

PAPER • OPEN ACCESS

## Determination of latent dimensionality in international trade flow

To cite this article: Duc P Truong *et al* 2020 *Mach. Learn.: Sci. Technol.* 1 045017

View the [article online](#) for updates and enhancements.



## PAPER

## Determination of latent dimensionality in international trade flow

## OPEN ACCESS

Duc P Truong<sup>1</sup> , Erik Skau<sup>2</sup> , Vladimir I Valtchinov<sup>3</sup> and Boian S Alexandrov<sup>4</sup> RECEIVED  
28 February 2020<sup>1</sup> Department of Mathematics, Southern Methodist University, Dallas, Texas, United States of America  
<sup>2</sup> Computer, Computational, and Statistical Sciences Division, Los Alamos National Laboratory, Los Alamos, NM, United States of AmericaREVISED  
30 May 2020<sup>3</sup> Department of Radiology and Center for Evidence based Imaging, Brigham and Women's Hospital and Department of Biomedical Informatics, Harvard Medical School, Boston, MA, United States of AmericaACCEPTED FOR PUBLICATION  
28 July 2020<sup>4</sup> Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM, United States of AmericaPUBLISHED  
15 October 2020E-mail: [boian@lanl.gov](mailto:boian@lanl.gov)**Keywords:** latent dimension determination, non-negative tensor factorization, international trade network

Original Content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

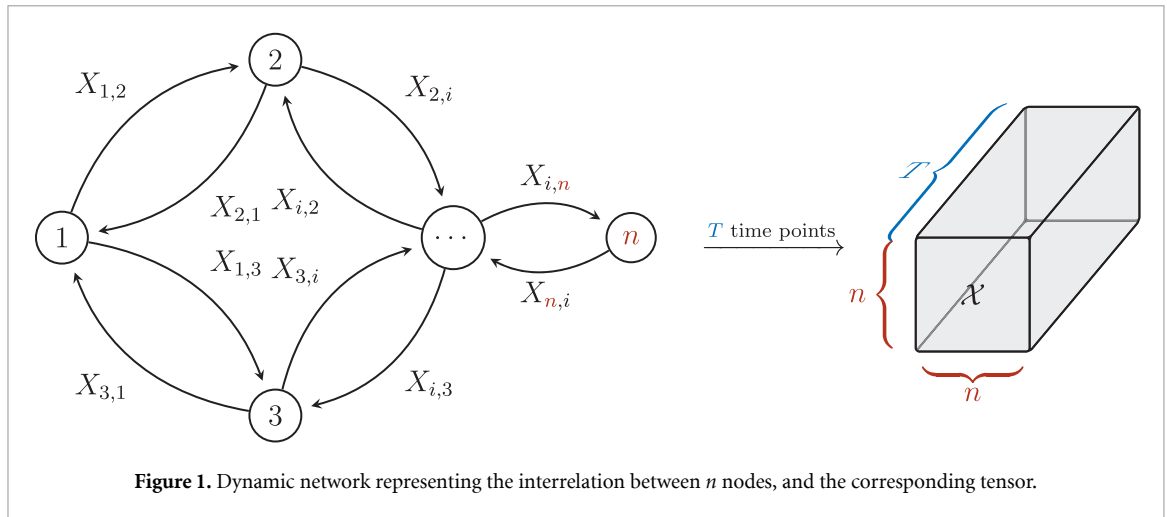
**Abstract**

Currently, high-dimensional data is ubiquitous in data science, which necessitates the development of techniques to decompose and interpret such multidimensional (aka tensor) datasets. Finding a low dimensional representation of the data, that is, its inherent structure, is one of the approaches that can serve to understand the dynamics of low dimensional latent features hidden in the data. Moreover, decomposition methods with non-negative constraints are shown to extract more insightful factors. Nonnegative RESCAL is one such technique, particularly well suited to analyze self-relational data, such as dynamic networks found in international trade flows. Particularly, non-negative RESCAL computes a low dimensional tensor representation by finding the latent space containing multiple modalities. Furthermore, estimating the dimensionality of this latent space is crucial for extracting meaningful latent features. Here, to determine the dimensionality of the latent space with non-negative RESCAL, we propose a latent dimension determination method which is based on clustering of the solutions of multiple realizations of non-negative RESCAL decompositions. We demonstrate the performance of our model selection method on synthetic data. We then apply our method to decompose a network of international trade flows data from *International Monetary Fund* and shows that with a correct latent dimension determination, the resulting features are able to capture relevant empirical facts from economic literature.

**1. Introduction**

Dynamic networks are commonplace in many fields, such as economics, neuroscience, biology, recommender systems, data mining, and others. In these networks, different entities are represented by nodes, and their interactions over time are tracked through their edges. Dynamic networks are interrelated with the statistical relational models [1] presented in artificial intelligence [2]. Statistical relational models often are characterized as: graphical models [3], latent class models [4] and tensor factorization models [5]. Interestingly, there is a natural interconnection between the concept of these models [6, 7]. The relational datasets are in the form of graphs, with nodes and edges representing, respectively, the various entities and their relationships [8], which can be understood as graph-structured knowledge data that contain information for the relationships between some entities.

The emerging large amounts of high-dimensional data, constantly generated all over the globe by: sensor networks; large-scale experiments; massive computer simulations; electronic communications; social-network activities, etc are formed only by directly observable quantities, while the underlying processes, variables or mechanisms remain often unobserved, hidden, or latent [9]. These hidden variables can be either impossible to measure/simulate directly or they are simply unknown. Deep learning and deep belief networks have been recently utilized for dimension reduction and feature extraction [10, 11], however, one of the most powerful unsupervised tools for extracting understandable latent features remains factor analysis. Tensor factorization is a cutting-edge approach for factor analysis [12, 13]. Tensor factorization



methods have been proposed long ago for analyzing relational datasets and dynamic networks as the dynamics can be fully represented as a three-dimensional tensor with the first two axes indexing the nodes, and the third axis indexing time, figure 1. It was shown that the RESCAL tensor decomposition can reveal notable interactions in dynamic asymmetric pairwise relationship tensors [5], especially, when restricting the factors to be non-negative, which makes the model part-based and highly interpretable [14].

One of the fundamental challenges in any model analysis is the determination of model hyperparameters [15]. For the tensor decomposition, this means selection of the correct latent dimension in the data that is related with the estimation of the rank of the analyzed tensor [16], which is an NP-hard problem [17]. The development of various procedures for model selection is an active research topic that include heuristics such as the Automatic Relevance Determination(ARD) [18] or generalized class of information criteria for tensors with specific properties [19]. Especially in economics extracting easily understandable latent variables impacts important questions about the hidden mechanisms, causes, propagating channels, and groups of any international trade and economic events [20, 21].

RESCAL can be considered as a specific Tucker-2 decomposition with two equal factors, because of the inherent symmetry of the data. Hence, in RESCAL we need to know the multi-rank rather the rank of the analyzed tensor. In the case of non-negative decomposition and even in presence of deficiency of the factors we can apply NMF to the corresponding unfoldings of the tensor, to find the minimal multi rank [22].

To determine the latent dimensionality in NMF, there exists a recent model determination technique, called NMFk, [23, 24] that is scalable [25], and has been used to decompose the biggest collection of human cancer genomes [26]. NMFk integrate the classical NMF with custom clustering and Silhouette statistics [27] and works on the principle of stability of the extracted latent variables, from several NMF-minimizations combined with the accuracy of the minimization in order to estimate the optimal number of latent features.

In this paper, we report the application of NMFk to determine latent dimensions, to the non-negative RESCAL tensor decomposition. We demonstrate the efficacy of our method on synthetic data and apply our model determination method to decompose a well-known international trade flow dataset. The result shows that with the correct latent dimension, the model can extract meaningful features which captures relevant established economic empirical findings.

### 1.1. Preliminaries and notation

Throughout this work vectors are denoted with lowercase bold letters,  $x$ , matrices are denoted with uppercase letters  $X$ , and tensors are denoted with uppercase script letters,  $\mathcal{X}$ . Mode-1 multiplication between an order three tensor and a matrix is defined  $(\mathcal{X} \times_1 Y)_{i,j,k} = \sum_{l=1} Y_{i,l} \mathcal{X}_{l,j,k}$ , and similarly for modes 2 and 3. The Frobenius norm of a matrix or tensor is the square root of the sum of the squares of the elements, or  $\|X\|_F = \sqrt{\sum_{i,j} X_{i,j}}$ ,  $\|\mathcal{X}\|_F = \sqrt{\sum_{i,j,k} \mathcal{X}_{i,j,k}}$ . A single subscript on a matrix, or tensor indicates the slice of the object along the last index, so  $X_i$  indicates the  $i^{th}$  column of the matrix, and  $\mathcal{X}_i$  indicates the  $i^{th}$  matrix along the third mode. Additionally, a superscript in parentheses is used to enumerate items in a set or ensemble,  $\{X^{(1)}, X^{(2)}, \dots, X^{(p)}\}$ .

### 1.2. Related works

Here we describe some relevant works that are working with a similar decomposition model. The first proposed model was the single domain Decomposition into Directional Components (DEDICOM)

model [28]. DEDICOM is capable of analyzing asymmetric data in marketing research and it has both matrix and tensor versions without constraints on its factors:

$$\begin{aligned} \text{two-way DEDICOM} & \quad X = ARA^T \text{ where } A \in \mathbb{R}^{n \times r}, R \in \mathbb{R}^{r \times r} \\ \text{three-way DEDICOM} & \quad \mathcal{X}_k = A\mathcal{D}_kR\mathcal{D}_kA^T \text{ for } k = 1, \dots, T \end{aligned}$$

In the three-way DEDICOM model,  $\mathcal{X}_k$  is the  $k^{\text{th}}$  frontal slice of the tensor  $\mathcal{X}$ , and  $\mathcal{D}_k \in \mathbb{R}^{r \times r}$  is a diagonal matrix. These models describe a *single domain* in the sense that they require the row space to be the same as the column space. Since there is no constraint on the factors, these decompositions can be estimated similarly as SVD.

Later, an alternating algorithm, Alternating Simultaneous Approximation Least Square and Newton (ASALSAN) was proposed to perform a three-way DEDICOM [29] that can be applied to large and sparse data. Moreover, the non-negative version of three-way DEDICOM was also introduced using a multiplicative update algorithm. The model was then applied to analyze email network data and export/import data. However, the latent dimension was selected without being justified, and the meaning of the extracted factors was not analyzed in details.

A relaxed version of three-way DEDICOM is the RESCAL model [5]. The model decomposes a three-way tensor  $\mathcal{X}$  as follows:

$$\mathcal{X}_k = A\mathcal{R}_kA^T \text{ for } k = 1, \dots, m$$

In [5], the factors  $A$  and  $\mathcal{R}_k$  minimized the  $l_2$ -regularized minimization problem, and can be estimated using a variation of ASALSAN:

$$\min_{A, \mathcal{R}_k} \frac{1}{2} \left( \sum_k \|\mathcal{X}_k - A\mathcal{R}_kA^T\|_F^2 \right) + \frac{1}{2} \lambda \left( \|A\|_F^2 + \sum_k \|\mathcal{R}_k\|_F^2 \right)$$

The model was then applied to different datasets, and the authors suggested to find latent dimension using cross-validation. Instead, they chose a rather large dimension ( $k = 20$ ), and then used k-means clustering method to cluster the matrix  $A$  into 6 groups.

Following the initial work, reference [30] introduced updating schemes for different variants of the non-negative RESCAL mode, including least-squares with  $l_2$  regularization, KL-divergence with  $l_1$  regularization, and other.

## 2. Methods and implementation

### 2.1. Nonnegative RESCAL

The RESCAL model is a tensor decomposition that takes advantage of the inherent structure of relational properties such as those expected in a dynamic network. More specifically, RESCAL decomposes an order three tensor by finding a common low dimensional latent space for the first two modes such that

$$\mathcal{X} = \mathcal{R} \times_1 A \times_2 A$$

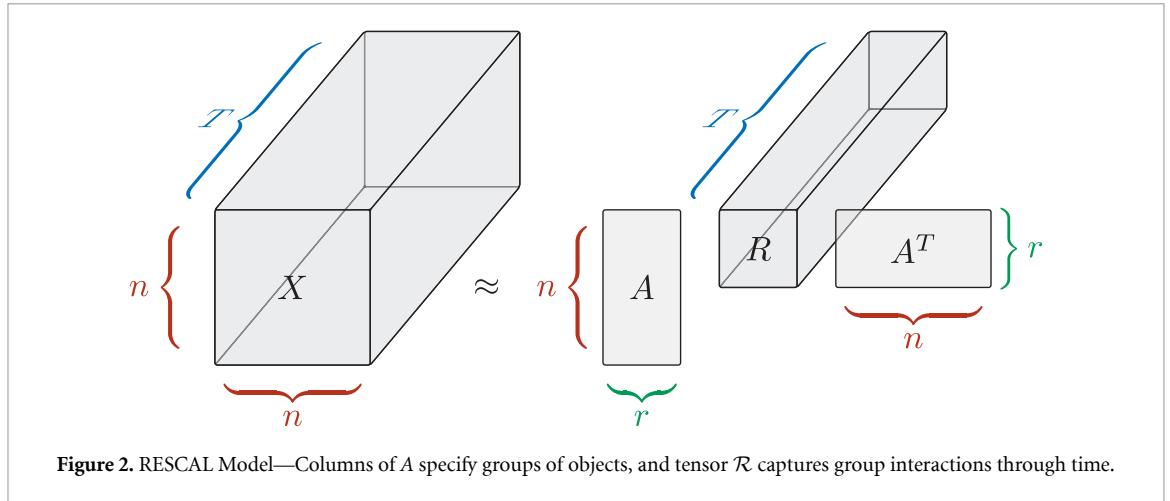
where  $A$  is an  $n \times r$  matrix containing the features,  $\mathcal{R}$  is an  $r \times r \times T$  tensor capturing the mixing relations between the features, and  $\times_i$  is the mode- $i$  product defined above. In practice, a decomposition is always approximated by solving the optimization problem

$$\operatorname{argmin}_{A, \mathcal{R}} \|\mathcal{X} - \mathcal{R} \times_1 A \times_2 A\|_F^2 \tag{1}$$

When applied to a dynamic networks, RESCAL is interpretable in the way that each column of  $A$  represents a group of objects or nodes, and  $\mathcal{R}_t$  represents the relations among these groups at time  $t$ . An equivalent problem statement demonstrates that RESCAL simultaneously decomposes each slice of an order three tensor with a rank  $r$  factorization,

$$\operatorname{argmin}_{A, \mathcal{R}_t} \sum_t \|\mathcal{X}_t - A\mathcal{R}_tA^T\|_F^2. \tag{2}$$

With this interpretation RESCAL attempts to find a common set of features that can be simultaneously used to represent both the row space and the column space of the matrices  $\mathcal{X}_t$ . To remove a scaling ambiguity, RESCAL also can constrain the columns of  $A$  to be unit norm,  $\|A_i\| = 1$  for  $1 \leq i \leq r$ .



In reality, we can easily find experimental data that are non-negative, such as signal/imaging data, trading amount (in currency) between countries, DNA microarray, etc quite often for the non-negative data, and especially with a correct latent dimension, a non-negative decomposition provides more physically meaningful and easier-to-interpret features [14]. Nonnegative RESCAL provides the same advantage by decomposing the data with non-negative features, and non-negative mixing matrices with the optimization

$$\begin{aligned} & \operatorname{argmin}_{A, \mathcal{R}_t} \sum_t \|\mathcal{X}_t - A\mathcal{R}_tA^\top\|_F^2 \\ & \text{subject to} \begin{cases} \sum_j A_{ij} = 1, \text{ for } 1 \leq j \leq r \\ A \geq 0 \\ \mathcal{R} \geq 0 \end{cases} \end{aligned} \quad (3)$$

### 2.2. Multiplicative update algorithm

To solve the non-negative constraint minimization problem in equation (3), we use the multiplicative update scheme similar to the one used for DEDICOM model [29], and for non-negative RESCAL [30]. Starting from non-negative random initialization of  $A$  and  $R$ , the following update steps are performed until the relative error converges to a predefined tolerance:

$$\begin{aligned} (\mathcal{R}_t)_{ij} &= (\mathcal{R}_t)_{ij} \frac{[A^\top \mathcal{X}_t A]_{ij}}{[A^\top A \mathcal{R}_t A^\top A]_{ij} + \epsilon} \text{ for } t = 1, \dots, T \\ A_{ij} &= A_{ij} \frac{\left[ \sum_{t=1}^T \mathcal{X}_t A \mathcal{R}_t^\top + \mathcal{X}_t^\top A \mathcal{R}_t \right]_{ij}}{\left[ A (\mathcal{R}_t A^\top A \mathcal{R}_t^\top + \mathcal{R}_t^\top A^\top A \mathcal{R}_t) \right]_{ij} + \epsilon} \end{aligned} \quad (4)$$

After the multiplicate update scheme converges,  $A$  and  $\mathcal{R}$  are appropriately scaled such that columns of  $A$  have a sum equal to one. Notice that the decomposition can be scaled without affecting the reconstruction error.

### 2.3. Model selection

To select an appropriate latent dimension, we adapt the clustering procedure that has found success in NMF [23]. For each explored latent dimension,  $k$ , our procedure applies three steps: (i) bootstrapping by resampling the data, (ii) decomposing the bootstrapped data, and then (iii) analyzing the cluster stability of the solutions. Figure 3 shows a diagram of the model selection scheme. To resample the data, we construct an ensemble of tensors  $\{\mathcal{X}^{(1)}, \mathcal{X}^{(2)}, \dots, \mathcal{X}^{(P)}\}$  sampled from  $\mathcal{X}_{i,j,k}^{(p)} \sim U(1 - \epsilon, 1 + \epsilon) * \mathcal{X}_{i,j,k}$  for  $1 \leq p \leq P$  where  $U(a, b)$  is a uniform distribution between  $a$  and  $b$ . Each resampling introduces some variability in the data which mitigates the possibility of overfitting.

We use a custom clustering algorithm designed to exploit the stability of the NMF’s solutions corresponding to the resampled data. The custom property of the clustering is that each cluster should contain precisely one feature vector from each  $A^{(p)}$ . Our algorithm is based on k-means but iterates over each  $A^{(p)}$  to assign each vector to an appropriate centroid. The assignment is done by a greedy algorithm applied to the cosine similarity between the columns of  $A^{(p)}$  and the current centroids.

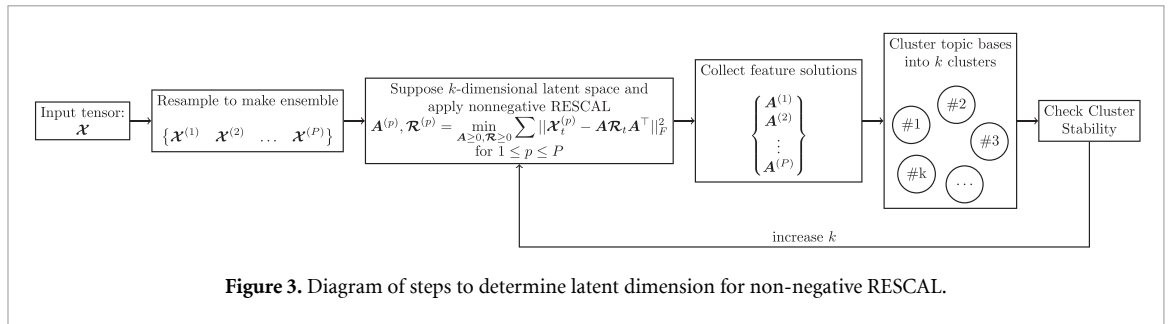


Figure 3. Diagram of steps to determine latent dimension for non-negative RESCAL.

To analyze the quality of the clustering, we use silhouette statistics [27]. The silhouette of a single point has a value between  $-1$  and  $1$  relating how close it is to other points in the same cluster and how far is this point to the closest of the other clusters. A high silhouette score indicates that the clusters are compact and well separated, while a low silhouette score indicates that the clusters are not well separated. We use both the mean silhouette score, as well as the minimum silhouette score of a group as cluster quality metrics.

The selection of the correct latent dimension is accomplished by considering both the silhouette statistics, which measures the stability of the solutions, and the relative error. We expect that with the correct latent dimension the solutions will cluster well with a small relative error. With a too small latent space, the relative error will be too large, and with too large latent space there will be latent features representing the noise in the system that are not stable and hence will not cluster well, resulting in a low silhouette score. We consider the correct latent dimension to be the largest dimension that generates low relative errors and high silhouette scores.

### 3. Experiments

#### 3.1. Synthetic data

Here we test the performance of our protocol for model selection in determining the ground truth latent dimension of synthetic datasets with different levels of noise. First, a synthetic data tensor  $\mathcal{X}$  with dimensions  $n \times n \times T$  and the latent dimension  $k$  is generated as follows:

- Elements of matrix  $A_{n \times k}$  are randomly sampled from  $U[0, 10)$ . The matrix  $A$  is only selected if  $rank(A) = k$ . Otherwise, it is regenerated.
- Elements of matrix  $A$  is sparsified by setting elements smaller than  $Thresh_A$  to zeros.
- The tensor  $\mathcal{R}$  is also generated from  $U[0, 10)$ .
- Tensor  $\mathcal{R}$  is also sparsified by setting elements smaller than  $Thresh_R$  to zeros.
- $\mathcal{X} = \mathcal{R} \times_1 A \times_2 A + \epsilon$  where elements of  $\epsilon$  are sampled from  $U[0, NoiseFactor)$ .
- The noise level of the data tensor can be adjusted by the value of  $NoiseFactor$ .
- The noise level is computed as  $\frac{\|\mathcal{X} - A\mathcal{R}A^T\|_F}{\|A\mathcal{R}A^T\|_F}$ .

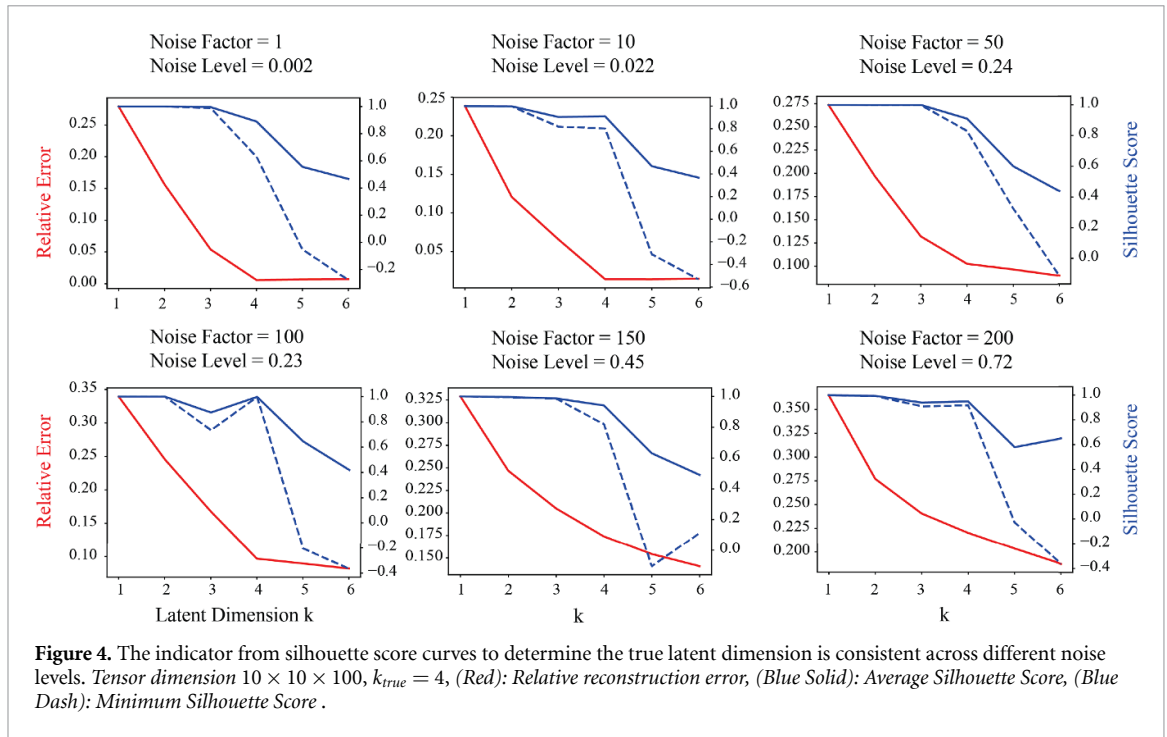
The latent dimension determination procedure described in 2.3 is then applied to the simulated data. More specifically, the sample perturbation is  $0.03$  and the number of iterations  $P = 50$ .

In these scenarios, for a range of  $NoiseFactor \in 1, 10, 50, 100, 150, 200$ , a tensor of dimension  $10 \times 10 \times 100$  is simulated with the ground truth dimension  $k_{true} = 4$ . The plots of relative reconstruction error, minimum and average silhouette scores are shown in figure 4. As the noise factor increases, the noise level also increases and gradually distorts the L-shape of the reconstruction error curves. However, the silhouette score curves are consistent up to  $k = 4$ , and after that, the silhouette score curves start diverging, demonstrating that the clusters are no longer compact and well separated.

#### 3.2. Economic application: decompose the international trade flows

International trade has been shown to generate mutual benefit between countries by allowing them to focus on specialization and exchange their produced goods and services. Unsurprisingly, it has increasingly contributed to the Gross Domestic Product (GDP). In fact, according to the World Bank, in 2017, the world GDP is about  $\$80$  trillion. It has been shown that the economic growth of a country is strongly associated with its role in the world trading network. Therefore, understanding the trade pattern is an essential step before we can discuss the effects of international trade, or even suggest policy changes.

Here we show that by applying the non-negative RESCAL model, with our latent dimension determination method, we can decompose the international trade network into different groups of countries



whose exports and imports are tightly linked together and their interactions over time are captured in the resulting interacting tensor. More interestingly, the groups' activities are also meaningful in the sense that they are consistent with stylized economic facts.

### 3.2.1. International trade flows data

We obtained the international trade flows data from *Direction of Trade Statistics, IMF* [31]. Specifically, the data contains monthly export amounts in U.S. dollars between 23 countries from January 1981 to December 2015 (420 months). Thus the data tensor has 23 by 23 by 420 dimensions, in which each entry,  $\mathcal{X}_{ijk}$ , represents how much country  $i$  exports to countries  $j$  in month  $k$ . For clarity, the list of countries includes Australia, Canada, China Mainland, Denmark, Finland, France, Germany, Hong Kong, Indonesia, Ireland, Italy, Japan, Korea, Malaysia, Mexico, Netherlands, New Zealand, Singapore, Spain, Sweden, Thailand, United Kingdom of Great Britain and Northern Ireland, United States.

This dataset has previously been analyzed with variations of the RESCAL model. In [32], Chen *et al* apply the RESCAL model, though not non-negative, and they do not have a procedure to determine the correct latent dimension, they simply chose a latent dimension of three for illustrative purposes of the model.

### 3.2.2. Determine the latent dimension & optimal factors

To determine the latent dimension for the model, we resampled as described in section 2.3 with the uniform distribution  $U(0.9, 1.1)$  so that each value in our ensemble has the measured value  $\pm 10\%$  error. We resampled from this distribution 50 times to construct our ensemble and calculated the silhouette statistic for a range of dimensions  $k \in \{3, \dots, 8\}$ . Figure 5 shows the relative reconstruction errors, the average and minimum silhouette statistics leading us to conclude that that the true latent dimension is 5.

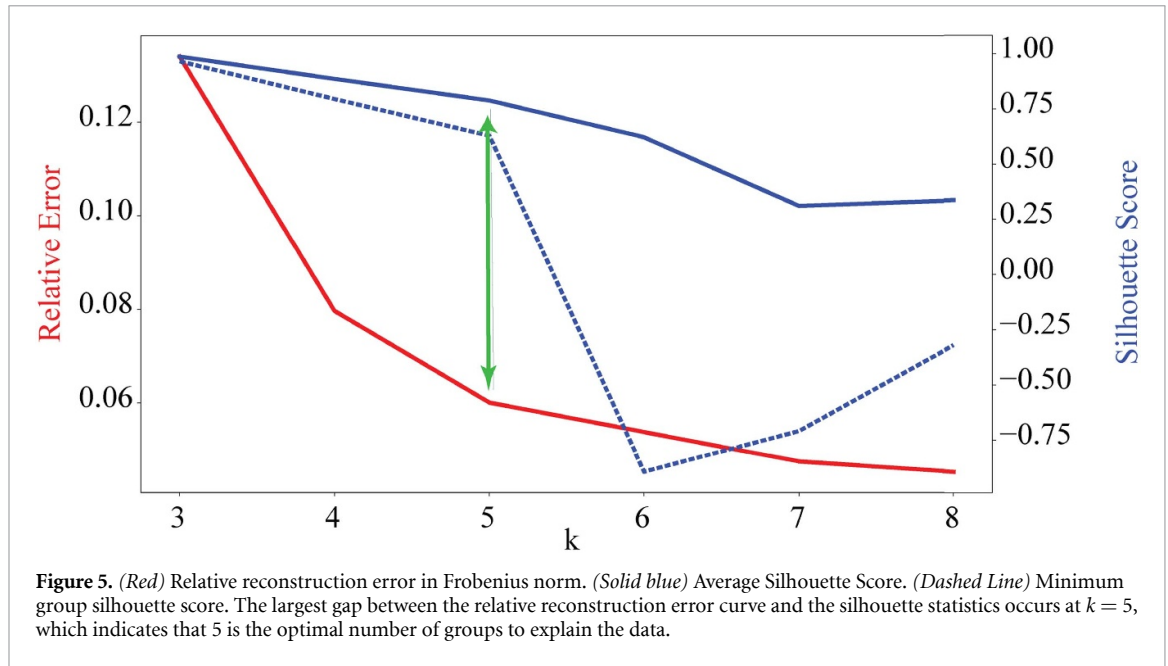
To determine the optimal factors for the analysis, we first ran 100 iterations from random initialization on the original data with the selected dimension  $k = 5$ , and select the decomposition, which provided the lowest reconstruction error. For both purposes, the stopping criterion is the relative convergence rate  $= 1 \times 10^{-8}$ .

### 3.2.3. Interpretation of the decomposition

Here we show the interpretation and analysis of the optimal decomposition. First, since the columns of  $A$  are normalized to have a sum equal to one, the entry  $A_{ij}$  can be interpreted as how much the country  $i$  contributes into group  $j$ . Figure 6 shows that the latent factors, which here are the country contribution profile in each group, approximately correspond to five economic regions: Asia and Pacific (without China), Europe, NAFTA (Canada, Mexico, and the U.S.), U.S., and China. These geo-economic regions are essential participants in the international trade, verifying that the model selection procedure extracts meaningful latent factors.

Next, we analyze the interacting tensor  $\mathcal{R}$  to determine its economic meanings. We first look at the aggregate export level for each economic group over time by summing across the rows of each  $\mathcal{R}_t$  without





the diagonal elements. By doing this, we exclude the contribution of the groups themselves to their aggregate export activities. We then check how well these approximated export activities agree with four global and local economic recession periods, which are identified by the National Bureau of Economic Research (NBER).

As shown in figure 7, the approximated activities match well with the international trading trends in all considered periods. For example, during the *Great Recession* (12/2007–06/2009), in which international trade was dramatically decreasing, the activities of all groups dropped substantially. Secondly, during *early 2000s Recession*, which was partially caused by the dotcom bubble and September 11, the activities represent the fact that the trading trend of all groups, except Asia, were dropping. Thirdly, during the *Asia Financial Crisis* (1997–1999), only the activities of the group Asia is going down, representing the fact that this Recession mostly affected Asian countries. Lastly, during the *European Debt Crisis*, which peaked in 2010–2012, while the economy of other regions was recovering from the Great Recession, only European countries struggled with their high government debt. This styled fact is replicated in the approximated activity of the Europe-group. Overall, the interacting tensor captured the connection between the export and economic health of different economic regions.

Finally, we analyze the interaction tensor  $R \in \mathbb{R}_{5 \times 5 \times 420}$  by summing the tensor  $R$  over time, which gives us the matrix  $S \in \mathbb{R}_{5 \times 5}$  describing how strongly these groups interact.  $S$  is then normalized by its maximum value for better visualization.

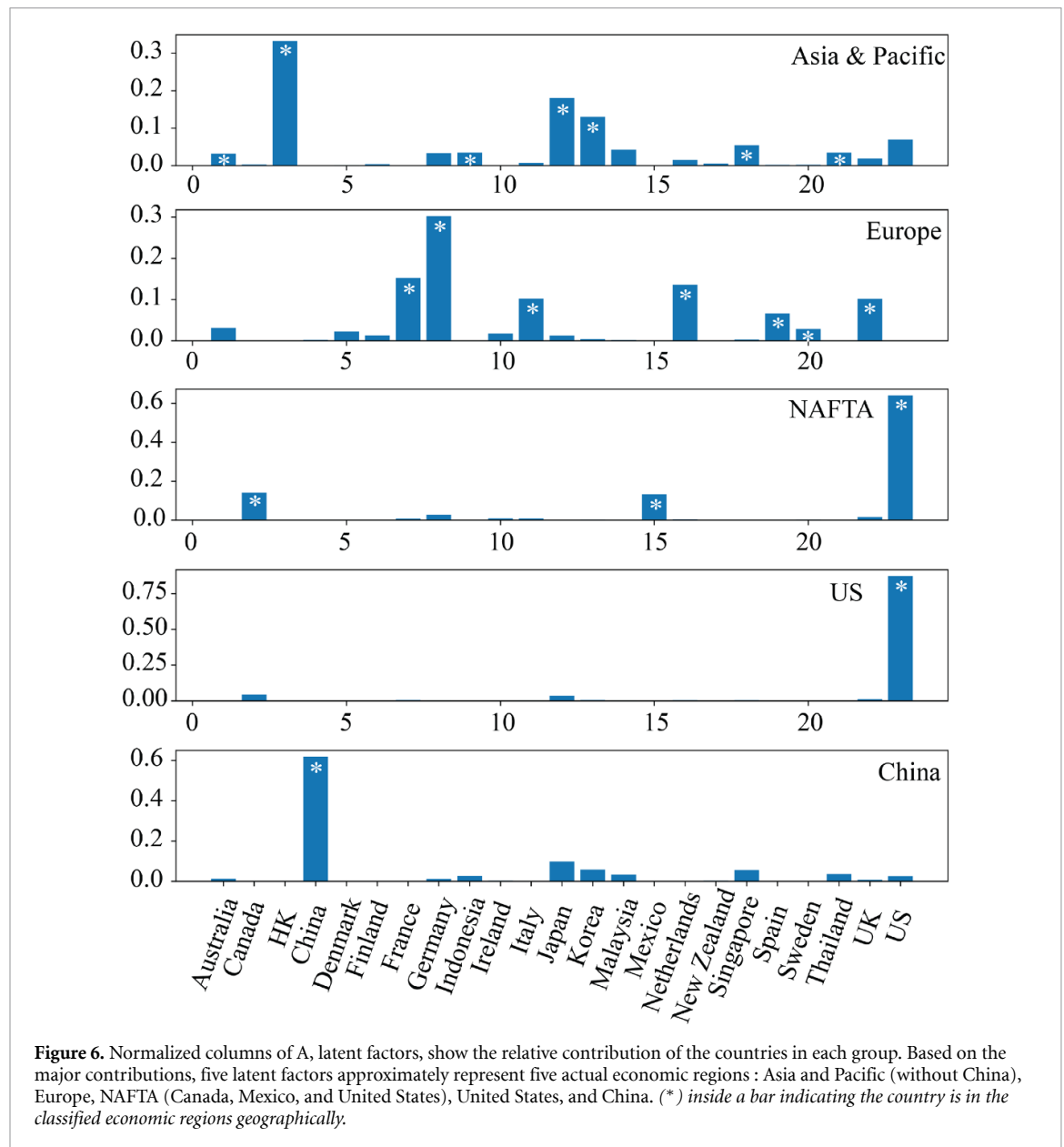
$$\bar{S} = S/\max(S) \text{ where } S_{i,j} = \sum_{t=1}^{420} R_{i,j,t}$$

We makes three observations from figure 8. First, the European group has the strongest interaction with itself, which makes sense since this is a group of advanced economies, and they have formed the European Union since 1995. Second, the strong two-way connection between NAFTA and the United States and between China and Asia, are also matched with statistics. Third, by comparing the row and column for China and United States, we can see that China is a trade surplus (the amount of exports is greater than the number of imports), and the United States is a trade deficit. Overall, the interacting tensor did extract meaningful information from the data in the sense that they agree with international economics empirical facts, indicating that our model selection procedure can capture meaningful activities from the data.

#### 4. Discussion

In this paper, we introduce a model selection protocol for determination of the dominant latent dimension of the non-negative RESCAL model. This method, which is based on nonnegativity assumptions on both factors  $A$  and  $\mathcal{R}$ , evaluates the stability of the latent factors  $A$ , or equivalently the quality of clusters generated by factorization of a set of different realizations of the input data. The method then selects the highest

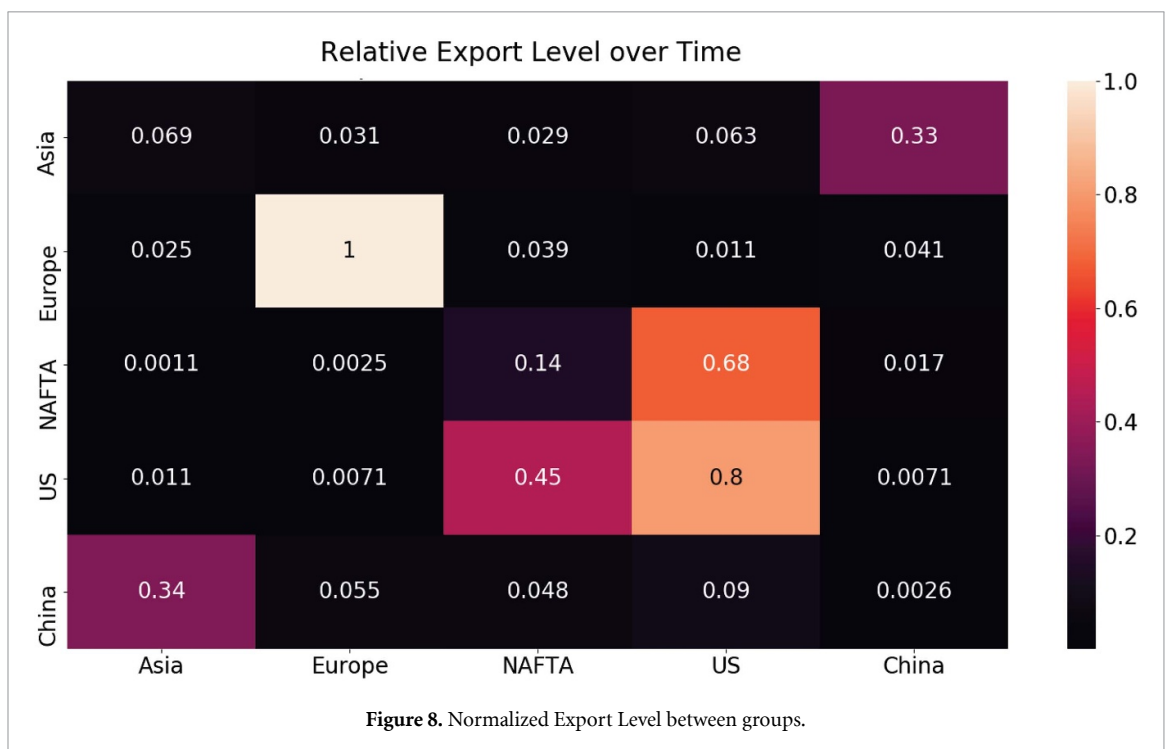
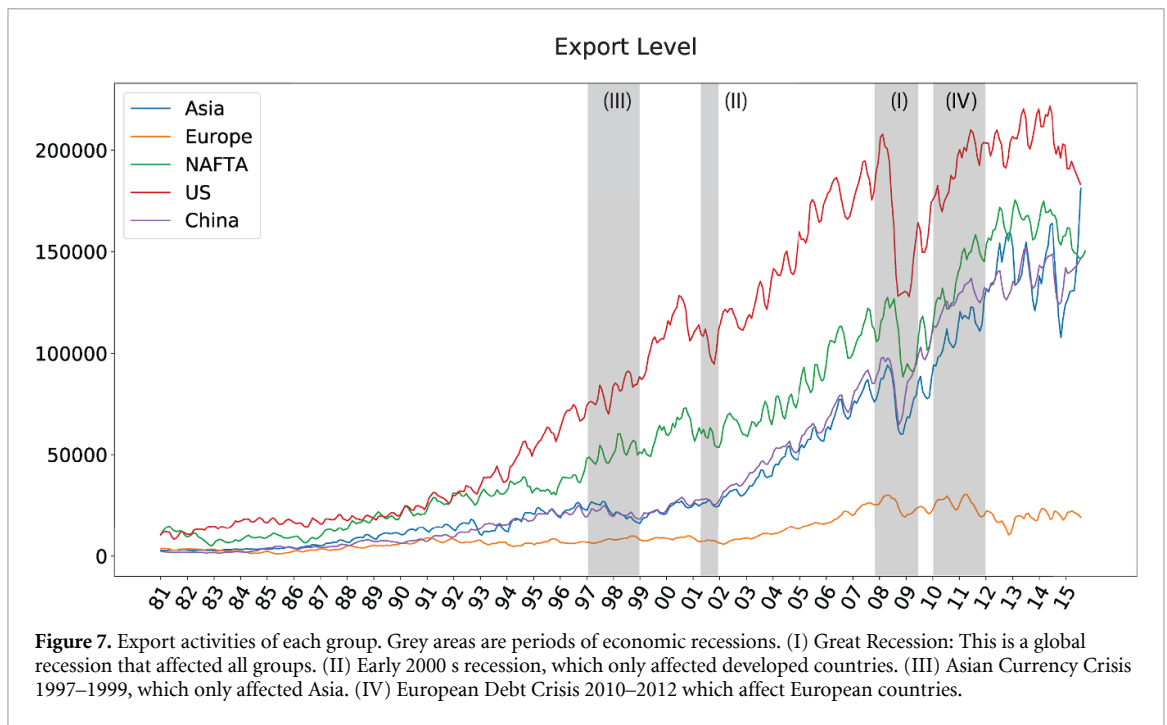




dimension at which the stability is still high. Our method performs well on sparse synthetic data with different noise levels. Moreover, when applied to a real dataset, the international trade flows data from IMF, the model was able to decompose considered countries into meaningful geo-economics regions; with interacting activities matching the trading characteristics of each region and consistent with what has been observed about different economic recessions.

One limitation of this method is that it requires some extent of uniqueness from the solution in order to significantly achieve the factors' consistency. During our testing with synthetic data, we did observe that the method occasionally produced poor cluster quality when the generating factors were dense. With increasing the sparsity of the generating factors, the occasional poor clustering disappeared as sparsity in the generating factors is likely to lead to boundary close and sufficiently spread data. Precisely how the sparsity of the data effects the method's performance is potentially an interesting opening question. Besides generating data with uniqueness, an alternative approach that could be employed in practice is to add regularization to the decomposition model that encourages uniqueness, such as minimum volume or sparsity. In many real-world applications though, severe non-uniqueness is atypical, so in practice this method of choosing the latent dimension with non-negative RESCAL performs well without additional regularization.

It would be of a particular interest to apply the method presented here to two common modeling challenges in business analytics and quantitative marketing: forecasting, and customer marketing segmentation. For example, traditional forecasting techniques rely almost exclusively on the time-series properties of the learning data set (usually called statistical forecasting methods). Another set of techniques



that have been developed introduces additional (external) variables, and a regression-like model was fitted with the additional requirement that the residuals are an ARIMA-distributed process; they are referred to as machine learning (ML) based forecasting methods [33]. However, it is unclear which set of techniques is the better one and would universally work for different types of forecasts: customers forecasting vs. revenue forecasting vs. i.e. inventory forecasting. Two recent papers have carried out an ad-hoc comparison of statistical vs. machine learning models and have arrived at precisely opposite conclusions using similar testing methodologies and goodness of fit metrics ([33] and [34]). Both statistical and machine learning techniques use the time-series nature of the data in the forecasting in a specific manner, whereas the method presented here treats the time dimension like all other dimensions of the multidimensional data set. Further, it is of interest to compare the performance of our method for forecasting of customers and revenue and contrast it against the statistical and the ML methodologies. Remarkably, Markidakis *et al* ([34]) have also compared the forecasting performance of several Deep Learning algorithms and have found it to be stacking

unfavorably against the statistical models—it would be instructive also to see how the Nonnegative RESCAL algorithm fares against deep learning models.

Further, the customer segmentation is a fundamental modeling exercise in quantitative and precision marketing. Customer segmentation has almost exclusively been done in two distinct, loosely connected steps: A) segmenting customers using static data (no time-resolved data, typically using clustering techniques), B) consider transitions to (slightly) different segmentation states, to simulate time-resolved behavior. We intend to apply the Nonnegative RESCAL methodology to this problem as the time dimension is treated in much more natural fashion here.

## Acknowledgments

The work performed in Los Alamos National Laboratory is supported by the U.S. Department of Energy National Nuclear Security

Administration under Contract No. DE-AC52-06NA25 396 and LANL laboratory directed research and development (LDRD) grant 20 190 020DR. Computations were supported by LANL Institutional Computing Program.

## ORCID iDs

Duc P Truong  <https://orcid.org/0000-0002-9806-3274>

Erik Skau  <https://orcid.org/0000-0002-7707-0838>

Boian S Alexandrov  <https://orcid.org/0000-0001-8636-4603>

## References

- [1] Sharan U and Neville J 2007 Exploiting time-varying relationships in statistical relational models *Proc. of the 9th Webkdd and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis* pp 9–15
- [2] Raedt L D, Kersting K, Natarajan S and Poole D 2016 *Synthesis Lectures on Artificial Intelligence and Machine Learning* **10** 1–189
- [3] Getoor L and Mihalkova L 2011 Learning Statistical Models From Relational Data *Proc. of the 2011 ACM SIGMOD Int. Conf. on Management of Data* pp 1195–8
- [4] Magidson J and Vermunt J K 2004 Latent Class Models *The Sage Handbook of Quantitative Methodology for the Social Sciences* (London, UK: SAGE Research) 175–98
- [5] Nickel M, Tresp V and Kriegel H P 2011 A three-way model for collective learning on multi-relational data *Icml* **11** 809–16
- [6] Ishteva M 2015 Tensors and latent variable models *Int. Conf. on Latent Variable Analysis and Signal Separation* Springer pp 49–55
- [7] Robeva E and Seigal A 2019 *Information and Inference: A Journal of the IMA* **8** 273–88
- [8] Koller D et al 2007 *Introduction to Statistical Relational Learning* (Cambridge, MA: MIT Press)
- [9] Everett B 2013 *An Introduction to Latent Variable Models* (Berlin: Springer Science & Business Media)
- [10] Yuan X, Zhou J, Huang B, Wang Y, Yang C and Gui W 2019 Hierarchical quality-relevant feature representation for soft sensor modeling: a novel deep learning strategy *IEEE Trans. Industrial Informatics*
- [11] Wang Y, Pan Z, Yuan X, Yang C and Gui W 2020 A novel deep learning based fault diagnosis approach for chemical process with extended deep belief network *ISA Trans.* **96** 457–67
- [12] Cichocki A, Zdunek R, Phan A H and Amari S i 2009 *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation* (New York: Wiley)
- [13] Kolda T G and Bader B W 2009 Tensor decompositions and applications *SIAM Rev.* **51** 455–500
- [14] Lee D D and Seung H S 1999 Learning the parts of objects by non-negative matrix factorization *Nature* **401** 788–91
- [15] MacKay D J 1994 Bayesian nonlinear modeling for the prediction competition *ASHRAE Trans.* **100** 1053–62
- [16] Anandkumar A, Ge R, Hsu D, Kakade S M and Telgarsky M 2014 Tensor decompositions for learning latent variable models *J. Machine Learning Res.* **15** 2773–832
- [17] Håstad J 1990 Tensor rank is NP-complete *J. Algorithms (Print)* **11** 644–54
- [18] MØrup M and Hansen L K 2009 Automatic relevance determination for multi-way models *J. Chemometrics: J. Chemometrics Soc.* **23** 352–63
- [19] Shi C, Lu W and Song R 2019 Determining the number of latent factors in statistical multi-relational learning *J. Machine Learning Res.* **20** 809–46
- [20] Eaton J, Kortum S, Neiman B and Romalis J 2016 Trade and the global recession *Am. Economic Rev.* **106** 3401–38
- [21] Bems R, Johnson R C and Yi K M 2013 The great trade collapse *Annu. Rev. Econ.* **5** 375–400
- [22] Alexandrov B, DeSantis D, Manzini G and Skau E 2019 Nonnegative Canonical Polyadic Decomposition with Rank Deficient Factors arXiv: 1909.07570
- [23] Alexandrov L B, Nik-Zainal S, Wedge D C, Campbell P J and Stratton M R 2013 Deciphering signatures of mutational processes operative in human cancer *Cell Rep.* **3** 246–59
- [24] Alexandrov B S and Vesselinov V V 2014 Blind source separation for groundwater pressure analysis based on nonnegative matrix factorization *Water Resour. Res.* **50** 7332–47
- [25] Chennupati G, Vangara R, Skau E, Djidjev H and Alexandrov B 2020 Distributed non-negative matrix factorization with determination of the number of latent features *J. Supercomputing* pp 1–31
- [26] Alexandrov L B et al 2013 Signatures of mutational processes in human cancer *Nature* **500** 415–21
- [27] Rousseeuw P J 1987 Silhouettes: a graphical aid to the interpretation and validation of cluster analysis *J. Comput. Appl. Math.* **20** 53–65

- [28] Harshman R A, Green P E, Wind Y and Lundy M E 1982 A model for the analysis of asymmetric data in marketing research *Market. Sci.* **1** 205–42
- [29] Bader B W, Harshman R A and Kolda T G 2006 Pattern analysis of directed graphs using DEDICOM: an application to Enron email *Technical Report* Sandia National Laboratories
- [30] Krompaß D, Nickel M, Jiang X and Tresp V 2013 Non-negative tensor factorization with rescal *Tensor Methods for Machine Learning, ECML workshop*
- [31] Marini M, Dippelsman R J and Stanger M 2018 New estimates for direction of trade statistics *New Estimates for Direction of Trade Statistics Working Paper* No. 18/16 International Monetary Fund
- [32] Chen E Y and Chen R 2017 Factor Models for High-Dimensional Dynamic Networks: with Application to International Trade Flow Time Series 1981-2015 arXiv: [1710.06325](https://arxiv.org/abs/1710.06325)
- [33] Nesreen K, Ahmed Amir F and Atiya N E G H E S 2010 An Empirical Comparison of Machine Learning Models for Time Series Forecasting *Econometric Rev.* **29** 594–621
- [34] Spyros Makridakis Evangelos Spiliotis V A 2018 Statistical and Machine Learning forecasting methods: Concerns and ways forward *PLoS One* **13** 1–27