

https://dspace.org/

Soft computing for the posterior of a matrix t graphical network

Jason Pillay; Andriette Bekker; Johannes Ferreira; Mohammad Arashi

ELSEVIER SCIENCE INC
https://hdl.handle.net/10539/46306
WOS

Downloaded from a DSpace repository. https://dspace.org/

International Journal of Approximate Reasoning 180 (2025) 109397

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

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

Soft computing for the posterior of a matrix t graphical network

Jason Pillay [a], Andriette Bekker [a,b,*], Johannes Ferreira [a,c], Mohammad Arashi [d]

[a] Department of Statistics, University of Pretoria, Pretoria, South Africa
[b] Centre for Environmental Studies, Department of Geography, Geoinformatics and Meteorology, University of Pretoria, Pretoria, South Africa
[c] School of Statistics and Actuarial Science, University of the Witwatersrand, Johannesburg, South Africa
[d] Department of Statistics, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, Iran

ARTICLE INFO

Keywords:
Adjacency matrix
Bayesian network
Gaussian graphical model
Matrix variate gamma distribution
Matrix variate t
Precision matrix
Stock price data

ABSTRACT

Modeling noisy data in a network context remains an unavoidable obstacle; fortunately, random matrix theory may comprehensively describe network environments. Noisy data necessitates the probabilistic characterization of these networks using matrix variate models. Denoising network data using a Bayesian approach is not common in surveyed literature. Therefore, this paper adopts the Bayesian viewpoint and introduces a new version of the matrix variate t graphical network. This model's prior beliefs rely on the matrix variate gamma distribution to handle the noise process flexibly; from a statistical learning viewpoint, such a theoretical consideration benefits the comprehension of structures and processes that cause network-based noise in data as part of machine learning and offers real-world interpretation. A proposed Gibbs algorithm is provided for computing and approximating the resulting posterior probability distribution of interest to assess the considered model's network centrality measures. Experiments with synthetic and real-world stock price data are performed to validate the proposed algorithm's capabilities and show that this model has wider flexibility than the model proposed by [13].

1. Introduction

1.1. Graphical networks and its reach

The explosion of information and significant advancements in computational speed have led to a notable increase in the capacity to gather substantial volumes of data, commonly called big data. This phrase encompasses a large number of sample observations and features/variables. In the current data-centric world, there is a significant generation and retention of vast quantities of high-dimensional or unstructured data [25]. Scientific progress increasingly relies on data, and researchers frequently perceive themselves as data consumers. The abundance of high-dimensional data presents both prospects and novel obstacles in data analysis, such as the need to investigate the latent patterns within each subset of the data. However, comprehending and deriving significant insights from the data may be challenging. One of the significant issues associated with big data is the introduction and accumulation of noise due to its high dimensionality, which can lead to numerous spurious correlations between features/variables. Big data also proves difficult to visually analyze and interpret succinctly, especially with the growing number of variables included in the exploratory

* Corresponding author.
E-mail address: andriette.bekker@up.ac.za (A. Bekker).

https://doi.org/10.1016/j.ijar.2025.109397
Received 13 February 2025; Accepted 19 February 2025

Available online 24 February 2025
0888-613X/© 2025 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

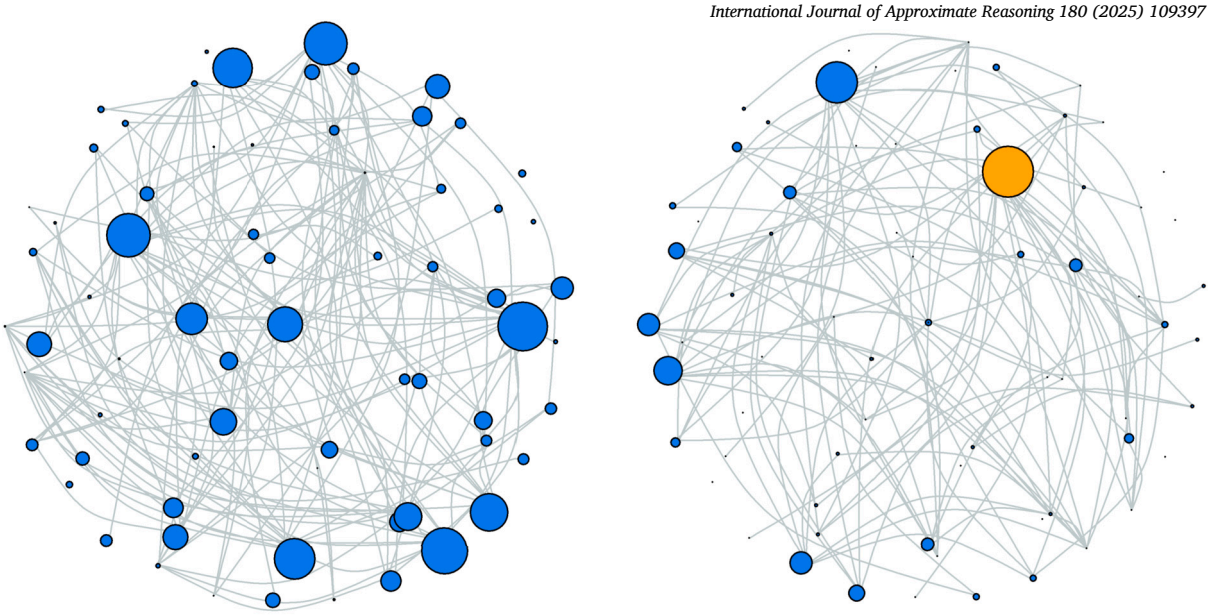


Fig. 1. The network on the left is visualized before denoising. The network on the right is visualized after denoising the network using the proposed method. The big orange dot on the right image is identified as the most influential node in the network. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

analysis and how such variables can influence each other. Naturally, this poses a problem in interpreting and communicating key findings. Hence, the importance of accurate statistical analysis to uncover latent correlations is increasing. For example, extensive structural analytics in high-dimensional big data can be found in [4] and [31].

Network analysis under a statistical framework is an area of research that shows potential as a valuable tool for analyzing big data. This approach is studied in depth in the mathematics domain under the term graph theory - where a graph in this context is synonymous with a network, that is, a collection of nodes and edges representing a graph. Applying graph theory to real scenarios assists in understanding how nodes interact; however, it is not often immediately apparent which components influence others. An example of a graphical network illustrates the visual advantage that graphs offer in high dimensional cases in Fig. 1 and highlights the noise in high-dimensional data. It is a visualization generated from stock price data from 70 European companies, where the blue dots are the so-called nodes in theoretical terms that represent a financial firm. Their size denotes the strength of their importance in the network - how influential they are. The grey lines between them represent the influence or connections between them. While the nodes in this instance denote a firm, nodes can represent any object or state of interest to the researcher.

A graph's information is comprehensively stored in a matrix and thus is typically characterized by the adjacency matrix, whose entries denote the number of edges from one node to another. Thus, estimating a graph itself is equivalent to estimating its adjacency matrix. Additionally, a series of networks is high dimensional and can be depicted as a set of adjacency matrices - paving the way to exploit random matrix theory to model a graphical network for investigating hidden relationships between objects (i.e., nodes). We use a set of nodes and edges for graph representation in probabilistic graphical modeling. To this end, nodes and edges are random variables and probabilistic dependencies, respectively. The relationships between nodes can be represented using adjacency metrics. Therefore, assigning a distribution to a matrix, like the Wishart distribution, is a way of identifying conditional dependencies. For some background on probabilistic graphical models, see [33,34,50].

In conjunction with graph theory, statistics, and machine learning are scientific areas that involve estimating the presence of edges between graph nodes, analyzing estimation errors, and calculating the estimated graph's structure. Evaluating the goodness-of-fit of any chosen edge estimation technique requires finding and evaluating statistics around its resulting error distribution, particularly the spread of errors about the population mean. The error's population covariance matrix is unknown and must, therefore, also be estimated. Frequentist estimation techniques, such as the method of moments, unconditional least squares, and maximum likelihood, exist but depend on the cleanliness of observed data to fill in most of the missing picture. This potentially poses a problem in practice as the variance of real data includes, among others, measurement errors and errors due to rounding. Misleading estimates and centrality measures are inevitable without incorporating additional information about the data's variation *apriori*. In other words, it is useful to have some prior information about the nature of these errors and update them via the information in the data. This realization incentivizes using the Bayesian statistical approach, which will be discussed in Section 4.

1.2. Contributions and outline

This paper proposes an improved model that denoises high-dimensional data. This model is an improvement of the previously considered model by [13] since the constant shape parameter of the Wishart can now assume any integer or non-integer value. We

also considered a prior on this shape parameter with surprising results when applying the centrality measures. We compared our proposal with the best of the existing matrix t graphical network by replacing the Wishart distribution with the more flexible matrix variate gamma distribution. The latter contains the former as a special case. Our theoretical results are thus an extension of the existing approach.

In particular, the contributions of the paper are to:

- Highlight the flexibility of choosing a more general prior distribution via the accuracy of centrality measures. In particular, the observation of the improvement in the accuracy measures, following the Bayesian estimation inference with matrix variate gamma and inverse matrix variate gamma as priors for covariance matrices;
- Scrutinise via detailed simulation experiments through a soft computing lens, representative of our empirical application, to better understand the best approach needed to improve estimation accuracy, and
- Extend the framework of denoising financial data in the context of network theory where the assumption of normality in the model is no longer sufficient to explain the variation.

Therefore, this paper combines the theory of probability, graphs, and the descriptive properties of matrices to append the literature surrounding denoising graphical network data. We should emphasize that there exist only two constructions for the matrix t graphical network: 1) using the Wishart distribution and 2) using that of the matrix gamma [26]. Thus, the only competitor for our proposal is the Wishart distribution, for which we extensively conducted numerical analysis to show our proposed model's superiority. Experimental results show that our proposed framework outperforms the work of [13].

The remainder of the paper is organized as follows: contextualizing literature is provided in Section 2 for a better context of the research's perspective. Section 3 offers some preliminary theory on the intersection between graph theory and multivariate distribution theory. Section 4 outlines and constructs the new matrix t graphical network. The computational algorithm to estimate the model is described in Section 5, followed by an illustration of the suggested methodology via an application to real-world stock price data and a simulation study. Section 6 concludes the paper.

2. Contextualizing literature

Within the global financial system, different parts of the financial world rely significantly on each other. It is demonstrated by [12] that network models explain how different items are connected in financial settings. Using a network model, the dynamics of crypto assets between different bitcoin exchanges were explored by, for example, by [22] and [17]. More recently, [3] evaluated, from a sustainable finance viewpoint, a network-based machine learning model implemented in a fintech platform aiming to assign credit ratings.

The approach of Bayesian graphical modeling in a variety of financial settings has been investigated, see [23,21,8,12]. [11] inferred that firm-level centrality does not correlate with market values and is positively linked to realized financial losses using probabilistic graphical models. By applying Bayesian time-varying networks, [14] detected different connectivity regimes and uncovered the role of covariates in the edge-formation process for financial networks. This paper focuses on Bayesian estimation for adjacency matrices and investigates the effect of prior information on its performance, as illustrated through network centrality measures. Particular attention is paid to the spread of error terms after a probabilistic graphical model fit and how it influences the estimated network's structure. Hence, this paper provides an enhanced Bayesian approach to estimating the noisy connectivity structure.

We propose a new matrix t graphical network that emanates from the two role players in the latest onset of the proposed superior performance of modeling the noise: the matrix variate gamma and the inverse matrix variate gamma distributions; this paper offers an extended view of the contribution as in [9]. The matrix variate gamma distribution, defined by [39], has the Wishart distribution as a nested model. This matrix variate gamma distribution received attention in the literature by, amongst others, [32,19,35,41]; for damping modeling and diffusion MRI [1,46]; for accurate signal reconstruction [42,29,16]; for image classification, see [40]. The interested reader is referred to [26] and [27] and references therein.

3. Preliminaries

In this paper, let G_t be a sequence of networks for $t = 1, \dots, \mathrm{T}$ for $\mathrm{T} \in \mathbb{N}$. Assume that there are a fixed number of nodes n concerning t, but the number of edges can change with respect to t. Assume further that each node bears a stationary time series of variables that characterizes said node. Using the nodes' time series, a sequence of networks G_t is estimated to represent the relationships between the nodes at time t. This also means an adjacency matrix is estimated for G_t at each time index t, say $\boldsymbol{Y}_t$. The inter-nodal relationships are represented by the graph edges. A node u that shares a relationship with another node v at time t results in a directed edge from u to v at time t so that the u, v^{th} entry of $\boldsymbol{Y}_t$ is non-zero. In practice, it would be redundant to consider an edge that loops around a node (i.e., an edge from node a to node a) as it would mean the node shares a relationship with itself. Hence, self-loops are excluded from this consideration. A stationary time series implies that the network structure at time t is a deviation from an underlying adjacency matrix $\boldsymbol{B}$, independent of t. It is then reasonable to view $\boldsymbol{Y}_t$ as 'noisy copy' of $\boldsymbol{B}$ given by:

$$\boldsymbol{Y}_t = \boldsymbol{B} + \boldsymbol{E}_t \quad \text{for } t = 1, \dots, \mathrm{T}, \tag{1}$$

where $\boldsymbol{E}_t$ is an $n \times n$ random error term, independent and identically distributed for all $t = 1, \dots, \mathrm{T}$. Ideally, $\boldsymbol{Y}_t$, $\boldsymbol{B}$, and $\boldsymbol{E}_t$ would all be matrices with discrete entries since they constitute an adjacency matrix from (1). Network estimation techniques rely on hypothesis

tests or suitable threshold values, and the likelihood of edges bases itself on probability, which is continuous and subject to additional error. Thus, a researcher must manually binarise it to resemble an adjacency matrix through a suitable transformation rule. It is clear that the ideal case is often unrealistic, and $\boldsymbol{E}_t$ would have to account for the errors in likelihood testing before the binarization - $\boldsymbol{E}_t$ must have continuous entries. Continuous error random variables should be centered around zero, assuming the model fitted explains the underlying pattern adequately. Alternative distributions explored in the surveyed literature focused more on skewness in univariate cases (see [7] for a comprehensive exploration). Observed data may contain additional noise, but there is insufficient motivation to consider the possibility that observations trail off too sharply to one region than others. Hence, unless there is intrinsic skewness to the data (e.g., discrete errors), a symmetric spread about zero is reasonable to assume.

4. Bayesian update of the matrix t graphical network

This Section provides a concise overview of the processes involved in edge determination and estimation, followed by an in-depth exploration of noise in graphical network data. Subsequently, it describes how this noisy data can be effectively incorporated into a Bayesian inference framework for practical application.

Extensive literature exists on a multivariate Gaussian distribution of errors. Articles that date back as early as the classical linear models [6] to relatively recent ones on engineering processes [5] provide an idea of how long researchers investigated the implications of this assumption. However, normality is useful when errors result from complete, unexplainable variation. That is, only randomness remains after a model's fit. However, data may also exhibit excess kurtosis (or eigenkurtosis), which implies that the observations have a large spread from their location. These factors suggest that the probabilistic nature of errors is symmetric but has excess kurtosis compared to normality [45]. Considering this additional error, the distribution of error terms is expected to be symmetric and have heavier tails. The latter is a meaningful and practical characteristic not captured by the normal distribution and motivates the practical advantages that would be inherited from the theoretical departure of this research.

A t-distribution thus seems a suitable choice to characterize error. Subsequently, $\boldsymbol{E}_t$ in (1) has a matrix variate t-distribution, $\boldsymbol{E}_t \sim t_{n,n}(\boldsymbol{0}, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2)$, with corresponding probability density function (pdf) [24]:

$$f(\boldsymbol{E}_t) = \frac{\Gamma_n(\frac{\nu+2n-1}{2})}{\pi^{(\frac{n^2}{2})}\Gamma_n(\frac{\nu+n-1}{2})} |\boldsymbol{\Sigma}_1|^{-\frac{n}{2}} |\boldsymbol{\Sigma}_2|^{-\frac{n}{2}} |\boldsymbol{I}_n + \boldsymbol{\Sigma}_1^{-1} \boldsymbol{E}_t \boldsymbol{\Sigma}_2^{-1} \boldsymbol{E}_t'|^{-\frac{\nu+2n-1}{2}}, \tag{2}$$

where $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2$, both $n \times n$ are positive definite matrices, denoted as $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2 > \boldsymbol{0}$. Throughout, $\boldsymbol{I}_n$ denotes the identity matrix of dimension n, $\boldsymbol{0}$ denotes the matrix of zeros respectively with size determined by the context, and $|\cdot|$ denotes the determinant. $\Gamma_p(\cdot)$ is the multivariate gamma function defined as

$$\Gamma_p(z) = \pi^{\frac{p(p-1)}{4}} \prod_{i=1}^{p} \Gamma\left(z - \frac{1}{2}(i-1)\right) \tag{3}$$

where p is a natural number and $\Gamma(\cdot)$ is Euler's gamma function. By the linearity property of a matrix variate t random matrix and since $\boldsymbol{B}$ is a constant matrix, equation (1) implies that $\boldsymbol{Y}_t \sim t_{n,n}(\nu, \boldsymbol{B}, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2)$ and consequently called the matrix variate t-model. Notice that $\boldsymbol{B}, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2$ and ν are unknown, which requires their estimation. Making use of Bayes' theorem, a posterior distribution of $\boldsymbol{B}, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2$, and ν is derived from which a sample can be simulated to estimate $\boldsymbol{B}$ using the method of moments estimator (mme) - this is explored in the next subsection.

4.1. Prior specification

Recall that model (1) must account for errors robustly, i.e., the spread of its error terms. Thus, it is reasonable to assume a Gaussian prior for $\boldsymbol{B}$, while $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are assessed in greater detail. That is, $\boldsymbol{B} \sim N_{n,n}(\boldsymbol{0}, \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2)$ with corresponding pdf [24]:

$$f(\boldsymbol{B}) = \frac{|\boldsymbol{\Omega}_1|^{-\frac{n}{2}} |\boldsymbol{\Omega}_2|^{-\frac{n}{2}}}{(2\pi)^{\frac{(n^2)}{2}}} \text{etr}\left(-\frac{1}{2}\left(\boldsymbol{\Omega}_2^{-1}(\boldsymbol{B})' \boldsymbol{\Omega}_1^{-1}(\boldsymbol{B})\right)\right), \quad \boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2 > \boldsymbol{0}$$

where $\text{etr}(\cdot)$ denotes the exponential trace of a matrix.

Since the expected value of $\boldsymbol{Y}_t$, has to be estimated and is only defined if $\nu > 1$, assume that ν follows a truncated gamma distribution, i.e., $\nu \sim \mathcal{TG}(a_\nu, b_\nu, 1)$, with corresponding pdf:

$$f(\nu) = \left(\frac{1}{1 - F_\nu(1)}\right) \frac{1}{a_\nu^{b_\nu} \Gamma(a_\nu)} \nu^{a_\nu - 1} e^{(-\frac{\nu}{b_\nu})}, \quad \text{for } \nu > 1 \text{ and 0 otherwise,} \tag{4}$$

where $F_\nu(1) = P(\nu \leq 1)$, and $a_\nu, b_\nu > 0$ respectively.

As for the probability distributions for the scale matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$, the literature focused on the Wishart and/or inverse-Wishart as priors. However, more control is added by considering generalized distributions, namely the matrix variate gamma and inverse matrix variate gamma distributions. Recall estimation robustness is characterized by the error term's spread, which is controlled by the scale matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$. This motivates the need to choose prior probability distributions for $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ that provide meaningful information to (1).

Furthermore, a useful prior should also allow the posterior pdf to depend on its information rather than potentially overpowered by the likelihood function. In other words, installing a weight parameter allowing prior information to propagate to the posterior is reasonable. To this end, let $\boldsymbol{\Sigma}_1|\gamma$ follow a matrix variate gamma distribution. That is, $\boldsymbol{\Sigma}_1|\gamma \sim \mathcal{MG}_n(\delta_1, \beta_1, (\gamma\boldsymbol{\Phi}_1)^{-1})$ with corresponding pdf [26]:

$$f(\boldsymbol{\Sigma}_1) = \frac{|(\gamma\boldsymbol{\Phi}_1)^{-\delta_1}|}{\beta_1^{n\delta_1}\Gamma_n(\delta_1)}\text{etr}\left\{-\frac{1}{\beta_1\gamma}\boldsymbol{\Phi}_1^{-1}\boldsymbol{\Sigma}_1\right\}|\boldsymbol{\Sigma}_1|^{\delta_1-\frac{n+1}{2}}, \tag{5}$$

and $\boldsymbol{\Sigma}_2|\gamma$ follow an inverse matrix variate gamma distribution. That is, $\boldsymbol{\Sigma}_2|\gamma \sim \mathcal{IMG}_n(\delta_2, \beta_2, (\gamma\boldsymbol{\Phi}_2)^{-1})$ with corresponding pdf [26]:

$$f(\boldsymbol{\Sigma}_2) = \frac{|(\gamma\boldsymbol{\Phi}_2)^{-\delta_2}|}{\beta_2^{n\delta_2}\Gamma_n(\delta_2)}\text{etr}\left\{-\frac{1}{\beta_2\gamma}\boldsymbol{\Phi}_2^{-1}\boldsymbol{\Sigma}_2^{-1}\right\}|\boldsymbol{\Sigma}_2|^{\delta_2-\frac{n+1}{2}} \tag{6}$$

where $\delta_2 > \frac{n-1}{2}$ and $\beta_2 > 0$. The common scale γ allows the prior dependence of the unconditional distribution of $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2$ to vary. Let $\beta_1 = \beta_2 = \beta$ for $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2$ to explore the effect of its added control. The weight parameter γ is unknown and thus must be estimated. Hence assume $\gamma \sim \mathcal{G}(a_\gamma, b_\gamma)$ with corresponding pdf:

$$f(\gamma) = \frac{1}{b_\gamma^{a_\gamma}\Gamma(a_\gamma)}\gamma^{a_\gamma-1}e^{(-\frac{\gamma}{b_\gamma})}, \tag{7}$$

where $a_\gamma, b_\gamma > 0$ respectively.

Notice that β is also unknown, but requires further discussion. The univariate scale parameter β could either be fixed or assumed unknown and estimated by imposing a prior on it, much like γ. This paper investigates the impact of including this scale parameter β in the set of unknown parameters. At this stage, not much prior information is known about it.

We include the notation $f(\beta)$ to act as a 'skeleton' prior for β, with the case that $f(\beta) = 1$ when β is fixed to a constant value. It is shown in Section 5 how we acquire more information through simulation utilizing the following cases:

Case 1. We fix β to a range of candidate values.
Case 2. To facilitate the transition into a Bayesian approach, we impose a non-informative Jeffrey's prior and examine how much information the likelihood contributes to estimating β.
Case 3. The information from cases 1 and 2 provides information to impose an informative prior - the inverse gamma distribution.

Subsequently, the joint prior pdf is:

$$\begin{aligned}\pi(\boldsymbol{B}, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \gamma, \nu) \propto\ & \gamma^{a_\gamma-1+n(\delta_1+\delta_2)}\nu^{a_\nu-1}\text{etr}\left\{-\frac{1}{2}\text{vec}(\boldsymbol{B})'(\boldsymbol{\Omega}_2\otimes\boldsymbol{\Omega}_1)^{-1}\text{vec}(\boldsymbol{B})\right\} \\ & \times\exp\left\{-\left(\frac{\gamma}{b_\gamma}+\frac{\nu}{b_\nu}\right)\right\}\beta^{-(n\delta_1+n\delta_2)}|\boldsymbol{\Sigma}_1|^{\delta_1-\frac{n+1}{2}}|\boldsymbol{\Sigma}_2|^{\delta_2-\frac{n+1}{2}} \\ & \times\text{etr}\left[-\gamma\left(\frac{1}{\beta}\boldsymbol{\Phi}_1\boldsymbol{\Sigma}_1+\frac{1}{\beta}\boldsymbol{\Phi}_2\boldsymbol{\Sigma}_2^{-1}\right)\right]\times f(\beta).\end{aligned} \tag{8}$$

The useful relationship between the vector variate and the corresponding matrix variate distribution is of value here: a random matrix $\boldsymbol{B}$ $n\times n$ is said to be matrix variate normal distributed with parameters $\boldsymbol{0}, \boldsymbol{\Omega}_1 > \boldsymbol{0}$, and $\boldsymbol{\Omega}_2 > \boldsymbol{0}$ denoted as $\boldsymbol{B} \sim N_{n,n}(\boldsymbol{0}, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) \iff \text{vec}(\boldsymbol{B}) \sim N_{nn}(\text{vec}(\boldsymbol{0}), \boldsymbol{\Omega}_2\otimes\boldsymbol{\Omega}_1)$, where $\otimes$ denotes the Kronecker product. That is, $\text{vec}(\boldsymbol{B})$ is a parsimonious reparameterization of $\boldsymbol{B}$ that still contains all the distributional properties of its matrix predecessor. Thus, where possible, $\text{vec}(\boldsymbol{B})$ is used.

4.2. Posterior approximation

Denote the random sample of estimated adjacency matrices and the collection of unknown parameters respectively, as follows:

$$\boldsymbol{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_\text{T}). \tag{9}$$

$$\boldsymbol{\theta} = (\boldsymbol{B}, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \gamma, \nu, \beta). \tag{10}$$

Then, the likelihood function has the following form:

$$\prod_{t=1}^{\text{T}}\frac{\Gamma_n(\frac{\nu+2n-1}{2})}{\pi^{(\frac{n^2}{2})}\Gamma_n(\frac{\nu+n-1}{2})}|\boldsymbol{\Sigma}_1|^{-\frac{n}{2}}|\boldsymbol{\Sigma}_2|^{-\frac{n}{2}}|\boldsymbol{I}_n+\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{Y}_t-\boldsymbol{B})\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{Y}_t-\boldsymbol{B})'|^{-\frac{\nu+2n-1}{2}}. \tag{11}$$

Now that the prior pdf (8) and the likelihood function (11) are established, the posterior pdf $f(\boldsymbol{\theta}|\boldsymbol{Y})$ is proportional to:

$$\prod_{t=1}^{\text{T}}\frac{\Gamma_n(\frac{\nu+2n-1}{2})}{\pi^{(\frac{n^2}{2})}\Gamma_n(\frac{\nu+n-1}{2})}|\boldsymbol{\Sigma}_1|^{-\frac{n}{2}}|\boldsymbol{\Sigma}_2|^{-\frac{n}{2}}|\boldsymbol{I}_n+\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{Y}_t-\boldsymbol{B})\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{Y}_t-\boldsymbol{B})'|^{-\frac{\nu+2n-1}{2}}$$

$$\times \gamma^{a_\gamma-1}\nu^{a_\nu-1}\text{etr}\left[-\frac{1}{2}\left(\text{vec}(\boldsymbol{B})'(\boldsymbol{\Omega}_2\otimes\boldsymbol{\Omega}_1)^{-1}\text{vec}(\boldsymbol{B})\right)\right]\exp\left[-\left(\frac{\gamma}{b_\gamma}+\frac{\nu}{b_\nu}\right)\right]$$
$$\times\exp\left\{-\left(\frac{\gamma}{b_\gamma}+\frac{\nu}{b_\nu}\right)\right\}\beta^{-(n\delta_1+n\delta_2)}|\boldsymbol{\Sigma}_1|^{\delta_1-\frac{n+1}{2}}|\boldsymbol{\Sigma}_2|^{\delta_2-\frac{n+1}{2}}$$
$$\times\text{etr}\left[-\gamma\left(\frac{1}{\beta}\boldsymbol{\Phi}_1\boldsymbol{\Sigma}_1+\frac{1}{\beta}\boldsymbol{\Phi}_2\boldsymbol{\Sigma}_2^{-1}\right)\right]\times f(\beta). \tag{12}$$

Notice (12) cannot be represented in closed form. Hence, pursuing an analytical expression for the expected value of $\boldsymbol{\theta}$ is not feasible (see also [38]). Instead, this paper follows the strategy that [13] and [48] employed, considering the data-augmented approach that approximates the likelihood function and produces closed-form full conditional posterior pdfs.

4.2.1. Augmented likelihood

The following proposition introduces a tractable likelihood function.

Proposition 4.1. *Let* $\boldsymbol{W}_t\sim\mathcal{IMG}_n(\frac{\nu+n-1}{2},2,\boldsymbol{\Sigma}_1)$ *and* $\boldsymbol{Y}_t|\boldsymbol{W}_t\sim N_{n,n}(\boldsymbol{B},\boldsymbol{W}_t,\boldsymbol{\Sigma}_2)$*, then* $\boldsymbol{Y}_t\sim t_{n,n}(\nu,\boldsymbol{B},\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)$.

Denote the collection of random matrices $\boldsymbol{W}=(\boldsymbol{W}_1,\dots,\boldsymbol{W}_\text{T})$. From Proposition 4.1, the augmented likelihood function $L_a(\boldsymbol{\theta}|\boldsymbol{Y})$ is given as:

$$\prod_{t=1}^{\text{T}}\frac{|\boldsymbol{\Sigma}_2|^{-\frac{n}{2}}|\boldsymbol{\Sigma}_1|^{\frac{\nu+n-1}{2}}}{\pi^{\frac{n^2}{2}}\left(2^{\frac{n(\nu+n-1)}{2}}\right)\Gamma_n(\frac{\nu+n-1}{2})}|\boldsymbol{W}_t|^{-\frac{1}{2}(\nu+3n)}$$
$$\times\text{etr}\left(-\frac{1}{2}\left(\boldsymbol{W}_t^{-1}(\boldsymbol{Y}_t-\boldsymbol{B})\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{Y}_t-\boldsymbol{B})'+\boldsymbol{W}_t^{-1}\boldsymbol{\Sigma}_1\right)\right)$$
$$\propto\left[2^{\frac{n(\nu+n-1)}{2}}\Gamma_n\left(\frac{\nu+n-1}{2}\right)\right]^{-n}|\boldsymbol{\Sigma}_2|^{-\frac{\text{T}n}{2}}|\boldsymbol{\Sigma}_1|^{\frac{\text{T}(\nu+n-1)}{2}}\left(\prod_{t=1}^{\text{T}}|\boldsymbol{W}_t|\right)^{-\frac{1}{2}(\nu+3n)}$$
$$\times\text{etr}\left(-\frac{1}{2}\left(\boldsymbol{\Sigma}_2^{-1}\sum_{t=1}^{\text{T}}(\boldsymbol{Y}_t-\boldsymbol{B})'\boldsymbol{W}_t^{-1}(\boldsymbol{Y}_t-\boldsymbol{B})+\boldsymbol{\Sigma}_1\sum_{t=1}^{\text{T}}\boldsymbol{W}_t^{-1}\right)\right). \tag{13}$$

Finally, from (13) and (8) the posterior pdf is approximated by:

$$f(\boldsymbol{B},\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2,\gamma,\nu)\propto L_a(\boldsymbol{\theta}|\boldsymbol{Y})\pi(\boldsymbol{B},\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2,\gamma,\nu)$$
$$\propto\left[2^{\frac{n(\nu+n-1)}{2}}\Gamma_n\left(\frac{\nu+n-1}{2}\right)\right]^{-n}|\boldsymbol{\Sigma}_1|^{\frac{\text{T}(\nu+n-1)}{2}}|\boldsymbol{\Sigma}_2|^{-\frac{\text{T}n}{2}}\left(\prod_{t=1}^{\text{T}}|\boldsymbol{W}_t|\right)^{-\frac{1}{2}(\nu+3n)}$$
$$\times\text{etr}\left(-\frac{1}{2}\left(\boldsymbol{\Sigma}_2^{-1}\sum_{t=1}^{\text{T}}(\boldsymbol{Y}_t-\boldsymbol{B})'\boldsymbol{W}_t^{-1}(\boldsymbol{Y}_t-\boldsymbol{B})+\boldsymbol{\Sigma}_1\sum_{t=1}^{\text{T}}\boldsymbol{W}_t^{-1}\right)\right)$$
$$\times\gamma^{a_\gamma-1}\nu^{a_\nu-1}\text{etr}\left[-\frac{1}{2}\left(\text{vec}(\boldsymbol{B})'(\boldsymbol{\Omega}_2\otimes\boldsymbol{\Omega}_1)^{-1}\text{vec}(\boldsymbol{B})\right)\right]$$
$$\times\exp\left[-\left(\frac{\gamma}{b_\gamma}+\frac{\nu}{b_\nu}\right)\right]\beta^{-(n\delta_1+n\delta_2)}|\boldsymbol{\Sigma}_1|^{\delta_1-\frac{n+1}{2}}|\boldsymbol{\Sigma}_2|^{\delta_2-\frac{n+1}{2}}$$
$$\times\text{etr}\left[-\gamma\left(\frac{1}{\beta}\boldsymbol{\Phi}_1\boldsymbol{\Sigma}_1+\frac{1}{\beta}\boldsymbol{\Phi}_2\boldsymbol{\Sigma}_2^{-1}\right)\right]$$
$$\times f(\beta). \tag{14}$$

Recall that $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ control the spread of $\boldsymbol{E}_t$ but do not control the kurtosis of $f(\boldsymbol{E}_t)$. However, if one of the scale matrices instead varies, it adds variance. Furthermore, if the probabilistic variation of the scale follows an informative distribution, it introduces 'modified copies' of $\boldsymbol{E}_t$ that increase the kurtosis of $f(\boldsymbol{E}_t)$.

4.2.2. Conditional posterior distributions

An observation from the distribution with pdf (14) estimates the adjacency matrix $\boldsymbol{B}$ in (1). However, the expression for the pdf (14) is not in closed form, but most of its full conditional distributions are, which are given below:

1. $f(\nu|\boldsymbol{Y},\boldsymbol{B},\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2)\propto\nu^{a_\nu-1}e^{-\frac{\nu}{b_\nu}}\left(\frac{\Gamma_n(\frac{\nu+2n-1}{2})}{\Gamma_n(\frac{\nu+n-1}{2})}\right)^{\text{T}}$, where

$$\overline{b_\nu} = \left(\frac{1}{b_\nu} + \frac{1}{2} \sum_{t=1}^{\mathrm{T}} ln(|\boldsymbol{I}_n + \boldsymbol{\Sigma}_1^{-1}(\boldsymbol{Y}_t - \boldsymbol{B})\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{Y}_t - \boldsymbol{B})'|) \right)^{-1}. \tag{15}$$

2. $f(\boldsymbol{W}_t|\boldsymbol{Y}, \boldsymbol{B}, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \nu) \propto |\boldsymbol{W}_t|^{-\frac{(\nu+2n-1+n+1)}{2}} \mathrm{etr}\left(-\frac{1}{2} \left((\boldsymbol{Y}_t - \boldsymbol{B})\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{Y}_t - \boldsymbol{B})' + \boldsymbol{\Sigma}_1 \right) \boldsymbol{W}_t^{-1} \right)$,

which is proportional to an inverse matrix variate gamma pdf with scale matrix $\overline{\boldsymbol{W}}$,

degrees of freedom p, where $\overline{\boldsymbol{W}} = (\boldsymbol{Y}_t - \boldsymbol{B})\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{Y}_t - \boldsymbol{B})' + \boldsymbol{\Sigma}_1$, $p = \frac{\nu + 2n - 1}{2}$, and $\beta = 2$. (16)

3. $f(\mathrm{vec}(\boldsymbol{B})|\mathrm{vec}(\boldsymbol{Y}), \mathrm{vec}(\boldsymbol{W}), \boldsymbol{\Sigma}_2 \otimes \boldsymbol{\Sigma}_1) \propto \exp\left(-\frac{1}{2} (\mathrm{vec}(\boldsymbol{B}) - \mathrm{vec}(\boldsymbol{M}))' \overline{\boldsymbol{\Omega}}^{-1} (\mathrm{vec}(\boldsymbol{B}) - \mathrm{vec}(\boldsymbol{M}))' \right)$,

which is proportional to a vector-variate normal pdf with mean $\mathrm{vec}(\boldsymbol{M})$ and covariance matrix $\overline{\boldsymbol{\Omega}}$ where,

$$\overline{\boldsymbol{\Omega}} = \left[\sum_{t=1}^{\mathrm{T}} (\boldsymbol{\Sigma}_2 \otimes \boldsymbol{W}_t)^{-1} + (\boldsymbol{\Omega}_2 \otimes \boldsymbol{\Omega}_1)^{-1} \right]^{-1} \text{ and } \mathrm{vec}(\boldsymbol{M}) = \overline{\boldsymbol{\Omega}} \sum_{t=1}^{\mathrm{T}} (\boldsymbol{\Sigma}_2 \otimes \boldsymbol{W}_t)^{-1} \mathrm{vec}(\boldsymbol{Y}_t). \tag{17}$$

4. $f(\boldsymbol{\Sigma}_1|\boldsymbol{W}, \gamma, \nu, \beta) \propto \mathrm{etr}\left\{ -\frac{1}{\beta} \left(\frac{\beta}{2} \sum_{t=1}^{\mathrm{T}} \boldsymbol{W}_t^{-1}\boldsymbol{\Sigma}_1 + \gamma\boldsymbol{\Phi}_1\boldsymbol{\Sigma}_1 \right) \right\} |\boldsymbol{\Sigma}_1|^{\frac{\mathrm{T}(\nu+n-1)}{2} + \delta_1 - \frac{n+1}{2}}$,

which is proportional to a matrix variate gamma pdf with scale parameters

$\overline{\boldsymbol{\Phi}_1}, \beta$, and shape parameter $\overline{\delta_1}$, where $\overline{\delta_1} = \delta_1 + \frac{\mathrm{T}(\nu + n - 1)}{2}$ and $\overline{\boldsymbol{\Phi}_1} = \left(\frac{\beta}{2} \sum_{t=1}^{\mathrm{T}} \boldsymbol{W}_t^{-1} + \gamma\boldsymbol{\Phi}_1 \right)^{-1}$. (18)

5. $f(\boldsymbol{\Sigma}_2|\boldsymbol{Y}, \boldsymbol{W}, \boldsymbol{B}, \boldsymbol{\Sigma}_1, \gamma) \propto |\boldsymbol{\Sigma}_2|^{-(\frac{Tn+n+1}{2} + \delta_2)} \mathrm{etr}\left\{ -\frac{1}{\beta} \left(\frac{\beta}{2} \sum_{t=1}^{\mathrm{T}} (\boldsymbol{B} - \boldsymbol{Y}_t)' \boldsymbol{W}_t^{-1} (\boldsymbol{B} - \boldsymbol{Y}_t) + \gamma\boldsymbol{\Phi}_2 \right) \boldsymbol{\Sigma}_2^{-1} \right\}$,

which is proportional to an inverse matrix variate gamma pdf with scale parameters $\overline{\boldsymbol{\Phi}}_2, \beta$

and shape parameter $\overline{\delta}_2$ where,

$$\overline{\delta}_2 = \delta_2 + \frac{\mathrm{T}n}{2} \text{ and } \overline{\boldsymbol{\Phi}}_2 = \left(\frac{\beta}{2} \sum_{t=1}^{\mathrm{T}} (\boldsymbol{B} - \boldsymbol{Y}_t)' \boldsymbol{W}_t^{-1} (\boldsymbol{B} - \boldsymbol{Y}_t) + \gamma\boldsymbol{\Phi}_2 \right)^{-1}. \tag{19}$$

6. $f(\gamma|\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) \propto \gamma^{\left(n\frac{\kappa_1 + \kappa_2}{2} + a_\gamma - 1 \right)} \exp\left(-\gamma \left(\frac{1}{b_\gamma} + \frac{1}{2} \mathrm{tr}\left(\boldsymbol{\Psi}_1\boldsymbol{\Sigma}_1 + \boldsymbol{\Psi}_2\boldsymbol{\Sigma}_2^{-1} \right) \right) \right)$,

which is proportional to a gamma pdf with shape and scale parameters $\overline{a_\gamma}$ and $\overline{b_\gamma}$, respectively,

where $\overline{a_\gamma} = n\frac{\kappa_1 + \kappa_2}{2} + a_\gamma$ and $\overline{b_\gamma} = \left(\frac{1}{b_\gamma} + \frac{1}{2} \mathrm{tr}\left(\boldsymbol{\Psi}_1\boldsymbol{\Sigma}_1 + \boldsymbol{\Psi}_2\boldsymbol{\Sigma}_2^{-1} \right) \right)^{-1}$. (20)

For cases 2 and 3, the full conditional posterior pdf of β is given as:

7. $f(\beta|\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \gamma) \propto \beta^{-(n\delta_1 + n\delta_2)} \mathrm{etr}\left\{ -\gamma \left(\boldsymbol{\Phi}_1\boldsymbol{\Sigma}_1 + \boldsymbol{\Phi}_2\boldsymbol{\Sigma}_2^{-1} \right) \frac{1}{\beta} \right\} \times f(\beta)$,

where $f(\beta)$ is the chosen prior. (21)

5. Application and computational experiments

Recall from (1) that $\boldsymbol{B}$ has to be estimated by the expected value of $\boldsymbol{\theta}$ from the posterior pdf in (12). Unfortunately, the latter is not in closed form, so finding an analytic expression for the expected value is not feasible. Instead, we take a sample average from the posterior pdf to estimate $\boldsymbol{B}$. Obtaining a sample from said pdf is accomplished via a soft computing method known as a Gibbs sampler, which is discussed next.

5.1. A Gibbs algorithm

This Section discusses the Gibbs sampler through hierarchical phases. Definitions of model evaluation metrics for diagnostics and comparisons are given in the Appendix.

An observation from the distribution with pdf (14) estimates the adjacency matrix $\boldsymbol{B}$ in (1). As mentioned before, the expression for the pdf (14) is not in closed form, but most of its conditional distributions are, as shown from (15) to (21). Hence, this Section outlines a useful algorithm via a soft computing technique to solve the complex computational problem (14). We approximate inference of the parameters via a Gibbs sampler [44], where Algorithm 1 is described to simulate sample of size n of a collection of observations $\boldsymbol{\theta} = \{\boldsymbol{B}, \boldsymbol{W}_1, \ldots, \boldsymbol{W}_{\mathrm{T}}, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \gamma, \nu\}$ (and $\boldsymbol{\theta} = \{\boldsymbol{B}, \boldsymbol{W}_1, \ldots, \boldsymbol{W}_{\mathrm{T}}, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \gamma, \nu, \beta\}$ for cases 2 and 3). Let $car(\boldsymbol{\theta})$ denote the cardinality of $\boldsymbol{\theta}$.

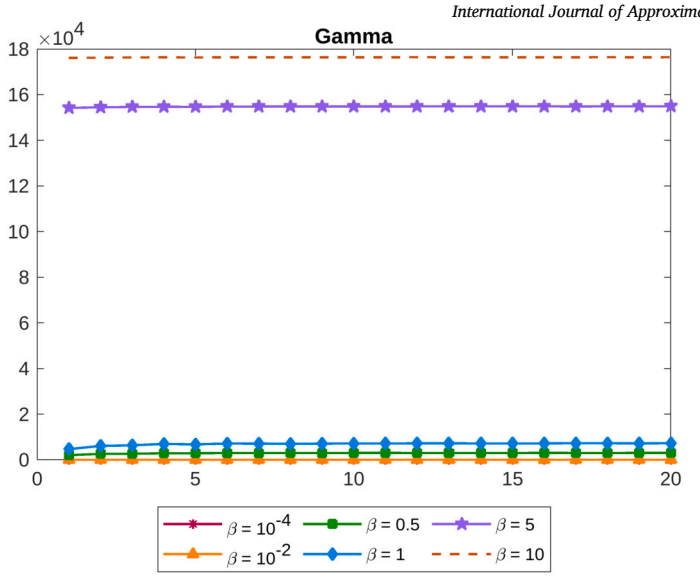


Fig. 2. The lines are the averages of samples obtained from Algorithm 1 for different fixed values of β.

Algorithm 1: Gibbs algorithm.

Initialize parameters, $\boldsymbol{\theta}^{(0)}$. Subscript elements as follows: $\boldsymbol{\theta}^{(0)} = \left(\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_{car(\boldsymbol{\theta})}^{(0)}\right)$.

for $i = 1$ **to** n **do**

  for $j = 1$ **to** $car(\boldsymbol{\theta})$ **do**

    from distribution with full conditional pdfs (15) to (21) :

    Sample $\theta_j^{(i)} \sim f(\theta_j | \boldsymbol{\Theta}^{(i-1)}, \boldsymbol{\Theta}^{(i)})$, where $\boldsymbol{\Theta}^{(i-1)} = \{\theta_k^{(i-1)} : k > j\}$ and $\boldsymbol{\Theta}^{(i)} = \{\theta_k^{(i)} : k < j\}$

  end

  Then $\boldsymbol{\theta}^{(i)} = \left(\theta_1^{(i)}, \theta_2^{(i)}, \ldots, \theta_{car(\boldsymbol{\theta})}^{(i)}\right)$, which is an observation from pdf (14).

end

5.2. Computational specifications

The simulations were run on MATLAB R2022b on the University of Pretoria Stats server. Runtime for simulations was 16 h, excluding time to compute associations between the nodes. Detailed specifications are as follows:

Server details	: Single power Server (SuperMicro, USA).
Cores	: 2 INTEL Skylake-e Central Processors (24 CPU-cores each).
CPU Clock base speed	: 2.6 GHz.
Hardware memory	: RAID-1 configured 800 GB SSD.
LINUX support	: SuSE, SLES, SP15.3.
Workspace support	: RAID-6 to support 10TB (Spindle drives).

5.3. Simulation study

To illustrate the proposed procedure in Section 4, we performed a simulation study. The methodology uses Algorithm 1 to estimate $\boldsymbol{B}$ for degrees of freedom ranging from 1 to 20. An attractive and practically crucial feature of graphical networks is the ability to measure which nodes are the most influential, referred to as centrality measures. Identifying influential nodes is what any further interpretation is built upon. It is reasonable to measure the performance of the proposed method by how close it gets to the true measures of centrality for the most influential nodes relative to how close the noisy data's measures are to the true centrality measures. We consider four popular centrality measures throughout the simulation experiments and in the application section: out-degree, out-closeness, eigencentrality, and betweenness. Their definitions can be found in the Appendix. To investigate the impact and role that β plays, we employ case 1 from Section 4 as a starting point. The methodology is applied with β set to a fixed value. The role of γ was explored by [13]. The following is observed from Figs. 2 and 3:

i. The roles of γ and β control the extent of influence the prior and likelihood information has on the posterior pdf.
ii. Larger values of β cause the likelihood to have a greater influence on the posterior pdf. For instance, refer to Fig. 3. For $\beta = 10$ (the brown-dashed line), the out-degree, out-closeness, and betweenness estimations are guided by the likelihood function almost entirely across the range of degrees of freedom.

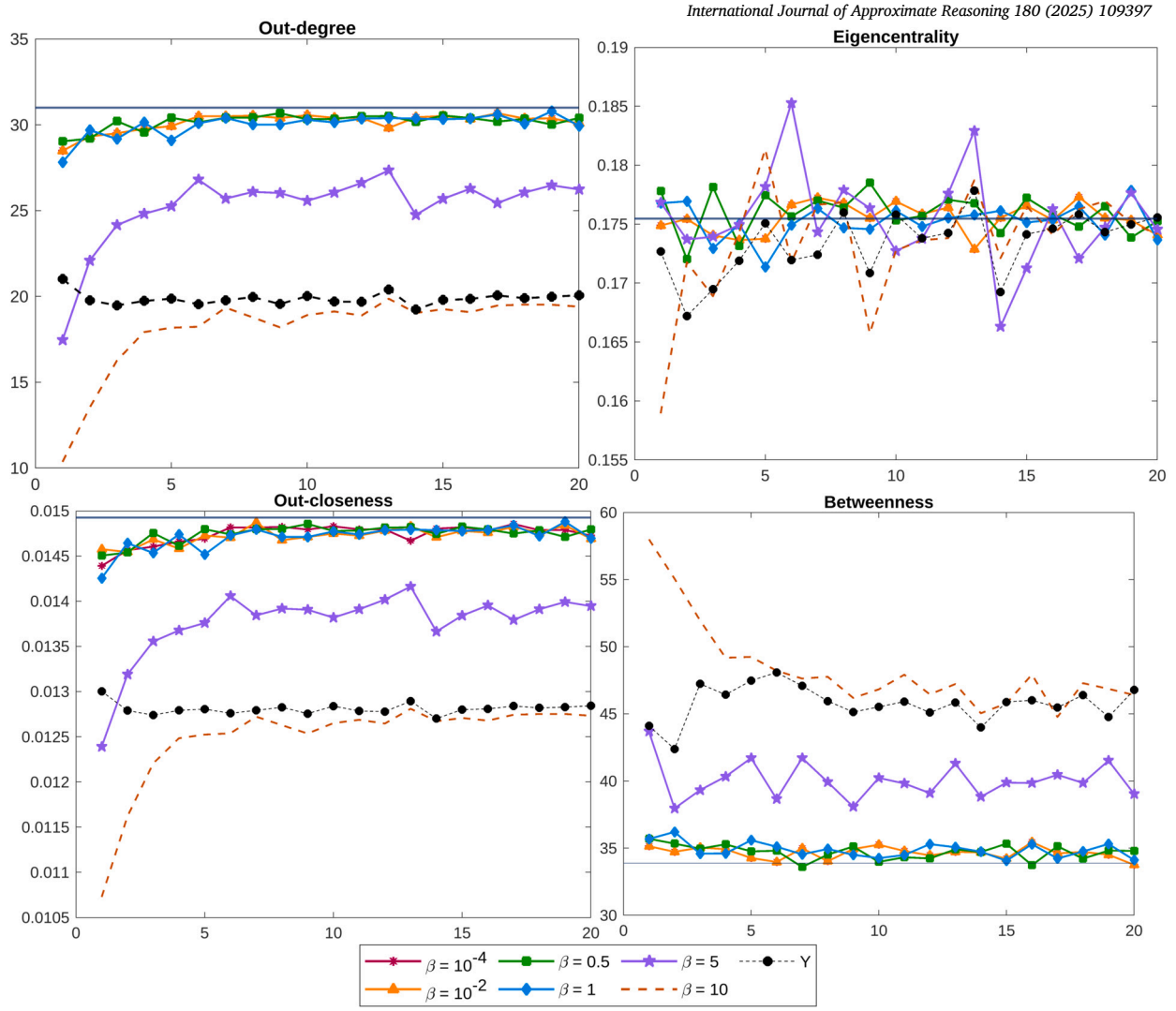


Fig. 3. The solid grey line represents the true value, the black bubble-dashed line represents the raw data averages, and the other marked lines are the averages of samples obtained from Algorithm 1 for different fixed values of β.

iii. Notice from Fig. 2 the average value of γ increases by 10^4 to compensate for the misleading impact $\beta = 10$ has on the estimation accuracy - i.e. there is no clear linear relationship between γ and β.

iv. Notice from Fig. 3 that the difference between the black dotted line and the actual value is large. Recall that the black dotted line represents the average of the raw data before any denoising procedure. This means the out-degree, out-closeness, and betweenness centrality measures are noise-sensitive. Overall, the results highlight how greatly misleading centrality measures could be based on raw data without proper denoising.

Fig. 3 implies that β directly affects the methodology's accuracy. It gives more control to weighing information that influences the updated distributions of scale matrices $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2$. Thus, it would be reasonable to consider it unknown and estimate it by imposing a prior distribution.

This paper now refines this control of β. First, it explores the likelihood's information on β. In other words, it imposes a non-informative prior on β, namely case 2. The strict and technical interpretation of a non-informative prior does not seem to have a consensus among various works in literature. We follow the heuristic approach that a non-informative prior should be invariant under monotone transformation. On this assumption, a suitable non-informative prior is proportional to the square root of Fisher Information, typically denoted $I_F(\beta)$ (otherwise known as Jeffrey's prior [10]). The Fisher information of β is the expected value of the second derivative of the natural log of the preceding pdf (8):

$$I_F(\beta) = -\mathbb{E}\left[\frac{\partial^2}{\partial^2 \beta} ln(\pi(\boldsymbol{\theta}))\right]. \tag{22}$$

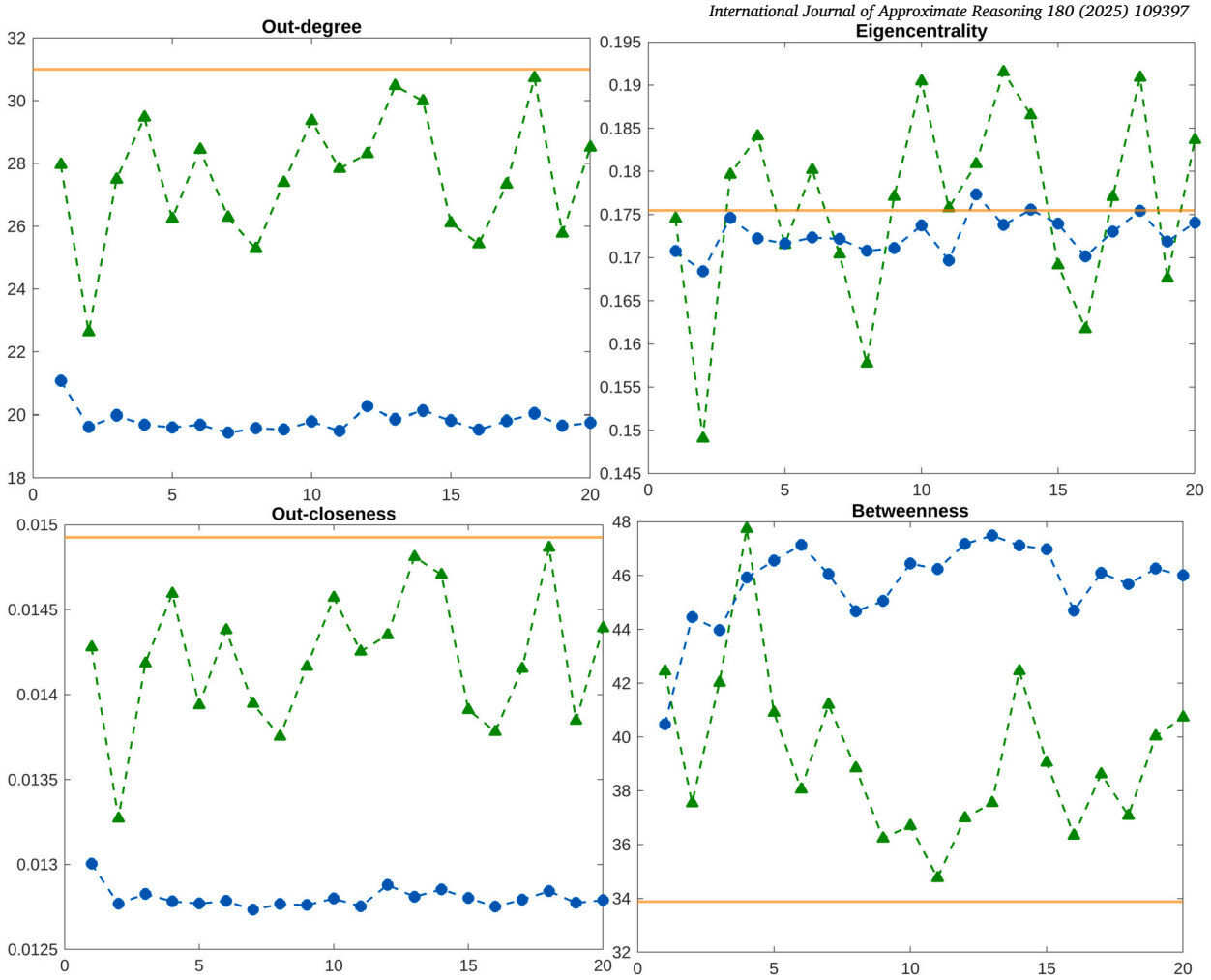


Fig. 4. The horizontal solid line represents the true value, the blue bubble-dashed line represents the raw data averages, and the triangle-dashed lines are the averages of samples obtained from Algorithm 1 using non-informative prior for β (23).

Thus, the non-informative prior pdf of β is:

$$f(\beta) \propto \sqrt{I_F(\beta)}. \tag{23}$$

Including pdf (23) to the prior pdf (8), and using Algorithm 1 with full conditional pdf (21) the results of the methodology are discussed below. The following is observed from Figs. 4 and 5:

i. For the non-informative prior pdf imposed on β, the proposed methodology's estimations on the centrality measures come closer to the true values compared to the data's average. In other words, the likelihood function does contain information to improve accuracy on the out-degree, out-closeness, and betweenness.
ii. In particular, there is an element of flexibility and robustness that a non-informative prior has on β. If prior information is not readily available, a non-informative prior still provides a great improvement to estimation accuracy compared to raw data averages as shown by the graphs in Fig. 4 - The green triangle-dashed line is consistently closer to the true values compared to the blue bubble-dashed line for the out-degree, out-closeness, and betweenness. This observation is consistent with the noise originating from a matrix t-distribution ranging from small to large degrees of freedom (the horizontal axis).
iii. from Fig. 5, there is an apparent direct relationship between the values γ and β assume. That is, an increase of γ occurs when β increases, and a decrease of γ occurs when β decreases.

Note that a non-informative prior did not improve the methodology's performance in estimating a graph's eigencentrality. The same is true for a graph's betweenness with adjacency matrices from a matrix variate t-distribution with degrees of freedom close to 1. Thus, it is worth investigating further the effect of prior knowledge on β. Guided by the form of (21), we impose a conjugate prior pdf known as an inverse gamma distribution with pdf:

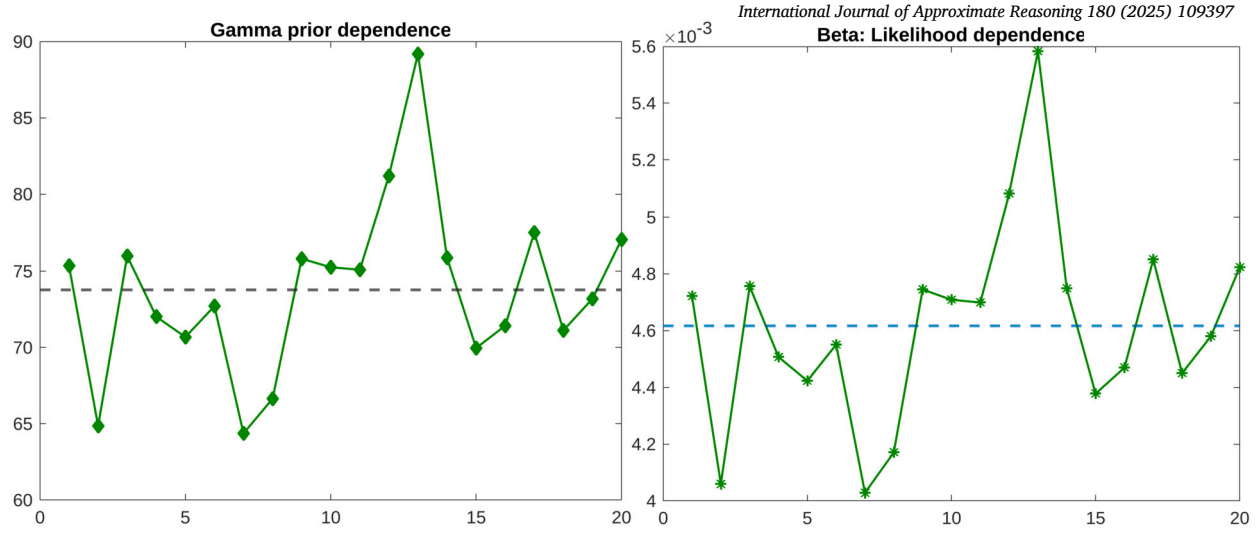


Fig. 5. Averages of γ (left) and β (right) for various degrees of freedom.

$$f(\beta) = \frac{1}{b_{\beta}^{a_{\beta}}\Gamma(a_{\beta})}\beta^{-(a_{\beta}+1)}e^{\left(-\frac{1}{b_{\beta}\beta}\right)}, \tag{24}$$

where the shape and scale parameters $a_{\beta}, b_{\beta} > 0$ respectively. Using pdf (24) to prior pdf (8) and using Algorithm 1 with full conditional pdf (21), the results of the methodology are given below (see Fig. 6). In Fig. 6, notice that the green triangle-dashed lines are much closer to the true values for all degrees of freedom. Notice also how much distance there is between the estimates of the proposed method and the data's average. Thus, there is an impressive improvement after imposing an informative prior on β. Particular attention to the methodology's accuracy concerning eigencentrality and the overall performance for observations from a matrix variate distribution with degrees of freedom close to one. Recall that the lower the degrees of freedom, the larger the kurtosis is for the t-distribution. Despite the excess kurtosis, there is a large difference between the raw data averages and the proposed methodology. From Fig. 7, there is a large difference in the magnitudes of γ and β compared to Fig. 5 - γ does not assume large values to correct the estimation. Hence, an informative prior makes a meaningful difference in how much information it gives to β.

One property to note of the $\mathcal{MG}_n(\delta_1, \beta, \mathbf{\Phi}_1)$ distribution is that if $\delta_1 = \frac{\kappa}{2}$ and $\beta = 2$ then the pdf simplifies to that of a Wishart distribution with κ degrees of freedom and scale matrix $\mathbf{\Phi}_1$. In other words it is $\mathcal{W}_n(\kappa, \mathbf{\Phi})$ distributed. Similarly, if $\delta_2 = \frac{\kappa}{2}$ and $\beta = 2$ then the pdf simplifies to that of an $\mathcal{IW}_n(\kappa, \mathbf{\Phi}_2)$ distribution. Thus, we compare the methodology's performance under the special case $\beta = 2$ versus the case β is estimated. We explore the added benefit that β has on improving the accuracy of the methodology. Fig. 7 shows that $\beta = 2$ is far from a suitable value to assume in the study. The blue-dashed line shows an average β around 5.1×10^{-6}. This motivates that there is added estimation gain to the role of β estimated in the methodology.

Refer to error plots in Fig. 8: The estimated error points for β are closer to the zero reference line among all centrality measures. In other words, the errors are consistently smaller across all centrality measures and for various degrees of freedom when β is estimated compared to assuming the special case that $\beta = 2$. An especially crucial highlight is an impressive error difference for degrees of freedom close to 1. In the presence of increased kurtosis, this extra parameter β improves estimation accuracy.

5.4. Data application

The methodology is applied to the weekly stock prices of 70 European firms from 4 January 2016 to 31 December 2019, resulting in 105 observations per firm. The data is comprised of a range of Global Industry Classification Standard sectors (or GICS sectors), namely: Communication Services (5), Consumer Discretionary (15), Consumer Staples (6), Energy (2), Financials (11), Health Care (6), Industrials (10), Information Technology (5), Materials (2), Food and Beverages (1), Real Estate (3), and Utilities (5). The firms also range over various countries in Europe: Germany (28), France (37), and Italy (5). Recall that the network on the left in Fig. 1 provides the network of this data, i.e., a noisy network. The associations between these firms are quantified pairwise through Granger causality hypothesis tests for the t^{th} week - a popular hypothesis test used in financial settings to test for statically significant associations between two time series [43]. The resulting test statistics belong to the observed $\boldsymbol{Y}_t$ matrix. Network modeling in finance is associated with identifying systemic risk [18]. For example, identifying the most important firms with a high "propagation power" and firms most susceptible to various kinds of economic shock improves the robustness of the financial system of concern. Of course, said identifications need to be accurate for any improvements to materialize. Thus, this Section illustrates that identifying firms with a special characteristic of interest, such as how important it is, how well it acts as a bridge between other firms, and how accessible it is between firms, requires more statistical analysis compared to using the average of a given noisy dataset.

We employ the well-known centrality measures that were considered in the simulation experiments. Comparing how similar or different the scores computed between raw data estimates and that of the proposed methodology can effectively showcase the

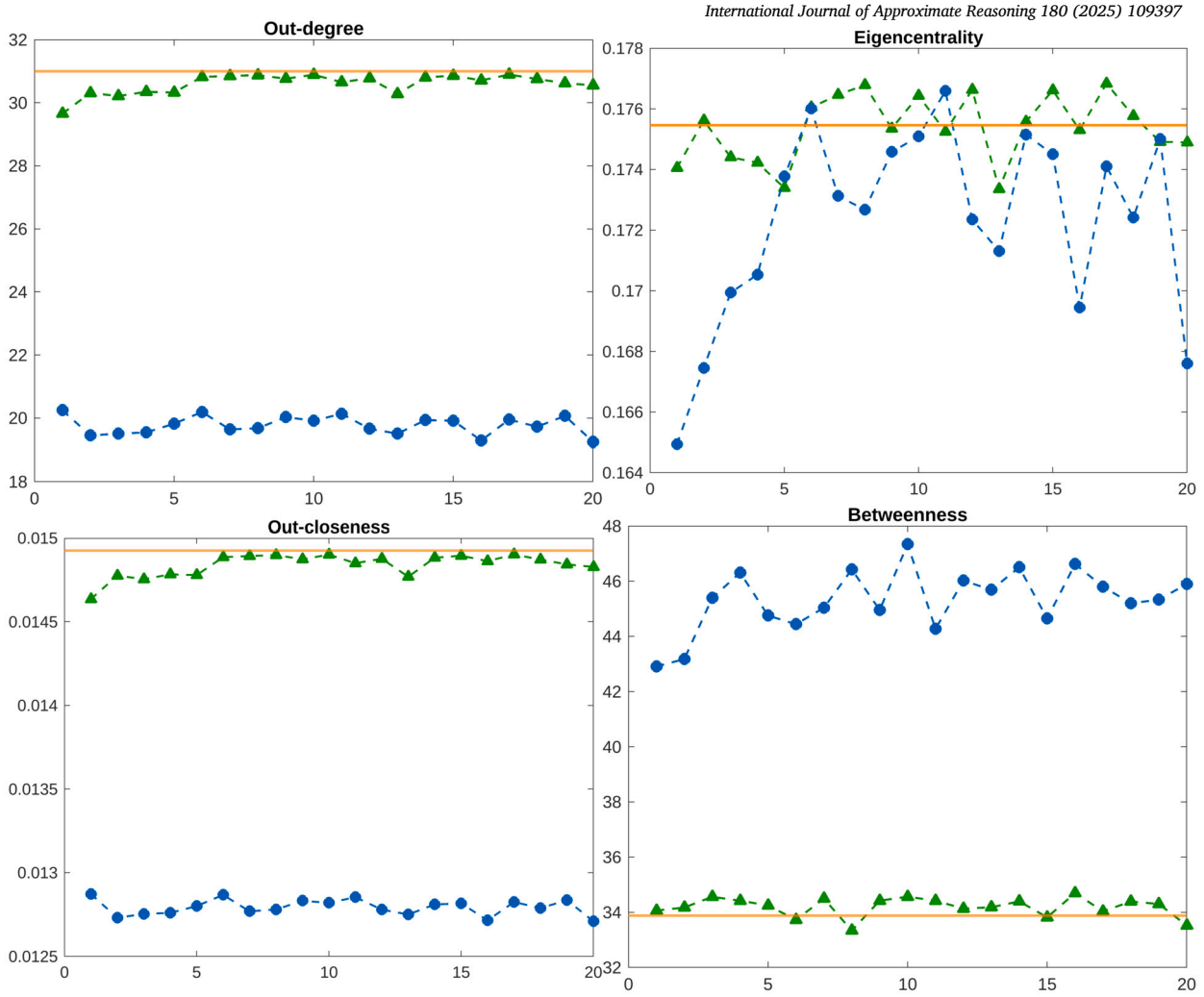


Fig. 6. Estimation results: The horizontal solid line represents the true value, the blue bubble-dashed line represents the raw data averages, and the green triangle-dashed lines are the averages of samples obtained from Algorithm 1 using informative prior (24) for β.

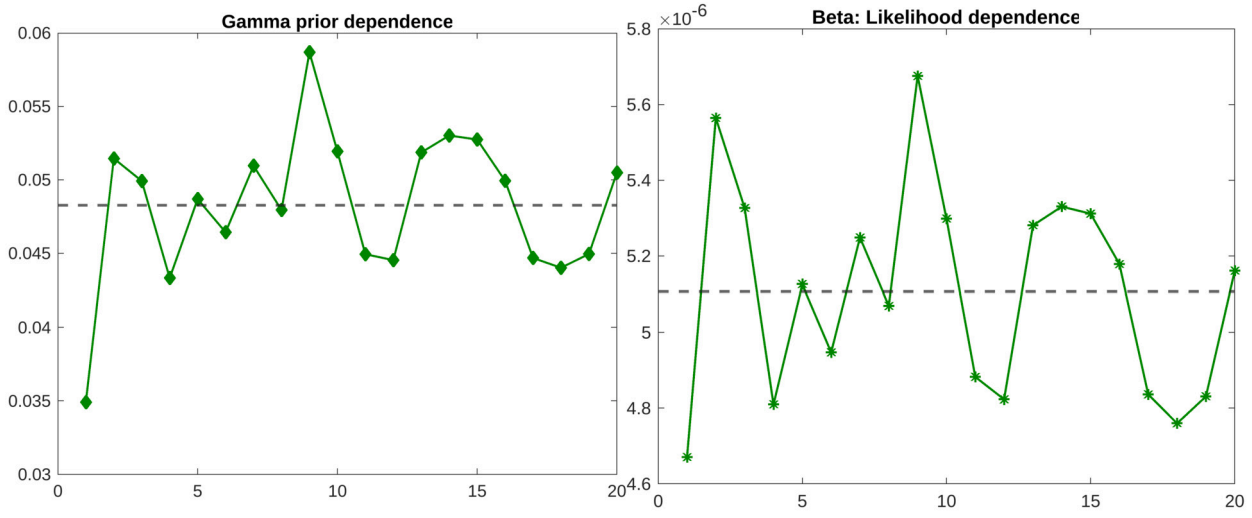


Fig. 7. Averages of γ (left) and β (right) for various degrees of freedom.

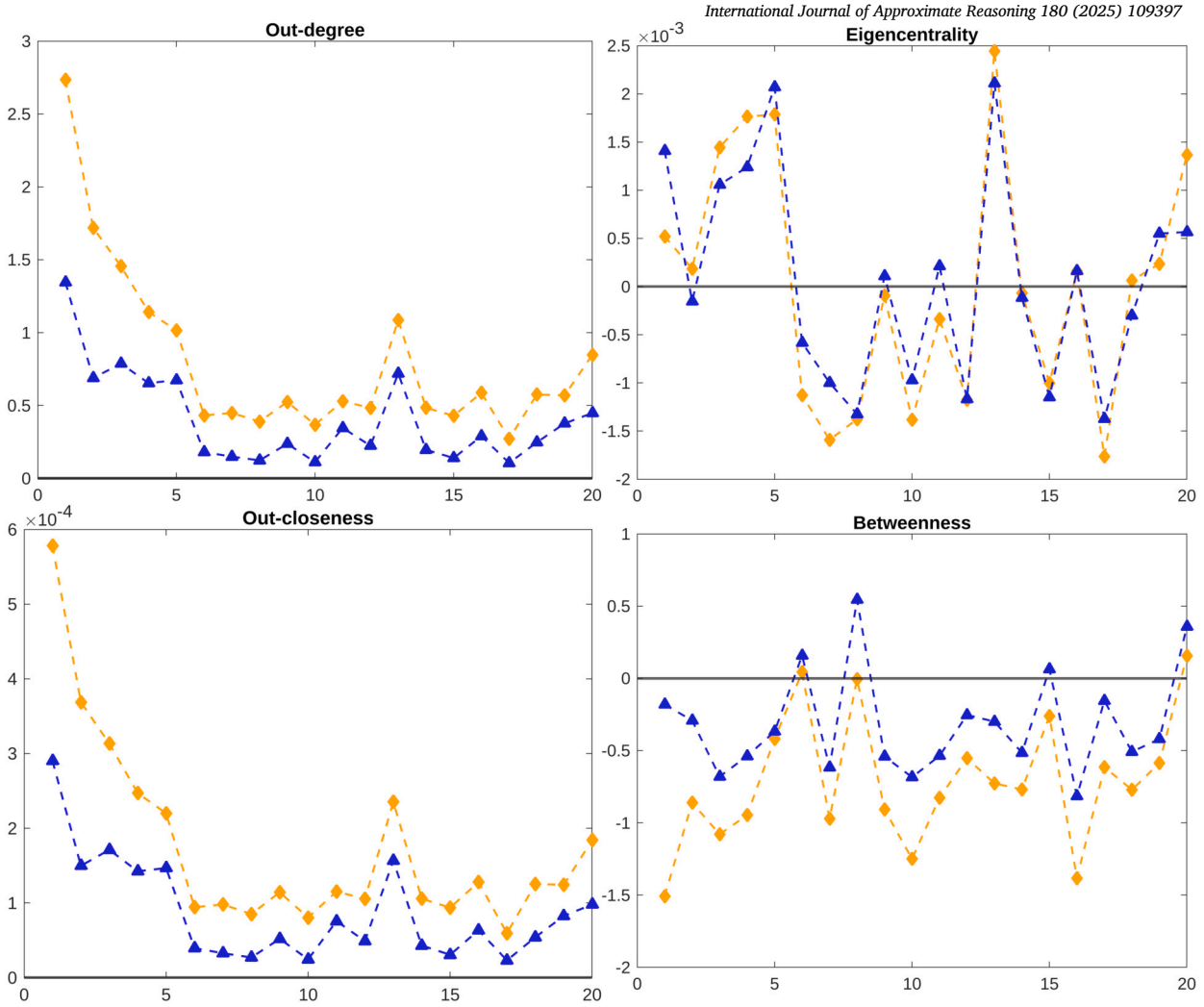


Fig. 8. Estimation error plots: Orange diamond-dashed lines and blue triangle-dashed lines represent errors when $\beta = 2$ and when β is estimated using informative prior pdf (24), respectively.

latter's significance. The results from the application are shown in Fig. 9. The sample obtained from Algorithm 1 has insignificant autocorrelations on a 5% level; see Fig. 10. Thus, it is reasonable to assume convergence of the algorithm [44]. It is observed that there are clear discrepancies between the different estimators, with particular attention to the out-degree, out-closeness, and eigencentrality. The raw data underestimates the general centrality measures. Fig. 9 shows that raw data averages have numerous nodes with similar high scores of each centrality, while the proposed method clearly highlights which node scores the highest. The most notable difference includes the performance of the raw data averages of the eigencentrality against the method's averages. The raw data heavily underestimates the true scores of the most influential nodes and leaves the researcher multiple nodes to consider as most influential (with scores seemingly in clusters). Still, the proposed method singles out the most influential node more clearly. In other words, the discrepancies highlight how noise left in data induces heavy bias in the estimation procedure. If not dealt with, the noise may jeopardize the validity and reliability of analysis built on data.

For example, an influential node according to the raw data concerning eigencentrality could be the 62^{nd} node, and any analysis and decision-making that is built on this will be flawed. The raw data incorrectly identifies the node with the highest degree as the 17^{th} node, whereas the proposed method identifies the 8^{th}. These examples emphasize that, without proper denoising, network analysis is rendered useless. The visualization of the proposed method is now compared to the raw data averages in Fig. 1. Notice in Fig. 1 that the influential nodes according to out-degree are identifiable after the denoising procedure.

6. Conclusion

This article introduces for the first time a flexible prior for the covariance structure of a matrix variate t-distribution for the noise process, following a Gaussian graphical model. A simulation experiment empirically demonstrated the results of the methodology in a way that is representative of the methodology's application to real-life stock data on 70 European firms. The results from the

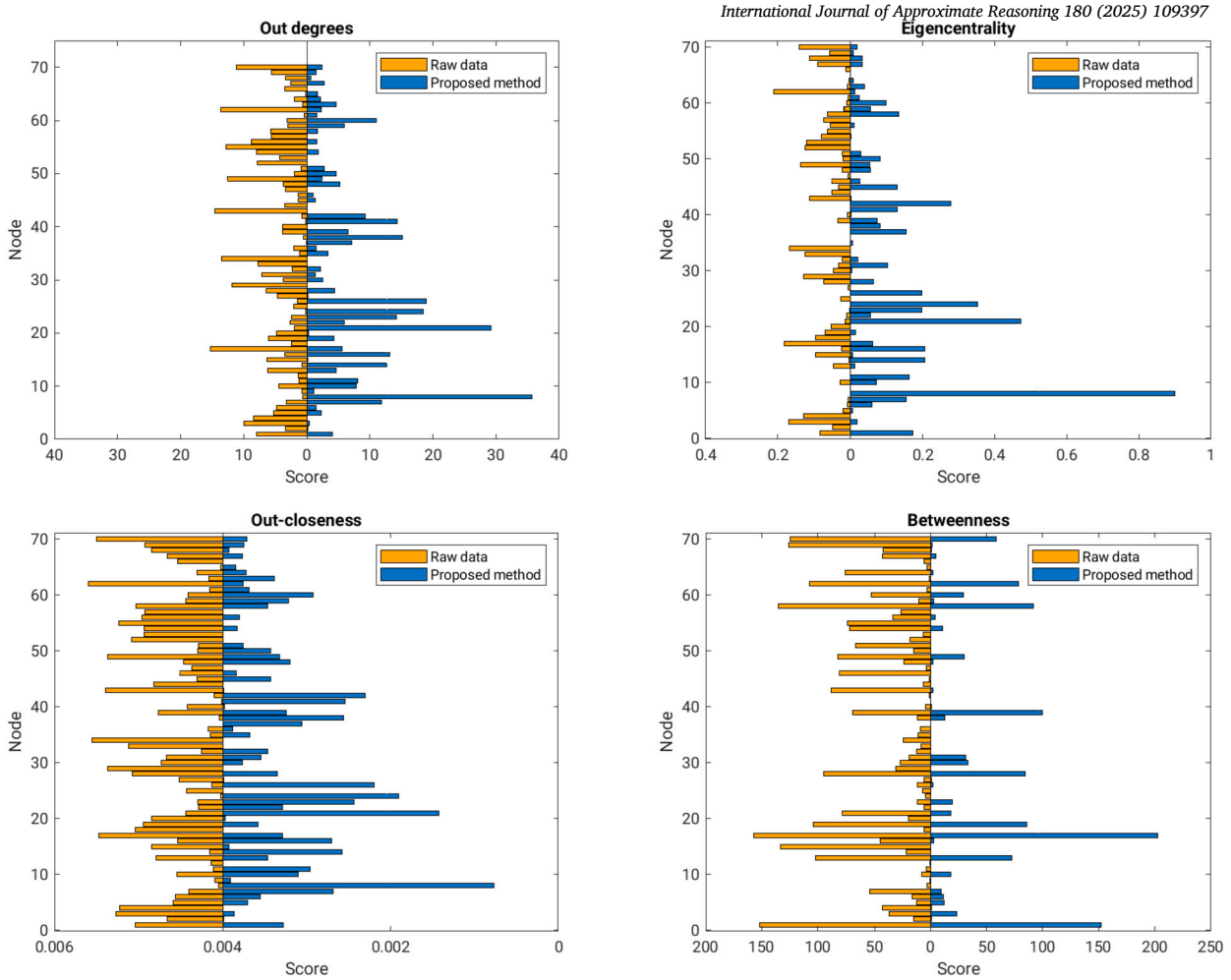


Fig. 9. Estimated centrality measures: The solid red line and dashed black lines represent the raw data averages and methodology, respectively.

simulations and empirical experiments highlight the added flexibility and control that a generalized prior has on the methodology and demonstrate that our proposed framework outperforms the previously considered work. The importance of flexible distributions is often emphasized in the literature [36,30,37]. The effect of prior information on better accommodating the nature of data on modeling variance robustly and flexibly has been overlooked in the literature. This study primarily focuses on the denoising of network data, specifically focusing on financial data. However, its applicability extends to several domains through the utilization of big data analytics. High-dimensional data, such as genome data, within the realm of big data, frequently encounters biases that arise from many unpredictable elements in the design of experiments [28]. Researchers must do extensive experiments to identify significant genes and proteins because of the high dimensionality. Neuroimaging methodologies, such as functional magnetic resonance imaging (fMRI), positron emission tomography (PET), and electrophysiological approaches, provide substantial potential to investigate functional brain networks. Gaining insight into the intricate and operational network architecture of the brain holds significant promise for nonintrusive methods to understand the impact of diseases on the brain [47]. An example of high-dimensional big data, geographical and noisy data is observed in the field of climate studies [15]. The suggested method can potentially improve the modeling of the relationships between areas and their greenhouse gas emissions.

Furthermore, this paper described assumptions about the nature of errors in the data and thus established their characterization and probabilistic nature. It extensively investigated the Bayesian approach to estimate the noisy connectivity structure. A range of centrality measures is applied to the estimated network to summarize the accuracy and robustness of the estimation comprehensively. In particular, it adds to the theoretical toolkit researchers may need for more robust and accurate network estimation. Future studies can extend the denoising method proposed in this paper to consider a generalized scalar Wishart distribution, the generalized diagonal Wishart distribution of [2], or the real version of the Wishart distribution of [20].

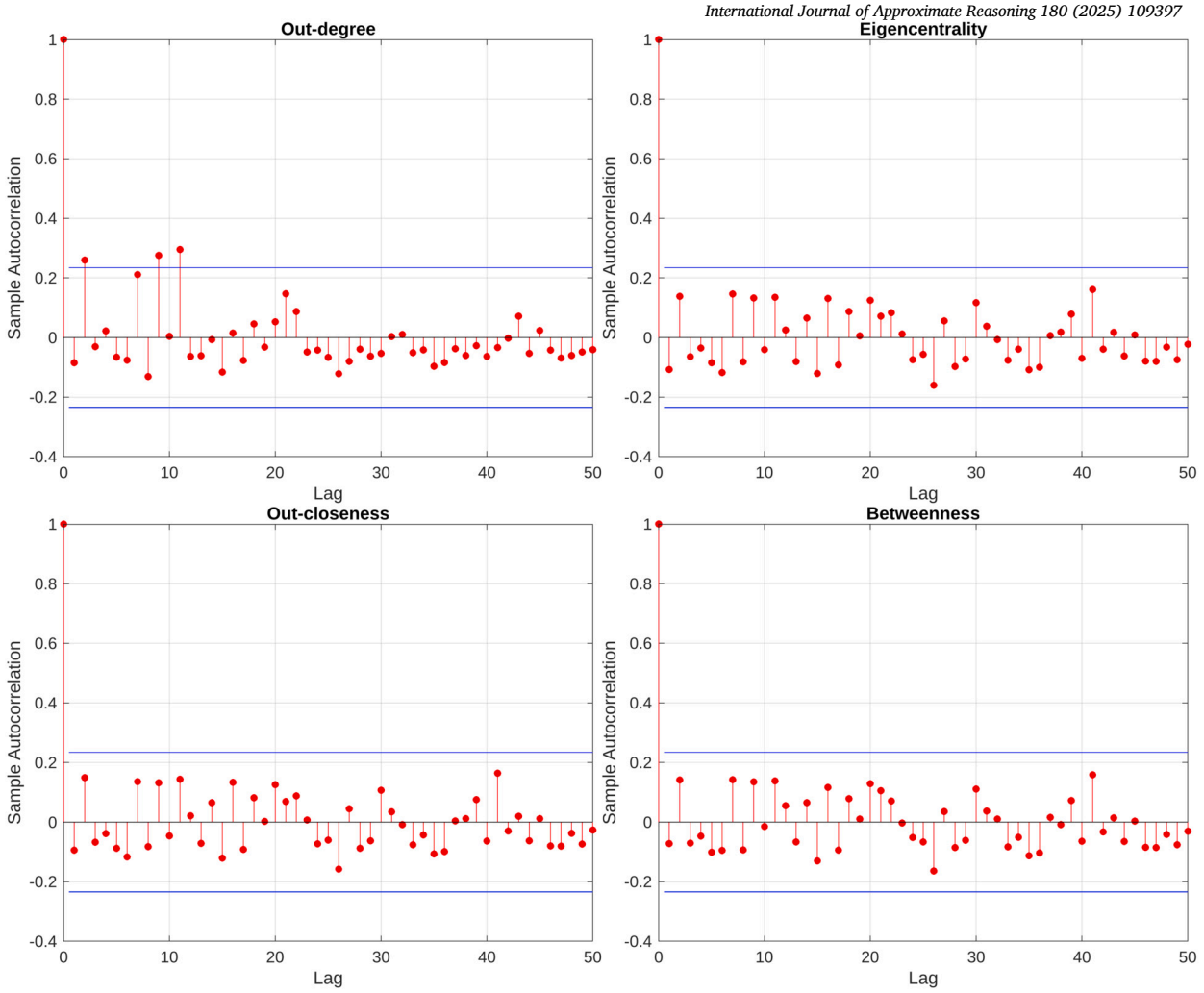


Fig. 10. Convergence check of Gibbs sampler: Autocorrelation plots of averaged centrality measures. The red pointed lines are the sample autocorrelation for each lag, and the horizontal solid blue lines are the 95% confidence intervals for insignificant autocorrelation.

CRediT authorship contribution statement

Jason Pillay: Formal analysis, Investigation, Methodology, Software, Writing – original draft. **Andriette Bekker:** Conceptualization, Formal analysis, Methodology, Project administration, Supervision, Visualization, Writing – review & editing. **Johannes Ferreira:** Formal analysis, Funding acquisition, Investigation, Project administration, Supervision, Writing – original draft, Writing – review & editing. **Mohammad Arashi:** Formal analysis, Investigation, Methodology, Supervision, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Johan Ferreira reports financial support was provided by National Research Foundation, grant RA201125576565 nr 145681. Andriette Bekker reports financial support was provided by National Research Foundation grant SRUG2204203865 nr 120839. Mohammad Arashi reports financial support was provided by National Research Foundation grant RA211204653274 nr 151035. Mohammad Arashi reports financial support was provided by Iran National Science Foundation grant 4015320. The opinions expressed, and conclusions arrived at are those of the authors and are not necessarily to be attributed to the NRF. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank the guest editor and two anonymous reviewers for their comments and expertise in contributing to a revised manuscript. This work was based upon research supported in part by the Centre of Excellence in Mathematical and Statistical Sciences, based at the University of the Witwatersrand (SA). The authors would like to thank Professor Monica Billio, University of Venice, for providing the data used in this paper. The authors acknowledge that a brief overview of preliminary results was published in the Classification and Data Analysis Group Conference Proceedings of 2023 titled 'Bayesian Analysis for a graphical t-model' [9].

Appendix A. Definitions of goodness-of-fit measures

Centrality is a comprehensive way to evaluate the goodness-of-fit of the proposed estimation techniques. No one measure is not a one-size-fits-all, as the interpretation of centrality varies greatly depending on the context of the field of study. Here, the following measures are reviewed to cover a broad perspective of node influence that can be used to evaluate the accuracy of the methodology discussed. For all definitions below [49], consider a graph with corresponding adjacency matrix $\boldsymbol{A}$. Let the ij^{th} entry of $\boldsymbol{A}$ be denoted $a_{i,j}$.

Definition Appendix A.1. Consider the number of edges that leave a node i. The more edges that leave i, the more connected it is to other nodes, thus making it more central in the graph. This is known as a node's out-degree. Mathematically let,

$$d^{(i)} = \sum_{j=1, j\neq i}^{n} a_{i,j}. \tag{A.1}$$

Then $d^{(i)}$ is the out-degree of node i.

Definition Appendix A.2. The path length from node i to node j is defined as the number of edges constituting the shortest path that connects i to j and is denoted as $l(i,j)$. If j is not reachable from i, $l(i,j) := 0$

Definition Appendix A.3. Consider path lengths from node i to nodes $j = 1, \ldots, n, j \neq i$. Heuristically, it is easy to see that a node with shorter lengths to other nodes implies it is easier for said node to reach other nodes. This is known as a node's out-closeness. Mathematically let,

$$c^{(i)} = \left\{ \sum_{j=1, j\neq i}^{n} l(i,j) \right\}^{-1}. \tag{A.2}$$

Then $c^{(i)}$ is the out-closeness of node i. Note that there is a normalized variation of $c^{(i)}$ given as

$$c^{(i)}_{norm} = \frac{R_i}{(n-1)} \left\{ \sum_{j=1, j\neq i}^{n} l(i,j) \right\}^{-1}, \tag{A.3}$$

where R_i is the number of reachable nodes from node i. $c^{(i)}_{norm}$ can be interpreted as the averaged shortest path from node i. The normalized variant is useful for comparing the closeness of nodes across graphs of different sizes and accounting for unreachable nodes.

Definition Appendix A.4. Let $n(i,j)$ denote the number of paths from node i to node j and let $n_v(i,j)$ denote the number of paths from i to j that pass through node v. Then, it is possible to measure the transmittable influence of node v as the proportion number of paths from i to j that passes through v. A large proportion indicates that node v acts as an influential transmitter between nodes. Mathematically let,

$$b^{(v)} = \sum_{i=1, v\notin\{i,j\}}^{n} \sum_{i=1, j\neq i}^{n} \frac{n_v(i,j)/n(i,j)}{(n-1)(n-2)}. \tag{A.4}$$

Then $b^{(v)}$ is the betweenness of node v where $(n-1)(n-2)$ represents the total number of node pairs to sum through.

Definition Appendix A.5. All definitions above measure a node's centrality without considering other nodes' centrality scores. It might be useful to measure how influential a node is relative to others. Hence, denote $x^{(i)}$ the relative influential score of node i and λ the largest eigenvalue of $\boldsymbol{A}$. Then,

$$x^{(i)} = \frac{1}{\lambda} \sum_{j=1}^{n} a_{i,j} x^{(j)}, \tag{A.5}$$

is known as the eigenvector centrality of node i.

Data availability

The authors do not have permission to share data.

References

[1] Sondipon Adhikari, Characterization of uncertainty in damping modeling, in: 48th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, 2007, p. 2151.
[2] Sondipon Adhikari, Generalized Wishart distribution for probabilistic structural dynamics, Comput. Mech. 45 (2010) 495–511.
[3] Daniel Ahelegbey, Paolo Giudici, Valentino Pediroda, A network based fintech inclusion platform, Socio-Econ. Plan. Sci. 87 (2023) 101555.
[4] S.E. Ahmed, Big and Complex Data Analysis: Methodologies and Applications, Springer, 2017.
[5] Amirhossein Amiri, Reza Ghashghaei, Mohammad Reza Maleki, On the effect of measurement errors in simultaneous monitoring of mean vector and covariance matrix of multivariate processes, Trans. Inst. Meas. Control 40 (1) (2018) 318–330.
[6] Steven F. Arnold, Linear models with exchangeably distributed errors, J. Am. Stat. Assoc. 74 (365) (1979) 194–199.
[7] Adelchi Azzalini, The skew-normal distribution and related multivariate families, Scand. J. Stat. 32 (2) (2005) 159–188.
[8] Matteo Barigozzi, Christian Brownlees, Nets: network estimation for time series, J. Appl. Econom. 34 (3) (2019) 347–364.
[9] Andriette Bekker, Johannes Ferreira, Jason Pillay, Mohammad Arashi, Bayesian analysis for a graphic t-model, in: Proceedings of the 14th Meeting of Cladag, September 2023, pp. 347–350.
[10] James O. Berger, José M. Bernardo, On the development of the reference prior method, Bayesian Stat. 4 (4) (1992) 35–60.
[11] Daniele Bianchi, Monica Billio, Roberto Casarin, Massimo Guidolin, Modeling systemic risk with Markov switching graphical SURs models, J. Econom. 210 (1) (2019) 58–74.
[12] Monica Billio, Mila Getmansky, Andrew W. Lo, Loriana Pelizzon, Econometric measures of connectedness and systemic risk in the finance and insurance sectors, J. Financ. Econ. 104 (3) (2012) 535–559.
[13] Monica Billio, Roberto Casarin, Michele Costola, Matteo Iacopini, A matrix-variate t model for networks, Front. Artif. Intell. 4 (2021) 49.
[14] Monica Billio, Roberto Casarin, Matteo Iacopini, Bayesian Markov-switching tensor regression for time-varying networks, J. Am. Stat. Assoc. (2022) 1–13.
[15] C. Bai, L. Zhou, M. Xia, C. Feng, Analysis of the spatial association network structure of China's transportation carbon emissions and its driving factors, J. Environ. Manag. (2020) 253.
[16] Emmanuel Caruyer, Numerical methods for white matter microstructure and brain structural connectivity in diffusion MRI, PhD thesis, Université de Rennes 1, 2022.
[17] Ying Chen, Paolo Giudici, Branka Hadji Misheva, Simon Trimborn, Lead behaviour in bitcoin markets, Risks 8 (1) (2020) 4.
[18] D.F. Alehegbey, The econometrics of Bayesian graphical models: a review with financial application, J. Netw. Theory Finance 2 (2016) 1–33.
[19] Dominic Edelmann, Donald Richards, Thomas Royen, Product inequalities for multivariate Gaussian, gamma, and positively upper orthant dependent distributions, Stat. Probab. Lett. 197 (2023) 109820.
[20] Johannes Ferreira, Andriette Bekker, Mohammad Arashi, Advances in Wishart-type modelling of channel capacity, REVSTAT Stat. J. 18 (3) (2020) 237–255.
[21] Paolo Giudici, Robert Castelo, Improving Markov chain Monte Carlo model search for data mining, Mach. Learn. 50 (2003) 127–158.
[22] Paolo Giudici, Gloria Polinesi, Crypto price discovery through correlation networks, Ann. Oper. Res. 299 (2021) 443–457.
[23] Paolo Giudici, Alessandro Spelta, Graphical network models for international financial flows, J. Bus. Econ. Stat. 34 (1) (2016) 128–138.
[24] Arjun K. Gupta, Daya K. Nagar, Matrix Variate Distributions, vol. 104, CRC Press, 1999.
[25] J. Fan, F. Han, H. Liu, Challenges of big data analysis, Natl. Sci. Rev. 1 (2014) 293–314.
[26] Anis Iranmanesh, M. Arashi, D.K. Nagar, S.M.M. Tabatabaey, On inverted matrix variate gamma distribution, Commun. Stat., Theory Methods 42 (1) (2013) 28–41.
[27] Anis Iranmanesh, Daya K. Nagar, Sara Shokri, Saralees Nadaraja, Generalized matrix t distribution based on new matrix gamma distribution, REVSTAT Stat. J. 20 (1) (2022) 33–50.
[28] J.T. Leek, J.D. Storey, Capturing heterogeneity in gene expression studies by surrogate variable analysis, PLoS Genet. 3 (2007) e161, https://doi.org/10.1371/journal.pgen.0030161.
[29] Bing Jian, Baba C. Vemuri, Evren Ozarslan, Paul Carney, Thomas Mareci, A continuous mixture of tensors model for diffusion-weighted MR signal reconstruction, in: 2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, IEEE, 2007, pp. 772–775.
[30] M.C. Jones, On families of distributions with shape parameters, Int. Stat. Rev. 83 (2) (2015) 175–192.
[31] A.K. Saleh, Md. Ehsanes, M. Arashi, R.A. Saleh, M. Norouzirad, Rank-Based Methods for Shrinkage and Selection: With Application to Machine Learning, Wiley and Sons, 2022.
[32] J. Tomasz Kozubowski, Stepan Mazur, Krzysztof Podgórski, et al., Matrix Gamma Distributions and Related Stochastic Processes, Örebro University School of Business, 2022.
[33] Steffen L. Lauritzen, Graphical Models, vol. 17, Clarendon Press, 1996.
[34] Steffen Lilholt Lauritzen, Nanny Wermuth, Graphical models for associations between variables, some of which are qualitative and some quantitative, Ann. Stat. (1989) 31–57.
[35] Han Na Lee, Armin Schwartzman, Inference for eigenvalues and eigenvectors in exponential families of random symmetric matrices, J. Multivar. Anal. 162 (2017) 152–171.
[36] Christophe Ley, Flexible modelling in statistics: past, present and future, J. Soc. Fr. Stat. 156 (1) (2015) 76–96.
[37] Christophe Ley, Slađana Babić, Domien Craens, Flexible models for complex data with applications, Annu. Rev. Stat. Appl. 8 (2021) 369–391.
[38] Chuanhai Liu, Donald B. Rubin, ML estimation of the t distribution using EM and its extensions, ECM and ECME, Stat. Sin. (1995) 19–39.
[39] Eugene Lukacs, Radha G. Laha, Applications of Characteristic Functions. Monographs, Hafner Publ. Co, 1964.
[40] Lei Luo, Jian Yang, Bob Zhang, Jielin Jiang, Heng Huang, Nonparametric Bayesian correlated group regression with applications to image classification, IEEE Trans. Neural Netw. Learn. Syst. 29 (11) (2018) 5330–5344.
[41] Arak M. Mathai, Serge B. Provost, On the singular gamma, Wishart, and beta matrix-variate density functions, Can. J. Stat. 50 (4) (2022) 1143–1165.
[42] Vladimir Ostashev, D. Keith Wilson, Matthew J. Kamrath, Chris L. Pettit, Modeling of signals scattered by turbulence on a sensor array using the matrix gamma distribution, J. Acoust. Soc. Am. 148 (4) (2020) 2476.
[43] Angeliki Papana, Catherine Kyrtsou, Dimitris Kugiumtzis, Cees Diks, Financial networks based on granger causality: a case study, Phys. A, Stat. Mech. Appl. 482 (2017) 65–73.
[44] Taeyoung Park, Seunghan Lee, Improving the Gibbs sampler, Wiley Interdiscip. Rev.: Comput. Stat. 14 (2) (2022) e1546.
[45] Gamini Premaratne, Anil K. Bera, Modeling asymmetry and excess kurtosis in stock return data, Illinois Research & Reference Working Paper No. 00-123, 2000.
[46] Alexis Reymbaut, Matrix moments of the diffusion tensor distribution and matrix-variate gamma approximation, J. Magn. Reson. Open 8 (2021) 100016.
[47] T.A. Henderson, M.J. Van Lierop, M. McLean, J.M. Uszler, J.F. Thornton, Y.H. Siow, P. Cohen, Functional neuroimaging in psychiatry—aiding in diagnosis and guiding treatment. What the American psychiatric association does not know, Front. Psychiatr. 11 (2020) 276.

[48] Geoffrey Z. Thompson, Ranjan Maitra, William Q. Meeker, Ashraf F. Bastawros, Classification with the matrix-variate-t distribution, J. Comput. Graph. Stat. 29 (3) (2020) 668–674.
[49] Stanley Wasserman, Katherine Faust, Social Network Analysis: Methods and Applications, Cambridge University Press, 1994.
[50] Joe Whittaker, Graphical Models in Applied Multivariate Statistics, Wiley Publishing, 2009.