




Bayesian Markov-Switching Tensor Regression for Time-Varying Networks

Monica Billio, Roberto Casarin & Matteo Iacopini

To cite this article: Monica Billio, Roberto Casarin & Matteo Iacopini (2022): Bayesian Markov-Switching Tensor Regression for Time-Varying Networks, Journal of the American Statistical Association, DOI: [10.1080/01621459.2022.2102502](https://doi.org/10.1080/01621459.2022.2102502)


To link to this article: <https://doi.org/10.1080/01621459.2022.2102502>

 [View supplementary material](#) 

 Accepted author version posted online: 25 Jul 2022.

 [Submit your article to this journal](#) 

 Article views: 76

 [View related articles](#) 

 [View Crossmark data](#) 

Bayesian Markov-Switching Tensor Regression for Time-Varying Networks

Monica Billio

Department of Economics, Ca' Foscari University of Venice

and

Roberto Casarin

Department of Economics, Ca' Foscari University of Venice

and

Matteo Iacopini

Department of Econometrics and Data Science, Vrije Universiteit Amsterdam

Abstract

Modeling time series of multilayer network data is challenging due to the peculiar characteristics of real-world networks, such as sparsity and abrupt structural changes. Moreover, the impact of external factors on the network edges is highly heterogeneous due to edge- and time-specific effects. Capturing all these features results in a very high-dimensional inference problem. A novel tensor-on-tensor regression model is proposed, which integrates zero-inflated logistic regression to deal with the sparsity, and Markov-switching coefficients to account for structural changes. A tensor representation and decomposition of the regression coefficients are used to tackle the high-dimensionality and account for the heterogeneous impact of the covariate tensor across the response variables. The inference is performed following a Bayesian approach, and an efficient Gibbs sampler is developed for posterior approximation. Our methodology applied to financial and email networks detects different connectivity regimes and uncovers the role of covariates in the edge-formation process, which are relevant in risk

and resource management. Code is available on GitHub. Supplementary materials for this article are available online.

Keywords: multidimensional data; sparsity; nonlinear time series; zero-inflated logit

1 Introduction

This paper provides a new, flexible model for time series of multidimensional binary arrays. Our model accounts for relevant data features emerging in many real-world applications, such as high dimension, sparsity, structural changes, and entry-specific covariates. The motivation for this paper relies on relevant stylized facts about dynamic multilayer networks, where the edge activation over time can be represented as a series of binary arrays (Holme and Saramäki, 2012; Boccaletti et al., 2014). Specifically, the investigation of predictors' impact and change detection are of paramount importance in the analysis of financial (Billio et al., 2012; Diebold and Yilmaz, 2014) and communication networks (Fox et al., 2016).

We illustrate some features of dynamic networks through the preliminary analysis of the datasets studied in Section 4. Figure 1 shows the snapshots, at different points in time, of a network among 61 European financial institutions (top) and an email network among 90 members of a large European research institution (bottom). The network structure, that is, the arrangement of nodes and edges (Diestel, 2012, ch. 8) considerably changed over time. The differences in terms of degree, path length, clustering coefficient, and number of hubs indicate deviations from standard random graph models (Diestel, 2012, ch. 11). More realistic models include edge-specific covariates, which make the inference and modeling tasks more challenging. Besides, the network structure changes can be sudden and dramatic, and the structures tend to cluster over time, resulting in the persistence of the node degree (compare graphs across columns and see Sections S.6–S.7 in the Supplement).

Existing models for dynamic networks do not account for heterogeneous effects of covariates and structural changes ([Smith et al., 2019](#); [Kim et al., 2018](#); [Krivitsky and Handcock, 2014](#); [Hanneke et al., 2010](#); [Xing et al., 2010](#); [Sarkar and Moore, 2006](#)), but they rely on simplifying assumptions, such as parameter pooling ([Durante and Dunson, 2014a](#)) and linearity ([Zhang et al., 2019](#); [Guhaniyogi et al., 2017](#); [Hoff, 2015](#); [Zhou et al., 2013](#); [Hoff, 2011](#)). This paper develops a new flexible tensor-on-tensor model to address these issues; the novelties of our contribution are discussed in the following.

One step toward more flexible network models was taken by [Durante and Dunson \(2014a\)](#), who proposed a logit specification with exogenous variables for the edges that assume the same coefficients over the network. For some real-world networks, this assumption is too restrictive. In financial networks, institution-specific variables, such as size and sector, and common risk factors, such as market volatility, are likely to have a heterogeneous impact on connections among institutions. Similarly, measuring the persistence of connectivity patterns in communication networks is crucial for understanding information spread. Users with distinct features, such as group membership, may have different persistence levels in their emailing activity. We aim to address these issues by including in the proposed model a covariate tensor to account for edge- and node-specific predictors and a coefficient tensor to allow for edge-specific effects. Similar patterns and dependence structures in the data are exploited to reduce the scale of the problem. Specifically, we introduce a tensor representation of the coefficients and propose a low-rank decomposition to achieve a parsimonious parametrization allowing for the partial pooling of the parameters.

Tensor algebra generalizes matrix algebra to multidimensional objects (for a review, see [Kolda and Bader, 2009](#); [Cichocki et al., 2016](#)) and permits the preservation of the intrinsic structure of the data, thus preventing data reshaping and manipulation. This is particularly important for dealing with the complex

structures of novel data, such as multilayer networks, three-way tables, and spatial panels with multiple series observed for each unit (e.g., municipalities, regions, and countries). Figure 2 provides a representation of a multilayer network as a three-dimensional binary tensor \mathcal{X} . Each frontal slice $\mathcal{X}_{::k}$ of the tensor describes the connectivity structure of the k th layer of the network. Another advantage of using tensors stems from their decompositions and approximations, which provide valuable representations in lower dimensional spaces (Hackbusch, 2012, ch.7-8). We exploit the parallel factor (PARAFAC) decomposition of the coefficient tensor for reducing the number of parameters to estimate, thus making inference on the models feasible.

Linear models for a real tensor-valued response have been proposed by Hoff (2011, 2015), who studied time series of relational data, and by Gerard and Hoff (2017), who introduced a regularized estimator for signal extraction from noisy tensor data. Linear tensor regression models have been used in neuroimaging for predicting scalar indicators with covariate tensor (e.g., Zhou et al., 2013; Zhang et al., 2019; Guhaniyogi et al., 2017). Li and Zhang (2017) defined a tensor-response linear regression on a covariate vector to study the relationship between brain activity and individual control variables and used an envelope method for the estimation. To the best of our knowledge, not much work has been done so far about developing nonlinear tensor-on-tensor models. We contribute to filling this gap by proposing a logit tensor regression with tensor-valued covariates that can find a direct application to multilayer network data. In addition, the model allows for dynamic coefficients to capture structural changes in the data. A hierarchical Bayesian approach based on prior regularization is adopted to shrink irrelevant coefficients. This has the advantage of not imposing restrictions on the tensor response and easing the interpretation compared to the envelope method.

Many real-world networks, such as those in Fig. 1, exhibit sparsity and time variation in the sparsity level. In finance, sparse networks characterize tranquil

states of the market and low levels of contagion ([Billio et al., 2012](#)), whereas, in behavioral and computer science, they correspond to periods of sporadic email activity and low probability of email cascades. The sparsity patterns result in a large proportion of zeros in the binary tensor representation of the network that the covariates cannot explain and could potentially cause the bias in coefficient estimation in standard logit models ([Harris and Zhao, 2007](#)). Motivated by these stylized facts, we propose a zero-inflated logit regression for the edge activation.

A visual inspection of the datasets in Fig. 1 reveals substantial time variation of the network structure, which has also been documented in other network data (e.g., [Billio et al., 2012](#); [Diebold and Yilmaz, 2014](#); [Bianchi et al., 2019](#)). The distinctive features of our data are sparsity clustering over time and abrupt structural changes. Discontinuities in time series are usually associated with dramatic events, such as financial crises or the initial propagation of viral emails, and call for Markov-switching processes. Our novel zero-inflated logit tensor regression with Markov-switching coefficients (ZIL-T-MS) enables us to study the impact of covariates in the different connectivity regimes.

To summarize, our logistic tensor model captures relevant features of the edge formation process in real-world networks. The results in the financial application are of great relevance for policymakers and private institutions interested in the assessment and monitoring of systemic risk. Moreover, detecting regimes in email and other communication networks is a valuable tool for their maintenance and optimization and helps the early detection of a cascade of messages between individuals in rapid succession (e.g., viral communications).

This article is organized as follows. Section 2 presents the model. Section 3 discusses the Bayesian inference procedure. Section 4 provides an application for financial network data. The Discussion summarizes our results and suggests possible future directions for this research. Further details and results are

provided in the supplementary material. Code is included in a MATLAB package available at <https://github.com/matteoiacopini/ZIL-T-MS>

2 A Markov-Switching Model for Networks

We propose a new model for edge activation probability in multilayer temporal networks, which accounts for the stylized facts illustrated in Section 1. First, we discuss the main model components for each edge; second, we provide a compact representation using tensors, which allows for partial pooling and a parsimonious parametrization of the coefficients.

2.1 A Multilayer Network Model

Let $G_t = (V_1, V_2, M, E_t)$ be a multilayer temporal network, where $V_1 = \{1, \dots, I\}$, $V_2 = \{1, \dots, J\}$ are two vertex sets, $M = \{1, \dots, K\}$ is the set of layers, and $E_t \subset (V_1 \times V_2 \times M)$ is the edge set at time $t = 1, \dots, T$. The network connectivity from node i to j on layer k at time t can be represented with the following sequence of binary variables

$$x_{ijk,t} = \begin{cases} 1 & \text{if } \{i, j, k\} \in E_t \\ 0 & \text{if } \{i, j, k\} \notin E_t. \end{cases} \quad (1)$$

This general definition includes undirected, directed, and undirected bipartite networks. It can be further extended to account for other types of networks ([Kivelä et al., 2014](#)).

The first important feature of real networks, such as those in Fig. 1, is sparsity. In random graph theory, sparsity is an asymptotic property of networks such that the number of edges grows subquadratically with the number of nodes ([Diestel, 2012, ch.7](#)). In finite graphs, sparsity occurs when there is an excess of zeros in the connectivity tensor, that is, when the degree distribution has a peak at 0. We account for this feature by assuming that the probability of an edge from node i to j on layer k at time t is a mixture of a Dirac mass at 0 and a Bernoulli distribution (ZIL specification):

$$x_{ijk,t} | \rho(t), \mathbf{g}_{ijk}(t) \sim \rho(t) \delta_{\{0\}}(x_{ijk,t}) + (1 - \rho(t)) \text{Bern}(x_{ijk,t} | \psi_{ijk}(t)), \quad (2)$$

where $\rho(t)$ and $\psi_{ijk}(t)$ are the time-varying mixing and Bernoulli probabilities, respectively. Regarding the proportion of edges whose formation can be explained by covariates, we specify a logistic link for the probabilities

$$\psi_{ijk}(t) = \frac{\exp(\mathbf{z}_{ijk,t}' \mathbf{g}_{ijk}(t))}{1 + \exp(\mathbf{z}_{ijk,t}' \mathbf{g}_{ijk}(t))} \quad (3)$$

where $\mathbf{z}_{ijk,t} \in \mathbb{R}^Q$ are edge-specific covariates and $\mathbf{g}_{ijk}(t) \in \mathbb{R}^Q$ are time-varying parameters. Motivated by the stylized facts that node degrees are persistent and cluster over time, we assume that the mixing probability and the coefficients are driven by an L -state hidden Markov chain $\{s_t\}_{t=1}^T$, that is $\rho(t) = \rho_{s_t}$ and $\mathbf{g}_{ijk}(t) = \mathbf{g}_{ijk,s_t}$, where ρ_l and $\mathbf{g}_{ijk,l}$ are state-specific parameters (MS specification). The transition matrix of the chain is assumed time-invariant and denoted by $\Xi = (\xi_1', \dots, \xi_L')$, where $\xi_l = (\xi_{l,1}, \dots, \xi_{l,L})$ is a probability vector and $\xi_{i,j} = p(s_t = j | s_{t-1} = i)$ is the transition probability from state i to state j .

2.2 Parsimonious Parametrization Based on Tensors

Covariates (risk factors in the application) are likely to have similar impacts on groups of edges (financial linkages), which calls for modeling them jointly. To this aim, we represent the binary multilayer networks and regression coefficients as tensors and propose a new tensor-on-tensor regression model. A low-rank decomposition of the coefficient tensor enables us to obtain a parsimonious parametrization, which induces partial pooling while preserving the heterogeneity in the impact of the covariates.

Let us define a real-valued D -order tensor $\mathcal{X} \in \mathbb{R}^{d_1 \times \dots \times d_D}$ as a D -dimensional array of size $(d_1 \times \dots \times d_D)$. A tensor can be thought of as the multidimensional extension of a matrix (i.e., a 2-order tensor), where each dimension is called mode. See the Supplement for some background material on tensors and [Hackbusch \(2012\)](#) for an introduction to tensor spaces. Following [Boccaletti](#)

et al. (2014) and Kivelä et al. (2014), we represent a multilayer temporal network with a 4-order tensor \mathcal{X} of size $(I \times J \times K \times T)$, and binary entries $x_{ijk,t}$. The following remarks provide two tensor representations of the ZIL-T-MS model in eq. (2)-(3) and show that it extends to tensors the latent and the switching regression models in McFadden (1974) and Frühwirth-Schnatter (2006), respectively (see Appendix A for proofs). First, let us introduce the mode- n product (Kolda and Bader, 2009; Hackbusch, 2012) between a D -order tensor $\mathcal{X} \in \mathbb{R}^{d_1 \times \dots \times d_D}$ and a vector $\mathbf{v} \in \mathbb{R}^{d_n}$, which is the $(D-1)$ -order tensor $\mathcal{Y} \in \mathbb{R}^{d_1 \times \dots \times d_{n-1} \times d_{n+1} \times \dots \times d_D}$ with entries

$$y_{i_1, \dots, i_{n-1}, i_{n+1}, \dots, i_D} = (\mathcal{X} \times_n \mathbf{v})_{i_1, \dots, i_{n-1}, i_{n+1}, \dots, i_D} = \sum_{i_n=1}^{d_n} x_{i_1, \dots, i_n, \dots, i_D} v_{i_n}. \quad (4)$$

Remark 2.1. Let $\mathcal{G}(t) \in \mathbb{R}^{I \times J \times K \times IJKQ}$ be a coefficient tensor and $\mathcal{Z}_t \in \mathbb{R}^{I \times J \times K \times Q}$ a covariate tensor. Define the map $\Psi: \mathbb{R}^{i_1 \times \dots \times i_4} \rightarrow \{0, 1\}^{i_1 \times \dots \times i_4}$ such that $\mathcal{X}^* \mapsto \mathcal{D}$ with $d_{i_1, \dots, i_4} = 1_{\mathbb{R}_+}(x_{i_1, \dots, i_4}^*)$, where $1_A(x)$ is the indicator function. Equations (2)-(3) are obtained as a special case of:

$$\begin{cases} \mathcal{X}_t = \mathcal{B}(t) \odot \Psi(\mathcal{X}_t^*) & b_{ijk}(t) \stackrel{iid}{\sim} \mathcal{Bern}(1 - \rho(t)) \\ \mathcal{X}_t^* = \mathcal{G}(t) \times_4 \text{vec}(\mathcal{Z}_t) + \mathcal{E}_t & \varepsilon_{ijk,t} \stackrel{iid}{\sim} \text{Logistic}(0, 1), \end{cases} \quad (5)$$

where \odot is the Hadamard product, $\mathcal{B}(t) \in \{0, 1\}^{I \times J \times K}$ and $\mathcal{E}_t \in \mathbb{R}^{I \times J \times K}$ are tensors with entries $b_{ijk}(t)$ and $\varepsilon_{ijk,t}$, respectively. Here, all the covariates in \mathcal{Z}_t have an impact on all entries of \mathcal{X}_t^* and the coefficients are entry-specific. Assuming the covariates $\mathbf{z}_{ijk,t} = (z_{ijk1,t}, \dots, z_{ijkQ,t})'$ have an impact only on the corresponding entry $x_{ijk,t}^*$ yields $\mathcal{X}_t^* = (\mathcal{G}(t) \odot \mathcal{Z}_t) \times_4 \mathbf{1}_Q + \mathcal{E}_t$, where $\mathbf{1}_Q$ is Q -dimensional vector of ones, and $\mathcal{G}(t) \in \mathbb{R}^{I \times J \times K \times Q}$ is the tensor with entries $g_{ijkq}(t) = \tilde{g}_{ijk}^\ell(t)$, where $\ell \neq 1 + (i-1) + (j-1)I + (k-1)IJ + (q-1)IJK$.

Remark 2.2. Assume $\mathbf{z}_{ijk,t} = \mathbf{z}_t$ for each i, j, k . Let us introduce a dummy coding for s_t through L binary variables $\zeta_{t,l} = 1_{\{l\}}(s_t)$, $l = 1, \dots, L$. Then model in Remark 2.1 can be written as:

$$\begin{cases} \mathcal{X}_t = \mathcal{B}(t) \odot \Psi(\mathcal{X}_t^*) & b_{ijk}(t) \stackrel{iid}{\sim} \text{Bern}(1 - \rho(t)) \\ \mathcal{X}_t^* = \mathcal{G} \times_4 (\zeta_t \otimes \mathbf{z}_t) + \mathcal{E}_t = \mathcal{G} \times_4 (\zeta_t', \zeta_t' \otimes \tilde{\mathbf{z}}_t')' + \mathcal{E}_t & \mathcal{E}_{ijk,t} \stackrel{iid}{\sim} \text{Logistic}(0,1) \\ \zeta_{t+1} = \Xi \zeta_t + \mathbf{u}_t & \mathbb{E}[\mathbf{u}_t | \mathbf{u}_1, \dots, \mathbf{u}_{t-1}] = 0 \end{cases} \quad (6)$$

which is a switching SUR (Bianchi et al., 2019), where \otimes denotes the Kronecker product, $\{\mathbf{u}_t\}_t$ is a martingale difference process, $\mathbf{z}_t = (1, \tilde{\mathbf{z}}_t)'$, and $\zeta_t = (\zeta_{t,1}, \dots, \zeta_{t,L})'$.

The representations in (5)-(6) allow us to propose a parsimonious parametrization of the regression coefficients based on a rank- R PARAFAC decomposition of the tensor $\mathcal{G}(t) = \mathcal{G}_{s_t}$:

$$\mathcal{G}(t) = \sum_{r=1}^R \gamma_1^{(r)}(t) \circ \gamma_2^{(r)}(t) \circ \gamma_3^{(r)}(t) \circ \gamma_4^{(r)}(t), \quad (7)$$

where the vectors $\gamma_h^{(r)}(t) = \gamma_{h,s_t}^{(r)}$, $h = 1, \dots, 4$, $r = 1, \dots, R$, are called the marginals of the PARAFAC decomposition and have length I , J , K and Q , respectively. Decomposing the coefficient tensor is more effective in achieving model parsimony compared to using a PARAFAC decomposition of the covariates. Our specification reduces the number of parameters from $IJKQ$ to $R(I + J + K + Q)$, for each state s_t . Sparsity in the coefficient tensor can be favored through a suitable prior distribution for the PARAFAC marginals.

In summary, the proposed ZIL-T-MS tensor regression provides a joint model for network edges with the following features: (i) a ZIL specification, to account for sparsity and covariates; (ii) MS coefficients, to deal with structural changes; (iii) a parsimonious parametrization, to address the curse of dimensionality.

3 Bayesian Inference

In this section, we discuss the prior assumptions, the data augmentation strategy, and a Markov chain Monte Carlo (MCMC) procedure to approximate the posterior distribution.

3.1 Prior Specification

We assume a global-local shrinkage prior for on $\gamma_{h,l}^{(r)}$

$$p(\gamma_{h,l}^{(r)} | \tau, \phi_r, w_{h,r,l}) \sim \mathcal{N}_{n_h}(\bar{\zeta}_{h,l}^{-r}, \tau \phi_r w_{h,r,l} \mathbf{I}_{n_h}) \quad (8)$$

for $r = 1, \dots, R$, each $h = 1, \dots, 4$, and each $l = 1, \dots, L$, where $n_1 = I$, $n_2 = J$, $n_3 = K$, $n_4 = Q$. The parameter τ represents the global component of the variance, common to all marginals, ϕ_r is the level component and $w_{h,r}$ is the local component. The choice of a global-local shrinkage prior, as opposed to a spike-and-slab distribution, is motivated by the reduced computational complexity and the capacity to handle high-dimensional settings. In what follows we denote with $p(\mathcal{G} | \mathcal{W}, \phi, \tau)$ the joint prior of the $\gamma_{h,l}^{(r)}$, where $\mathcal{W} = \{w_{h,r,l}\}_{h,r,l}$. We assume the following hyperpriors for the variance components¹:

$$p(\tau) \sim \mathcal{G}_a(\bar{a}^\tau, \bar{b}^\tau), \quad p(\phi) \sim \mathcal{Dir}(\bar{\alpha}), \quad p(w_{h,r,l} | \lambda_l) \sim \mathcal{Exp}\left(\frac{\lambda_l^2}{2}\right), \quad p(\lambda_l) \sim \mathcal{G}_a(\bar{a}_l^\lambda, \bar{b}_l^\lambda), \quad (9)$$

$\forall h, r, l$, where $\bar{a}^\tau = \bar{\alpha} R$, $\bar{\alpha} = \bar{\alpha} \mathbf{1}_R$, and $\mathbf{1}_n$ is the n -dimensional vector of ones. The further level of hierarchy for the local components $w_{h,r,l}$ is added with the aim of favoring information sharing across local components of the variance (indices h and r) within a given regime l . The specification of an exponential distribution for the local component of the variance of the $\gamma_{h,l}^{(r)}$ yields a Laplace (or Double

Exponential) distribution for each component of the vectors once the $w_{h,r,i}$ is integrated out, that is $\gamma_{h,i,i}^{(r)} | \lambda_i, \tau, \phi_r \sim \text{Laplace}(0, \lambda_i / \sqrt{\tau \phi_r})$ for all $i = 1, \dots, n_h$. The marginal distribution of each entry, integrating all remaining random components, is a generalized Pareto distribution, which favors sparsity. The properness of the posterior distribution of \mathcal{G}_i under the Markov-switching zero-inflated logit likelihood follows from prior properness in the following Lemma (proof in the Appendix).

Lemma 3.1 (Properness of the prior). *The marginals $\gamma_{h,i}^{(r)}$ are not separately identifiable, nonetheless the parameter of interest, \mathcal{G}_i , is exactly identified. It has well-defined prior distribution and finite absolute moments.*

For the state-specific mixing probability it is assumed a Beta prior distribution

$$p(\rho_l) \sim \mathcal{B}e(\bar{a}_l^\rho, \bar{b}_l^\rho) \quad \forall l. \quad (10)$$

A well-known identification issue for mixture models is the label switching problem (e.g., Frühwirth-Schnatter, 2006). When the specific application provides meaningful restrictions on the value of some parameters (e.g., from theory or interpretation), they can be used for identifying the regimes. Following this approach, we assume $\rho_1 > \rho_2 > \dots > \rho_L$, meaning that regime 1 represents the sparsest and regime L the densest. Finally, we assume each row ξ_l of the transition matrix Ξ follows a Dirichlet distribution

$$p(\xi_l) \sim \mathcal{D}ir(\bar{\mathbf{c}}_l) \quad \forall l. \quad (11)$$

Our hierarchical prior distribution is represented by the directed acyclic graph in Fig. 3.

3.2 Posterior Approximation

We introduce allocation variables for the mixture in eq. (2) and the Pólya-Gamma augmentation (see Polson et al., 2013; Wang et al., 2017; Holsclaw et al., 2017),

which allows for conjugate full conditional distributions and a better mixing of the MCMC chain. Define $\mathcal{X} = \{\mathcal{X}_t\}_{t=1}^T$, $\mathbf{s} = \{s_t\}_{t=0}^T$ and let θ denote the set of parameters. For each $l = 1, \dots, L$, we define $\mathcal{T}_l = \{t : s_t = l\}$. Define $\mathcal{D} = \{d_{ijk,t}\}_{ijk,t}$, $\Omega = \{\omega_{ijk,t}\}_{ijk,t}$ and $N_{gl}(\mathbf{s}) = \#\{s_{t-1} = g, s_t = l, t = 1, \dots, T\}$, $g, l = 1, \dots, L$, with $\#$ the cardinality of a set. The complete data likelihood of the model in eq. (5) is

$$L(\mathcal{X}, \mathcal{D}, \Omega, \mathbf{s} | \theta) = \left(\prod_{t=1}^T \prod_{i=1}^I \prod_{j=1}^J \prod_{k=1}^K p(\omega_{ijk,t}) \right) \left(\prod_{g=1}^L \prod_{l=1}^L \xi_{g,l}^{N_{gl}(\mathbf{s})} \right) \prod_{l=1}^L \prod_{t \in \mathcal{T}_l} \prod_{i=1}^I \prod_{j=1}^J \prod_{k=1}^K \left(\frac{2\rho_l \delta_{\{0\}}(x_{ijk,t})}{1 - \rho_l} \right)^{d_{ijk,t}} \frac{1 - \rho_l}{2} \exp\left(-\frac{\omega_{ijk,t}}{2} (\mathbf{z}_t' \mathbf{g}_{ijk,l})^2 + \kappa_{ijk,t} (\mathbf{z}_t' \mathbf{g}_{ijk,l})\right). \quad (12)$$

See Appendix A for details. In the following, we define $\mathcal{G} = \{\mathcal{G}_t\}_{t=1}^L$ and $\rho = \{\rho_t\}_{t=1}^L$, and let \mathbf{w}_l and $\mathbf{w}^{(r)}$ be the $(4 \times R)$ and $(4 \times L)$ matrices representing the l and r -th slices of \mathcal{W} , along the third and second mode, respectively. The complete data likelihood and the prior distributions yield a posterior sampling scheme consisting of four blocks (see the Supplement for the Gibbs sampler derivation).

In block (I) the sampler draws the latent variables from the full conditional distribution:

$$p(\mathbf{s}, \mathcal{D}, \Omega | \mathcal{X}, \mathcal{G}, \Xi, \rho) = p(\mathbf{s} | \mathcal{X}, \mathcal{G}, \Xi, \rho) p(\mathcal{D} | \mathcal{X}, \mathcal{G}, \rho, \mathbf{s}) p(\Omega | \mathcal{X}, \mathcal{G}, \rho, \mathbf{s}). \quad (13)$$

Samples of \mathbf{s} are drawn via the Forward Filter Backward Sampler (Frühwirth-Schnatter, 2006, ch. 13). The latent variables $\omega_{ijk,t}$ are sampled independently from

$$p(\omega_{ijk,t} | x_{ijk,t}, s_t, \mathcal{G}_{s_t}) \propto PG(1, \mathbf{z}_t' \mathbf{g}_{ijk,s_t}). \quad (14)$$

The latent variables $\omega_{ijk,t}$ are sampled in block for each t . The latent variables $d_{ijk,t}$ are sampled independently from

$$p(d_{ijk,t} = 1 | x_{ijk,t}, s_t, \mathcal{G}_{s_t}, \rho_{s_t}) \propto \rho_{s_t} \delta_{\{0\}}(x_{ijk,t})$$

$$p(d_{ijk,t} = 0 | x_{ijk,t}, s_t, \mathcal{G}_{s_t}, \rho_{s_t}) \propto (1 - \rho_{s_t}) \frac{\exp((\mathbf{z}_t' \mathbf{g}_{ijk,s_t}) x_{ijk,t})}{1 + \exp(\mathbf{z}_t' \mathbf{g}_{ijk,s_t})}. \quad (15)$$

The hyperparameters that control the variance of the PARAFAC marginals are sampled in block (II) from the full conditional distribution

$$p(\tau, \phi, \mathcal{W} | \{\gamma_{h,l}^{(r)}\}_{h,l,r}) = p(\phi | \{\gamma_{h,l}^{(r)}\}_{h,l,r}, \mathcal{W}) p(\tau | \{\gamma_{h,l}^{(r)}\}_{h,l,r}, \mathcal{W}, \phi) p(\mathcal{W} | \{\gamma_{h,l}^{(r)}\}_{h,l,r}, \phi, \tau). \quad (16)$$

We enable better mixing by blocking together the parameters ϕ . We set $\phi_r = \psi_r / (\psi_1 + \dots + \psi_R)$, where the auxiliary variables ψ_r are sampled independently for each r from

$$p(\psi_r | \{\gamma_{h,l}^{(r)}\}_{h,l}, \mathbf{W}^{(r)}) \propto \text{GiG}(2\bar{b}^{-\tau}, \sum_{h=1}^4 \sum_{l=1}^L \frac{\gamma_{h,l}^{(r)'} \gamma_{h,l}^{(r)}}{w_{h,r,l}}, \bar{\alpha} - n), \quad (17)$$

where $\text{GiG}(a, b, p)$ is Generalized Inverse Gaussian distribution with parameters $p \in \mathbb{R}$, $a > 0$ and $b > 0$, and $n = \sum_{h=1}^4 n_h$. The global variance parameter τ is drawn from

$$p(\tau | \{\gamma_{h,l}^{(r)}\}_{h,l,r}, \mathcal{W}, \phi) \propto \text{GiG}(2\bar{b}^{-\tau}, \sum_{r=1}^R \sum_{h=1}^4 \sum_{l=1}^L \frac{\gamma_{h,l}^{(r)'} \gamma_{h,l}^{(r)}}{\phi_r w_{h,r,l}}, (\bar{\alpha} - n)R). \quad (18)$$

The local variance parameters $w_{h,r,l}$ are independently drawn from

$$p(w_{h,r,l} | \gamma_{h,l}^{(r)}, \phi_r, \tau, \lambda_l) \propto \text{GiG}(\lambda_l^2, \frac{\gamma_{h,l}^{(r)'} \gamma_{h,l}^{(r)}}{\tau \phi_r}, 1 - \frac{n_h}{2}). \quad (19)$$

Finally, the hyperparameters λ_l are independently drawn from

$$p(\lambda_l | \mathbf{W}_l) \propto \lambda_l^{\bar{a}_l + 8R - 1} \exp(-\lambda_l \bar{b}_l - \frac{\lambda_l^2}{2} \sum_{r=1}^R \sum_{h=1}^4 w_{h,r,l}). \quad (20)$$

Block (III) concerns the marginals of the PARAFAC decomposition for the tensors \mathcal{G}_l . The vectors $\gamma_{h,l}^{(r)}$ are sampled independently from

$$p(\gamma_{h,l}^{(r)} | \mathcal{X}, \mathcal{W}, \phi, \tau, \mathbf{s}, \mathcal{D}, \mathbf{\Omega}) \propto \mathcal{N}_{n_h}(\zeta_{h,l}^r, \mathbf{\Lambda}_{h,l}^r). \quad (21)$$

The coefficient tensor in the regime l is obtained as $\mathcal{G}_l = \sum_{r=1}^R \gamma_{1,l}^{(r)} \circ \gamma_{2,l}^{(r)} \circ \gamma_{3,l}^{(r)} \circ \gamma_{4,l}^{(r)}$, and

the coefficients of the edge (i, j) on layer k are given by $\mathbf{g}_{ijk,l} = \sum_{r=1}^R \gamma_{1,l,i}^{(r)} \gamma_{2,l,j}^{(r)} \gamma_{3,l,k}^{(r)} \gamma_{4,l}^{(r)}$.

In block (IV) the mixing and the transition probabilities, ρ_l and ξ_l , are drawn from

$$p(\rho_l | \mathcal{D}, \mathbf{s}) \propto \mathcal{B}e(\tilde{a}_l^\rho, \tilde{b}_l^\rho), \quad p(\xi_l | \mathbf{s}) \propto \mathcal{D}ir(\tilde{\mathbf{c}}). \quad (22)$$

Blocks (I) and (II) are Rao-Blackwellized Gibbs steps: in block (I) we have marginalized over both (\mathcal{D}, Ω) in the full joint conditional distribution of the state \mathbf{s} and \mathcal{D} (together with ρ) in the full conditional of Ω , while in (II) we have integrated out τ from the full conditional of ϕ . The derivation of the full conditional distributions is given in Appendix A. Markov-switching models suffer from an identification issue arising from the invariance of the likelihood function to permutations of the regime labels. A common practice to achieve identification is to impose a constraint. As arbitrarily chosen constraints may perform poorly, they should have a clear interpretation according to the application at hand. Alternative approaches, such as the point process representation of the MCMC draws, can be used in the absence of prior information (Frühwirth-Schnatter, 2001).

The simulation studies in the Supplement illustrate the effectiveness of the inference method in recovering the true values of parameters and latent variables. In our experiments, all standard convergence statistics (e.g., autocorrelation function, numerical standard errors, relative numerical efficiency) suggest that the MCMC chain is mixing well, and a thinning rate of only 2 is needed to have MCMC samples close to i.i.d. draws. Overall, this implies that the Gibbs sampler is computationally efficient, and the Monte Carlo approximation of the posterior mean has low variance.

3.3 Model Comparison and Diagnostics

In the following, we discuss model selection criteria, misspecification diagnostics, and model uncertainty for the ZIL-T-MS model. See Sections S.5.2–S.5.3 in the Supplement for details.

In the model selection exercise, to assess the model fit and the statistical contribution of the tensor representation, we compare it with several competing models. As a baseline, we consider univariate logit and MS logit models, while for the tensor models, we assume PARAFAC and pooled parametrizations, where the latter imposes parameter pooling across edges for each covariate, that is $\mathbf{g}_{ijk,l} = \mathbf{g}_l$, for each i, j, k, l . We simulate the data according to the general specification in eq. (2)-(3) under different forms of heterogeneity in the true coefficient values across covariates and blocks of nodes. In all settings, the DIC_3 of [Celeux et al. \(2006\)](#) indicates that our model is able to exploit partial pooling along one or more directions of the tensor data to improve the estimation efficiency compared to the baseline.

Selecting the PARAFAC rank and making inference on the resulting ZIL-T-MS model on the same dataset leads to a post-selection inference problem. Following [Dukić and Peña \(2005\)](#), we address this issue via Bayesian model averaging (BMA), where the estimator is a weighted average of the estimators obtained for different values of R . We use the DIC to approximate the marginal likelihoods in the BMA weights, which accounts for prior information and penalizes the log-likelihood ([Robert, 2007](#)). In our simulations, as one model's DIC dominates the others, the BMA estimator of \mathcal{G}_l yields similar results as the Bayes estimator of the best model.

To check model fitness and detect potential misspecification, we apply posterior predictive checking (PPC). This approach consists of drawing simulated values from the joint posterior predictive distribution of the data and then comparing some summary statistics of the simulated and observed data ([Gelman et al., 2013](#), ch. 6). We generate synthetic datasets according to the general

specification with different sparsity patterns in each regime and, for each dataset, we computed the average density, $T(\mathbf{X}) = \frac{1}{IJT} \sum_{i,j,t} X_{ij,t}$, as a summary statistic.

Figure 4 reports the approximated posterior distribution of the difference between the test quantity computed on the simulated and the original data. As a distribution far away from zero signals model misspecification, we find that the tensor model has a good performance. In contrast, the baseline model cannot replicate the main features of the data, and its performance deteriorates as the network size increases. Similar results are obtained using different summary statistics and network sizes.

In conclusion, the ZIL-T-MS model performs well in terms of fit and diagnostics compared to a range of alternatives. Our findings suggest that it should be applied when prior beliefs, or preliminary results from univariate logit regressions, suggest that the impact of the covariates is similar across some of the response variables and potentially time-varying.

4 Real Data Applications

This section provides an illustration of the proposed ZIL-T-MS model on the financial and communication networks introduced in Section 1. The Supplement provides further details of the datasets and additional empirical results.

4.1 Financial Network

We apply the proposed methodology to a network among 61 European financial institutions, including 25 banks, 11 insurance companies, and 25 investment companies. The dataset consists of two sequences of 110 binary directed networks sampled monthly from December 2003 to January 2013. Following [Billio et al. \(2012\)](#), we extract each network using Granger causal analysis; therefore, an edge $x_{ijk,t} = 1$ represents a Granger-causal link from institution i to institution j on layer k at time t . The two layers are represented by the return (layer 1) and realized volatility (layer 2) linkages among the institutions.

The set of covariates $\mathbf{z}_{ijk,t}$ consists of a constant term, the realized covariance between the pair of institutions (edge-specific variable), and risk factors usually employed in empirical finance. Specifically, we include the credit spread (CRS), the term spread (TRS), the change of the VSTOXX index (DVX), the log-returns on the STOXX50 index, and the momentum factor. To account for linkage persistence, we add the network total degree. All covariates have been standardized and included with one lag, except DVX, which is contemporaneous, following the standard practice in volatility modeling.

The purpose of the study is to uncover the role of several risk factors as drivers of financial network connectivity across different regimes. This information is encoded by the coefficient tensor, \mathcal{G}_l , since each entry $g_{ijkq,l}$ measures the impact of the covariate $z_{ijkq,t}$ on the probability of the financial linkage $x_{ijk,t} = 1$ during regime l . The coefficients can be associated either to common covariates or to edge-specific covariates.

As is common practice, we impose an identification constraint following an economic interpretation of the regimes. The constraint $\rho_1 > \rho_2$ allows us to label states 1 and 2 as the sparse (low connectedness risk) and dense regime (high connectedness risk), respectively. We choose the tensor rank $R = 5$ and the number of regimes $L = 2$ according to the DIC₃ in [Celeux et al. \(2006\)](#). We estimated the ZIL-T-MS model and use the Gibbs sampler of Section 3 to obtain 5,000 draws from the posterior, after thinning and burn-in.

The estimated regimes (white and gray areas) and the return-layer total degree (black line) are given in the left plot of Fig. 5. The dense regime (gray) is active during periods of financial turmoil, such as the aftermath of the dot com bubble (2003), the global financial crisis (2008), and the European sovereign debt crisis (2010). The local and global connectivity measures (density, average clustering, minimum eigenvector centrality) increase during these periods, indicating that the dense regime is characterized by more complex and denser network structures.

The identification constraint permits recognizing low and high connectedness periods and is strongly supported by the data since the posterior distributions are well separated (middle plot). In the dense regime, the histogram of the $IJQ = 26047$ estimated coefficients (posterior means) across edges and covariates has fatter tails (right plot, light gray). This reveals a more substantial heterogeneity and a more considerable impact of the explanatory variables during periods of financial distress. Moreover, to get new and deeper insights into the impact of risk factors on the linkage formation across and within layers, one can inspect the posterior distribution of each entry of the coefficient tensor in the two regimes, which are the elements of $\mathbf{g}_{ijk,1}$ and $\mathbf{g}_{ijk,2}$. As it is hard to display all the posterior plots, we follow standard practice in global–local shrinkage prior literature and report the posterior mean of the coefficients.

The CRS and TRS are key reference variables to the policymaker, and investigating their impact on returns and volatility is crucial for designing more effective monetary policies aiming at price, financial, and macroeconomic stability. Figure 6 shows their estimated coefficients for each edge and layer in the dense regime, that are the entries of $\mathbf{g}_{ijk,2}$. In the dense regime and the return layer the TRS has a negative impact on all linkages to banks (blue color, BA columns) and a strong positive effect on connecting to insurance and investment companies (red color, IS and IV columns). Conversely, the CRS positively affects the probability of being connected to banks, especially from insurance companies (red color, IS rows and BA columns), and negatively impacts the edge probabilities among investment companies. Besides, higher CRS and TRS increase the probability of volatility linkages (layer 2) from insurance to investment companies and banks (red color, from IS rows to IV and BA columns). Most of our empirical findings cannot be captured with a pooled model, which, in contrast, leads to misleading conclusions about the role of risk factors on linkage formation (see Fig. S.33 in the Supplement).

We investigate the relationships between the estimated coefficients and the node degree (see Fig. 7). In the return layer, the linkages of the most central insurance companies and banks (squares and triangles) are strongly affected by the risk factors in the two regimes, while a weak positive relationship is found on the volatility layer for CRS and TRS. Instead, the pooled model under- or overestimates the relationships between covariates and linkages (dashed lines). In the return layer, increasing the TRS raises the edge probability from banks and the most central insurance company to investment companies. There is evidence of weak effects on the return linkages between banks and the insurance sector and a strong positive impact on the volatility linkages between them (see Fig. S.31 in the Supplement).

4.2 EUcore Email Network

The analysis of communication flows within an organization, such as emails, and its dynamics is of utmost importance for understanding the behavior of individuals and the functioning of the organization (Fox et al., 2016; Om et al., 2020). Representing the email flow among the employees within an organization by a dynamic network enables us to identify individuals' social roles and influence in information spreading and disentangle the factors driving communication intensity. Also, the growing volume of email traffic and its fluctuations pose several challenges to server performance and productivity, thus making the development of accurate models for email workload of primary importance.

The main concern of this study is to examine how email communication changed within two departments of a research institution and characterize better the information spread in the network, as sending an email can trigger a cascade of messages between individuals in rapid succession. Most existing models do not account for cross-sectional correlation and time variation, but they focus on the marginal distribution of network edges. In contrast, our proposed model can reproduce some relevant features of email time series, such as edge

persistence, structural changes, and temporal clustering in the flows. It also enables us to identify the influence of individuals on information spreading within the network.

Email datasets are difficult to find due to the many privacy concerns involved when making such data publicly available. The EUcore email corpus is one of the few public email datasets readily available for research and a reference for social network analysis of email traffic (Leskovec et al., 2007). We analyze the EUcore sender–receiver network among 90 researchers at a European research institution in department 3 (layer 1) and department 4 (layer 2). To investigate network persistence, we include the lagged degree of both layers as covariates. We obtain 5,000 draws from the posterior distribution after selecting rank $R = 5$ and $L = 3$ states based on the DIC_3 . We assume $\rho_1 > \rho_2 > \rho_3$ to identify the states as communication regimes characterized by different levels of email activity (Fox et al., 2016).

The estimated states identify long periods of intense communications (state 3, expected duration 1.66), alternated by shorter periods of regular activity (state 2, 1.24), and sporadic events of low traffic (state 1, 1.03). In Fig. 8, short periods of very sparse networks are in dark color (low clustering coefficient), gray and white shades indicate days of medium and large email flows. In both departments (layers), the email connections during ordinary times react very differently to changes in the covariates compared to the other communication regimes (larger dispersion in the scatterplots of Fig. 9). Reactions are synergistic (positive coefficients) or antagonistic (negative), and the covariates can either reinforce other covariates' effects (upward-sloping scatterplot, as layer 2 in state 2) or generate conflicting pressure on the communication (downward sloping, as layer 1 in state 2).

5 Discussion

Motivated primarily by some stylized facts about real-world dynamic multilayer networks, we introduce a new logit tensor-on-tensor regression with Markov-switching coefficients (ZIL-T-MS) for time series of binary arrays. A zero-inflated specification accounts for the excess of zeros in the data (sparsity), while the coefficient tensor (i) allows for tensor-valued covariates and (ii) captures the entry-specific effects of each covariate. A hidden Markov chain accommodates for structural changes in sparsity patterns and coefficients. We address the high-dimensionality issue by assuming partial pooling based on a PARAFAC decomposition of the coefficient tensor and a hierarchical shrinkage prior distribution. A Gibbs sampler is proposed for posterior approximation and its computational efficiency is tested in several simulation experiments. Model comparison and misspecification diagnostics suggest that our partial pooling framework is effective in coping with high dimensionality and overfitting while retaining flexibility compared to univariate logit and MS-logit benchmarks.

We illustrate the potential of the proposed model and inference with real-world datasets from two relevant fields: finance and social network analysis. The ZIL-T-MS model provided novel insights about the edge formation process, which were outside the scope of the previously existing literature. Notably, we find strong evidence of two financial connectivity regimes and heterogeneous effects of the covariates across linkages, layers, and regimes. Term spread and credit spread play an essential role in explaining the connectivity of central institutions. This result sheds new light on the return and volatility contagion mechanisms in the financial markets. In the email network, we find new evidence of different communication regimes and user behavior. The email connections react very differently across regimes and the reactions are either synergistic or antagonistic in both organizational units. Moreover, the covariates can either reinforce other covariates' effects or generate conflicting pressure on the communication. The findings of our motivating applications would help policymakers devise new strategies to stabilize financial connectedness and systemic risk. By identifying the social role of individuals, the results will also be relevant in providing new

opportunities for developing strategies to favor or counteract information spreading.

We believe the proposed approach is a step forward in the literature on tensor-valued time series and network modeling ([Durante and Dunson, 2014b](#); [Wang et al., 2017](#); [Chen et al., 2018](#)). Our model can be readily applied to a broad spectrum of datasets emerging in other fields, for example, to investigate the mechanisms driving the activation of brain cells or the occurrence of an event in a geographical area, text mining, and health data.

Methodological extensions and topics of our ongoing research include adapting the proposed framework to alternative specifications of the coefficients dynamics. For instance, smooth transition models would allow the parameters to switch between regimes smoothly, rather than with sudden jumps, and the change to be driven by exogenous factors. Another possibility consists in investigating specifications based on mode- n and contracted tensor products to define more parsimonious models for very high-dimensional applications. Bayesian nonparametric inference procedures combined with shrinkage prior can be used to cope with overfitting and account for clustering effects across coefficients. Finally, the proposed methods can be adapted to investigate nonnegative and count tensor-valued data, which are increasingly available in many fields, such as epidemic and climate studies.

Supplementary Materials

Background material on tensors, the derivation of the posterior, simulation experiments, and the description of the data, and further results are given in an online Supplement.

Acknowledgements

We thank for the comments the editors, the referees, and Federico Bassetti, Radu Craiu, Sylvia Frühwirth-Schnatter, Christian Gouriéroux, Rajarshi

Guhaniyogi, Raquel Prado, Christian P. Robert, Mark F. J. Steel, Lei Sun, Stefano Tonellato, and Mike West. MB and RC acknowledge financial support from the Italian MIUR under the PRIN project Hi-Di NET (grant agreement no. 2017TA7TYC). MI acknowledges financial support from the Université Franco-Italienne (grant Vinci 2016) and from the EU Horizon 2020 programme under the Marie Skłodowska-Curie scheme (grant agreement no. 887220).

A Proofs of the Results in the Paper

This appendix provides the derivation of the results. See the Supplement for further details.

Proof of Remark 2.1. Following the definition of Hadamard product, the first equation becomes $x_{ijk,t} = b_{ijk,t} d_{ijk,t}$, with $d_{ijk,t} = 1_{\mathbb{R}_+}(x_{ijk,t}^*)$, where $b_{ijk,t} \sim \text{Bern}(1 - \rho(t))$, $x_{ijk,t}$, and $x_{ijk,t}^*$ are the (i, j, k) -th elements of the tensors $\mathcal{B}(t)$, \mathcal{X}_t , \mathcal{X}_t^* , respectively. From the definition of mode- n product, the second equation is

$$x_{ijk,t}^* = \sum_{i'=1}^I \sum_{j'=1}^J \sum_{k'=1}^K \sum_{q'=1}^Q \tilde{g}_{ijk't}(t) z_{i'j'k'q',t} + \varepsilon_{ijk,t}$$

where $\ell = 1 + (i' - 1) + (j' - 1)I + (k' - 1)IJ + (q' - 1)IJK$. By imposing the restrictions $\tilde{g}_{ijk't}(t) = 0$ when $i' \neq i, j' \neq j$, and $k' = k$ and defining $\mathbf{g}_{ijk}(t) = (g_{ijk1}(t), \dots, g_{ijkQ}(t))'$ with $g_{ijkq}(t) = \tilde{g}_{ijk't}(t)$ when $\ell \neq 1 + (i - 1) + (j - 1)I + (k - 1)IJ + (q - 1)IJK$ yields

$$x_{ijk,t}^* = \sum_{q=1}^Q g_{ijkq}(t) z_{ijkq,t} + \varepsilon_{ijk,t} = \mathbf{z}_{ijk,t}' \mathbf{g}_{ijk}(t) + \varepsilon_{ijk,t}.$$

By the definitions of Hadamard and mode- n product, $\mathcal{X}_t^* = (\mathcal{G}(t) \odot \mathcal{Z}_t) \times_4 \mathbf{1}_Q + \mathcal{E}_t$. \square

Proof of Remark 2.2. First, define $\mathbf{z}_t = (1, \tilde{\mathbf{z}}_t)'$ and denote with \mathcal{G} the 4-order tensor obtained by stacking the 4-th order tensors $\mathcal{G}_1, \dots, \mathcal{G}_t$ along the 4-th dimension. For example, if each \mathcal{G}_t is a 4-order tensor of size $(I \times J \times K \times Q)$,

then \mathcal{G} is a 4-order tensor of size $(I \times J \times K \times QL)$. Finally, recall that $\mathcal{G}_{(t)} = \mathcal{G}_{s_t}$. Start by noting that $\mathcal{G}_{(t)} \times_4 \mathbf{z}_t = \mathcal{G}_{s_t} \times_4 \mathbf{z}_t$, hence $\mathcal{G}_{(t)} \times_4 \mathbf{z}_t = \mathcal{G}_l \times_4 \mathbf{z}_t$ if $s_t = l$. Let now introduce the L -dimensional binary vector $\zeta_t = (\zeta_{t,1}, \dots, \zeta_{t,L})'$ whose k th entry is $\zeta_{t,k} = \mathbb{R}_{(l)}(s_t)$, and let $\mathbf{1}_Q$ be a Q -dimensional column vector of ones. Therefore, by the properties of the Kronecker product one has

$$\mathcal{G}_{(t)} \times_4 \mathbf{z}_t = \mathcal{G}_{s_t} \times_4 \mathbf{z}_t = \mathcal{G} \times_4 (\zeta_t \otimes \mathbf{z}_t) = \mathcal{G} \times_4 (\zeta_t \otimes (\mathbf{1}, \tilde{\mathbf{z}}_t)') = \mathcal{G} \times_4 (\zeta_t', \zeta_t' \otimes \tilde{\mathbf{z}}_t')$$

The dynamics of the Markov chain can be expressed in a VAR-like form as follows. Assume the current state of the chain is $s_t = i$, then the conditional expectation of binary random vector ζ_t is $\mathbb{E}[\zeta_{t+1} | s_t = i] = \Xi_{i,:}$, where $\Xi_{i,:}$ is the i th column of Ξ . By stacking all states one obtains $\mathbb{E}[\zeta_{t+1} | \zeta_t, \zeta_{t-1}, \dots] = \mathbb{E}[\zeta_{t+1} | \zeta_t] = \Xi \zeta_t$, where the first equality follows from the Markov property. Thus, the process ζ_t takes on a finite set of values, is zero on average, and satisfies $\zeta_{t+1} = \Xi \zeta_t + \mathbf{u}_t$, where \mathbf{u}_t is a martingale difference sequence. \square

Proof of Lemma 3.1. For each $l = 1, \dots, L$, let $g_{ijk,l} = \sum_{r=1}^R g_{r,l}$, where $g_{r,l} = \gamma_{1,i,l}^{(r)} \gamma_{2,j,l}^{(r)} \gamma_{3,k,l}^{(r)} \gamma_{4,q,l}^{(r)}$, and let $m_1 = i, m_2 = j, m_3 = k$, and $m_4 = p$. The following Lemma is an adaptation of the results in [Springer and Thompson \(1970\)](#) using our notation.

Lemma A.1. *Under the prior specification in eq. (8), the distribution of $g_{r,l}$ has proper (conditional) probability density function:*

$$p(g_{r,l} | \tau, \phi, \mathbf{W}_l) = K_{r,l} \cdot G_{4,0}^{4,0} \left(g_{r,l} \prod_{h=1}^4 (2\tau\phi_r w_{h,r,l,m_h})^{-1} \mid a_0^0 \right),$$

with $G_{p,q}^{m,n}(x | a_a^b)$ a Meijer G -function

$$G_{p,q}^{m,n}(x | a_a^b) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{\prod_{j=1}^m \Gamma(b_j - s) \prod_{j=1}^n \Gamma(1 - a_j + s)}{\prod_{j=m+1}^p \Gamma(1 - b_j + s) \prod_{j=n+1}^q \Gamma(a_j - s)} ds,$$

with $\mathbf{a} = \mathbf{b} = \mathbf{0}$, $m = n = 4$, $p = q = 0$, and $K_{r,l} = (2\pi)^{-4/2} \prod_{h=1}^4 (2\tau\phi_r w_{h,r,l,m_h})^{-1}$. The absolute moments of $g_{r,l}$ are bounded:

$$\mathbb{E}[\|g_{r,l}\|^\ell] = \mathbb{E}[\|\gamma_{1,i,l}^{(r)}\|^\ell] \cdot \mathbb{E}[\|\gamma_{2,j,l}^{(r)}\|^\ell] \cdot \mathbb{E}[\|\gamma_{3,k,l}^{(r)}\|^\ell] \cdot \mathbb{E}[\|\gamma_{4,q,l}^{(r)}\|^\ell] < \infty, \quad \ell = 0, 1, 2, \dots$$

Owing to the conditional independent structure of the hierarchical prior, to prove that the joint prior distribution of $p(\mathcal{G}_l)$ is proper, it suffices to show that $p(g_{ijk,l})$ is proper:

$$\int p(g_{ijk,l} | \tau, \phi, \mathbf{W}_l) p(\tau, \phi, \mathbf{W}_l) dg_{ijk,l} < \infty.$$

Following Lemma A.1, the $g_{r,l}$ in the definition of $g_{ijk,l}$ has proper distribution $p(g_{r,l} | \tau, \phi, \mathbf{W}_l)$. To find the distribution of $g_{ijk,l}$, we define the following transformed vector \mathbf{z}_l , with elements $z_r = g_{r,l}$, $r = 1, \dots, R-1$, and $z_R = g_{1,l} + \dots + g_{R,l}$. The Jacobian of the transformation is one and the marginal distribution of \mathbf{z}_R has pdf

$$p(\mathbf{z}_R | \tau, \phi, \mathbf{W}_l) = \int_{\mathbb{R}^{R-1}} p(\mathbf{z}_{R,l} - \mathbf{z}_1 - \dots - \mathbf{z}_{R-1} | \tau, \phi, \mathbf{W}_l) \prod_{r=1}^{R-1} p(\mathbf{z}_r | \tau, \phi, \mathbf{W}_l) d\mathbf{z}_1 \dots d\mathbf{z}_{R-1}.$$

By Fubini's theorem one gets

$$\begin{aligned} \int_{\mathbb{R}} p(\mathbf{z}_R | \tau, \phi, \mathbf{W}_l) d\mathbf{z}_R &= \int_{\mathbb{R}^{R-1}} \prod_{r=1}^{R-1} p(\mathbf{z}_r | \tau, \phi, \mathbf{W}_l) \left(\int_{\mathbb{R}} p(\mathbf{z}_R - \sum_{s=1}^{R-1} \mathbf{z}_s | \tau, \phi, \mathbf{W}_l) d\mathbf{z}_R \right) d\mathbf{z}_1 \dots d\mathbf{z}_{R-1} \\ &= \int_{\mathbb{R}^{R-1}} \prod_{r=1}^{R-1} p(\mathbf{z}_r | \tau, \phi, \mathbf{W}_l) \left(\int_{\mathbb{R}} p(u | \tau, \phi, \mathbf{W}_l) du \right) d\mathbf{z}_1 \dots d\mathbf{z}_{R-1} = 1. \end{aligned}$$

The proof is completed by setting $\mathbf{z}_R = g_{ijk,l}$ and recalling that the prior distributions on τ , ϕ , and \mathbf{w}_l are proper and independent. Concerning the boundedness of the absolute moments, since $\mathbb{E}[\|g_{r,l}\|^\ell] < \infty$ for each r, l, ℓ , by Minkowski's inequality:

$$\left(\mathbb{E}[\|g_{ijk,l}\|^\ell] \right)^{1/\ell} = \left(\sum_{r=1}^R \mathbb{E}[\|g_{r,l}\|^\ell] \right)^{1/\ell} \leq \sum_{r=1}^R \left(\mathbb{E}[\|g_{r,l}\|^\ell] \right)^{1/\ell} < \infty, \quad \forall i, j, k, l, \ell.$$

□

A.1 Complete Data Likelihood

The data augmented likelihood of model (5) is $L(\mathcal{X}, \mathbf{s} | \boldsymbol{\theta}) = L(\mathcal{X} | \mathbf{s}, \boldsymbol{\theta}) L(\mathbf{s} | \boldsymbol{\theta})$, where

$$L(\mathcal{X} | \mathbf{s}, \boldsymbol{\theta}) = \prod_{l=1}^L \prod_{t \in \mathcal{T}_l} \prod_{i=1}^I \prod_{j=1}^J \prod_{k=1}^K \left(\frac{(1 - \rho_l) \exp(\mathbf{z}_t' \mathbf{g}_{ijk,l})}{1 + \exp(\mathbf{z}_t' \mathbf{g}_{ijk,l})} \right)^{x_{ijk,t}} \left(\rho_l + \frac{1 - \rho_l}{1 + \exp(\mathbf{z}_t' \mathbf{g}_{ijk,l})} \right)^{1 - x_{ijk,t}}$$

and $L(\mathbf{s} | \boldsymbol{\theta}) = \prod_{g=1}^L \prod_{l=1}^L \xi_{g,l}^{N_{g,l}(\mathbf{s})}$. We now introduce the latent allocation variable for the mixture in eq. (2), $d_{ijk,t} \in \{0, 1\}$, to get the conditional distribution

$$p(x_{ijk,t} | d_{ijk,t}, s_t = l, \mathcal{G}_l) = \left(\delta_{\{0\}}(x_{ijk,t}) \right)^{d_{ijk,t}} \frac{\left(\exp(\mathbf{z}_t' \mathbf{g}_{ijk,l}) \right)^{x_{ijk,t} (1 - d_{ijk,t})}}{\left(1 + \exp(\mathbf{z}_t' \mathbf{g}_{ijk,l}) \right)^{(1 - d_{ijk,t})}}. \quad (\text{A.1})$$

and the marginal distribution $p(d_{ijk,t} | s_t = l, \rho_l) = \rho_l^{d_{ijk,t}} (1 - \rho_l)^{1 - d_{ijk,t}}$. Then, we decompose the ratio in eq. (A.1) and obtain

$$p(x_{ijk,t} | d_{ijk,t}, \omega_{ijk,t}, s_t = l, \mathcal{G}_l) = \frac{\left(2 \delta_{\{0\}}(x_{ijk,t}) \right)^{d_{ijk,t}}}{2} \exp\left(-\frac{\omega_{ijk,t}}{2} (\mathbf{z}_t' \mathbf{g}_{ijk,l})^2 + \kappa_{ijk,t} (\mathbf{z}_t' \mathbf{g}_{ijk,l})\right)$$

where $\kappa_{ijk,t} = (1 - d_{ijk,t})(x_{ijk,t} - 1/2)$ and $\omega_{ijk,t} \sim PG(1, 0)$, with $PG(b, c)$ the Pólya-Gamma distribution with parameters $b > 0$ and $c \in \mathbb{R}$ (Polson et al., 2013, Theorem 1). Combining the latter equation with the marginal distribution of the latent variables and collecting all the data points, one gets the complete data likelihood in eq. (12).

A.2 MCMC Algorithm

Sampling ϕ_r

Let $n = n_1 + n_2 + n_3 + n_4 = I + J + K + Q$. The posterior full conditional is

$$\begin{aligned}
p(\phi | \mathcal{G}, \mathcal{W}) &\propto p(\phi) \int_0^\infty p(\mathcal{G} | \mathcal{W}, \phi, \tau) p(\tau) d\tau \\
&\propto \int_0^\infty \left(\prod_{r=1}^R (\tau \phi_r)^{\bar{a}-n-1} \right) \exp\left(-\frac{1}{2} \sum_{r=1}^R (2\bar{b}^\tau \tau \phi_r + \frac{1}{\tau \phi_r} \sum_{h=1}^4 \sum_{l=1}^L \frac{\gamma_{h,l}^{(r)'} \gamma_{h,l}^{(r)}}{w_{h,r,l}})\right) d\tau.
\end{aligned}$$

The integrand is the kernel of the GiG distribution given in eq. (17). Following [Guhaniyogi et al. \(2017\)](#), it is possible to sample from the posterior of ϕ_r , for each $r = 1, \dots, R$ by first sampling ψ_r , then setting $\phi_r = \psi_r / (\psi_1 + \dots + \psi_R)$.

Sampling τ and $w_{h,r,l}$

The full conditional distributions are GiG with kernels

$$\begin{aligned}
p(\tau | \mathcal{G}, \mathcal{W}, \phi) &\propto \tau^{\bar{a}^\tau - 1} \exp(-\bar{b}^\tau \tau) \prod_{r=1}^R \prod_{h=1}^4 \prod_{l=1}^L \left| \tau \phi_r w_{h,r,l} \mathbf{I}_{n_h} \right|^{-1/2} \exp\left(-\frac{1}{2} \frac{\gamma_{h,l}^{(r)'} \gamma_{h,l}^{(r)}}{\phi_r w_{h,r,l}}\right) \\
&\propto \tau^{\bar{a}^\tau - 4R - 1} \exp\left(-\frac{1}{2} (2\bar{b}^\tau \tau + \frac{1}{\tau} \sum_{r=1}^R \sum_{h=1}^4 \sum_{l=1}^L \frac{\gamma_{h,l}^{(r)'} \gamma_{h,l}^{(r)}}{\phi_r w_{h,r,l}})\right), \\
p(w_{h,r,l} | \gamma_{h,l}^{(r)}, \phi_r, \tau, \lambda_l) &\propto p(w_{h,r,l} | \lambda_l) p(\gamma_{h,l}^{(r)} | w_{h,r,l}, \phi_r, \tau) \propto \exp\left(-\frac{\lambda_l^2}{2} w_{h,r,l} - \frac{1}{2} \frac{\gamma_{h,l}^{(r)'} \gamma_{h,l}^{(r)}}{\tau \phi_r w_{h,r,l}}\right) w_{h,r,l}^{-n_h/2}.
\end{aligned}$$

Sampling λ_l

We apply Hamiltonian Monte Carlo ([Neal, 2011](#)) to sample from

$$p(\lambda_l | \mathbf{W}_l) \propto \lambda_l^{\bar{a}_l - 1} \exp(-\bar{b}_l \lambda_l) \prod_{r=1}^R \prod_{h=1}^4 \frac{\lambda_l^2}{2} \exp\left(-\frac{\lambda_l^2}{2} w_{h,r,l}\right) \propto \exp\left(-\lambda_l \bar{b}_l - \frac{\lambda_l^2}{2} \sum_{r=1}^R \sum_{h=1}^4 w_{h,r,l}\right) \lambda_l^{\bar{a}_l + 8R - 1}.$$

Full conditional distribution of $\gamma_{h,l}^{(r)}$

For deriving the full conditional distribution of PARAFAC marginals, $\gamma_{h,l}^{(r)}$, start by defining $u_{ijk,t} = \kappa_{ijk,t} / \omega_{ijk,t}$, $\mathcal{U}_t = (u_{ijk,t})_{ijk}$, $\mathbf{\Omega}_t = (\omega_{ijk,t})_{ijk}$, $\mathbf{u}_t = \text{vec}(\mathcal{U}_t)$ and $\bar{\mathbf{\Omega}}_t = \text{diag}(\text{vec}(\mathbf{\Omega}_t))$. Denoting with $p(\mathcal{G}_l)$ the joint prior distribution on $\{\gamma_{h,l}^{(r)}\}_{h,r}$, one gets

$$p(\mathcal{G}_l | \mathcal{X}_t, \mathcal{D}_t, \boldsymbol{\Omega}_t, s_t = l, \rho_l) \propto p(\mathcal{G}_l) \prod_{t \in \mathcal{T}_l} \prod_{i=1}^I \prod_{j=1}^J \prod_{k=1}^K p(\omega_{ijk,t}) \quad (\text{A.2})$$

$$\prod_{t \in \mathcal{T}_l} \exp\left(-\frac{1}{2}(\text{vec}(\mathcal{G}_l \times_4 \mathbf{z}_t) - \mathbf{u}_t)' \overline{\overline{\boldsymbol{\Omega}_t}} (\text{vec}(\mathcal{G}_l \times_4 \mathbf{z}_t) - \mathbf{u}_t)\right).$$

By the definitions of mode- n product and PARAFAC decomposition, denoting by $\langle \cdot, \cdot \rangle$ the standard inner product in the Euclidean space \mathbb{R}^n , we obtain

$$\text{vec}(\mathcal{G}_l \times_4 \mathbf{z}_t) = \sum_{r=1}^R (\boldsymbol{\gamma}_{1,l}^{(r)} \circ \boldsymbol{\gamma}_{2,l}^{(r)} \circ \boldsymbol{\gamma}_{3,l}^{(r)}) \langle \boldsymbol{\gamma}_{4,l}^{(r)}, \mathbf{z}_t \rangle = \sum_{r=1}^R \overline{\mathbf{g}}_{l,t}^{(-r)}. \quad (\text{A.3})$$

From eq. (A.3) we have:

$$\overline{\mathbf{g}}_{l,t}^{(-r)} = \langle \boldsymbol{\gamma}_{4,l}^{(r)}, \mathbf{z}_t \rangle \text{vec}(\boldsymbol{\gamma}_{1,l}^{(r)} \circ \boldsymbol{\gamma}_{2,l}^{(r)} \circ \boldsymbol{\gamma}_{3,l}^{(r)}) = \text{vec}(\boldsymbol{\gamma}_{1,l}^{(r)} \circ \boldsymbol{\gamma}_{2,l}^{(r)} \circ \boldsymbol{\gamma}_{3,l}^{(r)}) \mathbf{z}_t' \boldsymbol{\gamma}_{4,l}^{(r)} = \mathbf{A}_4 \boldsymbol{\gamma}_{4,l}^{(r)}, \quad (\text{A.4})$$

$$= \langle \boldsymbol{\gamma}_{4,l}^{(r)}, \mathbf{z}_t \rangle (\boldsymbol{\gamma}_{3,l}^{(r)} \otimes \boldsymbol{\gamma}_{2,l}^{(r)} \otimes \mathbf{I}_I) \boldsymbol{\gamma}_{1,l}^{(r)} = \mathbf{A}_1 \boldsymbol{\gamma}_{1,l}^{(r)} \quad (\text{A.5})$$

$$= \langle \boldsymbol{\gamma}_{4,l}^{(r)}, \mathbf{z}_t \rangle (\boldsymbol{\gamma}_{3,l}^{(r)} \otimes \mathbf{I}_J \otimes \boldsymbol{\gamma}_{1,l}^{(r)}) \boldsymbol{\gamma}_{2,l}^{(r)} = \mathbf{A}_2 \boldsymbol{\gamma}_{2,l}^{(r)}, \quad (\text{A.6})$$

$$= \langle \boldsymbol{\gamma}_{4,l}^{(r)}, \mathbf{z}_t \rangle (\mathbf{I}_K \otimes \boldsymbol{\gamma}_{2,l}^{(r)} \otimes \boldsymbol{\gamma}_{1,l}^{(r)}) \boldsymbol{\gamma}_{3,l}^{(r)} = \mathbf{A}_3 \boldsymbol{\gamma}_{3,l}^{(r)}. \quad (\text{A.7})$$

Setting $\overline{\mathbf{g}}_{l,t}^{(-r)} = \sum_{\substack{v=1 \\ v \neq r}}^R \overline{\mathbf{g}}_{l,t}^{(-v)}$ we get $\text{vec}(\mathcal{G}_l \times_4 \mathbf{z}_t) = \overline{\mathbf{g}}_{l,t}^{(-r)} + \overline{\mathbf{g}}_{l,t}^{(-r)}$. Thus, for each l we get

$$L(\mathcal{X}, \mathcal{D}, \boldsymbol{\Omega}, \mathbf{s} | \boldsymbol{\theta}) \propto \prod_{t \in \mathcal{T}_l} \exp\left(-\frac{1}{2}(\overline{\mathbf{g}}_{l,t}^{(-r)} + \overline{\mathbf{g}}_{l,t}^{(-r)} - \mathbf{u}_t)' \overline{\overline{\boldsymbol{\Omega}_t}} (\overline{\mathbf{g}}_{l,t}^{(-r)} + \overline{\mathbf{g}}_{l,t}^{(-r)} - \mathbf{u}_t)\right). \quad (\text{A.8})$$

We can now single out a specific component $\mathcal{G}_l^{(r)}$ of the PARAFAC decomposition of the tensor \mathcal{G} , which is incorporated in $\overline{\mathbf{g}}_{l,t}^{(-r)}$. For each $l = 1, \dots, L$ we obtain

$$L(\mathcal{X}, \mathcal{D}, \boldsymbol{\Omega}, \mathbf{s} | \boldsymbol{\theta}) \propto \prod_{t \in \mathcal{T}_l} \exp\left(-\frac{1}{2}(\overline{\mathbf{g}}_{l,t}^{(-r)} - \overline{\overline{\boldsymbol{\Omega}_t}} \overline{\mathbf{g}}_{l,t}^{(-r)} - 2(\mathbf{u}_t - \overline{\mathbf{g}}_{l,t}^{(-r)})' \overline{\overline{\boldsymbol{\Omega}_t}} \overline{\mathbf{g}}_{l,t}^{(-r)})\right). \quad (\text{A.9})$$

From eqs. (A.4) to eqs. (A.7) one obtains $\bar{\mathbf{g}}_{l,t}^{(-r)}$ written as $\mathbf{A}_h \boldsymbol{\gamma}_{h,l}^{(r)}$, with $h = 1, \dots, 4$, that is

$$\bar{\mathbf{g}}_{l,t}^{(-r)'} \bar{\boldsymbol{\Omega}}_t \bar{\mathbf{g}}_{l,t}^{(-r)} = \boldsymbol{\gamma}_{h,l}^{(r)'} (\bar{\boldsymbol{\Sigma}}_{h,l,t}^{(-r)})^{-1} \boldsymbol{\gamma}_{h,l}^{(r)} \quad (\text{A.10})$$

$$-2(\mathbf{u}_t - \bar{\mathbf{g}}_{l,t}^{(-r)})' \bar{\boldsymbol{\Omega}}_t \bar{\mathbf{g}}_{l,t}^{(-r)} = -2\boldsymbol{\mu}_{h,l,t}^{(-r)'} (\bar{\boldsymbol{\Sigma}}_{h,l,t}^{(-r)})^{-1} \boldsymbol{\gamma}_{h,l}^{(r)}. \quad (\text{A.11})$$

By Bayes' theorem and plugging eq. (A.10) and (A.11) into eq. (A.9) we get

$$p(\boldsymbol{\gamma}_{h,l}^{(r)} | \boldsymbol{\theta}) \propto \exp\left(-\frac{1}{2}(\boldsymbol{\gamma}_{h,l}^{(r)'} ((\bar{\boldsymbol{\Lambda}}_{h,l})^{-1} + \sum_{t \in \mathcal{I}_l} (\bar{\boldsymbol{\Sigma}}_{h,l,t}^{(-r)})^{-1}) \boldsymbol{\gamma}_{h,l}^{(r)} - 2(\bar{\boldsymbol{\zeta}}_{h,l}' (\bar{\boldsymbol{\Lambda}}_{h,l})^{-1} + \sum_{t \in \mathcal{I}_l} \boldsymbol{\mu}_{h,l,t}^{(-r)'} (\bar{\boldsymbol{\Sigma}}_{h,l,t}^{(-r)})^{-1}) \boldsymbol{\gamma}_{h,l}^{(r)}\right),$$

that is the kernel of a multivariate normal distribution.

Sampling $\omega_{ijk,t}$

Define $\psi_{ijk,t} = \mathbf{z}_t' \mathbf{g}_{ijk,s_t}$. Since $p(\omega_{ijk,t}) \sim PG(1, 0)$, by (Polson et al., 2013, Theorem 1) one gets the posterior full conditional distribution as

$$\begin{aligned} p(\omega_{ijk,t} | x_{ijk,t}, s_t, \mathcal{G}_{s_t}) &= \sum_{d_{ijk,t} \in \{0,1\}} \int p(\omega_{ijk,t}, d_{ijk,t} | x_{ijk,t}, s_t, \mathcal{G}_{s_t}, \rho_{s_t}) p(\rho_{s_t}) d\rho_{s_t} \\ &= \sum_{d_{ijk,t} \in \{0,1\}} \int \exp(\kappa_{ijk,t}^{(s_t)} \psi_{ijk,t}) \frac{\exp(\psi_{ijk,t} x_{ijk,t} (1 - d_{ijk,t}))}{(1 + \exp(\psi_{ijk,t}))^{1-d_{ijk,t}}} \exp\left(-\frac{\omega_{ijk,t}}{2} \psi_{ijk,t}^2\right) p(\omega_{ijk,t}) p(\rho_{s_t}) d\rho_{s_t} \\ &= \left(1 + \frac{1 + \exp(\psi_{ijk,t})}{\exp(\psi_{ijk,t} / 2)}\right) \exp(-\psi_{ijk,t}^2 \omega_{ijk,t} / 2) p(\omega_{ijk,t}) \propto \exp(-\psi_{ijk,t}^2 \omega_{ijk,t} / 2) p(\omega_{ijk,t}). \end{aligned}$$

Sampling $d_{ijk,t}$

The posterior full conditional posterior distribution is discrete

$$\begin{aligned} p(d_{ijk,t} = 1 | \mathcal{X}, \mathbf{s}, \mathcal{G}_{s_t}, \boldsymbol{\rho}_{s_t}) &\propto \rho_{s_t} \delta_{\{0\}}(x_{ijk,t}) \\ p(d_{ijk,t} = 0 | \mathcal{X}, \mathbf{s}, \mathcal{G}_{s_t}, \boldsymbol{\rho}_{s_t}) &\propto (1 - \rho_{s_t}) \frac{\exp((\mathbf{z}_t' \mathbf{g}_{ijk,s_t}) x_{ijk,t})}{1 + \exp(\mathbf{z}_t' \mathbf{g}_{ijk,s_t})}, \end{aligned}$$

which follows from

$$p(d_{ijk,t} | \mathcal{X}, \mathbf{s}, \mathcal{G}_{s_t}, \rho_{s_t}) \propto (\rho_{s_t} \delta_{\{0\}}(x_{ijk,t}))^{d_{ijk,t}} \left((1 - \rho_{s_t}) \frac{(\exp(\mathbf{z}_t' \mathbf{g}_{ijk,s_t}))^{x_{ijk,t}}}{1 + \exp(\mathbf{z}_t' \mathbf{g}_{ijk,s_t})} \right)^{1-d_{ijk,t}}.$$

Sampling ρ_l

Defining

$$N_1^l = \sum_{t \in \mathcal{T}_l} \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \mathbf{1}_{\{1\}}(d_{ijk,t}), \quad N_0^l = \sum_{t \in \mathcal{T}_l} \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \mathbf{1}_{\{0\}}(d_{ijk,t}),$$

one obtains the posterior full conditional distribution as the kernel of a Beta

$$p(\rho_l | \mathcal{X}, \mathcal{D}, \mathbf{s}) \propto \rho_l^{\bar{a}_l^\rho - 1} (1 - \rho_l)^{\bar{b}_l^\rho - 1} \prod_{t \in \mathcal{T}_l} \prod_{i=1}^I \prod_{j=1}^J \prod_{k=1}^K \rho_l^{d_{ijk,t}} (1 - \rho_l)^{1-d_{ijk,t}} = \rho_l^{N_1^l + \bar{a}_l^\rho - 1} (1 - \rho_l)^{N_0^l + \bar{b}_l^\rho - 1}.$$

Sampling ξ_l and \mathbf{s}_t

The posterior full conditional distribution of each row l is

$$p(\xi_l | \mathbf{s}) \propto \left(\prod_{k=1}^L \xi_{l,k}^{\bar{c}_k - 1} \right) \left(\prod_{g=1}^L \prod_{k=1}^L \xi_{g,k}^{N_{g,k}(\mathbf{s})} \right) \propto \prod_{k=1}^L \xi_{l,k}^{\bar{c}_k + N_{l,k}(\mathbf{s}) - 1},$$

which is the kernel of the Dirichlet in eq. (22), and $N_{i,j}(\mathbf{s}) = \sum_t \zeta_{t-1,i} \zeta_{t,j}$. We update \mathbf{s} via the Forward Filtering Backward Sampling algorithm (Frühwirth-Schnatter, 2006).

Note

¹We use the shape-rate formulation for the Gamma distribution, such that

$$\mathbb{E}(x) = \alpha / \beta, \text{Var}(x) = \alpha / \beta^2.$$

References

Bianchi, D., Billio, M., Casarin, R., and Guidolin, M. (2019). Modeling systemic risk with Markov switching graphical SUR models. *Journal of Econometrics*, 210(1):58–74.

Billio, M., Getmansky, M., Lo, A. W., 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.

Boccaletti, S., Bianconi, G., Criado, R., Del Genio, C. I., Gómez-Gardenes, J., Romance, M., Sendina-Nadal, I., Wang, Z., and Zanin, M. (2014). The structure and dynamics of multilayer networks. *Physics Reports*, 544(1):1–122.

Celeux, G., Forbes, F., Robert, C. P., Titterton, D. M., et al. (2006). Deviance information criteria for missing data models. *Bayesian Analysis*, 1(4):651–673.

Chen, X., Irie, K., Banks, D., Haslinger, R., Thomas, J., and West, M. (2018). Scalable Bayesian modeling, monitoring, and analysis of dynamic network flow data. *Journal of the American Statistical Association*, 113(522):519–533.

Cichocki, A., Lee, N., Oseledets, I., Phan, A., Zhao, Q., and Mandic, D. (2016). Tensor networks for dimensionality reduction and large-scale optimization: Part 1 low-rank tensor decompositions. *Foundations and Trends in Machine Learning*, 9(4-5):249–429.

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

Diestel, R. (2012). *Graph theory*, volume 173 of *Graduate Texts in Mathematics*. Springer.

Dukić, V. M. and Peña, E. A. (2005). Variance estimation in a model with Gaussian submodels. *Journal of the American Statistical Association*, 100(469):296–309.

Durante, D. and Dunson, D. B. (2014a). Bayesian dynamic financial networks with time-varying predictors. *Statistics & Probability Letters*, 93:19–26.

Durante, D. and Dunson, D. B. (2014b). Bayesian logistic Gaussian process models for dynamic networks. *Proceedings of the 17th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 33.

Fox, E. W., Short, M. B., Schoenberg, F. P., Coronges, K. D., and Bertozzi, A. L. (2016). Modeling e-mail networks and inferring leadership using self-exciting point processes. *Journal of the American Statistical Association*, 111(514):564–584.

Frühwirth-Schnatter, S. (2001). Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association*, 96(453):194–209.

Frühwirth-Schnatter, S. (2006). *Finite mixture and Markov switching models*. Springer.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian Data Analysis*. CRC Press.

Gerard, D. and Hoff, P. (2017). Adaptive higher-order spectral estimators. *Electronic Journal of Statistics*, 11(2):3703–3737.

Guhaniyogi, R., Qamar, S., and Dunson, D. B. (2017). Bayesian tensor regression. *Journal of Machine Learning Research*, 18(79):1–31.

Hackbusch, W. (2012). *Tensor spaces and numerical tensor calculus*. Springer Science & Business Media.

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

Harris, M. N. and Zhao, X. (2007). A zero-inflated ordered probit model, with an application to modelling tobacco consumption. *Journal of Econometrics*, 141(2):1073–1099.

Hoff, P. D. (2011). Separable covariance arrays via the Tucker product, with applications to multivariate relational data. *Bayesian Analysis*, 6(2):179–196.

Hoff, P. D. (2015). Multilinear tensor regression for longitudinal relational data. *The Annals of Applied Statistics*, 9(3):1169–1193.

Holme, P. and Saramäki, J. (2012). Temporal networks. *Physics Reports*, 519(3):97–125.

Holsclaw, T., Greene, A. M., Robertson, A. W., and Smyth, P. (2017). Bayesian non-homogeneous Markov models via Pólya-Gamma data augmentation with applications to rainfall modeling. *The Annals of Applied Statistics*, 11(1):393–426.

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

Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y., and Porter, M. A. (2014). Multilayer networks. *Journal of Complex Networks*, 2(3):203–271.

Kolda, T. G. and Bader, B. W. (2009). Tensor decompositions and applications. *SIAM Review*, 51(3):455–500.

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

Leskovec, J., Kleinberg, J., and Faloutsos, C. (2007). Graph evolution: Densification and shrinking diameters. *ACM transactions on Knowledge Discovery from Data*, 1(1):2–es.

Li, L. and Zhang, X. (2017). Parsimonious tensor response regression. *Journal of the American Statistical Association*, 112(519):1131–1146.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In Zarembka, P., editor, *Frontiers in Econometrics*, pages 105–142. Academic Press.

Neal, R. M. (2011). MCMC using Hamiltonian dynamics. In Brooks, S., Gelman, A., Galin, J. L., and Meng, X.-L., editors, *Handbook of Markov Chain Monte Carlo*. CRC.

Om, K., Boukoros, S., Nugaliyadde, A., McGill, T., Dixon, M., Koutsakis, P., and Wong, K. W. (2020). Modelling email traffic workloads with RNN and LSTM models. *Human-centric Computing and Information Sciences*, 10(1):1–16.

Polson, N. G., Scott, J. G., and Windle, J. (2013). Bayesian inference for logistic models using Pólya–Gamma latent variables. *Journal of the American Statistical Association*, 108(504):1339–1349.

Robert, C. P. (2007). *The Bayesian choice*. Springer.

Sarkar, P. and Moore, A. W. (2006). Dynamic social network analysis using latent space models. In *Advances in Neural Information Processing Systems*, pages 1145–1152.

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

Springer, M. D. and Thompson, W. E. (1970). The distribution of products of beta, gamma and Gaussian random variables. *SIAM Journal on Applied Mathematics*, 18(4):721–737.

Wang, L., Durante, D., Jung, R. E., and Dunson, D. B. (2017). Bayesian network–response regression. *Bioinformatics*, 33(12):1859–1866.

Xing, E. P., Fu, W., and Song, L. (2010). A state-space mixed membership blockmodel for dynamic network tomography. *The Annals of Applied Statistics*, 4(2):535–566.

Zhang, X., Li, L., Zhou, H., Zhou, Y., and Shen, D. (2019). Tensor generalized estimating equations for longitudinal imaging analysis. *Statistica Sinica*, 29:1977–2005.

Zhou, H., Li, L., and Zhu, H. (2013). Tensor regression with applications in neuroimaging data analysis. *Journal of the American Statistical Association*, 108(502):540–552.

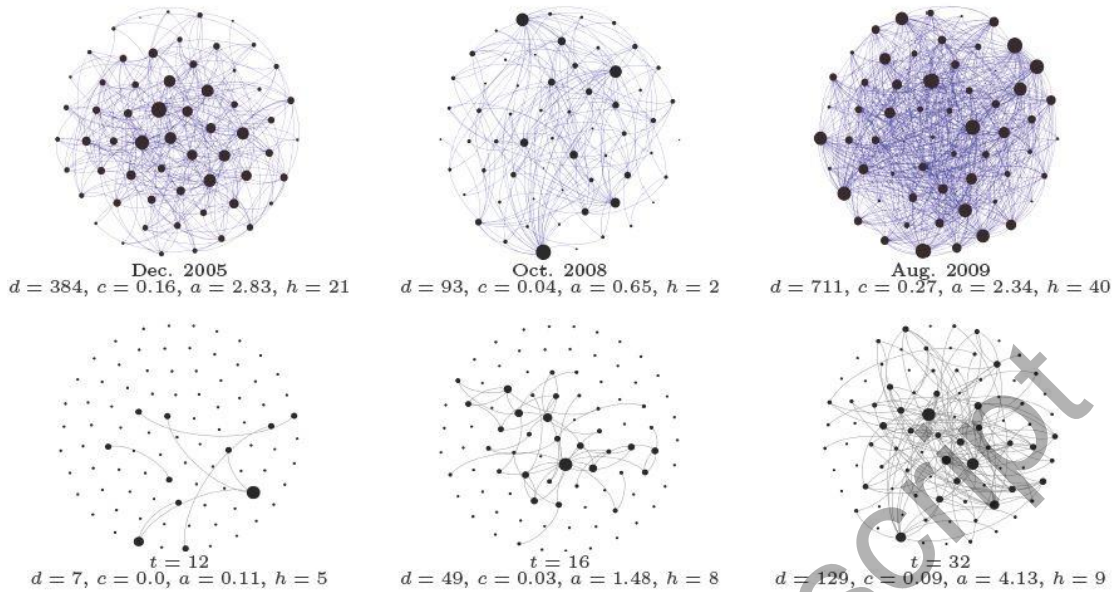


Fig. 1 *Top*: dynamic network of Granger-causal linkages (edges) among the log returns (layer 1) of 61 European financial institutions (nodes). *Bottom*: dynamic network of email exchanges (edges) among 90 researchers at a European institution (nodes), within department 1 (layer 1). Network structures and statistics: degree, d , average clustering coefficient, c , average path length, a , and number of hubs, h . Node size (black dots) is proportional to total degree. Edges (arcs) are clockwise directed (i.e., following the edge clockwise indicates the direction of the financial linkage).

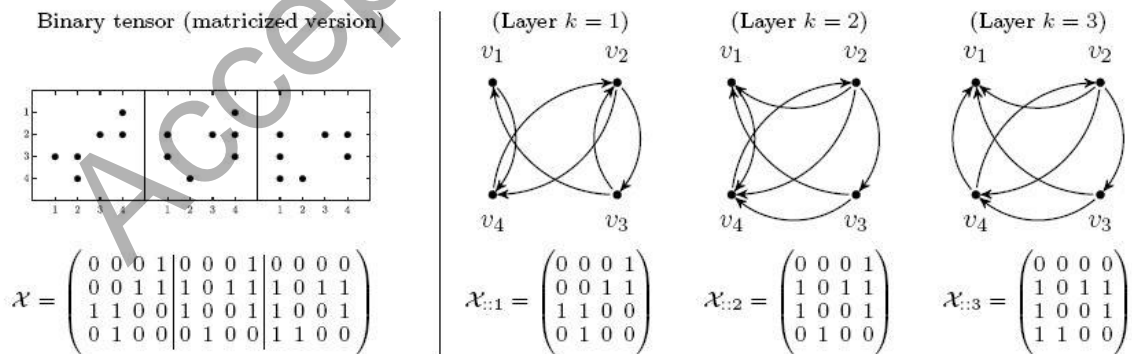


Fig. 2 Example of binary tensor \mathcal{X} of size $(4 \times 4 \times 3)$. Left column: matricized form. Right column, top row: multilayer network representation with $K=3$ layers

(in column). In each graph, node v_i represents subject i ; a clockwise-oriented edge from node j to node i on layer k indicates that the corresponding entry (i, j, k) in the binary tensor \mathcal{X} is non null. The vertex set is $V = \{v_1, v_2, v_3, v_4\}$. Right column, bottom row: matrix representation of the associated frontal slice of \mathcal{X} .

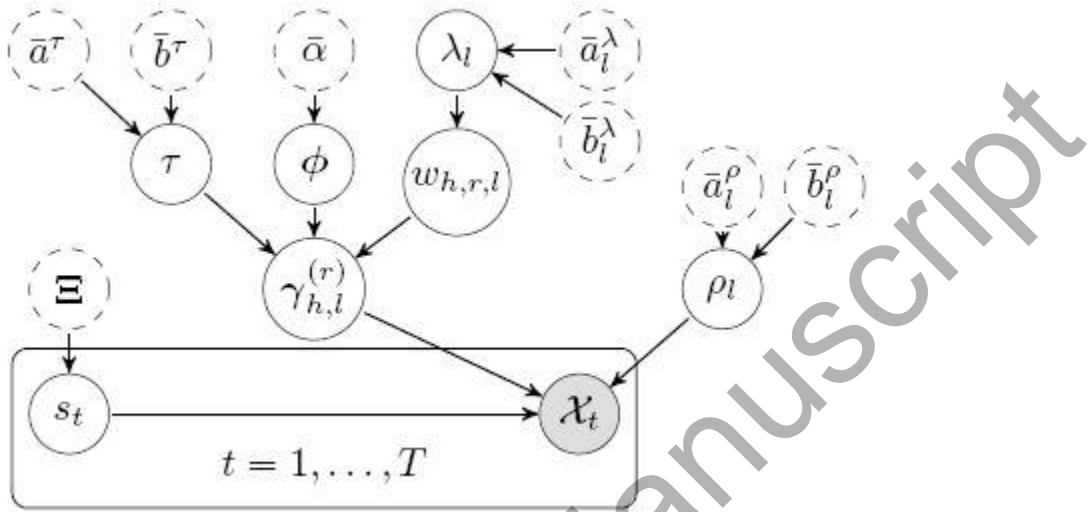


Fig. 3 Directed acyclic graph of the model in eq. (5), (8)-(11). Gray circles, white solid circles, and white dashed circles denote observables, parameters, and hyperparameters, respectively. Directed edges represent the conditional independence relationships.

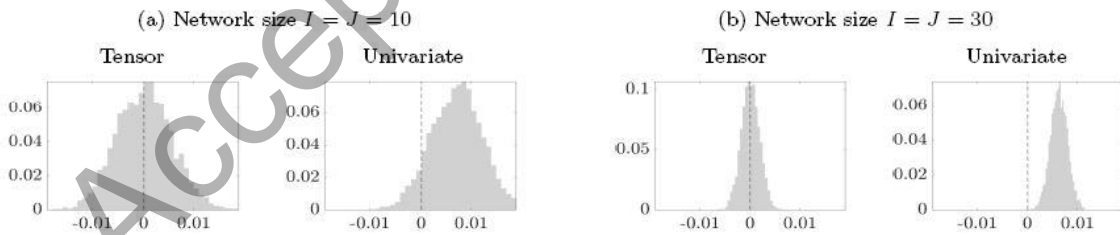


Fig. 4 Posterior predictive checking of the average network density for the ZIL-T-MS model with rank $R = 5$ (tensor) and the MS univariate models (univariate). Performance on simulated datasets of different sizes with $Q = 3$ covariates, $L = 2$ states, and $T = 100$ observations.

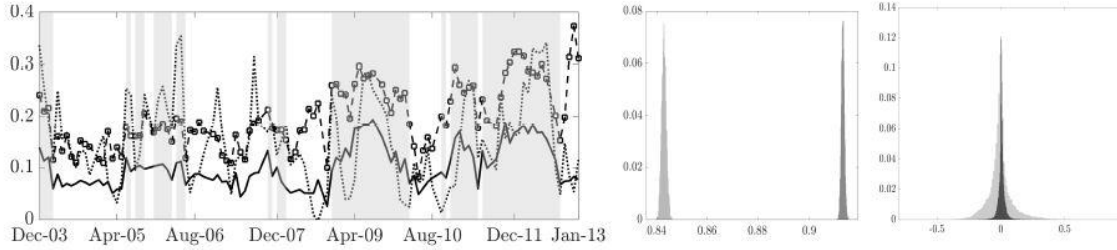


Fig. 5 *Left*: network density (*solid*), average clustering (*dashed*), and minimum eigenvector centrality (*dotted*) of the temporal network, and estimated regimes (*vertical bars*) over time (format *mm-yy*), where gray (white) vertical bars identify the dense (sparse) regime. *Middle*: posterior distribution of the sparsity parameters ρ_1 and ρ_2 . *Right*: histogram of the estimated coefficients across tensor entries. In the middle and right plots, light (dark) gray denotes the dense (sparse) regime.

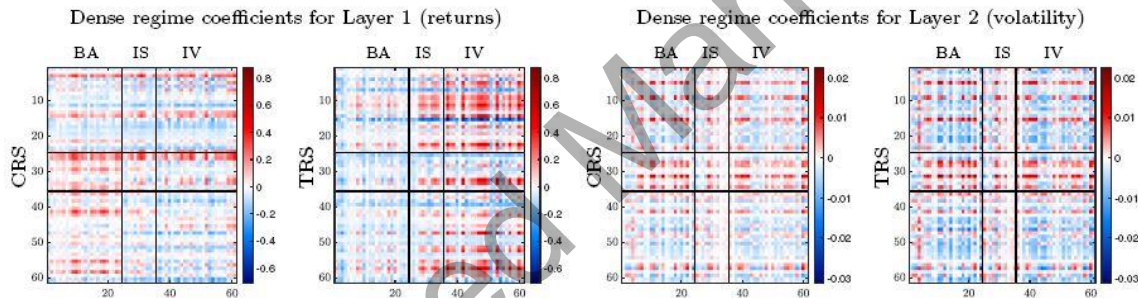


Fig. 6 Posterior mean of the CRS and TRS coefficients, in matricized form, across layers in the dense regime. Entry (i, j) represents the effect of the covariate on the probability of an edge from institution i to institution j . Black lines separate groups of institutions: banks (BA, i and j in $\{1, \dots, 25\}$), insurance (IS, $\{26, \dots, 36\}$), and investment companies (IV, $\{37, \dots, 61\}$). Red, blue and white colors indicate positive, negative and zero valued coefficients, respectively.

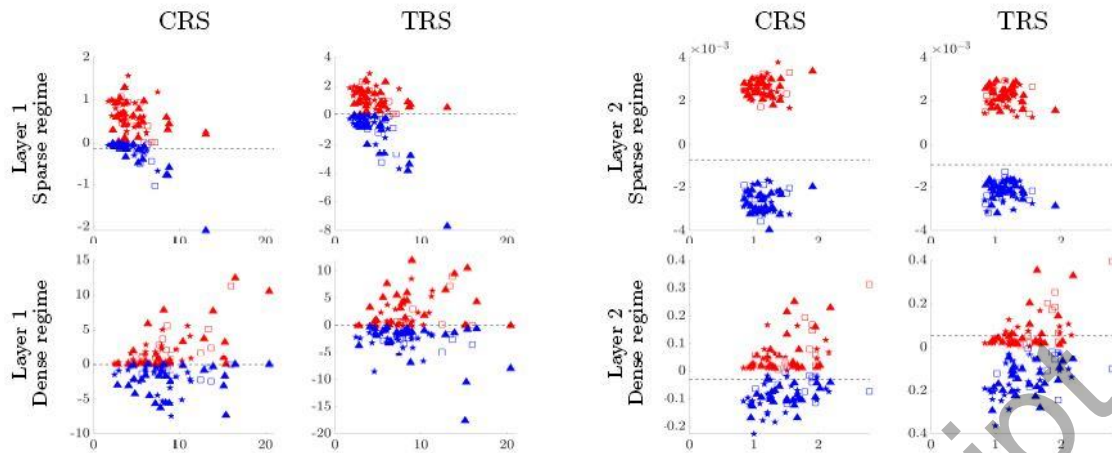


Fig. 7 Covariate coefficients (columns) for the incoming edge probabilities in the two regimes (rows). In each scatterplot: total node degree averaged over time within each regime (horizontal axis) versus the sum of the negative (blue) and positive (red) node coefficients of a given variable (vertical axis). Nodes: banks ($\blacktriangle, \blacktriangle$), insurance companies (\square, \square), and investment companies (\ast, \ast). Dashed line: pooled coefficient estimation.

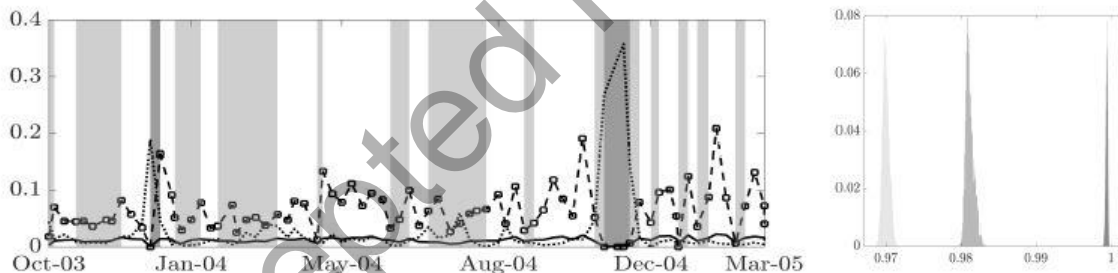


Fig. 8 *Left*: network density (*solid*), average clustering (*dashed*), and minimum eigenvector centrality (*dotted*) of the temporal network in department 3, and estimated regimes (*vertical bars*) over time. *Right*: posterior distribution of the sparsity parameters ρ . Vertical bars and colors identify the regime: dense (white), sparse (light white), very sparse (dark gray).

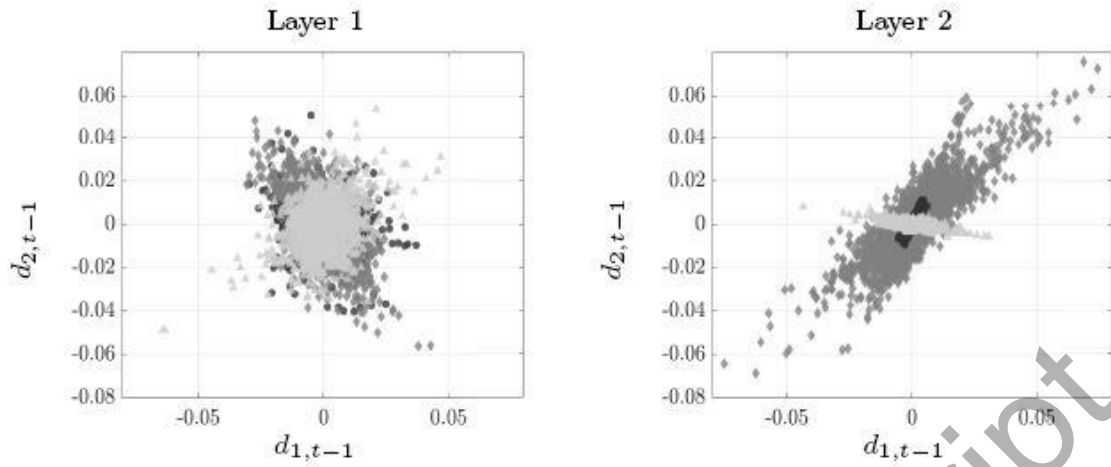


Fig. 9 Coefficients of the first covariate ($d_{1,t-1}$, x -axis) against those of the second one ($d_{2,t-1}$, y -axis) in layer 1 (left) and 2 (right). States: low (1, \circ), regular (2, \diamond), and intense activity (3, \blacktriangle).