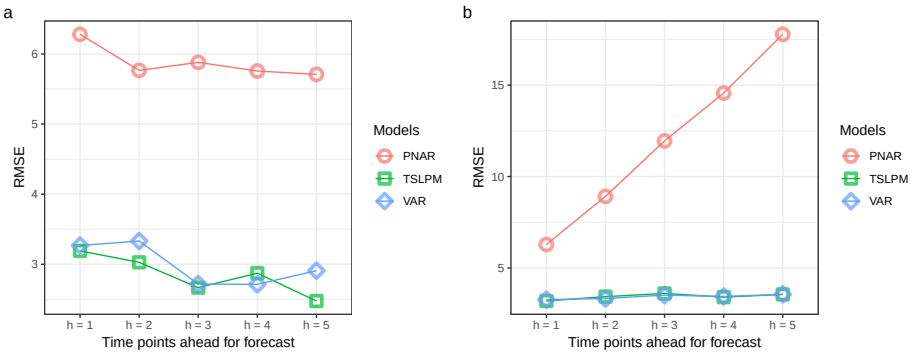

Fig. 3.8 Forecasting, case $N = 50$. RMSE of three models namely TSLPM, VAR, and PNAR in small-sized networks at each h-step ahead in the future using two different forecasting algorithms. a) One-step forecasting algorithm, and b) Multi-step forecasting algorithm.

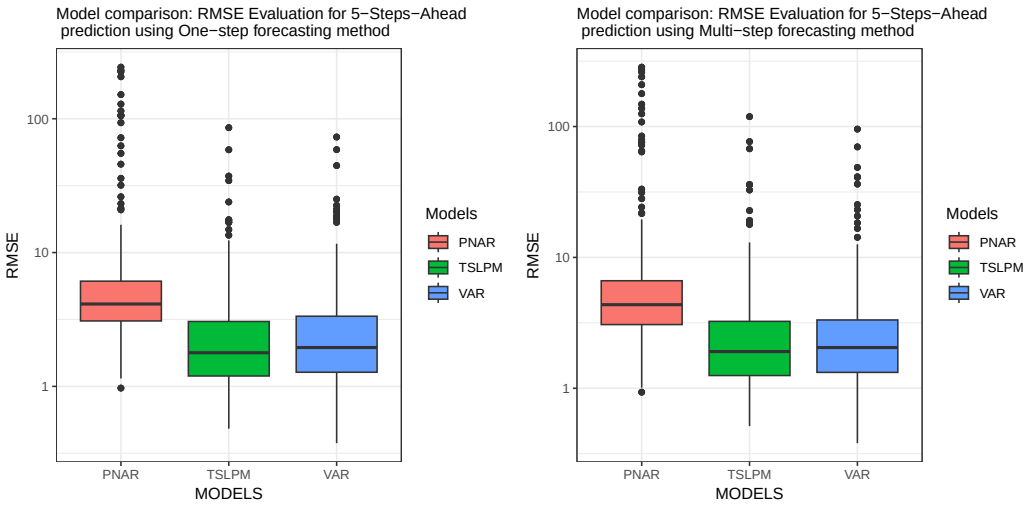

Fig. 3.9 Forecasting, case $N = 50$. RMSE evaluation in small-sized networks using TSLPM, VAR, and PNAR models across two future time points using two different forecasting algorithms. Left: 5-steps ahead prediction using a five 1-step ahead forecast (aggregated error). Right: 5-steps ahead prediction using a multi-step approach (aggregated error).

3.5 Chicago crime data

We consider now the analysis of burglary data within the south city of Chicago, Illinois, spanning from 2010 to 2015, utilizing crime data sourced from Clark and Dixon (2021)[1]. The dataset comprises counts of burglaries across $N = 552$ census block groups for 72 months. For each census block, we observe the monthly time series representing the number of burglaries recorded. In addition, we have geographical data from which we can create a network of neighbor census blocks: two blocks are linked by an undirected edge if they share a border.

The south side of Chicago is characterized by a relative racial and socio-economic homogeneity which is distinct from the rest of America. Although restricting the dataset to south Chicago reduces some socio-economic variability, residual differences persist. In order to understand these variations, we incorporate covariates into our model to reveal unique socio-economic and demographic characteristics for each region, while exploring potential relationships among census block groups to uncover underlying shared characteristics. Specifically, we consider population size, the number of young males (15-20 years old), and the mean percentage of unemployed individuals. The maximum number of burglaries within a given month in a census block is 17. Additionally, there is a clear seasonality trend in the data, as already documented in Clark and Dixon (2021) and Armillotta and Fokianos (2024).

3.5.1 Methodology

First, we construct a similarity matrix using the geographical location of the census block groups, resulting in a network adjacency matrix. This matrix describes which census blocks are geographically adjacent (i.e. they share a border). Through this process, we construct a 552×552 binary adjacency matrix, effectively mapping the direct neighborhood connections among the block groups. Since our method cannot handle 552 nodes due to computational limitations, we employ a network clustering algorithm to simplify the network, enabling us to examine the connections between a few macro-areas of Chicago that are comprised of clusters of block groups. For the sake of clarity in the subsequent discussion, these clusters of block groups will henceforth be referred to as “regions”.

This aggregation reduces the computational problems and it facilitates the interpretation of the results. Among the various clustering algorithms available, we choose the Girvan-Newman algorithm. Since this algorithm relies on edge-betweenness to capture and determine the importance of edges, we believe that the algorithm’s mechanism is coherent with the network topology that is induced by geographical information. While we choose to employ the Girvan-Newman algorithm, it is important to note that any clustering algorithm could be used in this

[1] https://github.com/nick3703/Chicago-Data

context. The specific choice of algorithm is not crucial to our goals; our primary objective is not to identify the most effective clustering but rather to demonstrate a methodology for quantifying and interpreting the relationships between the obtained regions. The clustering approach partitions the network into either five or eleven regions, as shown in Figure 3.10, using the modularity of the network as a measure to determine the optimal number of groups. Eleven regions achieve the highest modularity (0.78), but with only a slight increase from the modularity of five regions (0.72), making the latter a perfectly viable (and perhaps more parsimonious) alternative. Remarkably, the clusters found show excellent cohesion from a geographical point of view, in that none of the clusters are split into disconnected areas. The two partitions are shown in Figure 3.10.

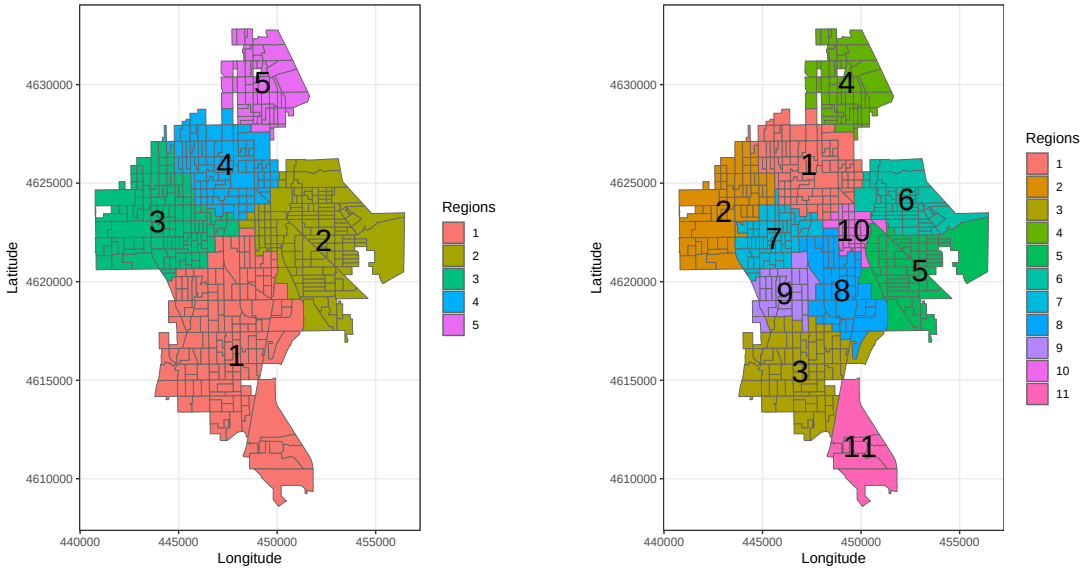

Fig. 3.10 Clustering results on the network of 552 census block groups in South Chicago, with each color representing a different aggregated region. Left: aggregation into 5 regions. Right: aggregation into 11 regions.

We proceed by aggregating the crimes for each block within a region to create an aggregated time series. These aggregated time series reveal a clear annual pattern in crime rates, as illustrated in Figure 3.11.

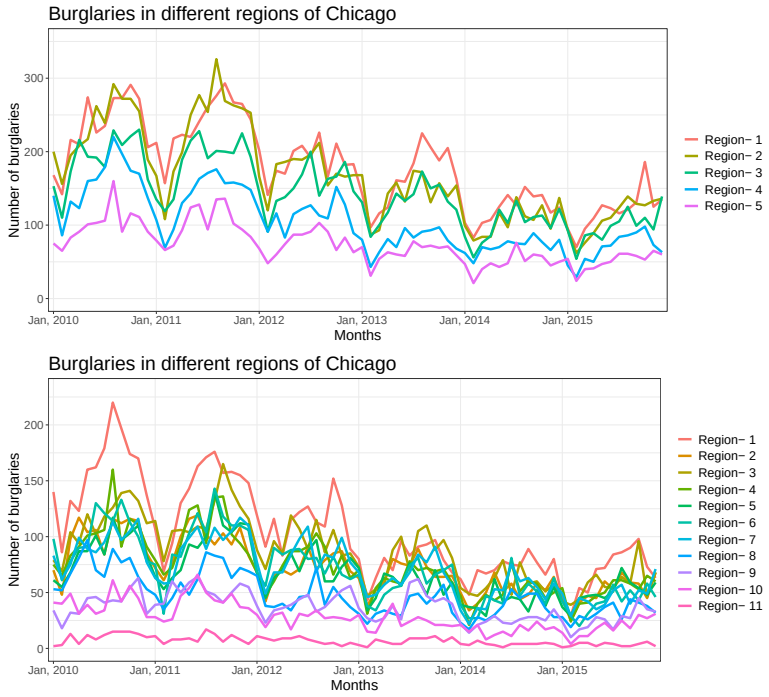

Fig. 3.11 Monthly number of cases of burglaries in the south side of Chicago (2010-2015). Left: aggregated over 5 regions. Right: aggregated over 11 regions.

3.5.2 TSLPM for the crime dataset

As a template model for the crime data, we extend the model of Eq. 3.6 as follows:

$$y_{it} \sim \text{Pois}(\lambda_{it})$$
$$\log(\lambda_{it}) = \alpha_i + \beta_i \log(y_{i(t-1)} + 1) + \sum_j \gamma_{ij} \log(y_{j(t-1)} + 1) + \eta_i \log(y_{i(t-12)} + 1) + \sum_k \delta_k x_{ik} \tag{3.12}$$

There are some differences with respect to the model of Eq. 3.6, namely: the intercept α may depend on the node, and we introduce a seasonality effect (with a lag of 12 months) with associated coefficients η. The parameters δ can obviously be different for each covariate, but they are the same parameters that are shared across different nodes. This is to simplify the interpretation of the results and to avoid potential non-identifiability issues.

This model essentially postulates that the average rate of change in the number of burglaries is determined by four main components. The first is the number of burglaries that happened last month measured through the parameter β, the second is the number of burglaries that happened 12 months before determined by coefficient η, the third is the number burglaries that happened in the other areas during the previous month quantified

through the latent space interactions γ, and, finally, the fourth is the effect given by the node-specific covariates, captured by the coefficients $\boldsymbol{\delta}$.

To enhance the predictive accuracy and interpretability of TSLPM, we have incorporated specific covariates that may highlight similarities across the regions. These covariates are the size of the population (hereafter, `pop`), the count of young males aged 15-20 years (`ym`), and the mean unemployment rate (`unemp`) within each region. These variables were chosen due to their potential impact on crime rate, and, in Eq. 3.12, they are indicated with $\mathbf{X} = \{x_{ik} : i = 1, \ldots, N \text{ and } k = 1, 2, 3\}$, respectively.

As regards seasonality, to effectively account for the annual seasonal effect, our analysis starts from Jan 2011 whereas the available data from January 2010 to December 2010 are employed just to calculate the seasonality effects from Jan 2011 onwards. This methodological approach allows us to utilize the initial year's data as a seed for the first observable year in our analysis, ensuring that the model adequately captures the annual seasonal trend from the outset. This strategy not only mitigates the issue of missing initial seasonal data but also enhances the predictive accuracy and relevance of the model by accurately reflecting the cyclical nature of crime rates.

All the computations were implemented in STAN and R. We used HMC for fitting TSLPM to obtain Markov chains for $10,000$ iterations with a burn-in period of $5,000$ iterations, and a thinning rate of 5, thus obtaining 1000 samples to represent the posterior distributions of the model parameters. For the model parameters $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\eta}, \mathbf{Z}$, and $\boldsymbol{\delta}$, we employ non-informative prior, specifying independent normal priors with a mean of zero and a standard deviation of 100 for each. This choice of high standard deviation for the priors is deliberate, aiming to minimize the influence of prior choices on the posterior distributions. This approach ensures that the analysis remains less biased by prior assumptions, allowing the data to play a more significant role in shaping the posterior densities.

We used the $\hat{R}$ statistic (also referred to as the potential scale reduction factor) and visual examination of the chains to check for their convergence (Gelman et al., 2013). To assess the quality and reliability of the samples of parameter estimates, the Effective posterior Sample Size (ESS) was checked. Divergence transitions for the Markov chains were also checked and eliminated. Sampling efficiency in STAN was managed by setting the tree depth parameter, which controls the simulation steps per iteration in the No-U-Turn Sampler (i.e. the sampler used by STAN). Adjusting this parameter may increase the computational load, so, we have carefully maintained a balance between allowing comprehensive exploration and managing computational resources. Our results all showed good convergence and reliable results for all models fitted.

3.5.3 Model selection

We proceed our analysis by fitting, on both datasets with 5 and 11 clusters, the model in Eq. 3.12 and a number of nested simpler models. In order to choose the best model overall, we use the Deviance Information Criterion (DIC), which stands out as particularly suitable for Bayesian multivariate time series models, offering a balanced approach to model selection. We also mention some viable alternatives such as the Widely Applicable Information Criterion (WAIC) proposed by Watanabe and Opper (2010) and leave-one-out cross-validation (LOO-CV) by Vehtari et al. (2017), although their applicability to models with time dependence is limited due to the potential for including future information in the analysis of past predictions. An alternative, Leave-Future-Out Cross-Validation (LFO-CV), specifically addresses this issue for univariate time series models, see Bürkner et al. (2020) for more detailed discussion. However, its use becomes computationally demanding with more intricate models, making DIC a more pragmatic option for many Bayesian time series applications, especially when considering the balance between accuracy and computational feasibility.

For the TSLPM selection process of Chicago's regional burglary network, we consider a set of $2^3 = 8$ alternative multivariate time series network models based on a different combination of the set of parameters $\{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\eta}\}$. Additionally, we integrate three covariates into these models in various configurations individually, in pairs, or all together—yielding a total of 50 alternative models. This approach allows us to examine how different parameters and covariate combinations influence our understanding of the patterns observed across the regions.

We initiated our analysis by applying the methodology outlined in Section 3.5.1 for each of the proposed model configurations, generating $1,000$ posterior samples of parameters for each model. Following the assessment of parameter estimates for quality and reliability, we examined potential collinearity to address any non-identifiability issues among variables, utilizing pairs scatter plots for this analysis. This examination revealed the presence of collinearity among some of the covariates, across all model configurations. In particular, the covariates $\mathtt{pop}$ and $\mathtt{ym}$ exhibited pronounced collinearity, compromising their simultaneous inclusion and thereby rendering the affected models unsuitable for fitting. Moreover, in models where the coefficients $\boldsymbol{\alpha}$ varied across different time series, collinearity among these coefficients was evident using pairs plot, signalling potential non-identifiability issues.

Given these challenges, the initial set of models was refined to a subset of 20 viable candidates, by making α constant across different series, and by removing one of the covariates. From there, we proceeded to calculate the DIC for each of the 20 models, for both the 5 clusters and 11 clusters datasets.

3.5.4 Crime data 5 clusters

We consider the analysis of the burglaries in five non-overlapping regions of Chicago from year 2011 to 2015. As regards model selection, the analysis of the DIC values from Table 3.1 reveals that the best model is the one where the set of parameters $\{\boldsymbol{\beta}, \boldsymbol{\eta}\}$ varies across nodes and α is constant. The covariates to include are $\{\texttt{pop}, \texttt{unemp}\}$. The model formulation

Table 3.1 DIC values for model comparisons for the 5 regions dataset.

Model	DIC for 5 regions
$\{\beta_i, \eta_i, \alpha, \delta_{\texttt{pop}}\}$	-46138.50
$\{\beta_i, \eta_i, \alpha, \delta_{\texttt{ym}}\}$	-45297.54
$\{\beta_i, \eta_i, \alpha, \delta_{\texttt{unemp}}\}$	-47787.59
$\{\beta_i, \eta, \alpha, \delta_{\texttt{pop}}\}$	-43940.26
$\{\beta_i, \eta, \alpha, \delta_{\texttt{ym}}\}$	-43104.12
$\{\beta_i, \eta, \alpha, \delta_{\texttt{unemp}}\}$	-47255.93
$\{\beta, \eta_i, \alpha, \delta_{\texttt{pop}}\}$	-43492.49
$\{\beta, \eta_i, \alpha, \delta_{\texttt{ym}}\}$	-44010.15
$\{\beta, \eta_i, \alpha, \delta_{\texttt{unemp}}\}$	-45590.13
$\{\beta, \eta, \alpha, \delta_{\texttt{pop}}\}$	-41448.41
$\{\beta, \eta, \alpha, \delta_{\texttt{ym}}\}$	-41420.87
$\{\beta, \eta, \alpha, \delta_{\texttt{unemp}}\}$	-45075.88
$\{\beta_i, \eta_i, \alpha, \delta_{\texttt{pop}}, \delta_{\texttt{unemp}}\}$	-48507.85
$\{\beta_i, \eta_i, \alpha, \delta_{\texttt{ym}}, \delta_{\texttt{unemp}}\}$	-48397.05
$\{\beta_i, \eta, \alpha, \delta_{\texttt{pop}}, \delta_{\texttt{unemp}}\}$	-45772.26
$\{\beta_i, \eta, \alpha, \delta_{\texttt{ym}}, \delta_{\texttt{unemp}}\}$	-47136.06
$\{\beta, \eta_i, \alpha, \delta_{\texttt{pop}}, \delta_{\texttt{unemp}}\}$	-46106.72
$\{\beta, \eta_i, \alpha, \delta_{\texttt{ym}}, \delta_{\texttt{unemp}}\}$	-45903.04
$\{\beta, \eta, \alpha, \delta_{\texttt{pop}}, \delta_{\texttt{unemp}}\}$	-44293.39
$\{\beta, \eta, \alpha, \delta_{\texttt{ym}}, \delta_{\texttt{unemp}}\}$	-43942.64

takes the form discussed in Eq. 3.12 with the sole distinction being that the parameter α is held constant across all regions.

Figure 3.12 shows the posterior summaries for all model parameters except the latent positions. These parameters characterise the individual behavior of nodes without taking into account the dependencies between nodes. We can see from these results that all $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ parameters tend to be positive, which indicates positive autocorrelations and a strong association due to seasonality. This can translate into a noticeable persistence of burglaries over time within the same region. The seasonality effect seems to be strongest for node 3, corresponding to the west region, and weakest (although still different from zero) for the south region (node 1).

The role of external factors, such as population and unemployment, are explored through the expression $\exp(\sum_k \delta_k x_{ik})$. This setup of the linear combination of terms allows for the

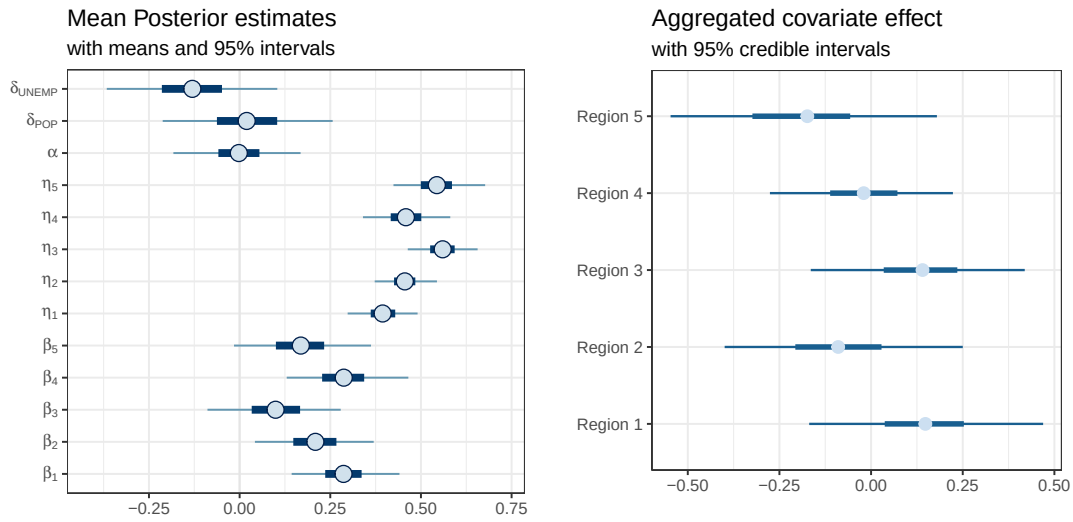

Fig. 3.12 Left: posterior averages with 95% credible intervals for the model parameters in the 5 clusters data, incorporating covariates. Right: aggregated mean effects of covariates (population and unemployment) on each region, shown with 95% credible intervals.

analysis of both collective and individual impacts of these factors on burglary rates. The model predicts that an increase in these covariates, assuming a simultaneous rise by their standard deviations, leads to proportional changes in burglary counts, either enhancing or reducing them based on the direction of change shown in the right panel of Figure 3.12.

We now delve into the detailed analysis of pairwise interactions between the regions, a primary aspect of this study. The integration of the interaction term γ_{ij} into our model facilitates this exploration of pairwise interactions among different regions, leveraging the spatial positioning $\mathbf{Z}$ of these regions as a fundamental element for interpreting their interactions. Through this methodology, our analysis offers both quantitative insights and visual representations to discern specific patterns of burglaries on the south side of Chicago, facilitating a deeper understanding of the geographical connections in the incidence of burglaries. First, we analyze the posterior distribution of the γ parameters, in Figure 3.13.

The credible intervals illustrate the posterior distribution of connections between these regions, where positive values indicate a positive relationship in burglary rates and vice versa.

As regards the latent positions, the model fit resulted in 1,000 posterior samples, representing the latent positions of each region within a latent space. To align these sampled coordinates for analysis, we applied a Procrustes transformation, as discussed in Section 4.4.2, using the MAP as a reference for orientation.

We show in Figure 3.14 the posterior samples made of 1,000 points from the posterior distribution of the estimated latent positions of each region post-Procrustes transformation.

Here, larger, solid points denote the posterior means, providing a clear visual representation of the estimated final locations of the regions.

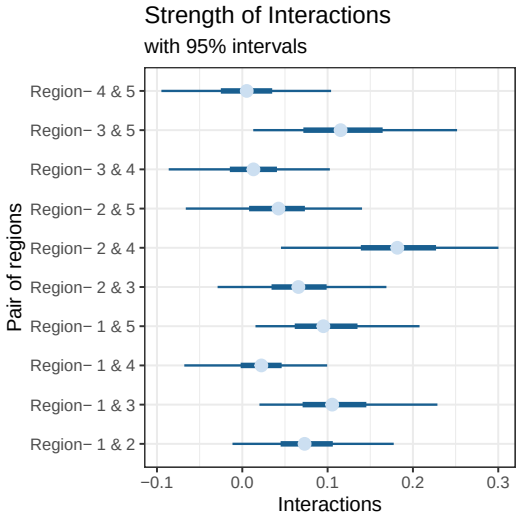

Fig. 3.13 Mean estimates with 95% credible intervals for pairwise interactions among the five different regions of Chicago.

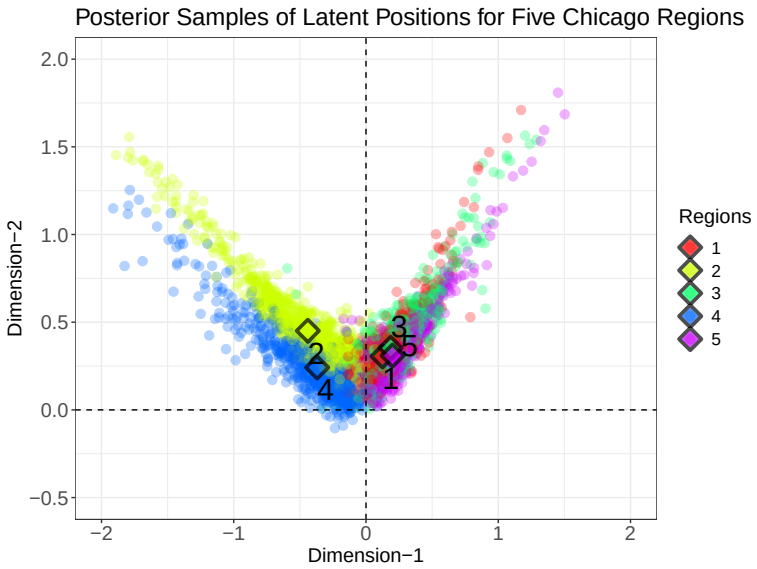

Fig. 3.14 Posterior samples of the latent positions for 5 different regions of Chicago. the posterior means are also shown in black and labeled in accordance to the region.

Additionally, we show on the left panel of Figure 3.15 the posterior average of the interactions matrix with pairwise connections (γs) in the non-diagonal entries and autoregressive effects (βs) on the diagonal.

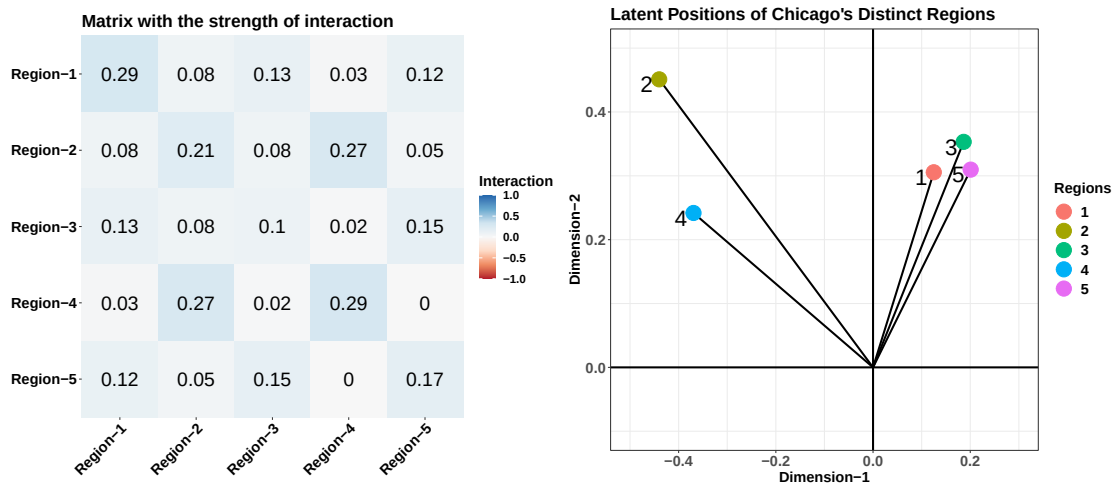

Fig. 3.15 Left: A matrix of connections between the five regions of Chicago. Right: latent space representation showing the posterior mean of locations for each region in Chicago. Each color represents a different region.

On the right panel of the same figure, instead, we show again the latent space representation using the posterior means and highlighting the directions they point at. This two plots allow for an appreciation of how the interdependencies between the series are captured via the latent space.

The left panel of Figure 3.15 reveals distinct patterns of burglary persistence over time, with the south and north-west regions of Chicago (regions 1 & 4) exhibiting the strongest autoregressive effects. In contrast, the north and eastern sides of Chicago (regions 2 & 5) show moderately weaker effects, with the weakest observed in the western region (region 3). This reflects the average varying degrees of historical burglary rates across regions, as reflected in the diagonal elements of the matrix of interactions indicating the strength of these autoregressive effects.

From the non-diagonal elements, and from the latent positions on the right panel, we can see that regions 2 and 4 appear aligned, and so are the regions 1, 3, and 5. This alignment, alongside the quantified interactions (ranging from 1 for strong positive to -1 for strong negative) reveals the nuanced connections between regions. Although interactions are generally modest, the adjacent regions 2 and 4 share the most pronounced connection. Further analysis of the regions 1, 3, and 5 reveals nuanced differences in their interconnections, with the strongest linkage between regions 3 (west) and 5 (northernmost), and a slightly lesser connection with region 1 (south). The interaction between regions 1 and 5, the geographically

most distant pairs, is found to be the weakest. These findings point to noteworthy connections in burglary occurrences across Chicago's varied geographic landscapes.

Finally, we employed posterior predictive checks to evaluate our TSLPM on these data, focusing on one-step-ahead predictions. This method generates new predictions from the posterior predictive distribution based on the fitted model and then compares these values against the observed data to identify any significant discrepancies. As illustrated in Figure 3.16, the observed time series data for burglaries are juxtaposed with the predictions, falling within the 95% confidence interval the vast majority of times.

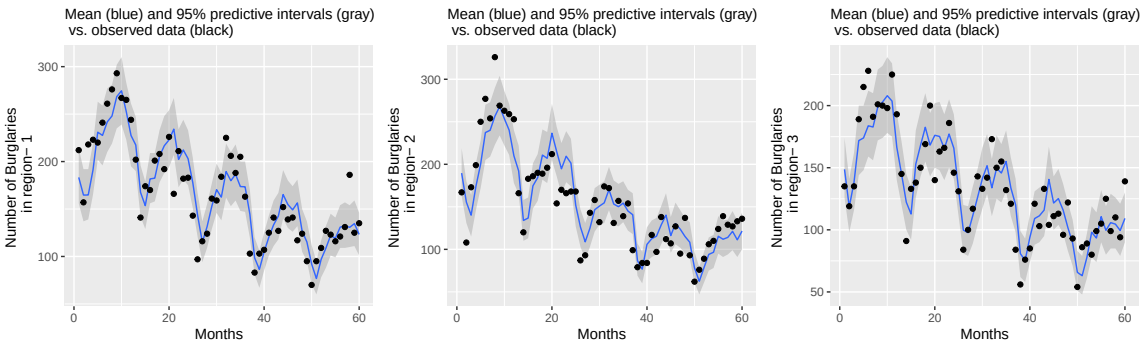

Fig. 3.16 Number of burglaries in the first three of five distinct regions of Chicago (2011-2015). The points are the observed number of burglaries. The blue line represents the prediction of the model with 95% prediction intervals shown in gray.

The findings detailed in Table 3.2 demonstrate the coverage of observed burglary counts by the posterior 95% credible intervals across the five regions. Overall, the model fit reveals minimal variance in the coverage of observed data demonstrating its effectiveness.

Table 3.2 Prediction coverage of the observed data using a 95% posterior credible interval for the 5 regions data.

Region	Percentage coverage {pop, unemp}
Region-1	80
Region-2	73
Region-3	78
Region-4	77
Region-5	85

3.5.5 Chicago's eleven-regional network data

We extend the analysis to the 11 regions dataset following the model selection strategy defined in Section 3.5.3. We select the final model by employing the DIC to identify the

optimal model. Table 3.3 reveals that the model exhibiting the lowest DIC has a variable $\boldsymbol{\beta}$ for each region, with α and η held constant across series and employing ($\texttt{pop}$, $\texttt{unemp}$) as a pair of covariates. The model adopts a the same form as the model of Eq. 3.12, with the only

Table 3.3 DIC values for model comparisons for the 11 regions dataset.

Model	DIC for 11 regions
$\{\beta_i, \eta_i, \alpha, \delta_{\texttt{pop}}\}$	-38711.20
$\{\beta_i, \eta_i, \alpha, \delta_{\texttt{ym}}\}$	-36858.17
$\{\beta_i, \eta_i, \alpha, \delta_{\texttt{unemp}}\}$	-35506.23
$\{\beta_i, \eta, \alpha, \delta_{\texttt{pop}}\}$	-40506.05
$\{\beta_i, \eta, \alpha, \delta_{\texttt{ym}}\}$	-40367.67
$\{\beta_i, \eta, \alpha, \delta_{\texttt{unemp}}\}$	-40592.83
$\{\beta, \eta_i, \alpha, \delta_{\texttt{pop}}\}$	-33705.27
$\{\beta, \eta_i, \alpha, \delta_{\texttt{ym}}\}$	-32448.32
$\{\beta, \eta_i, \alpha, \delta_{\texttt{unemp}}\}$	-30790.28
$\{\beta, \eta, \alpha, \delta_{\texttt{pop}}\}$	-35812.89
$\{\beta, \eta, \alpha, \delta_{\texttt{ym}}\}$	-35700.33
$\{\beta, \eta, \alpha, \delta_{\texttt{unemp}}\}$	-35713.67
$\{\beta_i, \eta_i, \alpha, \delta_{\texttt{pop}}, \delta_{\texttt{unemp}}\}$	-39529.08
$\{\beta_i, \eta_i, \alpha, \delta_{\texttt{ym}}, \delta_{\texttt{unemp}}\}$	-39115.18
$\{\beta_i, \eta, \alpha, \delta_{\texttt{pop}}, \delta_{\texttt{unemp}}\}$	-42335.00
$\{\beta_i, \eta, \alpha, \delta_{\texttt{ym}}, \delta_{\texttt{unemp}}\}$	-42242.19
$\{\beta, \eta_i, \alpha, \delta_{\texttt{pop}}, \delta_{\texttt{unemp}}\}$	-33243.90
$\{\beta, \eta_i, \alpha, \delta_{\texttt{ym}}, \delta_{\texttt{unemp}}\}$	-33520.29
$\{\beta, \eta, \alpha, \delta_{\texttt{pop}}, \delta_{\texttt{unemp}}\}$	-36016.88
$\{\beta, \eta, \alpha, \delta_{\texttt{ym}}, \delta_{\texttt{unemp}}\}$	-36390.81

difference being that the parameters α and η are constant across all regions. Thus, a key difference with previous models is that, in this optimal model, the parameter η is constant across all regions.

We can observe from the left panel of Figure 3.17, that the parameters $\boldsymbol{\beta}$ generally take positive values, although we notice more diverse values compared to the 5 clusters solution.

In particular, some regions exhibit small negative $\boldsymbol{\beta}$ values within their 95% credible intervals, suggesting that negative autocorrelations may also be possible. The $\boldsymbol{\beta}$ values show a diminished persistence of burglaries over time in the north-most and south-most regions of Chicago (regions 3, 11, and 4). The western, eastern, and central regions (regions 1, 2, 7, 8, 10 and 5) have comparatively strong autoregressive effects. Notably, the strongest effect is observed in the southwestern region (region 9) and the weakest in the northeast region (region 6). The parameter η takes a small positive value, suggesting that seasonal burglary patterns are likely to persist from one year to the next across all regions.

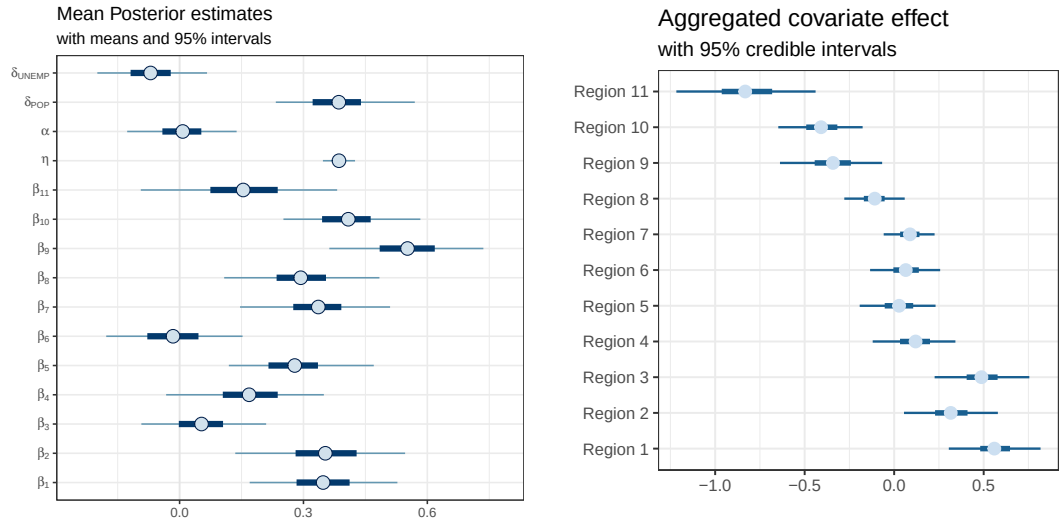

Fig. 3.17 Left: mean posterior estimates with 95% credible intervals for the model fitted to data across 11 regions, incorporating covariates. Right: aggregated mean effects of covariates (population and unemployment) on each region, shown with 95% credible intervals.

The model explores the impact of external factors like population and unemployment on burglary rates, with results shown in the right panel Figure 3.17. An increase in these covariates leads to corresponding changes in burglary counts, with regions 1, 2, and 3 notably experiencing an increase in burglary cases due to these combined effects. Additionally, if we look at the coefficients of the covariates, depicted in the left panel of Figure 3.17, we note a small coefficient for unemployment and a fairly large coefficient for population. These findings are consistent with previous studies on the same Chicago burglary data, such as those reported by Clark and Dixon (2021) and Armillotta and Fokianos (2024).

As regards the interaction parameters γ, the posterior distribution of connections between all regions are depicted in Figure 3.18.

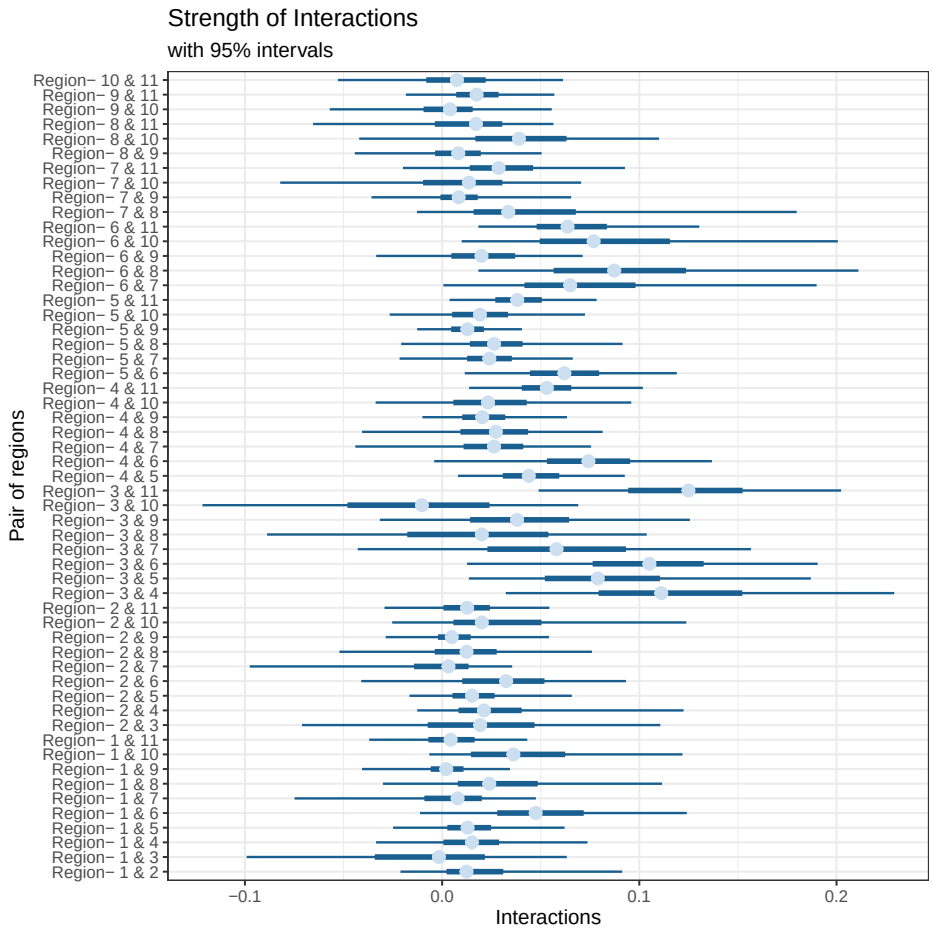

Fig. 3.18 Mean estimates with 95% credible interval of pairwise interactions between 11 regions of Chicago.

We note that some interactions are fairly strong so this suggests that a model capable of capturing these dependencies can be useful for these data.

Now, we show a latent space with all posterior samples for each of the 11 regions, post-Procrustes transformation, in Figure 3.19.

Here, the posterior means are denoted by larger, squared points, offering a clear visual representation of the estimated final locations of the regions. We also show, in Figure 3.20, the posterior averages of the latent positions besides the corresponding interactions and $\boldsymbol{\beta}$ parameters, to facilitate the interpretation of the latent space.

The latent space plot reveals a pattern of connections among the regions positioned in the northernmost, southernmost, south-eastern, and western areas (regions 3, 4, 5, 9, and, 11) due to their similar directions in the latent space. Similarly, regions in the eastern, western, and central regions of Chicago (regions 1, 2, 6, 7, 8, and, 10) also exhibit similar latent directions. Although overall interactions are modest, the strongest interactions are

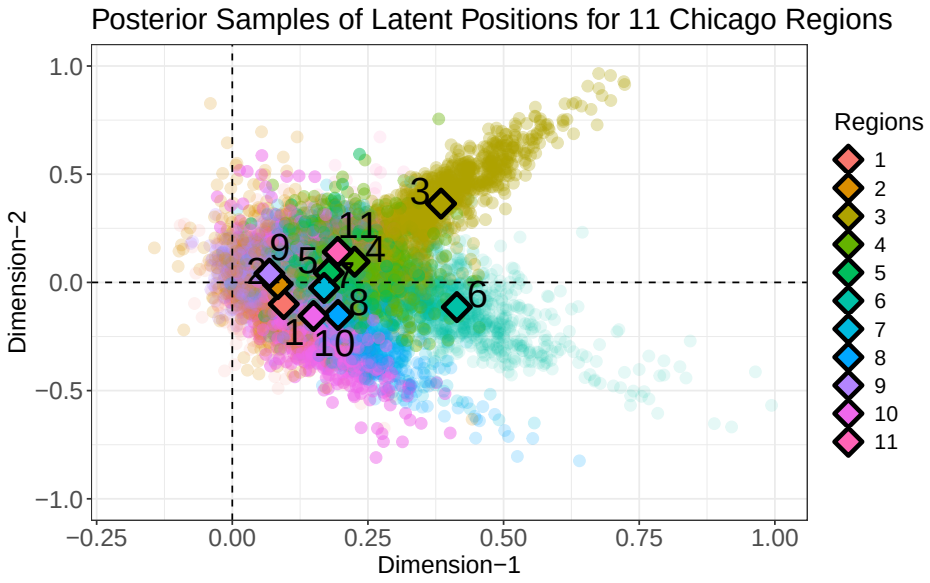

Fig. 3.19 Posterior samples of Latent positions for 11 different regions of Chicago. The large squared points represent the posterior means of the regions.

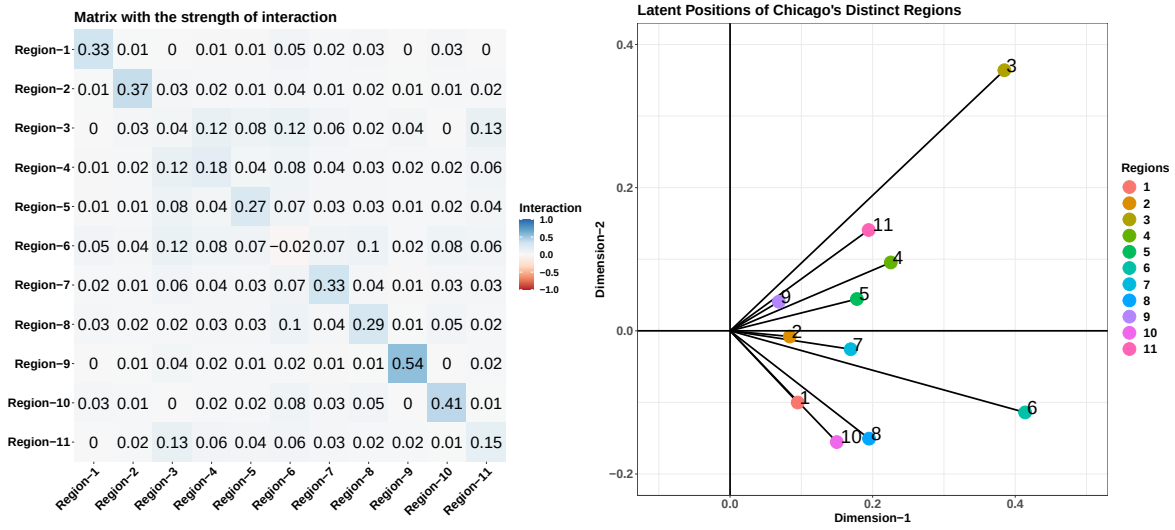

Fig. 3.20 Left: A matrix of connections between all 11 regions of Chicago. Right: latent space representation showing the posterior mean of locations for each region.

between regions 3 and 11 in the south, with significant connections also observed between regions 3 and 4 (north), and between region 6 in the northeast and region 8 in the center. The interaction matrix highlights another inter-group relation between northeastern region 6 with region 3 in the south. Overall, there are notable similarities observed among the northern, southern, and eastern regions, while the western regions appear to remain distinct.

Finally, Figure 3.21 presents the observed time series against predictions for the first three regions, including 95% credible intervals.

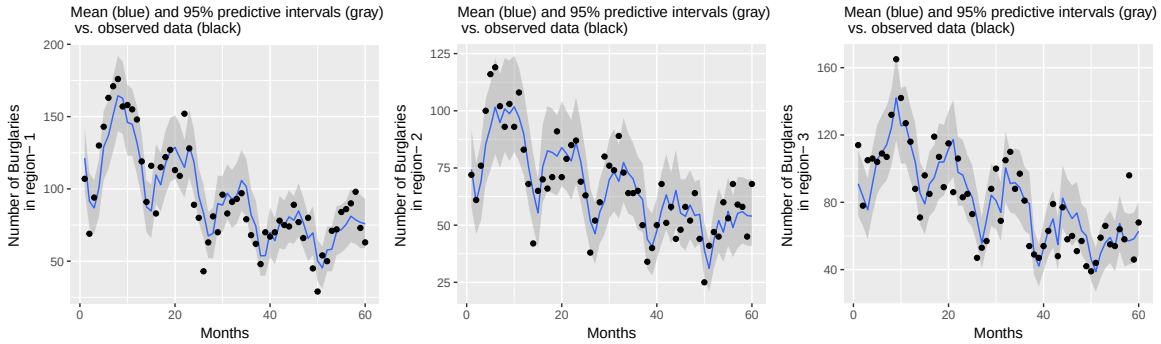

Fig. 3.21 Number of burglaries in the first three of eleven regions of Chicago. The points are the observed number of burglaries. The blue line represents the mean predictions of the model with 95% prediction intervals shown in gray.

These show good agreement and overall a very good model fit. This is also confirmed by Table 3.4, showing the extent to which the observed burglary counts are included in the posterior 95% credible intervals for each region.

The computational time for estimation using LBFGS is efficient. Specifically, the execution time for a single replication on a large network with 50 nodes is approximately two minutes of CPU time on an Intel Core i7 desktop computer with a Windows x64-based processor. In contrast, the computational time for Bayesian estimation of crime data is approximately 50 minutes with 5 clusters and 90 minutes with 11 clusters.

3.6 Conclusion and discussion

In this paper, we have introduced the TSLPM model, which provides a novel framework to analyse multivariate count time series data using a latent variable network approach. The TSLPM combines the flexibility of VAR models for time series, with the interpretable and useful representations that are obtained with latent variable network models. In the paper, we discussed model identifiability, stationarity, parameter interpretation, inference under a Bayesian setting, hence highlighting the main features of the model and how to use it. Through

Table 3.4 Prediction coverage of the observed data using a 95% posterior credible interval for the 11 regions data.

Region	Percentage coverage {pop, unemp}
Region-1	77
Region-2	87
Region-3	83
Region-4	87
Region-5	77
Region-6	85
Region-7	85
Region-8	90
Region-9	85
Region-10	85
Region-11	95

a series of simulation experiments, we have validated the TSLPM performance for parameter estimation and for prediction. The results indicate that the geometry of the estimated latent space closely matches with that of the true latent space, thus providing accurate modeling of the time series dependencies. When compared to other viable models in terms of forecasting accuracy, the TSLPM consistently performs as a very competitive approach, thus making it a suitable model for the analysis of multivariate count time series.

We showed an application of our new methodology on a dataset of crimes in the south city of Chicago. The model revealed connections in burglary occurrences across various geographic regions of Chicago. This reflects the idea that the frequency of crimes can follow rather complex patterns of interdependencies, which go beyond the simple geographical layout of the areas. Our results show a meaningful interaction between various areas of the city, with a some tendency to form clusters in the 5 regions case. The resulting inferred latent space is highly interpretable and offers a new model-based way to visualize the connections between the areas of the city. This latent space may serve as a starting point for further analysis, for example aiming at quantifying endogenous vs exogenous risk for each area, or for making future predictions regarding risk. Our crime data application very strongly suggests that the TSLPM methodology can be used in other contexts using an equivalent formulation: a natural application of these methods could be in finance and in the analysis of multivariate financial time series. In fact, extensions to multivariate time series data with continuous responses would be fairly straightforward. In the recent work by Zhu et al. (2017), a network vector autoregressive model (NAR) is proposed for continuous data. The model incorporates the information from the known network structure. An extension to this model is when the network structure is not known which is one of the assumptions in TSLPM. These two model

structures can be exploited to introduce a network model for multivariate time series with continuous response.

Count time series are non-negative and often exhibit overdispersion. One potential direction for extending our work is the incorporation of negative binomial models to account for overdispersion. Guo and Zhu (2024), in their recent study, present an integer-valued network autoregressive model that incorporates a non-random neighborhood structure. This model employs a negative binomial distribution for the conditional marginal distribution and utilizes the softplus function as the link function, providing greater flexibility in modeling multivariate time series of counts with overdispersion. An extension of this paradigm arises when the network structure is unknown. In our context, a possible research idea could be to combine a Poisson distribution with a log-linear model and the LPM, we propose combining the Latent Position Model (LPM) with the Softplus Negative Binomial Network AutoRegressive (SNBNAR) model introduced by Guo and Zhu (2024). The stationary conditions established by the authors for both fixed and increasing dimensions are crucial for ensuring the model's robustness across various scenarios. These conditions would also apply to the extended model and can be empirically validated.

Another potential extension is to consider a TSLPM of order greater than one. While we have partially done this using the seasonality effect in our real data application, a more formal and comprehensive approach may be considered to incorporate autoregressive coefficients of higher order. This inclusion serves to capture enduring dependencies and trends within the data, thus enhancing the precision and reliability of our forecasts, see Appendix 3.8.2 for brief discussion. As regards the model definition, another avenue for expansion is the integration of edge covariates. In numerous real network applications, data related to nodes and edges are readily accessible. By introducing these covariates into the model, we have the potential to significantly boost predictive performance.

Finally, in relation to the networks literature, a number of possible extensions can definitely be considered, including latent distance models and dynamic models. While the literature on dynamic network models has been fast growing, the domain of models addressing multivariate time series with underlying network structures remains relatively underrepresented. To our knowledge, the work by Kang et al. (2022) is one of the few papers in this sparse literature. Consequently, this area constitutes an open research question, awaiting further exploration.

3.7 Acknowledgement

This publication has emanated from research supported in part by a grant from the Insight Centre for Data Analytics which is supported by Science Foundation Ireland under Grant number $12/RC/2289\_P2$. We also extend our gratitude to Nicholas John Clark for providing the crime dataset analyzed in this paper.

3.8 Appendix

3.8.1 Supplementary tables

The results from fitting the TSLPM model to Chicago data, encompassing 5 regions, are presented in Table 3.5, and for 11 regions, in Table 3.6. These are the supplementary tables represent the quantified results of parameter estimates as given in the left panel of Figure 3.12 and 3.17.

Parameter	Mean	95% Credible Interval
β_1	0.29	(0.14,0.44)
β_2	0.21	(0.04,0.37)
β_3	0.10	(-0.09,0.28)
β_4	0.29	(0.13,0.47)
β_5	0.17	(-0.02,0.36)
η_1	0.39	(0.30,0.49)
η_2	0.46	(0.37,0.54)
η_3	0.56	(0.46,0.66)
η_4	0.46	(0.34,0.58)
η_5	0.54	(0.42,0.68)
α	0.00	(-0.18,0.17)
$\delta_{\texttt{pop}}$	0.02	(-0.21,0.26)
$\delta_{\texttt{unemp}}$	-0.13	(-0.37,0.10)

Table 3.5 Mean estimates with 95% credible interval for the model fitted to data across 5 regions with covariates.

3.8.2 TSLPM(p)

More generally, we introduce and study an extension of model 3.6 where y_{it} depends on past p lagged values. We can term this model as TSLPM(p) and it's defined analogously to 3.6 but with

$$\log(\lambda_{it}) = \alpha_i + \sum_{h=1}^{p} \zeta_h \left(\sum_{j=1}^{N} \gamma_{ij} \log(y_{j(t-h)} + 1) \right) + \sum_{k=1}^{K} \delta_k x_{ik}$$
$$\gamma_{ij} = \begin{cases} \beta_i & i = j \\ \mathbf{z}_i^\top \mathbf{z}_j & i \neq j \end{cases} \tag{3.13}$$

The interpretation of all the model parameters remains consistent with that of the TSLPM(1) with an additional parameter ζ_h that elevates or shrinks the impact of the log rate of series

Parameter	mean	95% Credible Interval
β_1	0.33	(0.15,0.52)
β_2	0.37	(0.16,0.56)
β_3	0.04	(-0.10,0.20)
β_4	0.18	(-0.03,0.38)
β_5	0.27	(0.08,0.44)
β_6	-0.02	(-0.18,0.16)
β_7	0.33	(0.14,0.52)
β_8	0.29	(0.11,0.48)
β_9	0.54	(0.36,0.73)
β_{10}	0.41	(0.23,0.58)
β_{11}	0.15	(-0.07,0.39)
η	0.39	(0.34,0.43)
α	-0.00	(-0.15,0.13)
$\delta_{\texttt{pop}}$	0.41	(0.24,0.58)
$\delta_{\texttt{unemp}}$	-0.04	(-0.18,0.09)

Table 3.6 Mean estimates with 95% credible interval for the model fitted to data across 11 regions with covariates.

i and the log rates of the counts of all other series on log rate of series i. Furthermore, the number of parameters to be estimated also increases as the order of the time series grows. The stationary conditions established for the TSLPM(1) model can be extended to the general case with minor modifications. Parameters can be estimated using the same inferential procedure opted for inference of TSLPM(1) i.e. HMC or plain MCMC.

3.8.3 Mean count for the time series

The plot displays the mean counts of the simulated series for node 1, generated during the data-generating process to assess the performance of the TSLPM. The results show that the counts are generally small (less than 40), highlighting the relevance of using the Poisson distribution to model the count time series. This observation supports the suitability of the Poisson distribution for capturing the characteristics of the simulated data, as it is well-suited for modeling non-negative integer-valued data with small counts.

3.8.4 Goodness of fit

An alternative approach to assess model fit on crime data across five regions involves the use of autocorrelation function (ACF) and cross-correlation function (CCF) plots. By comparing the autocorrelations and cross-correlations of the actual burglary data in each region with those of the posterior predictive time series, we can evaluate the goodness of fit of the model.

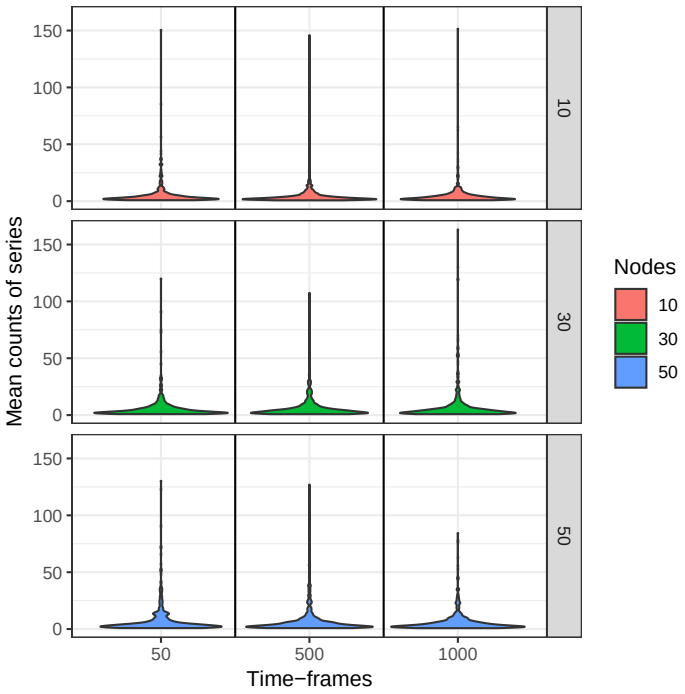

Fig. 3.22 Mean counts of series for node-1 across 500 different network datasets with N = {10, 30, 50} and time points T= {50, 500, 1000}.

Here, the posterior predictive time series refers to unobserved time series data conditioned on observed data and posterior samples of model parameters. We assume that the observed data points for the first 12 months (i.e., January 2010 to December 2010) are known to effectively account for annual seasonality when estimating the posterior predictive time series. We note that, with this generative process, the real dataset is used only to set the first 12 months, while from thereon the rest of the series is generated just iteratively using predictions of the fitted model. The ACF and CCF values that are obtained with this generative process are remarkably similar to those observed in the real data, as illustrated in Figures 3.24 and 3.25, confirming a good fit of the model.

Figure 3.23 illustrates the posterior predictive time series for Region 1, with the average posterior predicted burglaries represented in red. Figures 3.24 and 3.25 compare the ACF and CCF values of this posterior predictive series with those of the observed burglary time series in Region 1. The comparison demonstrates that the ACF and CCF values that are obtained with this generative process are remarkably similar to those observed in the real data, confirming a good fit for the model.

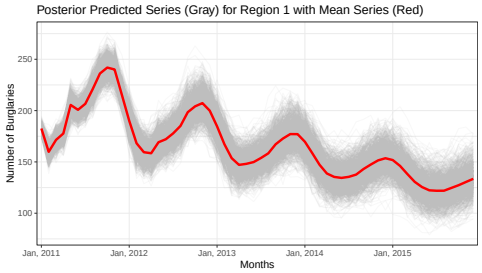

Fig. 3.23 Number of posterior predictive burglaries in region 1 with mean burglaries in region 1 shown in red.

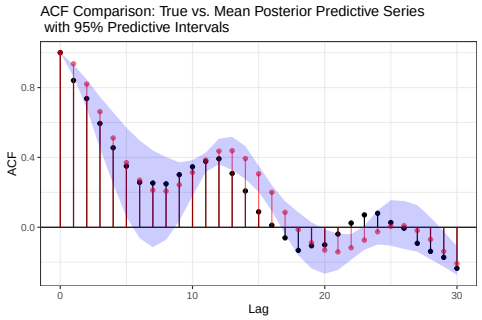

Fig. 3.24 ACF plots for burglaries in Region 1, with the true ACF shown in black, the ACF of the mean posterior predictive series in red, and the 95% prediction interval represented by blue shaded region.

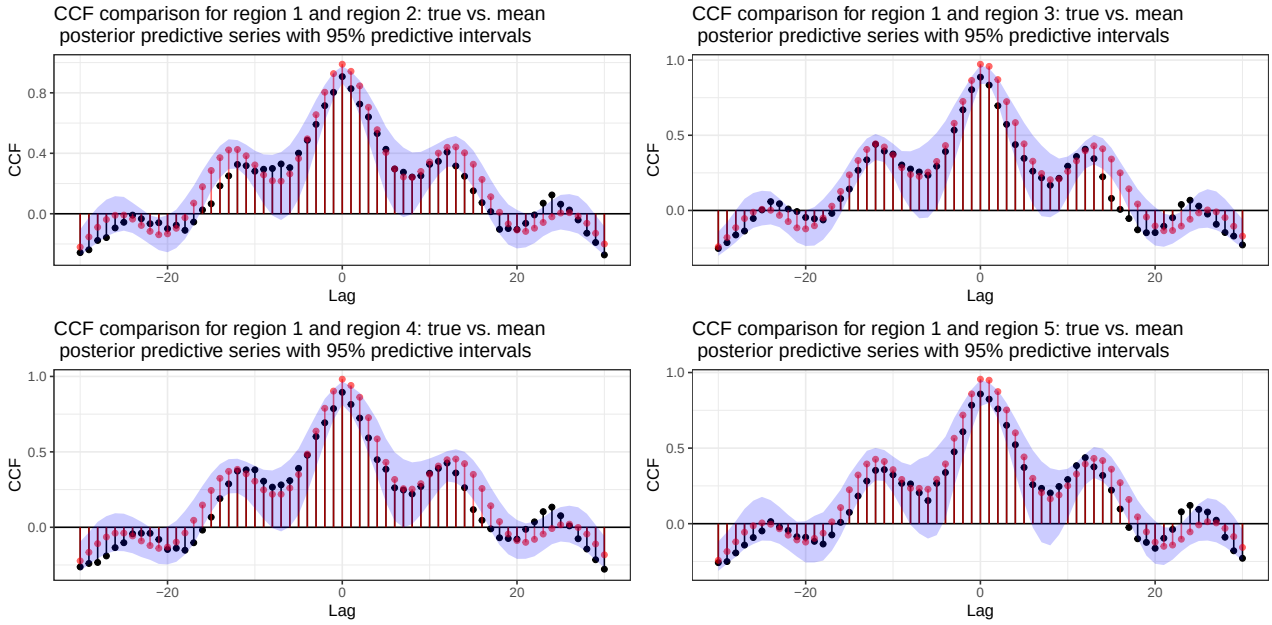

Fig. 3.25 CCF plots for burglaries in Region 1 with other regions, with the true CCF shown in black, the CCF of mean posterior predictive series in red and 95% predictive interval represented by blue shaded region.

References

Ahelegbey, D., Carvalho, L., and Kolaczyk, E. (2020). A bayesian covariance graphical and latent position model for multivariate financial time series. *SSRN Electronic Journal*, (181).

Aliverti, E. and Durante, D. (2019). Spatial modeling of brain connectivity data via latent distance models with nodes clustering. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 12(3):185–196.

Andreassen, C. (2013). *Models and inference for correlated count data*. PhD thesis, Department of Mathematics, Aarhus University.

Armillotta, M. and Fokianos, K. (2024). Count network autoregression. *Journal of Time Series Analysis*, 45(4):584–612.

Armillotta, M., Fokianos, K., and Krikidis, I. (2022). *Generalized Linear Models Network Autoregression*, pages 112–125. Springer International Publishing.

Barigozzi, M. and Brownlees, C. (2019). Nets: Network estimation for time series. *Journal of Applied Econometrics*, 34(3):347–364.

Bermúdez, L. and Karlis, D. (2011). Mixture of bivariate poisson regression models with an application to insurance. *Computational Statistics & Data Analysis*, 56(12):3988–3999.

Betancourt, M. and Girolami, M. (2015). *Hamiltonian Monte Carlo for Hierarchical Models*. Chapman and Hall/CRC.

Betancourt, M. and Stein, L. (2011). The geometry of hamiltonian monte carlo. *arXiv:1112.4118*.

Billio, M., Getmansky, M., Lo, A., and Pelizzon, L. (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics*, 104(3):535–559.

Box, G., Jenkins, G., Reinsel, G., and Ljung, G. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Bracher, J. and Held, L. (2022). Endemic-epidemic models with discrete-time serial interval distributions for infectious disease prediction. *International Journal of Forecasting*, 38(3):1221–1233.

Brown, E., Kass, R., and Mitra, P. (2004). Multiple neural spike train data analysis: state-of-the-art and future challenges. *Nature neuroscience*, 7(5):456–461.

Bürkner, P., Gabry, J., and Vehtari, A. (2020). Approximate leave-future-out cross-validation for bayesian time series models. *Journal of Statistical Computation and Simulation*, 90(14):2499–2523.

Chen, E., Fan, J., and Zhu, X. (2023). Community network auto-regression for high-dimensional time series. *Journal of Econometrics*, 235(2):1239–1256.

Clark, N. and Dixon, P. (2021). A class of spatially correlated self-exciting statistical models. *Spatial Statistics*, 43:100493.

Cox, D. R., Gudmundsson, G., Lindgren, G., Bondesson, L., Harsaae, E., Laake, P., Juselius, K., and Lauritzen, S. (1981). Statistical analysis of time series: Some recent developments [with discussion and reply]. *Scandinavian Journal of Statistics*, 8(2):93–115.

Dahlhaus, R. (2000). Graphical interaction models for multivariate time series1. *Metrika*, 51:157–172.

D'Angelo, S., Alfò, M., and Murphy, T. (2020). Modeling node heterogeneity in latent space models for multidimensional networks. *Statistica Neerlandica*, 74(3):324–341.

D'Angelo, S., Murphy, T., and Alfò, M. (2019). Latent space modelling of multidimensional networks with application to the exchange of votes in eurovision song contest. *Annals of Applied Statistics*, 13(2):900–930.

Darolles, S., Fol, G., Lu, Y., and Sun, R. (2019). Bivariate integer-autoregressive process with an application to mutual fund flows. *Journal of Multivariate Analysis*, 173(C):181–203.

Davis, R., Dunsmuir, W., and Wang, Y. (2000). On autocorrelation in a poisson regression model. *Biometrika*, 87(3):491–505.

Davis, R., Fokianos, K., Holan, S., Joe, H., Livsey, J., Lund, R., Pipiras, V., and Ravishanker, N. (2021). Count time series: A methodological review. *Journal of the American Statistical Association*, 116(535):1533–1547.

Davis, R. and Liu, H. (2012). Theory and inference for a class of observation-driven models with application to time series of counts. *Statistica Sinica*, 26(4):1673–1707.

Diebold, F. and Yılmaz, K. (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics*, 182(1):119–134.

Doukhan, P., Fokianos, K., Støve, B., and Tjøstheim (2020). Multivariate count autoregression. *Bernoulli*, 26(1):471 – 499.

Duane, S., Kennedy, A. D., Pendleton, B., and Roweth, D. (1987). Hybrid Monte Carlo. *Physics Letters B*, 195(2):216–222.

Durante, D. and Dunson, D. (2018). Bayesian inference and testing of group differences in brain networks. *Bayesian Analysis*, 13(1):29–58.

Durante, D., Dunson, D., and Vogelstein, J. (2017). Rejoinder: Nonparametric bayes modeling of populations of networks. *Journal of the American Statistical Association*, 112(520):1547–1552.

Durbin, J. and Koopman, S. (2000). Time series analysis of non-gaussian observations based on state space models from both classical and bayesian perspectives. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(1):3–56.

Ertekin, S., Rudin, C., and McCormick, T. (2015). Reactive point processes: A new approach to predicting power failures in underground electrical systems. *The Annals of Applied Statistics*, 9(1):122–144.

Ferland, R., Latour, A., and Oraichi, D. (2006). Integer-valued garch processes. *Journal of Time Series Analysis*, 27(6):923–942.

Fokianos, K. (2024). Multivariate count time series modelling. *Econometrics and Statistics*, 31:100–116.

Fokianos, K., Rahbek, A., and Tjøstheim, D. (2009). Poisson autoregression. *Journal of the American Statistical Association*, 104(488):1430–1439.

Fokianos, K., Støve, B., Tjøstheim, D., and Doukhan, P. (2020). Multivariate count autoregression. *Bernoulli*, 26(1):471 – 499.

Fokianos, K. and Tjøstheim, D. (2011). Log-linear poisson autoregression. *Journal of Multivariate Analysis*, 102(3):563–578.

Francq, C. and Zakoian, J.-M. (2019). *Multivariate GARCH Processes*, chapter 10, pages 272–316. John Wiley & Sons, Ltd.

Franke, J. and Rao, T. (1993). *Multivariate First-Order Integer-Valued Autoregressions*, volume 95 of *Berichte der Arbeitsgruppe Technomathematik*. Arbeitsgruppe Technomathematik, University.

Friel, N., Rastelli, R., Wyse, J., and Raftery, A. (2016). Interlocking directorates in irish companies using a latent space model for bipartite networks. *Proceedings of the National Academy of Sciences of the United States of America*, 113(24):6629–6634.

Frühwirth-Schnatter, S. and Wagner, H. (2006). Auxiliary mixture sampling for parameter-driven models of time series of small counts with applications to state space modelling. *Biometrika*, 93(4):827–841.

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., and Rubin, D. (2013). *Bayesian Data Analysis*. Chapman and Hall/CRC.

Gollini, I. and Murphy, T. (2016). Joint modeling of multiple network views. *Journal of Computational and Graphical Statistics*, 25(1):246–265.

Gormley, I. and Murphy, T. (2008). *A latent space model for rank data*, pages 90–102.

Guo, X. and Zhu, F. (2024). Softplus negative binomial network autoregression. *Stat*, 13(1):e638.

Gwee, X., Gormley, I., and Fop, M. (2023). A latent shrinkage position model for binary and count network data. *Bayesian Analysis*, 1(1):1 – 29.

Hall, E. and Willett, R. (2015). Online learning of neural network structure from spike trains. *International IEEE/EMBS Conference on Neural Engineering, NER*, 2015:930–933.

Handcock, M., Raftery, A., and Tantrum, J. (2007). Model-based clustering for social networks (with discussion). *Journal of the Royal Statistical Society. Series A: Statistics in Society*, 170(2):301–354.

Harvey, A. and Fernandes, C. (1989). Time series models for count or qualitative observations. *Journal of Business & Economic Statistics*, 7(4):407–417.

Heinen, A. and Erick, R. (2007). Multivariate autoregressive modeling of time series count data using copulas. *Journal of Empirical Finance*, 14(4):564–583.

Hledik, J. and Rastelli, R. (2023). A dynamic network model to measure exposure concentration in the austrian interbank market. *Statistical Methods & Applications*, 32(5):1695–1722.

Hoff, P. (2005). Bilinear mixed-effects models for dyadic data. *Journal of the American Statistical Association*, 100(469):286–295.

Hoff, P. (2007). Modeling homophily and stochastic equivalence in symmetric relational data. *NIPS'07: Proceedings of the 21st International Conference on Neural Information Processing Systems*, 20:657 – 664.

Hoff, P. (2011). Hierarchical multilinear models for multiway data. *Computational Statistics & Data Analysis*, 55(1):530–543.

Hoff, P. (2021). Additive and multiplicative effects network models. *Statistical Science*, 36(1):34–50.

Hoff, P., Raftery, A., and Handcock, M. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460):1090–1098.

Hoffman, M., Sountsov, P., Dillon, J., Langmore, I., Tran, D., and Vasudevan, S. (2019). Neutra-lizing bad geometry in hamiltonian monte carlo using neural transport. *arXiv: 1903.03704*.

Härdle, W., Wang, W., and Yu, L. (2016). Tenet: Tail-event driven network risk. *Journal of Econometrics*, 192(2):499–513.

Johnson, S. (2008). Repeat burglary victimisation: a tale of two theories. *Journal of Experimental Criminology*, 4:215–240.

Jørgensen, B., Christensen, S., Song, P.-K., and Sun, L. (1996). State-space models for multivariate longitudinal data of mixed types. *Canadian Journal of Statistics*, 24(3):385–402.

Jung, R., Liesenfeld, R., and Richard, J. (2011). Dynamic factor models for multivariate count data: An application to stock-market trading activity. *Journal of Business & Economic Statistics*, 29(1):73–85.

Kang, X., Ganguly, A., and Kolaczyk, E. (2022). Dynamic networks with multi-scale temporal structure. *Sankhya A: The Indian Journal of Statistics*, 84(1):218—260.

Kaur, H., Rastelli, R., and Friel, N. (2023). *Latent Position Network Models*, chapter 36, pages 526–541. SAGE Publications.

Knight, M., Leeming, K., Nason, G., and Nunes, M. (2020). Generalized Network Autoregressive Processes and the GNAR Package. *Journal of Statistical Software*, 96(5):1–36.

Knight, M., Nunes, M., and Nason, G. (2016). Modelling, detrending and decorrelation of network time series. *arXiv: 1603.03221*.

Krivitsky, P. and Handcock, M. (2008). Fitting position latent cluster models for social networks with latentnet. *Journal of Statistical Software*, 24(5):1–23.

Latour, A. (1997). The multivariate ginar(p) process. *Advances in Applied Probability*, 29(1):228–248.

Lee, Y., Lee, S., and Tjøstheim, D. (2018). Asymptotic normality and parameter change test for bivariate poisson ingarch models. *Test*, 27:52–69.

Livsey, J., Lund, R., Kechagias, S., and Pipiras, V. (2018). Multivariate integer-valued time series with flexible autocovariances and their application to major hurricane counts. *The Annals of Applied Statistics*, 12:408–431.

Luetkepohl, H. (2005). *The New Introduction to Multiple Time Series Analysis*. Springer Berlin, Heidelberg.

Lütkepohl, H. (1999). Vector autoregressions. *Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes*, (4).

Mark, B., Raskutti, G., and Willett, R. (2017). Network estimation via poisson autoregressive models. In *2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 1–5. IEEE.

Mohler, G., Short, M., Brantingham, P., Schoenberg, F., and Tita, G. (2011). Self-exciting point process modeling of crime. *Journal of the american statistical association*, 106(493):100–108.

Neal, R. (2012). MCMC using Hamiltonian Dynamics. *Handbook of Markov Chain Monte Carlo*.

Nickel, C. (2008). *Random dot product graphs a model for social networks*. PhD thesis, Johns Hopkins University.

Nocedal, J. and Wright, S. (2006). *Numerical Optimization*. Springer, New York, NY, USA, 2e edition.

Pandit, P., Sahraee-Ardakan, M., Amini, A., Rangan, S., and Fletcher, A. (2020). Generalized autoregressive linear models for discrete high-dimensional data. *IEEE Journal on Selected Areas in Information Theory*, 1(3):884–896.

Pedeli, X. and Karlis, D. (2013a). On composite likelihood estimation of a multivariate INAR(1) model. *Journal of Time Series Analysis*, 34:206–220.

Pedeli, X. and Karlis, D. (2013b). Some properties of multivariate INAR(1) processes. *Computational Statistics & Data Analysis*, 67:213–225.

Rastelli, R. and Corneli, M. (2023). Continuous latent position models for instantaneous interactions. *Network Science*, 11(4):560–588.

Ravishanker, N., Serhiyenko, V., and Willig, M. (2014). Hierarchical dynamic models for multivariate times series of counts. *Statistics and Its Interface*, 7:559–570.

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

Scotto, M., Weiß, C., Silva, M., and Pereira, I. (2014). Bivariate binomial autoregressive models. *Journal of Multivariate Analysis*, 125:233–251.

Sewell, D. and Chen, Y. (2015). Latent space models for dynamic networks. *Journal of the American Statistical Association*, 110(512):1646–1657.

Sewell, D. and Chen, Y. (2016). Latent space models for dynamic networks with weighted edges. *Social Networks*, 44:105–116.

Sewell, D. and Chen, Y. (2017). Latent space approaches to community detection in dynamic networks. *Bayesian analysis*, 12(2):351–377.

Shojaie, A. and Michailidis, G. (2010). Discovering graphical Granger causality using the truncating lasso penalty. *Bioinformatics*, 26(18):517–523.

Sim, C. (1980). Macroeconomics and reality. *Econometrica*, 48(1):1–48.

Stan Development Team (2024). RStan: the R interface to Stan. R package version 2.32.6.

Streett, S. (2000). *Some observation driven models for time series*. PhD thesis, Department of Statistics, Colorado State University.

Tafakori, L., Pourkhanali, A., and Rastelli, R. (2022). Measuring systemic risk and contagion in the european financial network. *Empirical economics*, 63(1):345–389.

Varga, R. (2011). *Geršgorin and His Circles*, volume 36 of *Springer Series in Computational Mathematics*. Springer Berlin Heidelberg.

Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and computing*, 27:1413–1432.

Watanabe, S. and Opper, M. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11(12):3571–3594.

Young, S. and Scheinerman, E. (2007). Random dot product graph models for social networks. In *International Workshop on Algorithms and Models for the Web-Graph*, pages 138–149. Springer.

Yuan, B., Li, H., Bertozzi, A., Brantingham, P., and Porter, M. (2019). Multivariate spatiotemporal hawkes processes and network reconstruction. *SIAM Journal on Mathematics of Data Science*, 1(2):356–382.

Zeger, S. (1988). A regression model for time series of counts. *Biometrika*, 75:621–629.

Zhou, J., Li, D., Pan, R., and Wang, H. (2020). Network garch model. *Statistica Sinica*, 30:1–18.

Zhu, X., Chang, X., Li, R., and Wang, H. (2019a). Portal nodes screening for large scale social networks. *Journal of Econometrics*, 209(2):145–157.

Zhu, X., Huang, D., Pan, R., and Wang, H. (2020). Multivariate spatial autoregressive model for large scale social networks. *Journal of Econometrics*, 215(2):591–606.

Zhu, X. and Pan, R. (2020). Grouped network vector autoregression. *Statistica Sinica*, 30(3):1437–1462.

Zhu, X., Pan, R., Li, G., Liu, Y., and Wang, H. (2017). Network vector autoregression. *The Annals of Statistics*, 45(3):1096 – 1123.

Zhu, X., Wang, W., Wang, H., and Härdle, W. (2019b). Network quantile autoregression. *Journal of Econometrics*, 212(1):345–358.

Chapter 4

A dynamic latent space time series model to assess the spread of mumps in England

4.1 Abstract

This work is motivated by an original dataset of reported mumps cases across nine regions of England, and focuses on the modeling of temporal dynamics and time-varying dependency patterns between the observed time series. The goal is to discover the possible presence of latent routes of contagion that go beyond the geographical locations of the regions, and instead may be explained through other non directly observable socio-economic factors. We build upon the recent statistics literature and extend the existing count time series network models by adopting a time-varying latent distance network model. This approach can efficiently capture across-series and across-time dependencies, which are both not directly observed from the data. We adopt a Bayesian hierarchical framework and perform parameter estimation using L-BFGS optimization and Hamiltonian Monte Carlo. We demonstrate with several simulation experiments that the model parameters can be accurately estimated under a variety of realistic dependency settings. Our real data application on mumps cases leads to a detailed view of some possible contagion routes. A critical advantage of our methodology is that it permits clear and interpretable visualizations of the complex relations between the time series and how these relations may evolve over time. The geometric nature of the latent embedding provides useful model based summaries. In particular, we show how to extract a measure of contraction of the inferred latent space, which can be interpreted as an overall risk for the escalation of contagion, at each point in time. Ultimately, the results highlight some possible critical transmission pathways and the role of key regions in driving

infection dynamics, offering valuable perspectives that may be considered when designing public health strategies.

Keywords: multivariate time series; vector autoregressive models; latent position models; dynamic networks; epidemic risk.

4.2 Introduction

Mumps is a viral infection that is transmitted primarily through respiratory droplets, similarly to colds and flu. Mumps may spread when infected droplets of saliva are released into the air and are either inhaled or transferred from contaminated surfaces to the mouth or nose. Mumps typically affects children aged 4 to 16 and young adults who did not receive the MMR (Measles, Mumps, and Rubella) vaccine during their childhood vaccination schedule. In vaccinated individuals, mumps tends to present with milder symptoms and fewer complications. The MMR vaccine was introduced in England in 1988, but also people born in the late 1990s and early 2000s who may have missed out on vaccination during that period remain at higher risk.

Outbreaks of mumps in England tend to occur in settings where individuals live or work in close contact, amplifying the risk of the spread. The most notable outbreak in recent history took place in 2005, primarily caused by a shortage of MMR vaccines, leading to an unusually high percentage of the population remaining unvaccinated. Outbreaks can lead to spillovers between different geographical areas, potentially leading to local or global epidemics. For this reason, it is essential that the risk of contagion (and spillover) is carefully assessed and taken into account when making decisions on health policies and interventions. While the geography of England definitely plays a role in determining the spread of disease, it would be naive to consider this as the only relevant factor to take into account. Indeed, as mentioned, uptake of vaccines and other socio-economical factors also play a key role in determining the prevalence and spread of the disease. Social networks, professional networks, and mobility networks connect individuals in a number of non-trivial ways, which are very difficult to observe and objectively measure. This suggests that a modeling framework based on latent factors may represent a viable choice to explain the data and its complex patterns, without necessarily relying on a large number of covariates that can only be used as proxies at best.

Building on the recognition of these complex, interrelated factors influencing the spread of mumps, our study focuses on the analysis of reported mumps cases across nine regions of England, specifically examining the period from 2008 to 2020. To address the challenges posed by the intricate transmission patterns and latent factors at play, we introduce an original time series model that combines a log-linear time series model and latent position network models. Our final goal is to analyze the temporal dynamics and transmission patterns of mumps, providing a detailed view of how the disease has been spreading across different

regions. This framework also allows us to explore broader statistical relationships between the regions, treating the spread of disease as part of a larger, interconnected system.

From the statistical methodological point of view, this paper makes a number of contributions which advance the recent literature on network count time series modeling. Multivariate time series may be seen under the lenses of network analysis, with time series being represented as the nodes of a network. The connections between the nodes of the network, represented as edges between the nodes, describe instead the pairwise relationships among the time series. Network-based time series models have become a popular approach for analyzing multivariate data, since they provide a structured framework to uncover underlying patterns and dependencies. The network framework permits a more parsimonious model specification, where the additional structured layers tend to favor a more efficient use of model parameters. Critically, the embedding of data permits a more clear and interpretable visualization of the data, which can lead to new perspectives and results.

A point to note regarding our contribution is that we take care of the network framework using both static and dynamic models. The temporal feature can be essential for a statistical model in order to capture the inherent dynamic behavior of the time series. Time-evolving networks, unlike their static counterparts, offer a more robust way to understand the associations between variables by reflecting their true relationships as they change over time. Thus, we directly extend our previous framework (Kaur and Rastelli, 2024) and present here a development into studying dynamic time series network models, with a focus on the analysis of mumps transmission patterns in England.

Similarly to our previous work (Kaur and Rastelli, 2024),we employ the Latent Position network Model (LPM) to give a structure to the interactions between the time series. A key novelty of this paper is that we focus on the distance model, which uncovers the strength of connections between the nodes based on the Euclidean distance that separates them. The advantage of the distance model, compared to the projection model used in Kaur and Rastelli (2024) is that the distance model can give more interpretable visualizations, while maintaining a comparable flexibility and inferential setup.

In fact, we illustrate via various simulated experiments that our analysis method can tackle and accurately infer a number of complex patterns from the data, thus being able to represent a multitude of realistic scenarios. As regards the output of our modeling, the geometric nature of our LPM permits the creation of model-based summaries that can illustrate some new perspectives about the dataset. In the context of our original mumps dataset, we introduce a measure of expansion and contraction of the latent space: this has an immediate and straightforward interpretation since it provides an aggregate and model-based measure of the possibility of spillovers and contagion. Thus, this model can be used to quantify the evolution of epidemic risk and, ultimately, it can be helpful in making critical decisions regarding vaccination programs and other related medical interventions.

We emphasize that our model hinges on a latent variable specification, whereby both the time series interactions and the underlying network structure are not observed, unknown, and completely inferred from the data. As such, the role of the latent variable is to capture not only the relevance of some information that may potentially be observed (e.g. the geographical location of the nodes); but also capture many other non-observed, non-observable, and perhaps abstract socio-economic factors that possibly play a central role for this dataset.

4.2.1 Related Work

The study of multivariate time series models using network approaches has attracted considerable attention over the last few years. The traditional time series models, such as one-dimensional autoregressive models ((Box et al., 2015)), can fall short in capturing the dependencies between the time series. Vector Autoregressive (VAR) models ((Luetkepohl, 2005)), provide a very flexible structure and are largely used in econometrics and other research fields. However, these models can often be over-parametrized, especially when considering model orders larger than one. This serves as a primary motivation to consider VAR extensions using networks, which enforce more structure into the model, thus promoting parsimony.

Other recent related works include that of Knight et al. (2016), who introduced the Network AutoRegressive (Integrated) Moving Average (NARIMA) models, which define multivariate continuous time series coupled with a network structure. Their model successfully tracked the dynamics of mumps outbreaks in UK in 2005 by directly modeling the network of interactions based on the movements of infected individuals between different areas of the UK using NARIMA processes. While our approach shares similarities with Knight et al. (2016), as both works involve network time series to model the spread of mumps, our method differs by assuming that the network structure is not known and evolves dynamically over time. Zhu et al. (2017) introduce instead the Network Vector AutoRegressive (NAR) model, an extension of the traditional VAR model, which incorporates a structured network framework. In this model, a node's value is influenced not only by its own history, but also by the average historical values of its neighboring nodes. The NAR model also includes node-specific covariates, reducing complexity by using a constrained set of parameters, which improves both manageability and interpretability. Inference is performed using least squares in two asymptotic settings: (a) increasing the time sample size ($T \to \infty$) with a fixed number of nodes N, and (b) both N and T increasing simultaneously. Knight et al. (2020) introduce the Generalized Network Autoregressive (GNAR) model for continuous random variables, building on the model proposed by Zhu et al. (2017). This extension incorporates node-specific effects based on the sizes of neighboring networks. Chen et al. (2023) propose the community network vector autoregressive model, enhancing the flexibility and generality of model formulated by Zhu et al. (2017) by allowing heterogeneous effects across different network communities.

Nason et al. (2024) presents a community-α GNAR model that extends the GNAR framework, designed to identify and model community structures within network time series, which can reveal interesting dynamics, especially in the context of presidential elections in the USA. This model is particularly useful for high-dimensional settings. Our work extends the foundational study by Zhu et al. (2017) and its extensions on network time series by modeling count time series with an unknown, dynamic network structure, unlike the aforementioned models that assume a known fixed network and continuous data.

A related strand of literature focuses on continuous time series data where, similar to our approach, the network structure is unknown. However, unlike our work, these studies typically assume a fixed network structure rather than modeling it changes over time. The work of Bolstad et al. (2011) addresses the problem of inferring sparse causal networks using multivariate autoregressive processes, by establishing conditions under which a group-lasso method can consistently estimate network structures. Basu et al. (2015) introduces Granger causal models to learn a network structure from temporal panel data. The authors introduce a group-lasso regression regularization framework to facilitate this learning process. Additionally, they explore a threshold-based variant to tackle issues related to group misspecification. Barigozzi and Brownlees (2019) introduce a network-based approach for analyzing multivariate time series, defining a network for multivariate time series as a graph where the vertices represent the components of the time series, and the edges are inferred through the use of the long-run partial correlation matrix between multiple time series. Ahelegbey et al. (2020) presents a Bayesian hierarchical model combining a VAR model, Covariance Graphical Model (CGM), and a LPM model, their work centers on continuous financial time series. Additionally, their approach employs a CGM to model the network and an LPM framework to uncover the spatial position of the nodes.

Although there is extensive literature on modeling and inference for time series with continuous responses, research on multivariate count time series models for network data remains limited. Recently, work by Armillotta and Fokianos (2024) has significantly contributed to this area. They expanded the Network AutoR egressive Model (NAR) to specifically accommodate count data, addressing the complexities associated with non-random neighborhood structures. This extension resulted in the development of the Poisson Network Autoregression (PNAR) model. Armillotta and Fokianos (2024) detail the PNAR(p) model specification for both linear and log-linear cases, including discussions on stability properties, which are essential for understanding the model's behavior. Their research comprehensively examines two asymptotic regimes discussed in Zhu et al. (2017) and proposes a robust theoretical framework for asymptotic inference. In their work, Armillotta et al. (2022) introduce a statistical framework that unifies the findings of Zhu et al. (2017) and Armillotta and Fokianos (2024). This framework is designed to handle both continuous and count responses over time for each node within a known network, showcasing its practical applicability. Liu and Nason

(2023) reviews existing count time series models and presents two new models, GNARI (Generalized Network Autoregressive Integrated) and NGNAR (Negative Generalized Network Autoregressive) for count network time series.Yin et al. (2024) presents the Functional Coefficients Network Autoregressive (FCNAR) model, which extends existing count time series network models by incorporating both nonlinear autoregressive and network effects in a multivariate context. This model allows for a more flexible approach to understanding complex relationships within data sets. The above studies assume a fixed network structure across time. In contrast, our work contributes to the literature on network count time series by introducing a time series modeling framework inspired by the work of Armillotta and Fokianos (2024), with the key distinction that, in our setting, the network structure is unknown and evolves dynamically over time.

Another strand of literature that is related to our work relates to networks that evolve over time. In the dynamic networks literature, various models have been developed to extend models from the static network literature. The dynamic extension of the Latent Position Model (LPM) (Hoff et al., 2002) can be found in a number of works including Sarkar and Moore (2005), Sewell and Chen (2015), Friel et al. (2016), Durante and Dunson (2016) and, Rastelli and Corneli (2023). Other related dynamic networks models are extensions of the stochastic blockmodel, by Matias and Miele (2017), Ludkin et al. (2018), and Pensky (2019).

More recently, some works have considered dynamic network models to study the dependency structure in multivariate time series. Kang et al. (2022) propose a network autoregressive model that accounts for a time-varying network structure by combining multi-scale modeling with network-based neighborhood selection. The model is designed to capture local temporal structures and detect significant changes in the network over time, all while ensuring sparsity in the interactions. Krampe (2019) further extents the work of Knight et al. (2016) and Zhu et al. (2017) by modeling random network structure. Krampe (2019) introduces a multivariate doubly stochastic time series framework that allows for the separate modeling of dynamic attributes and the underlying network structure. The paper describes a vertex-labeled dynamic network represented by a weighted time-dependent adjacency matrix. The methodological framework of Zhu et al. (2017) is expanded by Knight et al. (2020), who propose an extension of the GNAR model, by accommodating time variations in the network structure using time-dependent weights and neighborhood structures. Another relevant contribution in this context is of Castro et al. (2012) who study count time series using embedded network structure. The authors combine spatial and temporal dependencies within the count data modeling. This work effectively incorporates spatial dependencies by applying a spatial structure to the latent continuous variables. The framework also accommodates both time-stationary and time-varying temporal correlation patterns. Tjøstheim et al. (2023) review of some recent developments in the literature of time series and dynamic networks, noting

that the literature on time series with dynamic network embeddings is much more sparse than the static case.

Our work builds on LPMs, popularized by Hoff et al. (2002) as a statistical model for social networks. This highly influential work resulted in a significant increase in the research on LPMs, yet, to our knowledge, no studies have applied LPMs to uncover temporal network structures in multivariate count time series. This gap in the literature is a key focus of our contribution to network time series research.

The rest of the paper is organized as follows: Section 4.3 describes our proposed model. Section 4.4 outlines the inferential process. Section 4.5 contains simulation studies exploring the performance of our methodology on multivariate count time series with underlying non-random and random network structures. Section 4.6 includes the application of the proposed model using the original dataset on mumps cases in England. Section 4.8 concludes the paper with a conclusion and discussion on the possible extensions.

4.3 Dynamic Time series Latent Position model (DTSLPM)

4.3.1 Data

The observed dataset consists of a multivariate time series $\mathbf{y}_t = \{y_{it}, t = 1, 2, \ldots, T$ and $i = 1, 2, \ldots, N\}$, for N different time series and T time-points. We assume that the data is Poisson distributed, and let $\boldsymbol{\lambda}_t = \{(\lambda_{it}), t = 1, 2, \ldots, T$ and $i = 1, 2, \ldots, N\}$ be the corresponding intensity process. Define $\mathscr{F}_t$ be the σ-field generated by $\mathbf{y}_{t-1}$. According to the model specification, we make the assumption that $\lambda_t = E(\mathbf{y}_t|\mathscr{F}_{t-1})$. A network time series $\mathcal{X} = (\mathbf{y}_t, \mathcal{G})$ is a stochastic process composed of a multivariate time series $\mathbf{y}_t$ and an underlying network $\mathcal{G} = (\mathcal{K}, \mathcal{E})$ where $\mathcal{K} = \{1, \ldots, N\}$ is the set of nodes on which time series are observed, $\mathcal{E} \subset \mathcal{K} \times \mathcal{K}$ is the set of weighted edges, and $\mathcal{G}$ is an undirected graph with N nodes. This article focuses on the situation where the y_{it} multivariate time series are counts, that is integers greater than or equal to zero accounting for a time-varying network structure with a fixed number of nodes and varying edge weights varying across time denoted as $\mathcal{E}_t$.

4.3.2 Model Specification

We present a multivariate time series framework characterized by a log-linear autoregressive model structure for count data, popularized through a number of works which include Doukhan et al. (2020), Fokianos (2024), and Davis et al. (2021). The proposed model exhibits a hierarchical structure and we name it the Dynamic Time Series Latent Position Model

(DTSLPM) of order 1, defined as follows:

$$
\begin{aligned}
y_{it}|\mathscr{F}_{t-1} &\sim \text{Pois}(\lambda_{it}) \\
\log(\lambda_{it}) &= \alpha + \sum_j \gamma_{ij(t-1)} \log(y_{j(t-1)} + 1) \\
\gamma_{ijt} &= \begin{cases} \beta_i & i = j \\ \frac{2}{1+\exp(d_{ijt})} & i \neq j \end{cases}
\end{aligned} \tag{4.1}
$$

where, $d_{ijt} = d(\mathbf{z}_{it}, \mathbf{z}_{jt}) = \sqrt{\sum_{i=1}^{K}(z_{itk} - z_{jtk})^2}$ for $i, j = 1, \ldots, N$ and t= $1, \ldots T$. The parameters of the model that need to be estimated are $\alpha \in \mathbb{R}$, $\boldsymbol{\beta} = \{\beta_1, \ldots, \beta_N\} \in \mathbb{R}^N$, $\mathcal{Z}_t = \{\mathbf{z}_{it} \in \mathbb{R}^2 | i = 1, \ldots, N \text{ and } t = 1, \ldots T\}$. In the proposed model, the latent space that we consider is two-dimensional so that we facilitate the visualization of network structure.

The model states that the log rate of the mean of the i-th series at time t is influenced by the log values of the i-th series itself, using an autoregressive framework. The effect of this term is quantified by the parameter β_i. In addition, the i-th series is also driven by the log values of the other series, each weighted by the coefficient γ_{ijt}. The presence of the unit value within the logs is to facilitate numerical stability. The parameters of the model $\boldsymbol{\beta}$ and γ may be zero or any positive or negative value, and their interpretation is analogous to the coefficients of a VAR time series model.

The association between any two series at any time-step t is described by the interaction parameter γ_{ijt} through a logistic function on d_{ijt} (Sarkar and Moore, 2005). This choice is motivated by the fact that it enables the quantification of the connection weight based on distances with a fairly simple interpretation. Whenever two nodes are close to each other, the interaction term has a higher value, capped at 1, which is attained when the two nodes have exactly the same position. Imposing this upper bound is reasonable since higher values would yield a non-stationary process, as discussed later in Section 4.3.3. If two nodes are far from each other, then the interaction term gets small, tending to zero when the distance increases. This can be interpreted as a situation of independence between the time series. We note that negative values of the interaction terms cannot be obtained: this is a model assumption that we make, since negative interactions typically lead to less clear interpretations, and because they are difficult to represent with the geometric nature of our model based on pairwise distances. The parameter α is the intercept of the model and it is constant over time. This facilitates the convergence of the series to a non-zero mean process. The series-specific autoregressive parameter β_i is also constant over time, and its interpretation is equivalent to that of a autoregressive model parameter: a positive value of β leads to sustained trends and positive autocorrelations, whereas a negative value leads to more erratic behavior, unclear trends, and, possibly, alternating signs on the autocorrelation plots. More information on

the interpretation on of the model parameters can be found in the previous work (Kaur and Rastelli, 2024).

The latent positions $\mathcal{Z}_t$ are a central feature of interest for our model. Since this latent space evolves over time, the latent nodes are characterized by a latent trajectory across the Euclidean space. We assume that, apriori, these latent trajectories are random walks, thus, similarly to Friel et al. (2016); Sarkar and Moore (2005); Sewell and Chen (2015, 2016), we specify a Markov process characterized by the following initial distribution:

$$\pi(\mathcal{Z}_1|\boldsymbol{\phi}) = \prod_{i=1}^{N} \mathcal{N}(\mathbf{z}_{i1}|\mathbf{0}, \rho^2\mathbf{I}), \tag{4.2}$$

and transition equation:

$$\pi(\mathcal{Z}_t|\mathcal{Z}_{t-1}, \boldsymbol{\phi}) = \prod_{i=1}^{N} \mathcal{N}(\mathbf{z}_{it}|\mathbf{z}_{i(t-1)}, \sigma^2\mathbf{I}) \tag{4.3}$$

for $t = 2, 3, \ldots, T$. For a latent space $\mathbb{R}^2$, $\mathbf{z}_{it}$ is the vector of the i^{th} actor's latent position at time t, and $\mathcal{Z}_t$ is the $N \times 2$ matrix whose i^{th} row is $\mathbf{z}_{it}$, where $\mathbf{I}$ is the 2×2 identity matrix, $N(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the normal probability density function with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ evaluated at $\mathbf{x}$, and $\boldsymbol{\phi}$ is a vector of parameters.

In the definition of our model, we achieved the desired property that the strength interaction between two nodes i and j at time t increases as the distance between their respective latent positions $\mathbf{z}_{it}$ and $\mathbf{z}_{jt}$ in the latent space decrease. Consequently, the underlying networks $G = (\mathcal{K}, \mathcal{E})$ that describe the conditional dependencies among the time series across the system at different time points are determined by the estimated weighted interconnections. In order to preserve the information across successive time steps, we make a key assumption within this study: we assume that a node can transition within latent space across time intervals, but significant transitions are improbable. This is achieved by setting the parameter σ^2 to a small value. This choice facilitates more clear visualizations since the movement of nodes tends to be the minimum movement that is required to explain the data. Besides, there could be an argument to bring up if large transitions are allowed, in that the model will try to overfit the data by having fewer constraints on its temporal evolution.

Finally, we briefly note that in this paper we arbitrarily fix the number of dimensions (K) in the latent space to 2. This decision is especially common in the literature on LPMs, albeit not necessarily well founded from a purely statistical perspective. Indeed, a clear goal would be to try to infer, perhaps automatically, the ideal number of dimensions from the given dataset. We do not pursue this goal here, as we feel that this is a problem that is currently being tackled by the literature on LPMs, and for which we do not have a clear solution yet. So, we

make a pragmatic decision of choosing two latent dimensions, since this allows for very clear representations in $\mathbb{R}^2$, while not compromising the flexibility of the model or its nature.

4.3.3 Stability of network time series process

To ensure the stability of the network time series process denoted by the pair $\mathcal{X} = (\mathbf{y}_t, \mathcal{G})$, it is critical to establish the stability of the multivariate count time series $\mathbf{y}_t$ modeled using log-linear model which takes the form:

$$\mathbf{y}_t \sim \mathsf{Pois}(\boldsymbol{\lambda}_t) \quad \log(\boldsymbol{\lambda}_t) = \alpha + \Gamma \log(\mathbf{y}_{t-1} + \mathbf{1}_N) \tag{4.4}$$

Building on the discussion by Doukhan et al. (2020), and further elaborated in our previous work (Kaur and Rastelli, 2024), a key sufficient condition required to ensure the stability and stationarity of the joint process $(\mathbf{y}_t, \boldsymbol{\lambda}_t)$ is:

$$-1 < \min_i(\beta_{ii} - r_i) < |\Omega_{\mathsf{max}}(\Gamma)| < \max_i(\beta_{ii} + r_i) < 1$$

where $\Omega_i(\Gamma)$ denotes the eigenvalues of $N \times N$ matrix $\mathbf{\Gamma}$, β_{ii} is the diagonal entry in row i, and $r_i = \sum_{i \neq j} |\gamma_{ij}|$ is the sum of the absolute values of the off-diagonal elements in row i of $\mathbf{\Gamma}$. The above inequality gives a sufficient condition on the interaction matrix of the model to ensure the stationarity of the process. More details can be found in Kaur and Rastelli (2024).

4.3.4 Prior specification

We adopt a Bayesian framework to estimate the model parameters. In particular, we allocate independent and non-informative priors to the parameters α, $\boldsymbol{\beta}$, and $\boldsymbol{\mathcal{Z}}_t$. Both α and $\boldsymbol{\beta}s$ are assigned independent normal priors with a mean of zero and a standard deviation of 10. The choice of a large standard deviation is intentional, as it reduces the impact of prior assumptions on the posterior distributions.

As mentioned, we assign independent Gaussian priors to the increments of the parameters $\boldsymbol{\mathcal{Z}}_t$. The initial latent positions $\boldsymbol{\mathcal{Z}}_1$ arise from a zero mean Gaussian prior with high standard deviation ρ. At each time frame t, the set of two-dimensional latent coordinate matrix $\boldsymbol{\mathcal{Z}}_t$ are generated using the positions at the previous time frame $\boldsymbol{\mathcal{Z}}_{t-1}$ as a mean, and a small standard deviation σ is used to penalize large displacements over time (Sarkar and Moore, 2005). The parameters ρ and σ may be either be gamma distributed with shape and rate hyperparameters $\{a_\rho, b_\rho, a_\sigma, b_\sigma\}$, respectively; or they may be considered hyperparameters themselves and thus be arbitrarily set. A regards the gamma prior approach, we follow previous works which include Friel et al. (2016); Sewell and Chen (2015, 2016). In those applications where we fix the values of these standard deviations, we set them as $\rho = 10$ and $\sigma = 0.05$.

4.4 Model Fitting

4.4.1 Estimation

We consider two primary methodologies for inference on the parameters of a DTSLPM, implemented through the STAN framework (Stan Development Team, 2024). One of these methodologies uses the Limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm (Nocedal and Wright, 2006) to optimize the likelihood function, which is particularly effective for large-scale problems like DTSLPM. The algorithm iteratively updates parameter estimates by combining information about gradients and curvature, efficiently converging even when the number of parameters increases with the size of the time series.

In addition to the optimization-based approach, we also employed Hamiltonian Monte Carlo (HMC) (Betancourt and Girolami, 2015; Duane et al., 1987; Neal, 2012), a sophisticated Markov Chain Monte Carlo (MCMC) technique that uses Hamiltonian dynamics to sample from the posterior distribution of model parameters. In HMC, the model parameters represent the position in a high-dimensional space, while auxiliary momentum variables, typically following a multivariate normal distribution, are introduced to aid in sampling. This combination of position and momentum allows the algorithm to explore the parameter space more efficiently by leveraging gradients of the log-density function, facilitating better transitions, and reducing random walk behavior.

Following standard practices of Bayesian model fitting, we primarily rely on the sampling approaches to fit our models, and run the sampling until we observe satisfactory convergence to a stationary distribution through trace plots and other related convergence diagnostics that are available in STAN.

4.4.2 Post-processing of posterior samples

A key challenge in fitting Bayesian LPMs is that the likelihood function associated to the model will only depend on the latent positions through the pairwise distances. This means that the model is not identifiable up to any transformation of the latent space that preserves the pairwise distances between nodes. These transformations can be rotations, reflections, or translations of all positions simultaneously for all time frames. While this may not be necessarily an issue under an optimization setting, under a Bayesian context the circumstances become more problematic. The samples that we use to calculate posterior summaries are sensitive to rotations, reflections, and translations of the latent space, and thus should be suitably transformed before the aggregations.

For this, we post process the posterior samples of latent locations employing the so-called Procrustes matching (Gower and Dijksterhuis, 2004), adjusting the procedure of Hoff et al. (2002) to a dynamic network setting, as already done by Friel et al. (2016). Since we are

interested in transforming the complete trajectories, we stack the N latent positions for each of the T times frames, thus obtaining latent spaces made of TN points for all nodes at all times. We then proceed with standard Procrustes analysis on these stacked matrices. As a reference set, we take the Maximum-a-Posteriori (MAP) latent positions $\hat{\mathcal{Z}}$ which is a $(TN) \times 2$ matrix. Then, for each latent space $\mathcal{Z}^{(1)}, \ldots, \mathcal{Z}^{(M)}$ that we obtain during the sampling, we perform a Procrustes transformation on each of these matrices (Friel et al., 2016; Sewell and Chen, 2015, 2016) using the MAP as a reference. As a final operation, we un-stack each of the matrices used to obtain the transformed collection of points for every N, T and sample.

4.5 Simulation study

We examine the performance of our DTSLPM framework in the estimation of the model parameters, using a number of realistic scenarios where we try to mimic some typical dependency structures that are observed in real applications.

4.5.1 Static time series distance model

Initially, we consider a simpler version of our model whereby the latent positions remain constant over time. As a consequence, the interactions between series do not change over time, and so the model may be defined as "static". In the static version of the DTSLPM model, the form remains the same as in Eq. 4.1, with the only difference being that γ_{ij} has no time dependence.

We conduct a simulation study across different network dimensions with varying sample sizes. The data generation process involves the simulation of the time series using three main parameters: the intercept parameter α, the autoregressive coefficient for each series $\boldsymbol{\beta}$, and the latent positions $\mathcal{Z}$. The parameter α is generated from a uniform distribution between -3 and 3. In a similar manner, the parameter β is also generated from a uniform distribution, ranging from -1 to 1. We note that the value of α does not have any meaningful effect on the stationarity of the process, whereas the autoregressive parameter necessarily has to be between -1 and 1. It is important to note that no constraints are imposed on the parameters during the inferential procedure in either version of the proposed model. As a result, there is a possibility that the inference obtained through both methods may lead to a non-stationary series.

In this section, the latent positions remain fixed over time, denoted as, $\mathbf{z}_1, \ldots, \mathbf{z}_N$, and are generated independently from a multivariate normal distribution, ensuring that the series are organized into at least two clusters. The rationale behind the presence of clustering is that we aim to obtain interaction values that are appreciably different, from very strong to weak, which can lead to reasonable and realistic settings. If all nodes are positioned too far apart in

the latent space, they may become independent from other nodes, making the estimation of their positions particularly challenging. By contrast, if all nodes are close to each then we obtain too many strong positive interactions and unstable behavior. The clustering framework represents then some middle ground that that can correspond to a realistic setting, with both dependence and independence patterns being present.

Each cluster is centered in the latent space with a unique mean vector, which determines its location, and all clusters share the same covariance matrix. Initially, the covariance matrix is diagonal with a variance of 0.1, which leads to fairly strong interactions for nodes that are in the same cluster. However, our goal is to obtain datasets whereby the interactions between series are meaningful, so that we can accurately estimate the interaction parameters and, as a consequence, the latent positions. For this reason, once an initial set of latent positions is created, we iteratively adjust the latent space by expanding the latent positions by a small factor. This iterative process continues recursively as long as the stability conditions are satisfied as outlined in Section 4.3.3. Through this expansion procedure, we ensure that the stability conditions for the time series remain satisfied, while at the same time we can obtain a compelling dataset with meaningful interactions.

Additionally, the number of clusters increases with the number of series. This choice is driven by the need to maintain stable series, as having too many nodes within a single cluster increases the likelihood that the interaction matrix will fail to meet the stability conditions. In consideration of this, we selected 2 clusters for datasets of 5 to 10 nodes, and as the number of nodes increased to 15, we increased the number of clusters to 3.

4.5.2 Large scale simulation study using optimization

We consider datasets with $N = 5, 10, 15$ and time frames $T = 100, 500, 1000$. Specifically, we simulate multivariate time series datasets with $N = 5, 10$ in two clusters located at $(0, 0)$ and $(3, 3)$, and for $N = 15$, we use three clusters located at $(0, 0)$, $(-1.8, 3)$, and $(-3, -3)$ in the latent space. This choice is driven by the need to maintain stable series, as having too many nodes within a single cluster increases the likelihood that the interaction matrix will fail to meet the stability conditions. All covariance matrices are defined as $\Sigma = diag(0.1, 0.1)$, where this choice of mean vectors and covariance matrix leads to strong within-cluster connections and weak between-cluster connections. For each different pairs of values of N and T, we generate 100 different realizations of multivariate time series data with varying random seeds. We employ the L-BFGS optimization on each dataset to obtain point estimates by maximizing the joint posterior distribution from the model. Then, we compare our estimated parameters with the true parameter values that have generated the data.

We illustrate the estimation error for the parameters α and $\boldsymbol{\beta}$ using the density plots in Figure 4.1.

Similarly, we report a comparison between true and estimated values in Figure 4.2. The

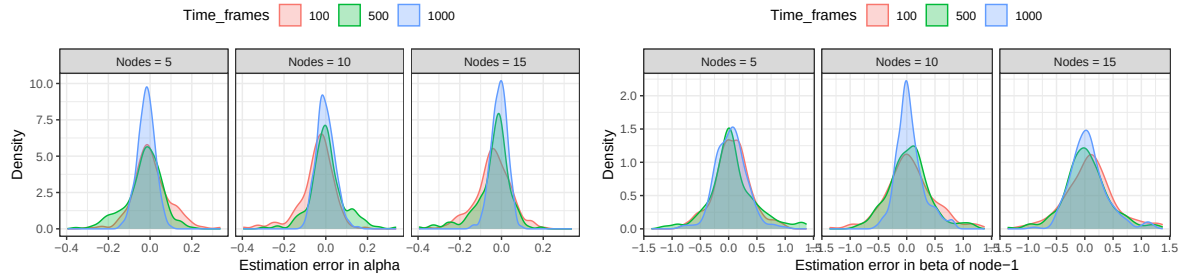

Fig. 4.1 Static model. Estimation errors across 100 different network datasets: Left panel shows the error for the parameter α, and the right panel shows the error for the parameter β, with node counts $N = \{5, 10, 15\}$ and time points $T = \{100, 500, 1000\}$.

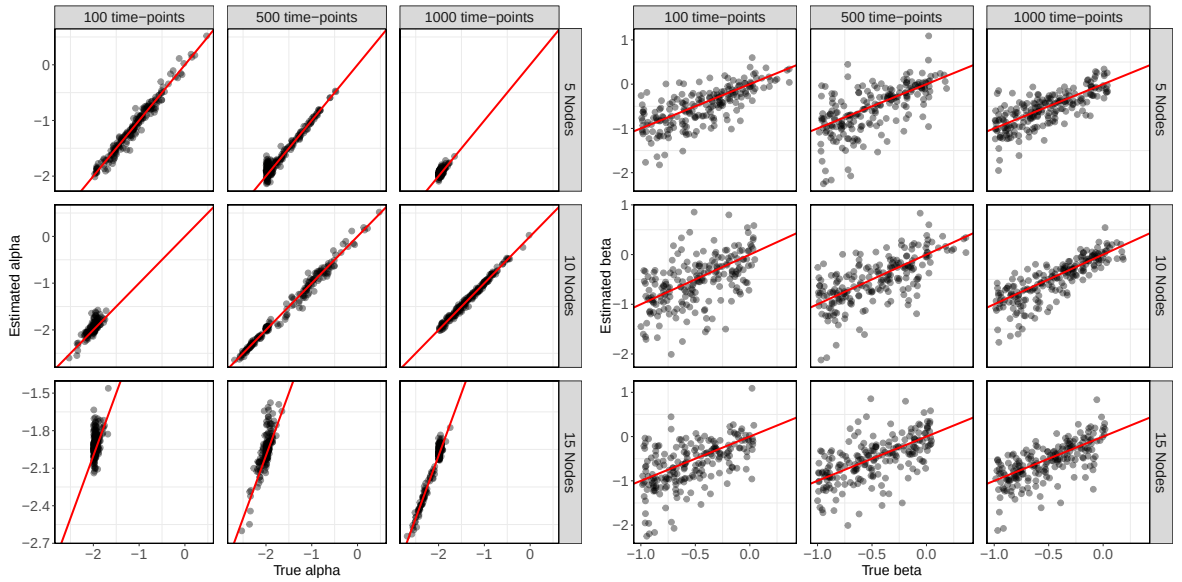

Fig. 4.2 Static model. True values (x-axis) versus estimated values (y-axis) for the parameters across 100 different network datasets with nodes $N = \{5, 10, 15\}$ and time points $T = \{100, 500, 1000\}$. Left: α, Right: β

simulated results lead to several key conclusions. As shown in the density plots in Figure 4.1, there are observable trends as N and T increase. Specifically, the estimation errors for both parameters α and $\boldsymbol{\beta}$ decrease as the number of time frames grows grows. Similar insights can be gathered from the scatter plots in Figure 4.2, where smaller sample sizes result in greater dispersion of the estimated values, while larger sample sizes yield estimates that are closer to the true values. In all plots, we observe that the error decreases when the number of time frames increases. For α, the error decreases also when the number of nodes increase, or when the number of time frames and number of nodes jointly increase. For β, we do not notice any particular improvement when the number of nodes increases. This is reasonable since each β parameter is node-specific, and so adding more nodes also increases the number of parameters to estimate. Overall, the results demonstrate the effectiveness of the L-BFGS optimizer in providing reliable parameter estimates, even with relatively limited data. Our results are in agreement with the recent evidence also shown for a similar model in Kaur and Rastelli (2024).

Similarly to Kaur and Rastelli (2024), we assess the estimation accuracy for the latent positions using the ratios between estimated distances and true distances. The plot in Figure 4.3 presents the distributions of the ratios separately for interactions that happen between-cluster and within-cluster. This method of plotting ratios is also employed by Sewell and Chen

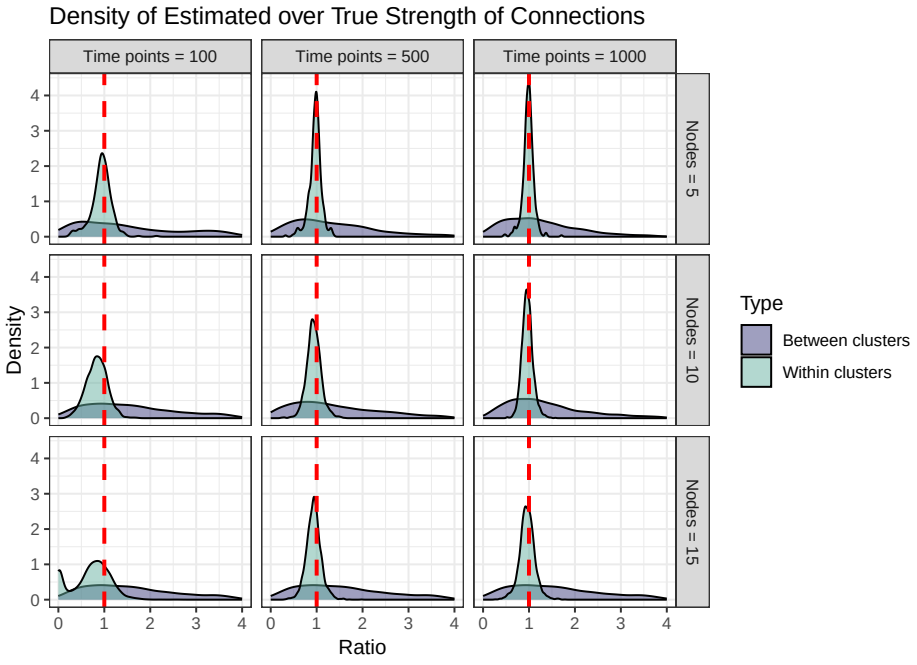

Fig. 4.3 Distribution of pairwise interaction ratios between and within clusters, comparing estimated latent positions with true latent positions among 100 different network datasets with nodes $N = \{5, 10, 15\}$ and time points $T = \{100, 500, 1000\}$.

(2015, 2016) to compare the estimated latent space with the true latent space, bypassing any potential rotation or translation. A narrow distribution plot centered around 1 suggests that the estimated latent space closely matches the true latent space, up to a scaling factor.

The distribution of ratios for within-cluster interactions is consistently centered around 1, which is indicated by the red dotted line in the plot. Additionally, as the sample size increases, these distributions become more tightly concentrated around 1. This pattern is consistent across different network sizes. Conversely, the distribution of ratios for between-cluster interactions appears relatively flat. These results emphasize a critical feature of the model, whereby dependency information is required to accurately estimate the latent positions; conversely, when series tend to be independent, the model struggles to recover the latent space because there is no information to be used.

4.5.3 Dynamic time series distance model

In this section, we apply our methodology to simulated data in order to assess the temporal performance of the DTSLPM model through a series of experiments. Our primary objective is to assess the model's capability in capturing network connections. It is important to note that the goal is not to recover the exact trajectories but rather to estimate trajectories that yield approximately similar interconnections between the nodes.

In this section, since we consider one individual dataset at a time, we proceed to apply a fully Bayesian procedure using HMC to obtain estimates of the latent node positions $\mathcal{Z}_t$ and other model parameters. With regards to the random walk process prior on the latent trajectories, the initial and subsequent standard deviations of the increments are modeled using

$$\rho \sim \text{Gamma}(2, 2)$$
$$\sigma \sim \text{Gamma}(5, 10)$$

similarly to Friel et al. (2016); Sewell and Chen (2015, 2016). For each dataset, we assess convergence through visual inspection of the trace plots, along with the examination of the $\hat{R}$ statistic, also known as the potential scale reduction factor (Gelman et al., 2013). To ensure the quality and reliability of the parameter estimates, we also consider the Effective Posterior Sample Size (ESS).

Experiment 1

In this initial experiment, the data-generating mechanism is designed to capture the time-varying interrelationships between nodes within clusters. We note that, under a dynamic

latent space setting, the definition of cluster becomes more blurred, since nodes can move along trajctories and thus change community or exhibit other types of dynamic behaviors. We consider a network of four nodes observed over 200 time points, and we let the nodes move in non-trivial ways, but nonetheless following clear regular patterns.

The data generation process is as follows:

$$\alpha \sim \text{Uniform}(-3, 3)$$
$$\boldsymbol{\beta} \sim \text{Uniform}(-1, 1)$$

The initial positions of nodes are generated such that they are grouped up in two pairs, with nodes changing their relative positions while remaining within their cluster (see top panels of Figure 4.4). Thus, the objective of this experiment is to track node-specific activity within the clusters.

In this experiment, we obtain a total of $5,000$ iterations, with a burn-in period of $3,000$ iterations and a thinning rate of 2, resulting in $1,000$ samples to represent the posterior distributions of the model parameters. Table 4.1 presents the posterior mean estimates for two parameter values alongside their true values.

Table 4.1 Posterior means with 95% credible intervals for experiment 1.

Parameter	True	Mean	95% credible interval
α	-0.31	-0.20	(-0.33,-0.07)
β_1	-0.86	-0.87	(-1.20,-0.57)
β_2	-0.25	-0.12	(-0.33,0.07)
β_3	-0.84	-0.97	(-1.30,-0.65)
β_4	-0.74	-0.64	(-0.94,-0.35)

In Figure 4.4, we display two separate plots, each comparing the true and estimated node-specific cluster activities. The left plot corresponds to the first half of the time period, while the right plot represents the second half, showing the corresponding latent positions. These plots highlight the strong alignment between the estimated and true node-specific trajectories.

Another key objective of this experiment is to capture the intensity of interconnections between nodes within the clusters. Figure 4.5 demonstrates that the estimated interactions closely follow the true interactions.

Snapshots of the true and estimated latent positions are also provided in Figure 4.6, where we see a good agreement thus indicating accurate estimation.

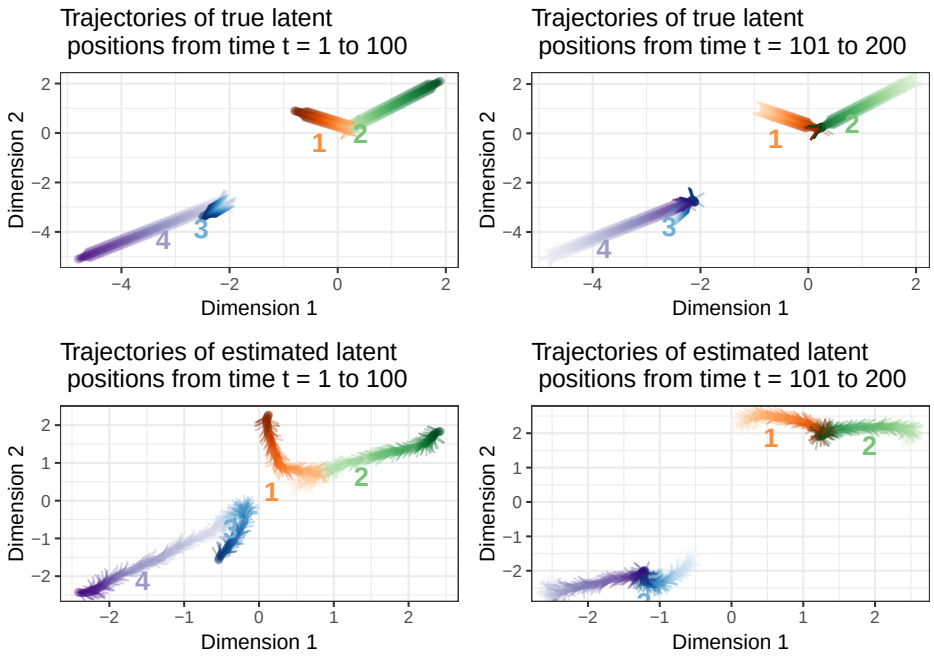

Fig. 4.4 Posterior means of latent node positions, with each color representing a different node and arrows indicating the temporal evolutions of the trajectories. The left panel shows the trajectories of each node during the first half of the time period (t = 1 to 100), while the right panel displays the trajectories during the second half (t = 101 to 200).

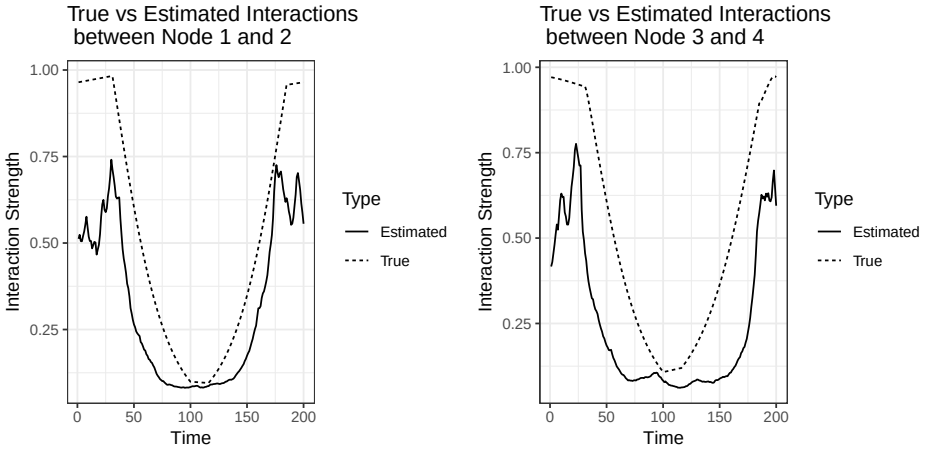

Fig. 4.5 Plot of the posterior means of the estimated pairwise connections between nodes, shown as solid lines, and compared with the true pairwise connections, represented as dotted lines. The left panel compares the connection between nodes 1 and 2, while the right panel compares the connection between nodes 3 and 4.

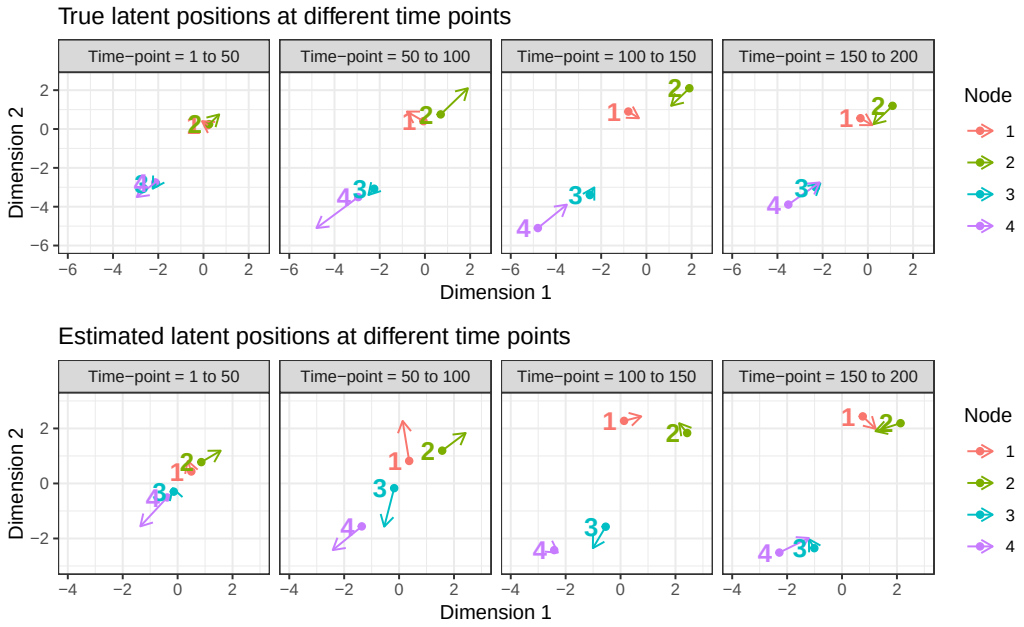

Fig. 4.6 Posterior means of true and estimated latent node positions at different time points, with each color representing a different node and arrows indicating the temporal evolutions of the trajectories.

Experiment 2

In this second experiment, we utilize a data-generating method that is designed to measure the temporal interactions of a node as it transitions between clusters. The experiment involves a network of four nodes observed over 200 time points. The data-generating process for this experiment is identical to that used in Experiment 1, however we create different trajectories in this case. The primary objective is to track the temporal interactions and trajectory of node 1 as it transitions between two clusters.

The HMC chains were run for $10,000$ iterations, with a burn-in period of $5,000$ iterations and a thinning rate of 5, resulting in $1,000$ samples to represent the posterior distributions of the model parameters. The results for the posterior mean estimates of two parameter values, as well as the true parameter estimates, are presented in Table 4.2. Figure 4.7 displays two distinct plots, respectively displaying the actual and the estimated trajectories particular to each node.

Table 4.2 Posterior means with 95% credible intervals for experiment 2.

Parameter	True	Mean	95% credible interval
α	-0.005	0.16	(0.04,0.28)
β_1	-0.53	-0.67	(-0.79,-0.54)
β_2	0.01	-0.04	(-0.14,0.06)
β_3	-0.67	-0.74	(-0.92,-0.57)
β_4	-0.15	-0.33	(-0.46,-0.19)

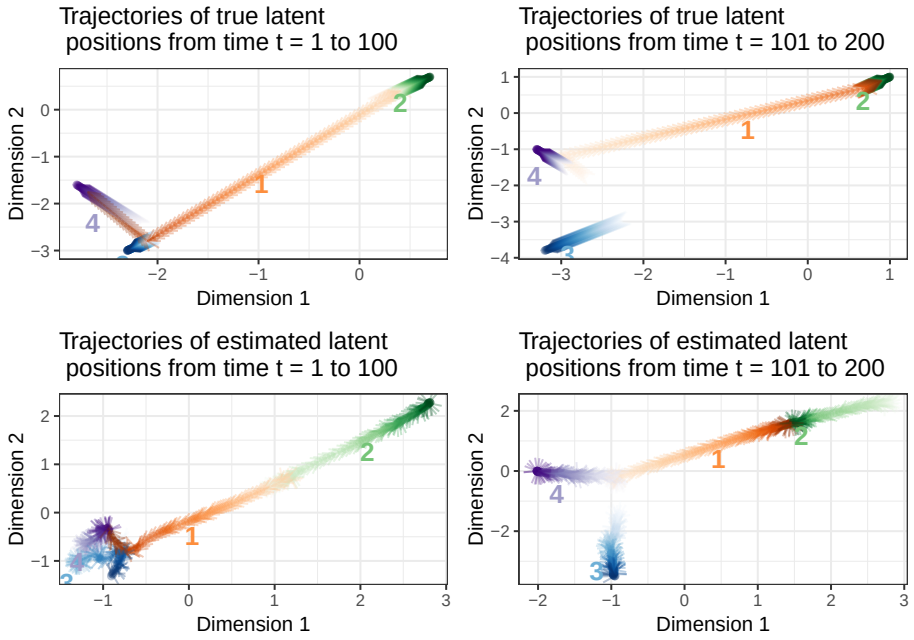

Fig. 4.7 Posterior means of latent node positions, with each color representing a different node and arrows indicating the temporal evolutions of the trajectories. The left panel shows the trajectories of each node during the first half of the time period (t = 1 to 100), while the right panel displays the trajectories during the second half (t = 101 to 200).

The fitted model demonstrates a robust ability to capture the trajectories for all nodes. The results, as depicted in Figure 4.7, highlight that the trajectory of node 1, which is of particular interest, aligns closely with the true trajectory. Figure 4.9 further illustrates that the estimated movement of node 1 is well aligned with the its true movement.

The trajectories of the other nodes are also well recovered by the model. Indeed, the model performs well in recovering the interaction intensities between node 1 and the other nodes, as shows in Figure 4.8.

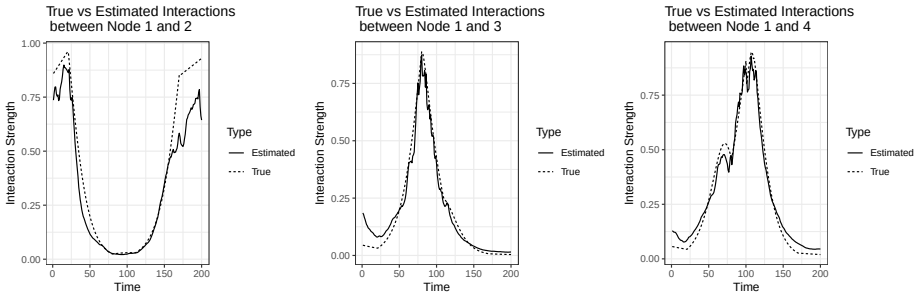

Fig. 4.8 Plot of the posterior means of the estimated pairwise connections between nodes, shown as solid lines, and compared with the true pairwise connections, represented as dotted lines. The left plot compares the connection between nodes 1 and 2, the middle plot compares the connection between nodes 1 and 3, and the right panel compares the connection between nodes 1 and 4.

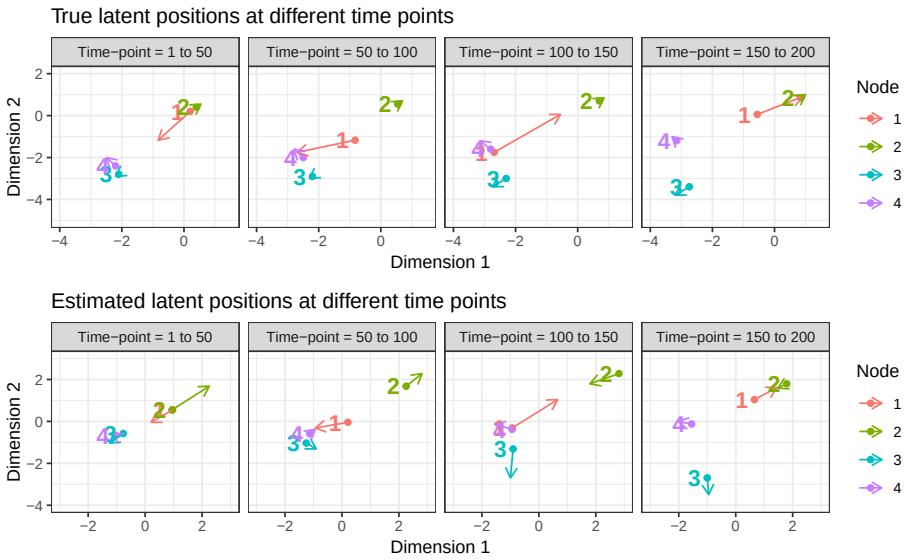

Fig. 4.9 Posterior means of true and estimated latent node positions at different time points, with each color representing a different node and arrows indicating the temporal evolutions of the trajectories.

Experiment 3

In the final experiment, the data-generating mechanism is intended to capture the temporal interactions between nodes as they simultaneously switch clusters. We analyze dynamic networks consisting of 4 nodes observed over 200 equally spaced time points. The data generation process in this experiment is the same as that used in Experiment 1, albeit with different trajectories. Initially, the nodes are divided into two groups, with nodes 1 and 4 then

exchanging their positions. The main objective is to track how the connections between the nodes change as a result of these shifts in cluster membership.

The HMC algorithm was run for $50,000$ iterations, with the first $30,000$ iterations discarded as burn-in. A thinning rate of 20 was applied, resulting in $1,000$ samples used to approximate the posterior distributions of the model parameters. Table 4.3 shows strong correspondence in the true parameter estimates and the posterior mean estimates of model parameters α and $\boldsymbol{\beta}$.

Table 4.3 Posterior means with 95% credible intervals for experiment 3.

Parameter	True	Mean	95% credible interval
α	0.22	0.37	(0.17,0.56)
β_1	-0.67	-0.70	(-0.81,-0.58)
β_2	-0.15	-0.19	(-0.33,0.07)
β_3	-0.85	-1.10	(-1.30,-0.90)
β_4	-0.04	-0.09	(-0.16,-0.01)

Figure 4.10 presents the actual and estimated trajectories for each node. The fitted model shows a high level of accuracy in capturing the trajectories of nodes as they change clusters, as illustrated in the figure. Figure 4.12, which depicts the latent trajectories of each node at specific time points, further confirms that the estimated trajectories closely reflect the true trajectories, reinforcing the model's accuracy in tracking the movement of nodes over time.

In addition, the model performs well in recovering the connections between nodes with shifting cluster memberships. Figure 4.11 demonstrates that the estimated interconnections closely align with the true interconnections.

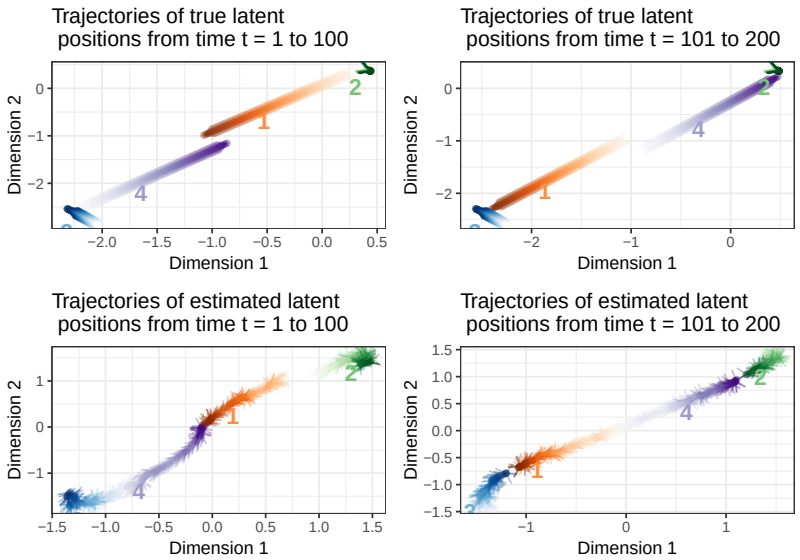

Fig. 4.10 Posterior means of latent node positions, with each color representing a different node and arrows indicating the temporal evolutions of the trajectories. The left panel shows the trajectories of each node during the first half of the time period (t = 1 to 100), while the right panel displays the trajectories during the second half (t = 101 to 200).

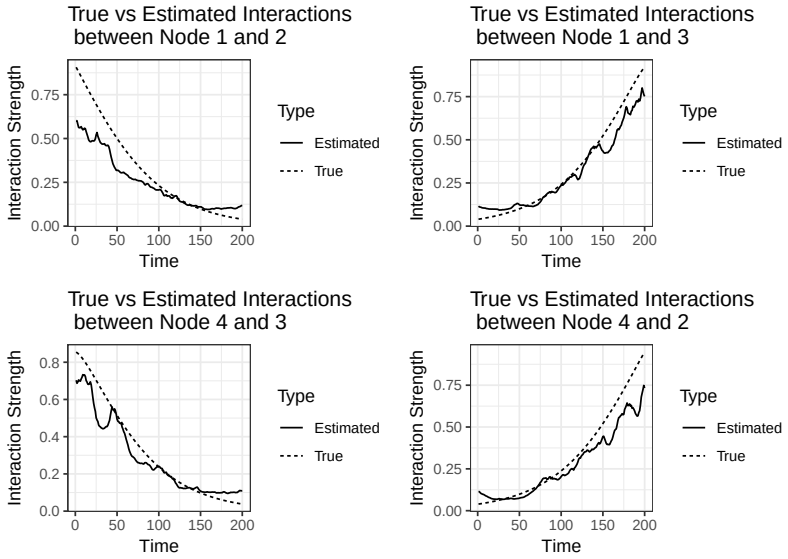

Fig. 4.11 Plot of the posterior means of the estimated pairwise connections between nodes, shown as solid lines, and compared with the true pairwise connections, represented as dotted lines. The top panels compares the connections of node 1 with nodes 2(left) and 3 (right), while the bottom panel compares the connections of node 4 with nodes 2 (left) and 3 (right).

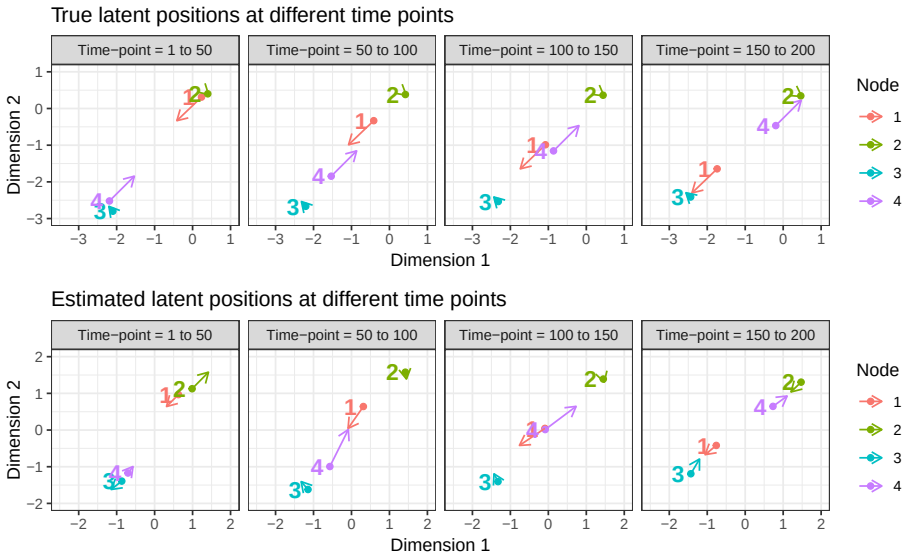

Fig. 4.12 Posterior means of true and estimated latent node positions at different time points, with each color representing a different node and arrows indicating the temporal evolutions of the trajectories.

4.6 Analysis of mumps cases

We study the number of cases of mumps in $N = 9$ regions of England recorded quarterly from $2008(Q1) - 2020(Q2)$, for a total of 50 quarters. The data were collected from the Public Health England website[1]. Cases include those confirmed by oral fluid IgM antibody tests and by routine laboratory reports.

These observed data can be written as y_i^t for regions $i = 1, \ldots, 9$ over quarters $t = 1, \ldots 50$: an illustration of it is provided in the left panel of Figure 4.13. In the right panel, we present a

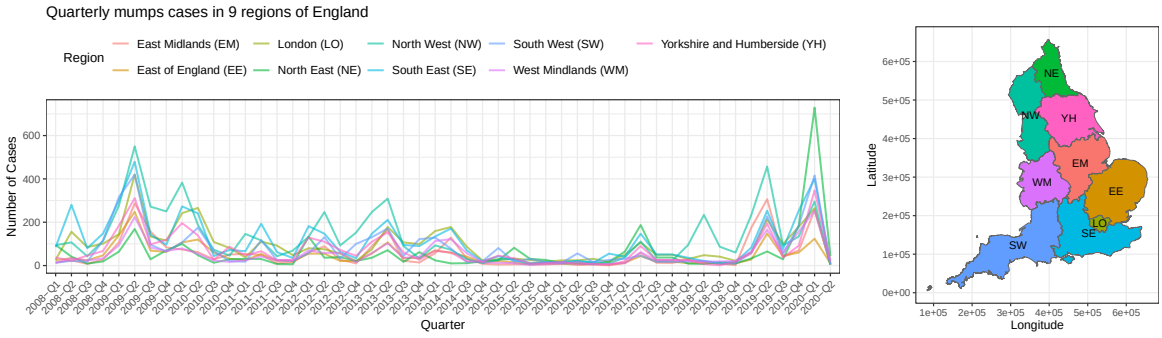

Fig. 4.13 Left: quarterly number of mumps cases in 9 different regions of England. Right: regional breakdown of England.

[1] https://www.gov.uk/government/organisations/public-health-england

geographical representation of England and its regions. Snapshots for 4 selected time points are shown in Figure 4.14, highlighting the periods with the highest peaks of mumps cases, with the number of cases being color-coded: deep red representing a high number of cases and yellow indicating fewer cases. The goal of this analysis is to employ the DTSLPM model

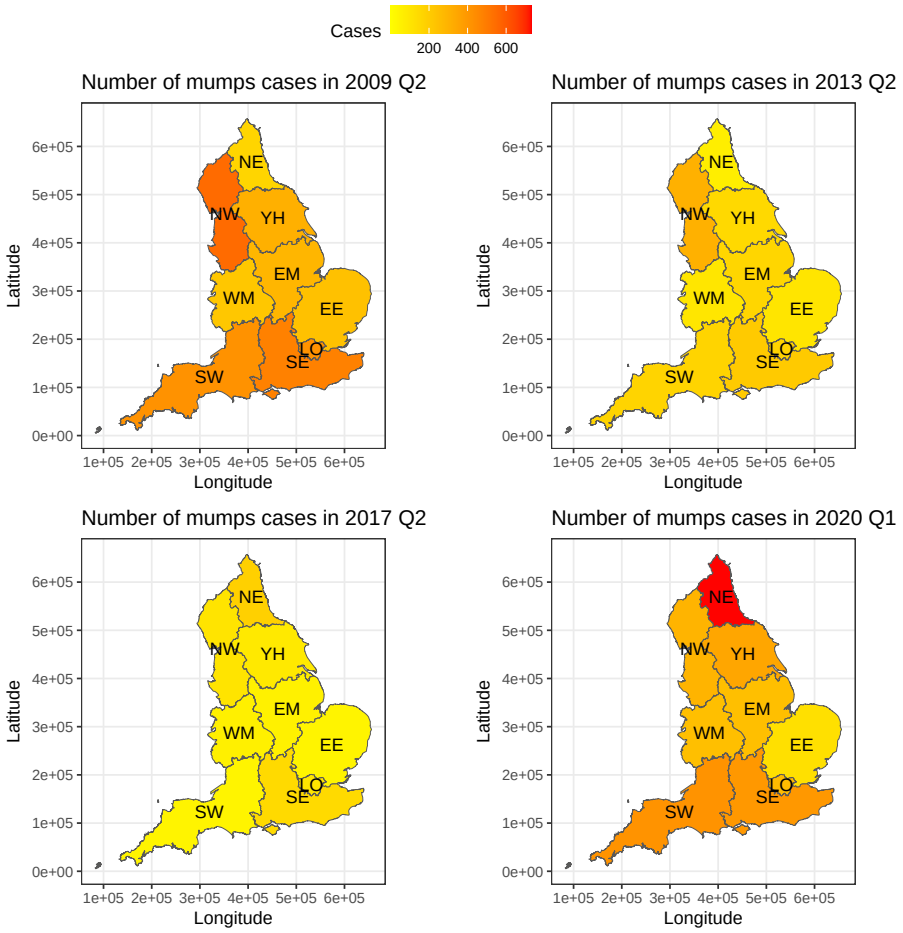

Fig. 4.14 Distribution of mumps cases across regions of England: 2009 Q2 (top-left), 2013 Q2 (top-right), 2017 Q2 (bottom-left), and 2020 Q1 (bottom-right).

to highlight patterns of interactions between the number of mumps cases in different regions, at different times. These latent patterns can provide critical insights on possible contagion routes, and, ultimately, on the risk of an escalation of the spread of disease.

For completeness, we fit both the static and the dynamic version of our model. We note that, in either model, we do not make use of any covariates, or geographical information on the regions. We point out that this additional information may definitely be used in our model, similarly to the work of Kaur and Rastelli (2024). However, we do not consider this additional information within the scope of this work based on a latent variable framework. Indeed, the role of the latent positions that we estimate is to capture all hidden patterns, caused by

geographical factors or otherwise. Additionally, we also note that the geographical information per se, is not necessarily the driving factor for contagion. Indeed, human mobility and human interactions may not necessarily be determined purely by the geographical distance, but rather, at least intuitively, they may be more related to transportation networks and other socio-economic factors that are difficult to observe and measure.

4.6.1 Static model for the mumps cases

We proceed by fitting the DTSLPM model as discussed in Section- 4.5.1, to analyze the underlying static network that represents the interactions between regions. We use HMC for fitting the DTSLPM and thus to obtain the posterior samples. A total of $30,000$ iterations were performed, with a burn-in period of $20,000$ iterations and a thinning rate of 10, resulting in $1,000$ samples to represent the posterior distributions of the model parameters. For the model parameters $\alpha, \boldsymbol{\beta}$, and $\mathcal{Z}$, we employ non-informative priors, specifying independent normal priors with a mean of zero and a standard deviation of 10 for each.

We use the $\hat{R}$ statistic and visual examination of the trace plots to check for convergence (Gelman et al., 2013). To assess the quality and reliability of the samples of parameter estimates, the effective sample sizes were checked. The results demonstrate good convergence, and the posterior summaries for all model parameters, except the latent positions, are presented in Table 4.4. We observe that all the $\boldsymbol{\beta}$ parameters are positive, indicating positive

Table 4.4 Posterior means with 95% credible interval for the fitted static model.

Parameter	mean	95% credible interval
α	2.18	(2.12,2.23)
β_{EM}	0.27	(0.22, 0.33)
β_{EE}	0.32	(0.16,0.46)
β_{LO}	0.34	(0.28,0.41)
β_{NE}	0.47	(0.43,0.51)
β_{NW}	0.51	(0.48,0.56)
β_{SE}	0.40	(0.33,0.46)
β_{WM}	0.32	(0.25,0.39)
β_{YH}	0.18	(0.13,0.24)
β_{SW}	0.29	(0.24,0.33)

autocorrelations. The North West (NW) region shows the strongest persistence in mumps cases over time within the same region, while Yorkshire and Humberside (YH) exhibit the weakest persistence.

We now turn to a detailed analysis of the pairwise relationships between the regions, which constitutes a key aspect of this study. The inclusion of the interaction term γ_{ij} in our model enables us to investigate these regional interactions, using the spatial positioning $\mathcal{Z}$ as

a central element for interpreting these connections. This approach provides both quantitative data and visual representations, allowing us to identify specific patterns in the spread of mumps.

The HMC output resulted in 1000 posterior samples of latent positions $\mathcal{Z}$ for each region of England. To post-process these latent positions for analysis, we applied a Procrustes transformations, as discussed in Section 4.4.2, using the MAP as a reference for orientation.

The left panel of Figure 4.15 shows the estimated regional positions, derived by calculating the posterior means of the latent positions for each region, following the application of the Procrustes transformation. In the right panel, the posterior average of the interactions matrix

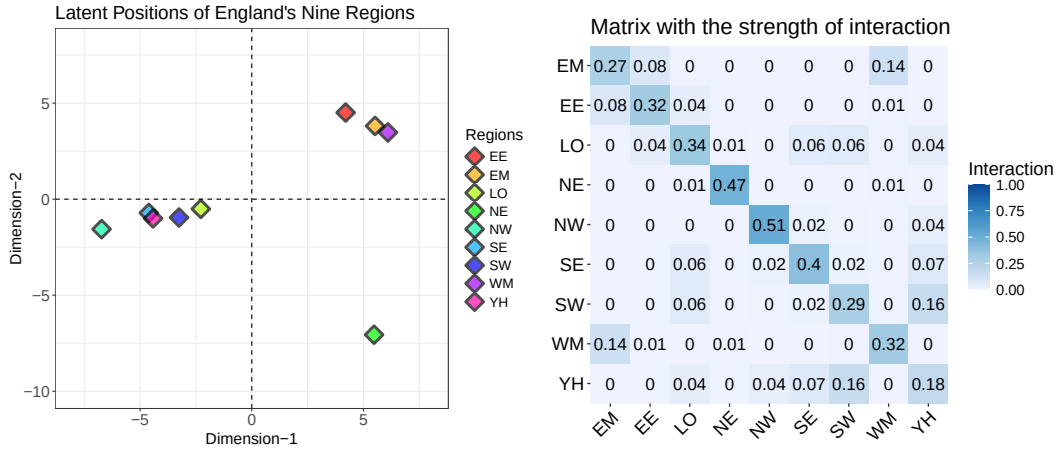

Fig. 4.15 Left: latent space representation showing the posterior mean of locations for each region in England. Right: A matrix of connections between the different regions of England.

is presented, where the off-diagonal elements represent the pairwise interaction terms (γs), and the diagonal elements capture the autoregressive effects (βs). Together, these plots provide insights into how the latent space models the interdependencies among the time series.

The right panel of Figure 4.15 reveals distinct patterns, with the North West (NW), North East (NE), and South East (SE) exhibiting the strongest autoregressive effects, while London (LO), East of England (EE), East Midlands (EM), West Midlands (WM), and the South West (SW) exhibit moderate to weaker effects. Yorkshire and Humberside (YH) show the weakest autoregressive effect. These patterns may reflect the differing historical levels of mumps cases across regions.

The latent positions shown in the left panel of the same Figure reveal that Yorkshire and Humberside (YH), as well as the South West (SW), are positioned closely in the latent space, as are East Mainlands (EM) and West Mainland (WM). This proximity reflects the interconnection between these regions, also quantified by the non-diagonal elements on the right panel of Figure 4.15. Although the overall connections are modest, the proximity of the

nodes indicates subtle interactions in the spread of mumps between these regions, quantified on a scale from 1 (strong positive connection) to 0 (independence). The geographical proximity of East Mainlands (EM) and West Mainland (WM) corresponds to one of the strongest connections observed. Interestingly, despite being geographically distant, Yorkshire and Humberside (YH) and the South West (SW) show pronounced connections, too. These findings highlight spreading patterns that are not necessarily directly affected by geographical proximity.

4.6.2 Results for dynamic time series distance model

We extend now the analysis by fitting our full DTSLPM model in its dynamic version, in order to uncover the underlying temporal network and gain insights on mumps cases dynamics. The HMC algorithm was run for $300,000$ iterations, with a burn-in period of $250,000$ iterations and a thinning rate of 25, resulting in $2,000$ posterior samples for the model parameters. Non-informative priors, as discussed in Section 4.3.4, were used for the model parameters α, $\boldsymbol{\beta}$, and $\mathcal{Z}_t$. The remaining parameters related to latent positions are the hyperparameters ρ and σ for the movement of nodes that are user-specified. The hyperparameters are specified as follows: $\rho = 10$ and $\sigma = 0.05$.

The posterior summaries for all model parameters, except the latent positions, are presented in Table 4.5. We observe in this case that the majority of the $\boldsymbol{\beta}$ parameters are

Table 4.5 Posterior means with 95% credible intervals for the fitted dynamic model.

Parameter	Mean	95% credible interval
α	1.45	(1.02,1.83)
β_{EM}	-0.31	(-0.49,-0.13)
β_{EE}	-0.60	(-0.79,-0.17)
β_{LO}	-0.21	(-0.34,-0.08)
β_{NE}	-0.12	(-0.25,0.03)
β_{NW}	0.02	(-0.09,0.14)
β_{SE}	-0.03	(-0.14,0.07)
β_{WM}	-0.46	(-0.66,-0.12)
β_{YH}	-0.03	(-0.16,0.10)
β_{SW}	-0.22	(-0.34,-0.08)

negative, indicating alternating positive and negative autocorrelations, which contribute to more unpredictable behavior. This result on the $\boldsymbol{\beta}$ parameters can be justified by the absence of any consistent upward or downward trend in mumps cases across the regions of England. Instead, the number of mumps cases in each region fluctuates, increasing and decreasing over time, which can be explained with a negative autoregressive coefficient.

We now turn to a crucial aspect of the analysis: examining the pairwise relationships between the regions of England based on the estimated spatial positions $\mathcal{Z}_t$, which evolve over time t. The model fit generated 2,000 posterior samples of the latent positions $\mathbf{z}_{it}$ for each i^{th} region in England. For analysis, these latent positions were post-processed using Procrustes transformation, as outlined in Section 4.4.2, with the MAP estimate serving as the orientation reference. Figure 4.16 illustrates the posterior means of the latent locations of each region in England in each quarter from 2008 to 2020. The figure provides a visual representation of

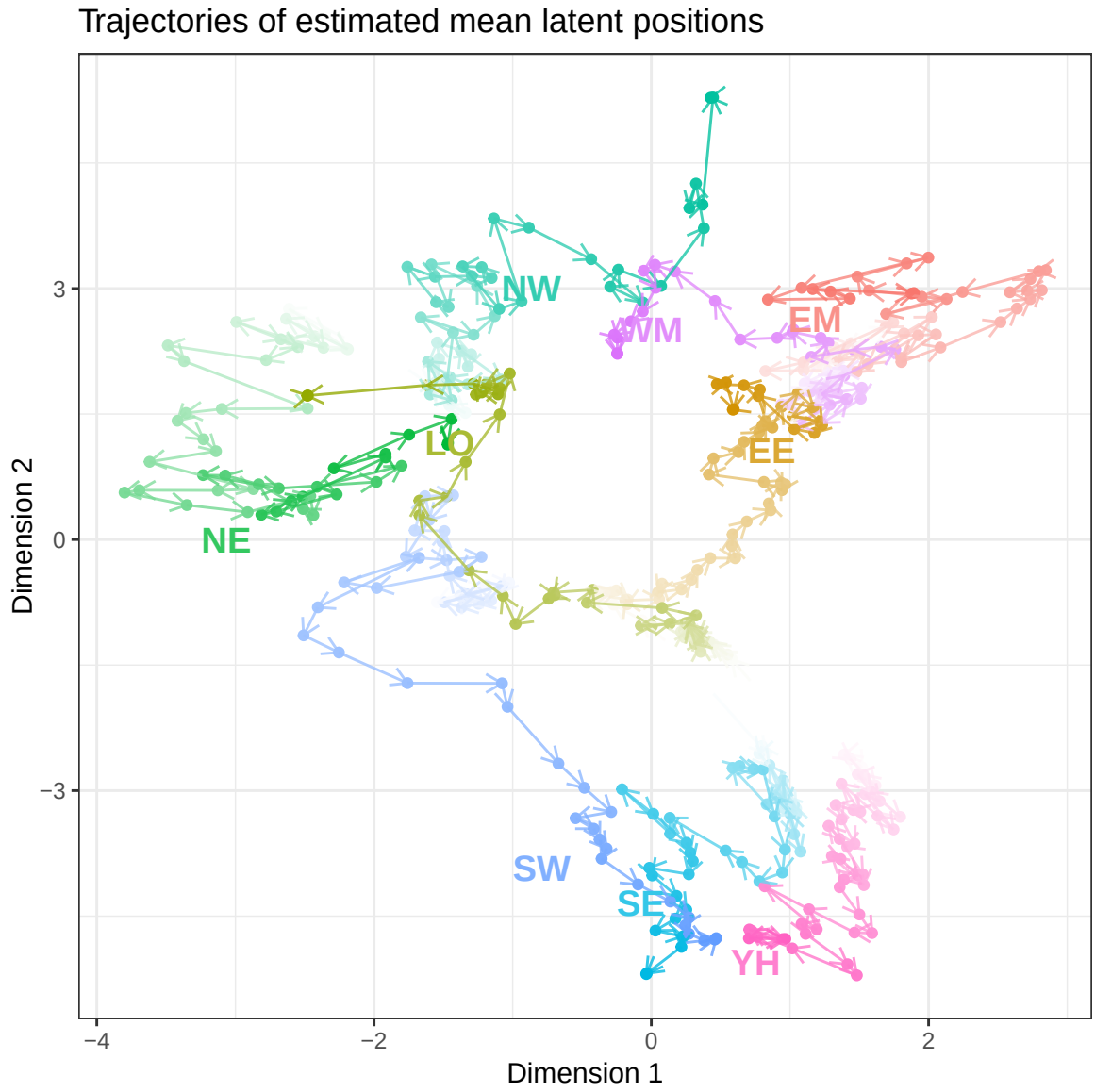

Fig. 4.16 Posterior means of latent positions for each region in England. Each color represents a different region and arrows indicating the temporal evolutions of the trajectories.

the hidden patterns of interconnections between the nodes and how they evolve with time

based on the proximity of their locations in the latent space. The trajectories obtained by each region reveal some geographical patterns. We break down the information contained in this plot by focussing on four selected time points, and show the corresponding latent spaces in Figure 4.17. Similarly, and for completeness, the latent interactions are also show in Figure 4.18, but using the actual geographical positions of the regions. Based on the data from 2008 to 2020, there is evidence of a strong connection in the spread of mumps infection between the East Midlands (EM) and West Midlands (WM) and their neighboring region in England. During the initial phase, the transmission dynamics between these regions were notably strong, indicating a significant level of infectious spread. However, a decline in direct transmission or spread between these regions is observed as time progresses. Additionally, a chain of transmission is observed, where mumps infections appear to be indirectly linked between West Midlands (WM) and East of England (EE) through the East Midlands (EM) region. Notably, the connections between East Midlands (EM) and East of England (EE), as well as West Midlands (WM) and East of England (EE), have been strengthening over time, highlighting the dynamic nature of mumps spread in these areas.

Despite not being geographically connected, there is a significant link in the spread of mumps cases between the South East (SE) region and Yorkshire and Humberside (YH) over the observed time period. This suggests indirect transmission pathways or shared risk factors driving the infection dynamics across these regions. Furthermore, from 2016 to 2020, Yorkshire and Humberside (YH), South West (SW), and South East (SW) formed a cluster of interconnected regions, with each influencing the spread of mumps in the others.

Before 2016, the South East (SE) region was notably affected by its neighboring region London (LO). However, the South West (SW) region exhibited significant connections with multiple regions, even those that were not its direct geographical neighbors. These connections evolved over time, linking the South West (SW) with East of England (EE), London (LO), North East (NE), and North West (NW). This pattern indicates a broader and dynamic network of mumps transmission, highlighting the role of indirect pathways in the spread of the infection across these diverse regions.

Additionally, the South East (SE) region has been notably affected by its neighboring areas, London (LO) and the South West (SW). The transmission dynamics indicate a strong epidemiological link with London (LO) until 2015, which then shifted towards a stronger connection with the South West (SW) in the subsequent years reflecting the shifting transmission routes.

An interesting pattern for the distribution of mumps cases across England is that the highest number of cases consistently occurs in the northernmost regions, specifically in the North East (NE) and North West (NW). Figure 4.14 illustrates this trend, revealing that during the four peak periods depicted, the initial two periods exhibit the highest mumps cases in the North West (NW), while the last two peak periods show the highest cases concentrated in

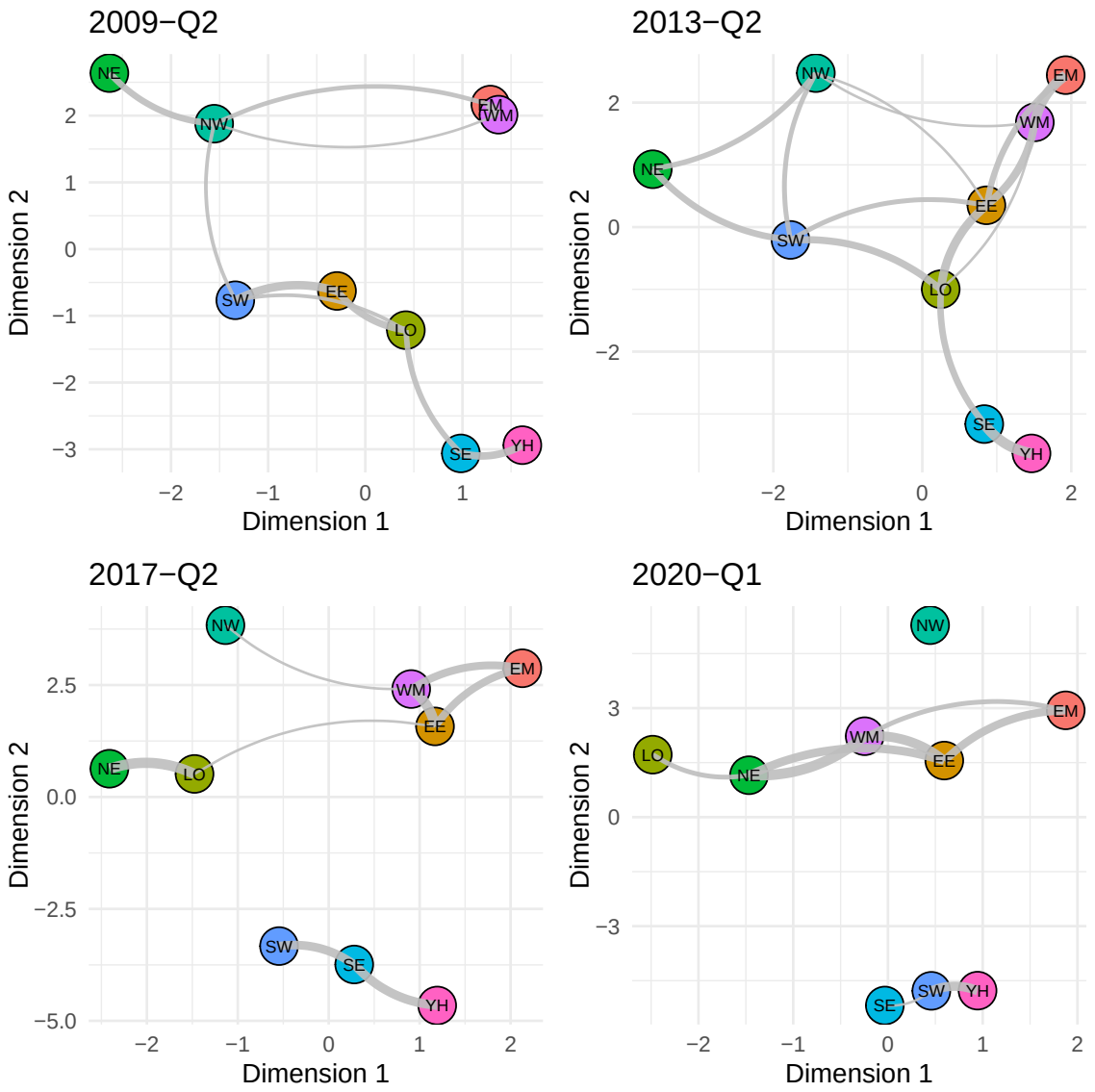

Fig. 4.17 England's mumps cases time-stamped network: 4 networks representing the posterior mean of latent location of each region and size of the edge represents the strength of the connection between the regions. The colors of the nodes represent different regions.

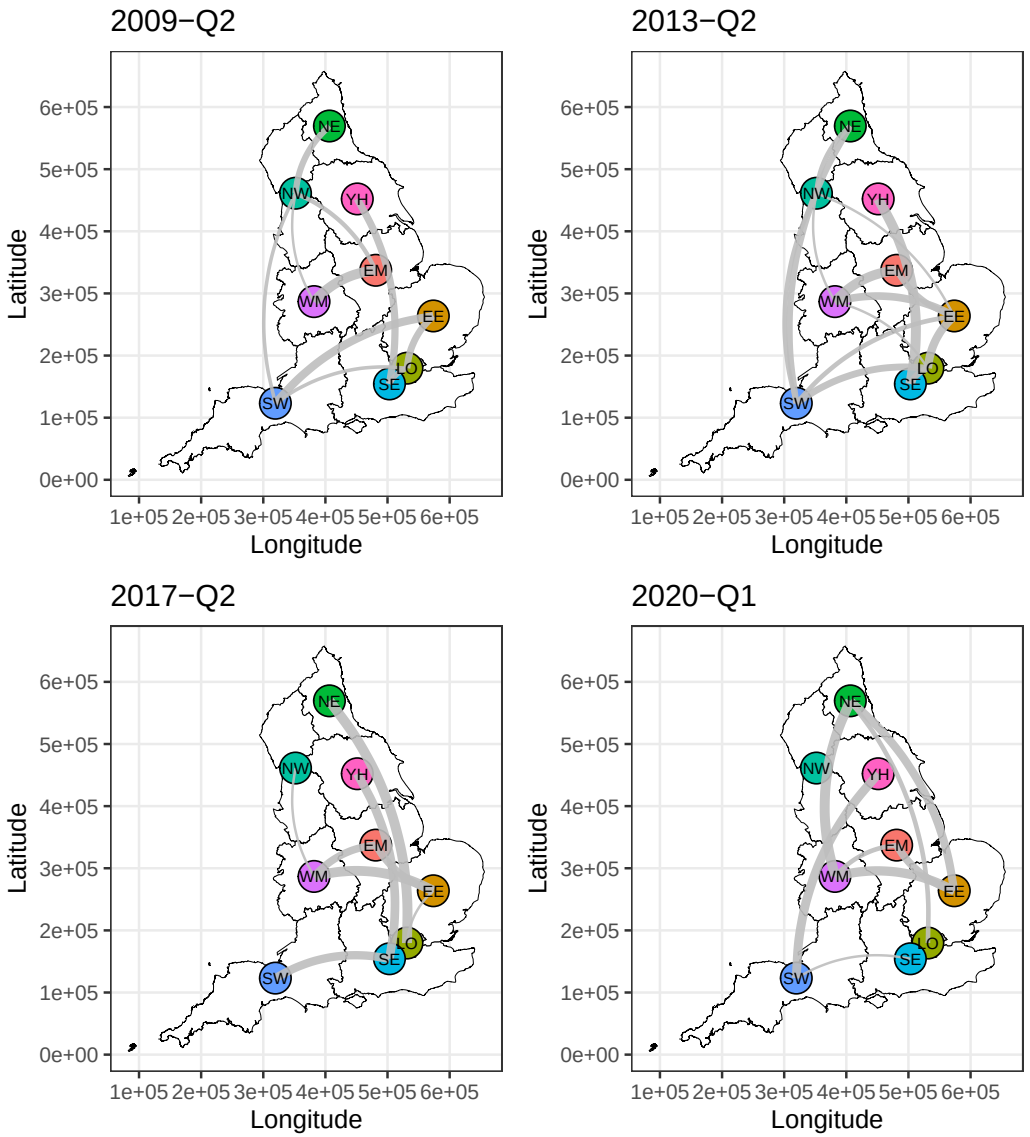

Fig. 4.18 England's mumps cases time-stamped network: 4 networks display each region at its true geographic position, with edge thickness representing the strength of connections between regions. The colors of the nodes represent different regions.

the North East (NE). This spatial distribution highlights the persistent vulnerability of these northern regions to mumps outbreaks over time.

Additionally, distinct patterns emerge in their connections with other regions during these peak periods. When the North West (NW) recorded the highest number of cases, it displayed significant connections with several regions, including North East (NE), East Midlands (EM), West Midlands (WM), and South West (SW), facilitating the spread of the infection over time. In contrast, during its peak periods, the North East (NE) primarily showed strong connections with the West Midlands (WM) and had a particularly notable link to London (LO), indicating focused transmission pathways between these areas.

London emerges as the most central node in the latent space overall, as depicted in Figure 4.16, illustrating its evolving connections with various regions over time. This centrality is further emphasized in Figure 4.19, which shows how London's connections with other regions changes throughout the observed period. As the central node, London (LO) both

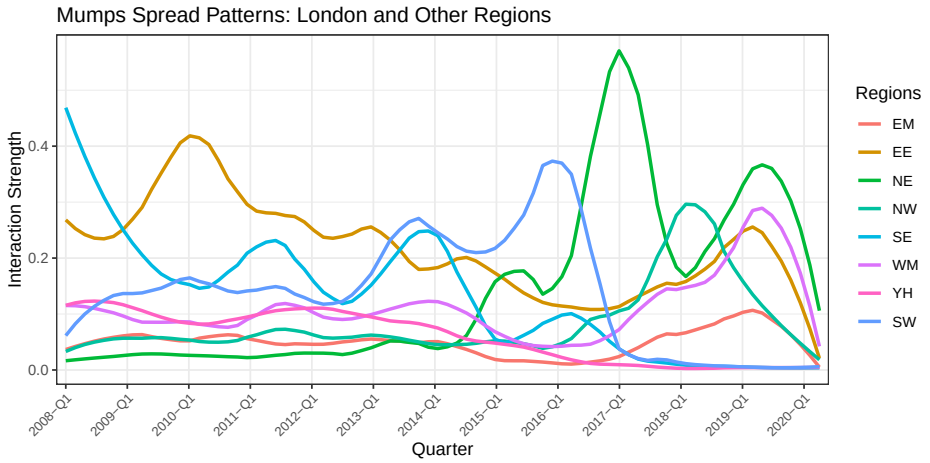

Fig. 4.19 Smoothed temporal dynamics of interconnections between London and other regions

influences and is influenced by different regions within the network. It consistently maintains a connection with its neighboring region, East of England (EE), throughout the entire time span. Furthermore, from 2015 onward, London (LO) develops significant links with the North East (NE) and North West (NW), particularly during periods when these regions experience peak mumps cases. This highlights London's crucial role in both driving and responding to mumps transmission dynamics across the network.

Figure 4.17 illustrates that the network estimated during the last peak period (2020 Q1) exhibits distinct clustering patterns among the regions. Specifically, regions such as London (LO), East Midlands (EM), East of England (EE), West Midlands (WM), and the North East (NE) form a cohesive cluster, indicating strong inter-regional connections and potential pathways for mumps transmission. In contrast, the South East (SE), South West (SW), and Yorkshire and Humberside (YH) are grouped into a separate cluster, highlighting

their localized interaction in the spread of the infection. Notably, the North West (NW) region appears isolated, showing minimal connections with other regions during this period. This clustering pattern underscores the geographical dynamics in mumps transmission, where certain regions exhibit tightly knit connections, while others, like the North West (NW), seem to operate independently.

4.6.3 Risk of contagion

One major advantage of using LPMs is that they permit various useful model-based summaries to be created. These measures can be extracted as summaries of the latent space, thus being themselves dynamic over time. The empirical variance concept outlined by Friel et al. (2016) is a prime example of this. The measure enables us to quantify the level of expansion or contraction of the latent space. By measuring the empirical variance of the latent positions at each time, we can express numerically how close to each the nodes tend to be, at each point in time. This is a very valuable summary in this application, since it offers valuable insights into whether the risk of spread of infection strengthens or weakens over time. Indeed, a large empirical variance signals that the number of cases tend to evolve independently, suggesting that there is no spillover and contagion. By contrast, a contracted latent space is suggestive of strong interactions between the series and thus the possible presence of frequent contagion.

The empirical variance is calculated from the output as follows:

$$\mathsf{S}_t = \frac{1}{2N} \sum_{i=1}^{N} (\mathbf{z}_{it} - \bar{\mathbf{z}}_t)^\top (\mathbf{z}_{it} - \bar{\mathbf{z}}_t) \tag{4.5}$$

where $\bar{\mathbf{z}}_t = \frac{\sum_{i=1}^{N} \mathbf{z}_{it}}{N}$ for $t = 1, \ldots, T$ and "$\top$" represents the transpose. The above computation is performed for each HMC iteration. Figure 4.20 illustrates the evolution of the empirical variance, displaying an upward trend over the observed period. This trend indicates that as time progresses, the interconnections between regions decrease, suggesting that the spread of mumps infection becomes less pronounced as the latent space expands and regions move farther apart from each other in the latent space. This pattern is also highlighted in Figure 4.17, where edges between the regions seem to decrease as time progresses indicating lesser interconnection among regions. This growing distance between latent positions reflects a weakening of inter-regional connections, suggesting that the transmission of mumps cases among regions has become weaker over time.

The computational time required for parameter estimation using either the LBFGS or Bayesian methods is relatively low for simulation experiments conducted on both the static and dynamic versions of the proposed model. Specifically, the execution time for a single replication on a large network with 30 nodes using LBFGS ranges from approximately 2 to

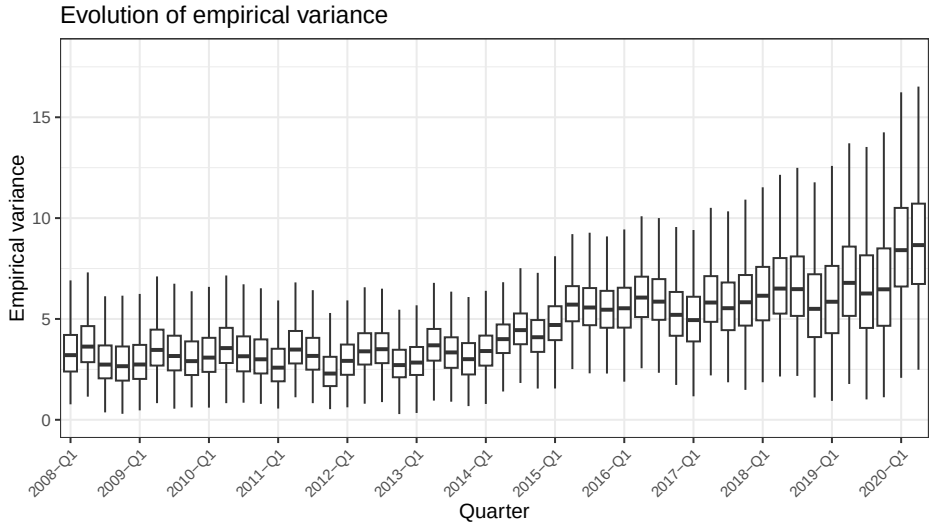

Fig. 4.20 Temporal evolution of the posterior distribution of empirical variance for the period 2008 (Q1) to 2020 (Q2).

3 minutes, while Bayesian experiments with 4 nodes and 200 time points require less than 60 minutes of CPU time on an Intel Core i7 desktop computer with a Windows x64-based processor. In contrast, the computational time for Bayesian estimation applied to mumps case data is approximately 2 hours for the static model and around 3 days for the dynamic distance time series model. The latter is particularly computationally intensive.

There is a clear need to identify potential methods to accelerate this estimation procedure. Alternative approaches could be developed to improve the efficiency of the inferential process, making the model more feasible for small-sized networks of approximately 20–30 nodes with 100–200 time points. However, it is crucial to note that the inferential procedure becomes increasingly computationally demanding for moderate to large-sized networks, as the computational time scales quadratically with the number of nodes and linearly with the number of time points. Additionally, as the number of nodes increases, a greater proportion of cross-connections become weak, making it more challenging to accurately estimate the relative positions of the nodes.

4.7 Goodness of model fit

To assess the goodness of fit of the proposed DTSLPM, one useful test is to evaluate and compare the autocorrelation function (ACF) and cross-correlation function (CCF) plots of true mumps cases observed in each region of England with the posterior predicted mumps cases evaluated using the fitted model. This will bring a clear picture of how well our model fits the

data. To generate the posterior predictive series, the series is generated iteratively using predictions of the fitted model.

We compare the ACF values of the mean posterior predictive series to the true ACFs of the actual mumps cases in East Mainlands as shown in Figure 4.21. This comparison demonstrates a strong alignment between ACF values. The posterior predictive interval demonstrates minimal variance in its coverage, supporting the model's effectiveness.

Furthermore, we examine cross-correlations across regions, with a particular focus on East Mainland and its relationships with other strongly connected regions, namely East of England, North West, and West Mainland. Figure 4.22 demonstrates a high degree of agreement between the CCF values of observed mumps cases in East Mainland and North West and their respective CCF values derived from the mean posterior predictive series, further supporting the model's accuracy and reliability.

Additionally, Figure 4.22 shows a fairly good comparison between the CCFs of observed mumps cases in East Mainland and East of England, as well as between East Mainland and West Mainland, with their respective CCFs from the mean posterior predictive series. However, the predictive interval observed is slightly wider at some lags, indicating some degree of uncertainty in the model's estimates.

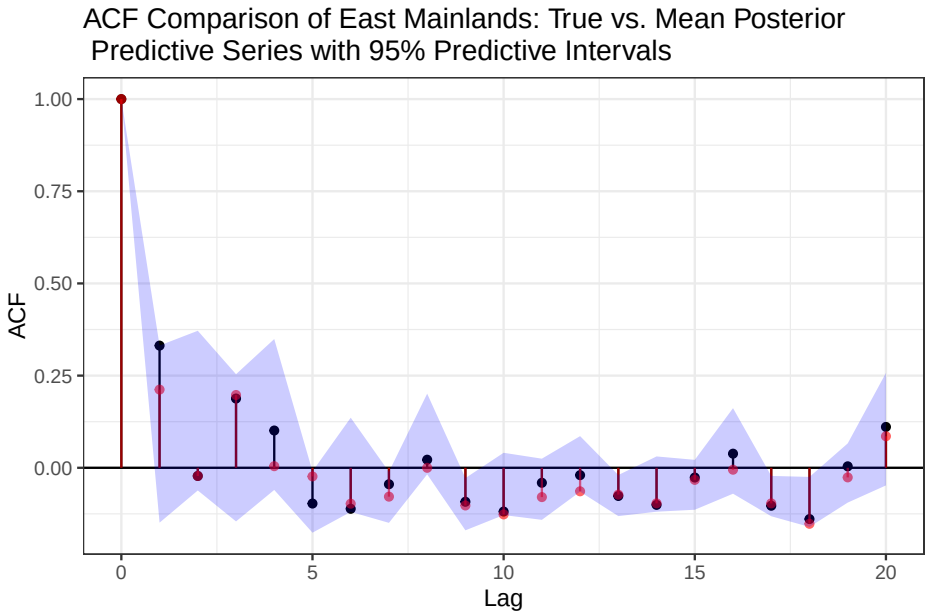

Fig. 4.21 ACF plots for time series of the number of mumps cases in East Mainlands, with the true ACF shown in black, the ACF of mean posterior predictive series in red and 95% predictive interval represented by blue shaded region.

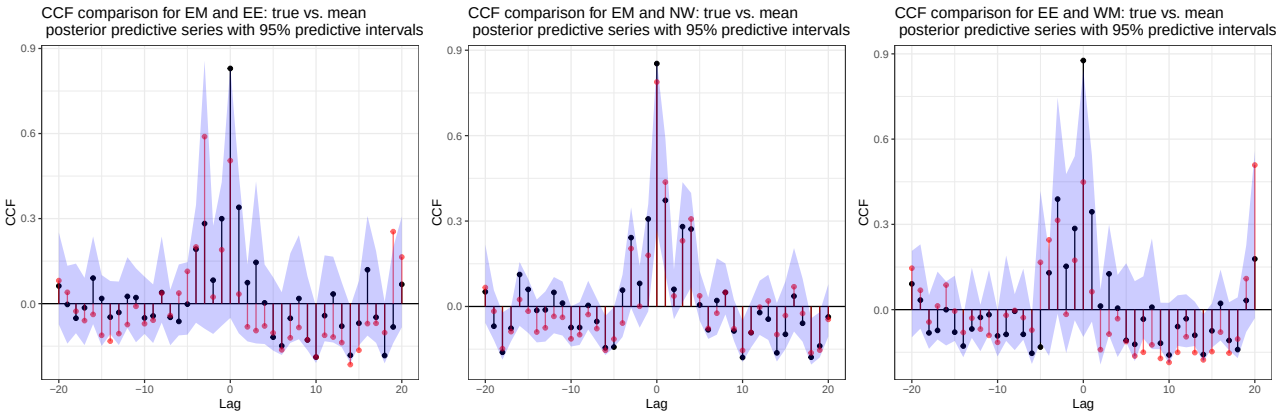

Fig. 4.22 CCF plots for number of mumps cases in EM with other regions, with the true CCF shown in black, the CCF of mean posterior predictive series in red and 95% predictive interval represented by blue shaded region.

4.8 Conclusions

We have introduced a new methodology for the analysis of time series of counts, which relies on a multi-layer hierarchical model, whereby we infer both the interactions between the time series, and the network that generates those interactions. The model has a clear advantage over other autoregressive time series models, in that it combines their flexibility with more clear parameter interpretations. The network formulation allows us to create a parsimonious structure, which can help disentangle some of the complex interaction patterns that determine the observed datasets.

We also proposed an in-depth analysis of a dataset of mumps cases in England, using both a static and a dynamic models. Our objective was to estimate an underlying network of interactions between the regions of England, and how this network may change over time. An important aspect of our analysis is that we bypass the geographical information, to focus on a latent variable formulation that can directly capture other non-observed or non-observable factors. Our model provides a clear model based measure of the risk of contagion, which is directly created from the estimated latent space by measuring its expansions and contractions over time. The results from our model suggest that the level of contagion has been constantly decreasing over the last few years, possibly thanks to a higher uptake of the relevant vaccines. In addition, our results consistently highlight that the northern regions are the most vulnerable areas, and that the London area emerges as a central hub in both propagating and mitigating mumps transmission across the network.

In this paper, we have proposed both a static and dynamic variants of our model, with results suggesting some relevant differences and use for them. The dynamic framework provides the most flexibility, but also a much higher number of model parameters. Through a

number of simulations we have shown that either model can be accurately estimated on small datasets, thus recovering a variety of dependency patterns that we believe to be realistic.

Our work leaves a number of viable directions for future research. A main critical question that remains unanswered is definitely the computational scalability of our methodology. In this paper, we only focused on relatively small datasets, primarily due to long computing times. While STAN provides an accurate solution to the model estimation task, there are definitely other possible approaches that may be explored to speed up inference.

Secondly, a key question for latent position models regards the number of latent dimensions to choose. We did not pursue this research question here, and chose two dimensions as is common in many other works. However, selecting an ideal number of groups that is specific to the application can definitely lead to a superior methodology, especially considering that one main motivation for this paper was to address the over-parametrization that can occur in time series models.

A critical challenge often encountered when working with count time series is addressing overdispersion. This issue was also observed in the data application we worked on. While the proposed model demonstrated good performance and achieved satisfactory results, we did not investigate the potential advantage of using negative binomial models over Poisson models for handling overdispersion. A promising direction for future research could involve incorporating negative binomial models to better account for overdispersion. The modeling framework proposed by Guo and Zhu (2024), which offers flexibility in modeling multivariate count time series with overdispersion, could be extended to scenarios where the network structure is both unknown and dynamic over time.

Another comment can be made for the autoregressive order of a DTSLPM. In this paper, we have only considered processes of order one. However, it would be of significant interest to explore how network structures can be utilized for processes of higher orders. One possible way to propose a generalized DTSLPM model of order p, where y_{it} depends on past p lagged values can be defined as:

$$\begin{aligned} \log(\lambda_{it}) &= \alpha + \sum_{h=1}^{p} \sum_{j=1}^{N} \gamma_{ij(t-h)} \log(y_{j(t-h)} + 1) \\ \gamma_{ij(t-h)} &= \begin{cases} \beta_{i(t-h)} & i = j \\ \frac{2}{1+\exp(d_{ij(t-h)})} & i \neq j \end{cases} \end{aligned} \tag{4.6}$$

where, $d_{ijt} = d(\mathbf{z}_{it}, \mathbf{z}_{jt}) = \sqrt{\sum_{i=1}^{K}(z_{itk} - z_{jtk})^2}$. The interpretation of the model parameters remains consistent with that of the DTSLPM(1). The stationary conditions established for the DTSLPM(1) model can be extended to the generalized case. Finally, differently from the recent work (Kaur and Rastelli, 2024), we have not included covariates or seasonality

in the DTSLPM formulation. However, we note that this may done in an analogous and straightforward way for those applications that would strongly require these features.

References

Ahelegbey, D., Carvalho, L., and Kolaczyk, E. (2020). A bayesian covariance graphical and latent position model for multivariate financial time series. *SSRN Electronic Journal*, (181).

Armillotta, M. and Fokianos, K. (2024). Count network autoregression. *Journal of Time Series Analysis*, 45(4):584–612.

Armillotta, M., Fokianos, K., and Krikidis, I. (2022). *Generalized Linear Models Network Autoregression*, pages 112–125. Springer International Publishing.

Barigozzi, M. and Brownlees, C. (2019). Nets: Network estimation for time series. *Journal of Applied Econometrics*, 34(3):347–364.

Basu, S., Shojaie, A., and Michailidis, G. (2015). Network Granger Causality with Inherent Grouping Structure. *Journal of Machine Learning Research*, 16(1):417–453.

Betancourt, M. and Girolami, M. (2015). *Hamiltonian Monte Carlo for Hierarchical Models*. Chapman and Hall/CRC.

Bolstad, A., Van Veen, B., and Nowak, R. (2011). Causal Network Inference Via Group Sparse Regularization. *IEEE Transactions on Signal Processing*, 59(6):2628–2641.

Box, G., Jenkins, G., Reinsel, G., and Ljung, G. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Castro, M., Paleti, R., and Bhat, C. (2012). A latent variable representation of count data models to accommodate spatial and temporal dependence: Application to predicting crash frequency at intersections. *Transportation Research Part B: Methodological*, 46(1):253–272.

Chen, E., Fan, J., and Zhu, X. (2023). Community network auto-regression for high-dimensional time series. *Journal of Econometrics*, 235(2):1239–1256.

Davis, R., Fokianos, K., Holan, S., Joe, H., Livsey, J., Lund, R., Pipiras, V., and Ravishanker, N. (2021). Count time series: A methodological review. *Journal of the American Statistical Association*, 116(535):1533–1547.

Doukhan, P., Fokianos, K., Støve, B., and Tjøstheim (2020). Multivariate count autoregression. *Bernoulli*, 26(1):471 – 499.

Duane, S., Kennedy, A. D., Pendleton, B., and Roweth, D. (1987). Hybrid Monte Carlo. *Physics Letters B*, 195(2):216–222.

Durante, D. and Dunson, D. (2016). Locally adaptive dynamic networks. *Annals of Applied Statistics*, 10(4):2203–2232.

Fokianos, K. (2024). Multivariate count time series modelling. *Econometrics and Statistics*, 31:100–116.

Friel, N., Rastelli, R., Wyse, J., and Raftery, A. (2016). Interlocking directorates in irish companies using a latent space model for bipartite networks. *Proceedings of the National Academy of Sciences of the United States of America*, 113(24):6629–6634.

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., and Rubin, D. (2013). *Bayesian Data Analysis*. Chapman and Hall/CRC.

Gower, J. and Dijksterhuis, G. (2004). *Procrustes Problems*. Oxford University Press.

Guo, X. and Zhu, F. (2024). Softplus negative binomial network autoregression. *Stat*, 13(1):e638.

Hoff, P., Raftery, A., and Handcock, M. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97(460):1090–1098.

Kang, X., Ganguly, A., and Kolaczyk, E. (2022). Dynamic networks with multi-scale temporal structure. *Sankhya A: The Indian Journal of Statistics*, 84(1):218—260.

Kaur, H. and Rastelli, R. (2024). A latent space model for multivariate count data time series analysis. *arXiv: 2408.13162*.

Knight, M., Leeming, K., Nason, G., and Nunes, M. (2020). Generalized Network Autoregressive Processes and the GNAR Package. *Journal of Statistical Software*, 96(5):1–36.

Knight, M., Nunes, M., and Nason, G. (2016). Modelling, detrending and decorrelation of network time series. *arXiv: 1603.03221*.

Krampe, J. (2019). Time series modeling on dynamic networks. *Electronic Journal of Statistics*, 13(2):4945 – 4976.

Liu, H. and Nason, G. (2023). New methods for network count time series. *arXiv: 2312.01944*.

Ludkin, M., Eckley, I., and Neal, P. (2018). Dynamic stochastic block models: parameter estimation and detection of changes in community structure. *Statistics and Computing*, 28(6):1201 – 1213.

Luetkepohl, H. (2005). *The New Introduction to Multiple Time Series Analysis*. Springer Berlin, Heidelberg.

Matias, C. and Miele, V. (2017). Statistical clustering of temporal networks through a dynamic stochastic block model. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 79(4):1119–1141.

Nason, G., Salnikov, D., and Borja, M. (2024). Modelling clusters in network time series with an application to presidential elections in the usa. *arXiv: 2401.09381*.

Neal, R. (2012). MCMC using Hamiltonian Dynamics. *Handbook of Markov Chain Monte Carlo*.

Nocedal, J. and Wright, S. (2006). *Numerical Optimization*. Springer, New York, NY, USA, 2e edition.

Pensky, M. (2019). Dynamic network models and graphon estimation. *The Annals of Statistics*, 47(4):2378 – 2403.

Rastelli, R. and Corneli, M. (2023). Continuous latent position models for instantaneous interactions. *Network Science*, 11(4):560–588.

Sarkar, P. and Moore, A. (2005). Dynamic social network analysis using latent space models. In *Proceedings of the 18th International Conference on Neural Information Processing Systems*, NIPS'05, pages 1145–1152. MIT Press.

Sewell, D. and Chen, Y. (2015). Latent space models for dynamic networks. *Journal of the American Statistical Association*, 110(512):1646–1657.

Sewell, D. and Chen, Y. (2016). Latent space models for dynamic networks with weighted edges. *Social Networks*, 44:105–116.

Stan Development Team (2024). RStan: the R interface to Stan. R package version 2.32.6.

Tjøstheim, D., Jullum, M., and Løland, A. (2023). Some recent trends in embeddings of time series and dynamic networks. *Journal of Time Series Analysis*, 44:686–709.

Yin, H., Safikhani, A., and Michailidis, G. (2024). A functional coefficients network autoregressive model. *arXiv: 2402.07373*.

Zhu, X., Pan, R., Li, G., Liu, Y., and Wang, H. (2017). Network vector autoregression. *The Annals of Statistics*, 45(3):1096 – 1123.

Chapter 5

Conclusion

5.1 Summary

Network-integrated time series models are still in their early stages of development, yet they hold significant potential as the collection and analysis of time series data continues to expand across various fields. The work outlined in this thesis represents a significant addition to the available literature on network-integrated time series models. This work specifically contributes toward bridging the gap in the literature of count time series with network structures, where network structure is completely unknown. We introduced a novel statistical framework that combines the widely-used log-linear time series model for count data with the latent position model, a prominent network modeling approach. The core objective of this combination is to uncover the underlying patterns in the connections between the time series and provide a model-based visualization of these relationships for clearer interpretation.

Chapter 2 delves into the core theory of the latent position model, which has gained prominence for offering a geometric representation of networks that is both intuitive and insightful. We explore two key variants: the latent distance model and the latent projection model, both of which enable visualization of relationships between nodes based on their positions in an underlying latent space. This approach is essential for revealing hidden structures within network data and serves as a foundation for the subsequent work discussed in Chapter 3 and Chapter 4. In chapter 2, we examine the two main inferential procedures for these models, addressing issues of parameter identifiability in a Bayesian context and employing Procrustes matching to mitigate these challenges. Additionally, we outline key extensions of the model, showcasing its versatility and broad applicability. Using an example of a coauthorship network, we demonstrate how the latent position model and its extensions effectively capture significant features of real-world networks.

In Chapter 3, we leverage the strengths of the latent position model to study multivariate count time series data. By exploiting the model's ability to visualize relationships between nodes based on their positions, we gain deeper insights into the connections within the multivariate time series. We introduce the TSLPM by integrating LPM with a Poisson log-linear multivariate autoregressive model, enabling us to uncover the dependencies between time series by representing them as nodes in a latent network. Inference in the model is performed using a Bayesian approach with Hamiltonian Monte Carlo (HMC), while addressing identifiability issues inherent in latent position models. To resolve these, we employ Procrustes analysis, which aligns latent space configurations by removing arbitrary rotations and reflections. We apply the TSLPM to a burglary dataset from Chicago, uncovering geographical patterns of crime in the city. The model effectively captures how burglary incidents in one area influence or are influenced by neighboring regions, revealing clusters of high interdependence.

Chapter 4 extends work in Chapter 3 on static network count time series by introducing a model to uncover temporal interconnection patterns in dynamic network structures of count data. Unlike the prior study, which utilized a latent projection model, we now adopt a latent distance model, another LPM model, to explore the evolving relationships between nodes. The latent distance model provides a more nuanced understanding of the strength of connections between time series based on their proximity in latent space, which helps to capture both spatial and temporal dependencies in the count data. For inference, we employ a Bayesian approach, using HMC to estimate the model parameters. One of the challenges encountered is the issue of identifiability, as latent positions are non-identifiable due to invariance to rotation, translation, and reflection. To resolve this, we employ Procrustes analysis to align the latent configurations over time, ensuring that the latent space dynamics can be interpreted meaningfully. We apply the model to a dataset on mumps cases across nine regions of England, uncovering temporal patterns in the spread of the disease. The findings uncovered both geographical and indirect pathways in the spread of the infection, which shifted over time. Notably, the northernmost regions consistently experienced the highest number of cases, while London emerged as the most central node, both influencing and being influenced by other regions in the network. Additionally, the model-based analysis revealed a weakening of inter-regional connections over time, suggesting that the transmission of mumps between regions has diminished as time progressed.

Overall, this dissertation expands the range of modeling approaches for analyzing count time series data with unknown network structures, whether fixed or time-varying. By utilizing latent position models to explore hidden connection patterns in time series data, it offers a structured method for uncovering these relationships through clear, straightforward, and interpretable visualizations. This approach enables the generation of model-based summaries of the time series data, producing results with a high degree of interpretability. This, in turn, provides a distinct advantage over existing models for count time series analysis.

5.2 Future Work

There are several promising avenues for extending the applications, theory, and computational aspects of the TSLPM and DTSLPM models. Both the models uncover weighted but undirected cross-connections between the time series using LPM. A future direction could explore more complex types of network structures underlying the time series system, such as weighted and directed edges, which would better capture both the intensity and directionality of relationships between time series. Additionally, incorporating more flexible forms of non-linearity in the count data could enhance the model's applicability across a wider range of domains.

In this thesis, we primarily focused on relatively small datasets due to the extended computation times involved. Future research could prioritize enhancing its computational scalability. Although Stan offers an accurate solution for model estimation, alternative approaches could be explored to accelerate the inference process, making the methodology more feasible for larger datasets.

A promising direction for extending our work is to incorporate the negative binomial distribution into both proposed models to better address overdispersion in multivariate count time series data. Overdispersion, where the variance exceeds the mean, is a common issue in such data. In these cases, the negative binomial distribution often performs better than the Poisson distribution. In their recent study, Guo and Zhu (2024) introduced an integer-valued network autoregressive model that uses a negative binomial distribution for the conditional marginal distribution and applies the softplus function as the link function. This approach provides greater flexibility for handling overdispersed multivariate count data. A valid research idea could be to extend this framework to cases where the network structure is unknown. Building on the TSLPM, which integrates a Poisson distribution with a log-linear model and the LPM, we suggest combining the LPM with the Softplus Negative Binomial Network AutoRegressive (SNBNAR) model introduced by Guo and Zhu (2024). The stationary conditions established in their study, for both fixed and growing dimensions, ensure the model's robustness and can be empirically validated. As part of this proposed extension, the network structure could also be made dynamic, allowing connections between nodes to change over time. This further enhances the model's applicability to real-world scenarios where network structures are not static.

In terms of extending the methodology, a promising direction would be to develop models for multivariate time series data with continuous responses. These models have broad applications in fields like finance and econometrics. A recent example is the network vector autoregressive (NAR) model proposed by Zhu et al. (2017), which integrates information from known network structures. A natural extension of this model would involve cases where the network structure is unknown, which aligns with one of the core assumptions of the proposed

models. These two model frameworks could be combined to develop a network model for multivariate time series with continuous responses.

Another possible direction for future research would be to develop models for multivariate time series with categorical data. Similar to count time series, categorical time series have received relatively little attention in the literature. Some recent articles have focused on problems of statistical inference and modeling of categorical time series, see Cadez et al. (2003), García-Magariños and Vilar (2015), Fulcher (2018), Li et al. (2021), López-Oriona et al. (2023), and López-Oriona and Vilar (2024) for their contribution. To our knowledge, no existing work has explored the analysis of categorical time series data using latent position models. A potential extension of this work could involve integrating the modeling frameworks from these references with the latent position model, leveraging a network-based approach to uncover latent structures within categorical time series data.

Another promising avenue for future research is the application of our framework to species abundance data, which has significant relevance in ecology and biology. The study of species abundance data through network analysis has gained increasing attention in recent years (see, for instance, Delmas et al. (2019); Chiquet et al. (2021); Faisal et al. (2010); Vázquez et al. (2005)). A potential research direction could involve applying the latent position model to infer species interaction networks, offering new insights into ecological relationships and biodiversity patterns.

Further extension of our work could be explored in the field of biology, particularly in the study of brain activity. The time-varying dynamics of the brain are often analyzed using various network-based approaches. For instance, Dimitriadis et al. (2010)applied well-established measures of neural synchrony to short segments of brain activity time series, enabling a time-dependent characterization of brain connectivity through the investigation of functional segregation and integration. The authors developed a parameter-free method for deriving adjacency matrices from connectivity graphs, simplifying the identification of significant edges in brain networks and enhancing the understanding of their scale-free properties. Similarly, Vidaurre et al. (2017) employed Hidden Markov Models (HMMs) to analyze brain network dynamics, providing a framework that explicitly captures time-dependent information and improves our understanding of network evolution over time. Additionally, Lehnertz, K. et al. (2017) utilized various time series analysis techniques—including univariate, bivariate, and multivariate methods—depending on whether the data originated from a single site or multiple sites. Their approach leveraged multiple realizations of transient dynamics, replacing temporal averages with ensemble averages, and incorporated concepts from the network and graph theory to analyze the statistical and spectral properties of brain networks. A promising direction for extending these studies is the incorporation of latent position models (LPMs) in the analysis of brain activity. This approach would offer a novel perspective on capturing

brain dynamics by modeling the underlying latent structure of neural interactions, potentially providing new insights into the functional organization of the brain.

An intriguing extension of both the proposed TSLPM and DTSLPM models would be the incorporation of autoregressive coefficients from lags greater than one. While we have partially addressed this in TSLPM by incorporating a seasonality effect, a more formal and comprehensive approach could further enhance the model. It would be valuable to explore how network structures can be leveraged for higher-order processes in DTSLPM. This addition would allow the model to capture long-term dependencies and trends, thereby improving forecast accuracy and reliability. Additionally, integrating node and edge covariates, commonly available in real network applications, could significantly enhance the model's predictive performance. While we introduced node covariates in the TSLPM model, they were not included in DTSLPM. Furthermore, investigating the potential improvement in model accuracy through the inclusion of both node and edge covariates could offer significant insights.

Another avenue to explore would be the inclusion of community structure information in the network based time series model. Extensive literature exists on the identification of communities within network nodes. Notably, Bennett et al. (2022) proposes a method for the detection of lead-lag clusters of time series in multivariate systems. This approach reveals that the network of pairwise lead-lag relationships between time series with continuous responses can be aptly viewed as a directed network. Nason et al. (2024) also work in a similar direction, presenting a model that allows for the analysis of dynamic interactions across communities in time series data with embedded network structure. Surprisingly, the realm of community detection has received limited attention when it comes to counting or integer-valued time series. This gap underscores the imperative need to develop dedicated models and methodologies in this specialized domain.

Another future direction for both the proposed models involves addressing the important question of determining the number of dimensions in the latent space to visualize the network of connections. In our work, we did not explore this issue in depth, instead opting for the commonly used two-dimensional latent space, as is standard in many studies. However, this is an arbitrary choice and may lead to either an incomplete or overly complex representation of the network. Identifying the optimal number of latent dimensions tailored to specific applications could result in a more refined and effective modeling framework. Future research focused on selecting the ideal number of dimensions could significantly enhance the flexibility and accuracy of these models.

References

Bennett, S., Cucuringu, M., and Reinert, G. (2022). Lead–lag detection and network clustering for multivariate time series with an application to the us equity market. *Machine Learning*, 111:4497 – 4538.

Cadez, I., Heckerman, D., Meek, C., Smyth, P., and White, S. (2003). Model-based clustering and visualization of navigation patterns on a web site. *Data mining and knowledge discovery*, 7:399–424.

Chiquet, J., Mariadassou, M., and Robin, S. (2021). The poisson-lognormal model as a versatile framework for the joint analysis of species abundances. *Frontiers in Ecology and Evolution*, 9:588292.

Delmas, E., Besson, M., Brice, M.-H., Burkle, L., Dalla Riva, G., Fortin, M.-J., Gravel, D., Guimarães Jr., P., Hembry, D., Newman, E., Olesen, J., Pires, M., Yeakel, J., and Poisot, T. (2019). Analysing ecological networks of species interactions. *Biological Reviews*, 94(1):16–36.

Dimitriadis, S., Laskaris, N., Tsirka, V., Vourkas, M., Micheloyannis, S., and Fotopoulos, S. (2010). Tracking brain dynamics via time-dependent network analysis. *Journal of neuroscience methods*, 193(1):145–155.

Faisal, A., Dondelinger, F., Husmeier, D., and Beale, C. (2010). Inferring species interaction networks from species abundance data: A comparative evaluation of various statistical and machine learning methods. *Ecological Informatics*, 5(6):451–464.

Fulcher, B. (2018). Feature-based time-series analysis. In *Feature engineering for machine learning and data analytics*, pages 87–116. CRC press.

García-Magariños, M. and Vilar, J. (2015). A framework for dissimilarity-based partitioning clustering of categorical time series. *Data mining and knowledge discovery*, 29(2):466–502.

Guo, X. and Zhu, F. (2024). Softplus negative binomial network autoregression. *Stat*, 13(1):e638.

Lehnertz, K., Geier, C., Rings, T., and Stahn, K. (2017). Capturing time-varying brain dynamics. *EPJ Nonlinear Biomedical Physics*, 5:2.

Li, Z., Bruce, S., and Cai, T. (2021). Classification of categorical time series using the spectral envelope and optimal scalings. *arXiv: 2102.02794*.

López-Oriona, Á. and Vilar, J. (2024). Analyzing categorical time series with the r package ctsfeatures. *Journal of Computational Science*, 76:102233.

López-Oriona, Á., Vilar, J., and D'Urso, P. (2023). Hard and soft clustering of categorical time series based on two novel distances with an application to biological sequences. *Information Sciences*, 624:467–492.

Nason, G., Salnikov, D., and Borja, M. (2024). Modelling clusters in network time series with an application to presidential elections in the usa. *arXiv: 2401.09381*.

Vidaurre, D., Smith, S., and Woolrich, M. (2017). Brain network dynamics are hierarchically organized in time. *Proceedings of the National Academy of Sciences*, 114(48):12827–12832.

Vázquez, D., Poulin, R., Krasnov, B., and Shenbrot, G. (2005). Species abundance and the distribution of specialization in host-parasite interaction networks. *Journal of Animal Ecology*, 74(5):946–955.

Zhu, X., Pan, R., Li, G., Liu, Y., and Wang, H. (2017). Network vector autoregression. *The Annals of Statistics*, 45(3):1096 – 1123.