Higher-order hierarchical spectral clustering for multidimensional data

Giuseppe Brandi[1] and T. Di Matteo[1,2,3]

[1] Department of Mathematics, King's College London, The Strand, London, WC2R 2LS, UK
giuseppe.brandi@kcl.ac.uk
[2] Complexity Science Hub Vienna, Josefstaedter Strasse 39, A 1080 Vienna, Austria
[3] Centro Ricerche Enrico Fermi, Via Panisperna 89 A, 00184 Rome, Italy

Abstract. Understanding the community structure of countries in the international food network is of great importance for policymakers. Indeed, clusters might be the key for the understanding of the geopolitical and economic interconnectedness between countries. Their detection and analysis might lead to a bona fide evaluation of the impact of spillover effects between countries in situations of distress. In this paper, we introduce a clustering methodology that we name Higher-order Hierarchical Spectral Clustering (HHSC), which combines a higher-order tensor factorization and a hierarchical clustering algorithm. We apply this methodology to a multidimensional system of countries and products involved in the import-export trade network (FAO dataset). We find a structural proxy of countries interconnectedness that is not only valid for a specific product but for the whole trade system. We retrieve clusters that are linked to economic activity and geographical proximity.

Keywords: Tensor decomposition, Factor analysis, Clustering analysis, Spectral clustering, Multidimensional data, Multilayer networks, Food networks

1 Introduction

In this paper, we propose a clustering methodology that can be used by policymakers to analyse and extrapolate community structures in multidimensional datasets, such as the FAO data network. Food networks have been widely studied in the literature [26,20,10,21,11,17,31,12]. The interest on the subject comes from different perspectives, such as the study of trade networks [17,12,1], the fragility of the food network [20,31], health-related shocks which can flow from contaminated food through the trade network [10] and the connection between the food import-export and economic development of countries [21]. Other studies tried to detect common pattern of countries in relation to specific products through clustering techniques [26,11]. However, all these papers analyse the food networks product by product without exploiting the multidimensionality of the data. In this paper, we instead introduce a methodology that is able to produce a synthetic proxy of geopolitical and economic interconnectedness by applying a multidimensional data consistent approach.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77977-1_31

Several dimensionality reduction techniques have been introduced in the literature to synthesize datasets. In particular, factor analysis is a dimensionality reduction technique which has been extensively applied in time series and cross-sectional data [23,3]. The main objective of factor analysis is to reduce the full dataset to a set of few relevant factors which explain most of the information contained in the original dataset [3,24]. These models have been also extensively employed in multivariate analysis to inspect latent features of the data [23,25] and has then been later extended to the analysis of multidimensional data, i.e. tensors [13,27]. Tensors, also known as multiway or multidimensional data [16,15] arise in several research fields, e.g. economics, 3D tomographic images, psychometrics, and factor analysis applied to these systems is commonly known as tensor decomposition [15]. In this paper we use the higher order Tucker decomposition [27], which is an extension of the bilinear factor analysis to multidimensional data. With respect to community detection, several clustering algorithms have been proposed in the literature. In this paper, we apply the Directed Bubble Hierarchical Tree (DBHT), which is a hierarchical filtering algorithm able to retrieve clusters, without the need of choosing the number of clusters a priori or a threshold parameter. This clustering is an unsupervised learning technique that can be applied to multivariate data and in particular, when applied to bilateral economic networks, can be employed to tailor economic and political interventions [26,11] or to create synthetic factors combining intra-cluster components [28]. Often when applied to multidimensional data, clustering techniques are commonly employed on single slices of data, e.g. layers in a multiplex resulting in a computationally intensive procedure and do not synthesize the dataset, resulting in different clustering outputs for each layer. In this paper we propose a methodology which overcomes this issue by implementing a tensor decomposition analysis in combination with a hierarchical clustering technique. The application of this methodology to the Food and Agriculture Organization of the United Nations (FAO) network is able to produce a proxy granted by geographic and economic interpretation. The paper is structured as follows. In Section 2 we describe the methodology, in Section 3 we report the results of the application of this methodology to the network of the FAO while Section 4 concludes.

2 Clustering multidimensional data

The higher-order hierarchical spectral clustering method is based on the combination of tensor decomposition [27,15] and the DBHT clustering tool [22,28] by means of a 2-steps approach. In the first step, we decompose the multidimensional dataset using the Tucker decomposition [27,15] from which we obtain a set of factor loadings matrices that projects the higher dimensional dataset in a low-dimensional space which compresses the information on common factors. Then, the DBHT algorithm [22,28] is performed on such matrices to obtain clusters of specific dimension, e.g. countries or product. In the next subsections, we provide a brief review of the methods applied in the 2-steps approach.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77977-1_31

2.1 Tensor decomposition

Tensors are a generalization of vectors and matrices and are ideal instruments to study multidimensional data. Tensors, like matrices, can be decomposed into smaller (in terms of rank) objects [15]. Among several tensor decomposition methods, we employ the Tucker decomposition, which is mainly used for factor analysis or dimensionality reduction and extends the bilinear factor analysis to the higher dimensional case [27,6,4,7]. Throughout this work, the notation follows the standard convention introduced in [15]: x is a scalar, $\mathbf{x}$ is a vector, $\mathbf{X}$ is a matrix and $\boldsymbol{\mathcal{X}}$ is a tensor. Take an n-th order tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times I_2 \cdots \times I_N}$, the Tucker decomposition of $\boldsymbol{\mathcal{X}}$ can be written as a *n-mode* product, i.e.:

$$\boldsymbol{\mathcal{X}} \approx \boldsymbol{\mathcal{F}} \times_1 \mathbf{\Lambda}^{(1)} \times_2 \mathbf{\Lambda}^{(2)} \cdots \times_N \mathbf{\Lambda}^{(N)} = \boldsymbol{\mathcal{F}} \times \{\mathbf{\Lambda}^{(n)}\}, \tag{1}$$

where $\mathbf{\Lambda}^{(n)}$ are the factor loading matrices and $\boldsymbol{\mathcal{F}}$ is the *core* tensor and it is usually of smaller dimension than $\boldsymbol{\mathcal{X}}$.

2.2 DBHT clustering algorithm

Directed Bubble Hierarchical Tree (DBHT) is a machine learning hierarchical clustering method that exploits the topological property of the Planar Maximally Filtered Graph (PMFG) in order to find the clusters [22,28]. The PMFG is a generalization of the Minimum Spanning Tree (MST), that allows for loops and more edges by preserving all hierarchical properties of the MST. This is constructed following the same procedure of the MST, except that the non-loop condition is replaced with the weaker condition of planarity (i.e. each added link must not cut a pre-existent link). Thanks to this more relaxed topological constraint, the PMFG is able to retain a larger number of edges, hence of information. In particular, it can be shown that each PMFG contains exactly $3(N-2)$ edges for a system of N nodes. The key elements of a PMFG are the three-cliques elements, subgraphs made of three nodes all reciprocally connected (i.e., triangles). The DBHT exploits this topological structure, and in particular the distinction between separating and non-separating three-cliques, to identify a clustering partition of all nodes in the PMFG. A complete hierarchical structure is then obtained for both inter-clusters and intra-clusters by following a traditional agglomerative clustering procedure. The algorithm requires as inputs a distance matrix D and a similarity matrix S.

2.3 Higher-order hierarchical spectral clustering (HHSC)

After presenting the 2-steps approach, we here introduce the HHSC used to extract the clusters from the multidimensional dataset. In the first step, HHSC extracts a set of factor matrices $\mathbf{\Lambda}^{(n)}$ by means of Equation 1 and then it computes a distance and a similarity matrix between the factor loadings corresponding to each element, i.e. country or product. In the second step, by inputting the two matrices in the DBHT algorithm, we identify the clusters.[4] This approach follows

[4] The algorithm's time complexity is described in Appendix A.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77977-1_31

the same spirit of spectral clustering [30] by not directly performing the clustering procedure on the original dataset but rather, on the dimensionally reduced system. This avoids the over-dispersion of information in the original dataset, which can make the analysis very noisy. Some papers in the tensor literature use the values of the factor loadings matrices to directly cluster the data (nodes in the case of tensor networks) by identifying to which factors they are more related to [2]. Even if this is a sensible approach, it neglects the distribution of the factor loadings by focusing only on the maximum value for each node. Conversely, the use of the DBHT algorithm in our procedure, takes into account the full distribution of the factor loadings. Despite it is clear that the maximum loading of each node will weight more, they will not be the only drivers of the community detection in our procedure because all factor loadings are considered. Indeed, the DBHT has been proved to outperform standard factor model analyses [28].

3 FAO trade network

In this section, we apply the HHSC described in Section 2 to an economic network, i.e. the FAO trade network. In particular, we show how HHSC can be used to extrapolate relevant structures from a multidimensional dataset and synthesize the information in geographically and economically meaningful clusters.

3.1 Data

The dataset is collected from the Food and Agriculture Organization of the United Nations (FAO) website.[5] The FAO trade matrix is an economic network in which nodes correspond to importing and exporting countries, layers represent the products and the last dimension is related to the time. Edges at each layer represent the trade relationships of a specific product between countries in a specific time period. We have collected yearly data between 1986 and 2018 for 123 countries and 137 products. We represent this data by a 4-th order tensor $\mathcal{Y}_t$ of dimension $128 \times 128 \times 137 \times 33$.[6]. In order to mitigate the difference in magnitude between the data and avoid the model to only fit high data values, we apply the log transformation which is commonly used in the literature for bilateral trades, i.e. $\bar{\mathcal{Y}}_t = log(1 + \mathcal{Y}_t)$. Finally, to ensure data stationary, we use the following first-order difference of the log-transformed trade tensor, i.e.:

$$\mathcal{X}_t = \bar{\mathcal{Y}}_t - \bar{\mathcal{Y}}_{t-1},$$

where $\mathcal{X}_t$ is a tensor of dimension $128 \times 128 \times 137 \times 32$ and t is the time index. This transformation represents the rate of change of the original dataset.

[5] http://www.fao.org/faostat/en/data/TM.

[6] We filtered out some data from the full dataset. We report the data filtering methodology in Appendix B.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77977-1_31

3.2 Step-by-step HHSC methodology

In the multivariate time series and panel data literature, the dynamic factor model [3,24] starts from a function of the data of the following form:

$$\mathbf{X}_t = \mathbf{\Lambda}\mathbf{F}_t + \mathbf{E}_t, \tag{2}$$

where $\mathbf{X}_t$ is the data, $\mathbf{\Lambda}$ is the factor loading matrix, $\mathbf{F}_t$ is the factor matrix and $\mathbf{E}_t$ is the error term. By assuming Gaussianity of $\mathbf{F}_t$ and $\mathbf{E}_t$, $\mathrm{Cov}(\mathbf{X}_t) = \mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Sigma}^2$, where $\mathbf{\Sigma}^2$ is a diagonal matrix representing the idiosyncratic error of each component of $\mathbf{X}_t$ incorporated in $\mathbf{E}_t$ while $\mathbf{\Lambda}\mathbf{\Lambda}'$ is the factor covariance matrix with rank equal to the number of factors in $\mathbf{\Lambda}$. This is equivalent to the n-mode product formulation of the Tucker decomposition in Equation 1, i.e.:

$$\mathbf{X}_t = \mathbf{F}_t \times_1 \mathbf{\Lambda} + \mathbf{E}_t. \tag{3}$$

This model can be easily extended to the multidimensional case by using the higher-order Tucker representation [14] as:

$$\mathfrak{X}_t = \mathfrak{F}_t \times_1 \mathbf{\Lambda}^{(1)} \times_2 \mathbf{\Lambda}^{(2)} \cdots \times_{N-1} \mathbf{\Lambda}^{(N-1)} + \mathfrak{E}_t, \tag{4}$$

where now $\mathfrak{X}_t$ is a tensor representing the multidimensional dataset at time t, $\mathfrak{F}_t$ is the dynamic core factor tensor while each $\mathbf{\Lambda}^{(i)}$ is a factor loading matrix for the i-th mode. This formulation corresponds to the '$N-1$' Tucker decomposition where the time dimension is not factorized but rather used to estimate the factor components of the other modes. As for the multivariate case, each mode covariance matrix is assumed to be of the form $\mathrm{Cov}(\mathfrak{X}_t^{(n)}) = \mathbf{\Lambda}^{(n)}\mathbf{\Lambda'}^{(n)} + \mathbf{\Sigma}^{(n)}\mathbf{\Sigma'}^{(n)}$, where $\mathbf{\Lambda}^{(n)}$ is estimated through the Higher Order Singular Value Decomposition (HOSVD) [9] for the Tucker model, and $\mathbf{\Sigma}^{(n)}$ is estimated through the flip-flop algorithm applied to the residuals of the model [6,14]. This algorithm is based on the assumption that the residuals follow the array Normal distribution [14] and iteratively estimate the covariance matrix of each mode considering the others as fixed. The Algorithm 1 is reported below:

Algorithm 1 Flip-flop algorithm for covariance estimation

1: **Initialize the algorithm to some** $\mathbf{\Sigma}^{(1)} \cdots \mathbf{\Sigma}^{(N)}$
2: Compute $\mathfrak{E}_t = \mathfrak{X}_t - \widehat{\mathfrak{F}}_t \times_1 \widehat{\mathbf{\Lambda}}^{(1)} \times_2 \widehat{\mathbf{\Lambda}}^{(2)} \cdots \times_N \widehat{\mathbf{\Lambda}}^{(N)}$
3: **for** $n = 1, \ldots, N$
4: Compute $\mathfrak{E}_t^{(n)} = \mathfrak{E}_t \times_1 \mathbf{\Sigma}^{(1)} \cdots \times_{n-1} \mathbf{\Sigma}^{(n-1)} \times_n \mathbf{I}^{(n)} \times_{n+1} \mathbf{\Sigma}^{(n+1)} \cdots \times_N \mathbf{\Sigma}^{(N)}$
5: Compute $\widehat{\mathbf{\Sigma}}^{(n)}\widehat{\mathbf{\Sigma}'}^{(n)} = \mathbb{E}[\mathbf{E}_{(n)}\mathbf{E}_{(n)}^T]$
6: **Return** $\frac{\widehat{\mathbf{\Sigma}}^{(1)}\widehat{\mathbf{\Sigma}'}^{(1)}}{Tr(\widehat{\mathbf{\Sigma}}^{(1)}\widehat{\mathbf{\Sigma}'}^{(1)})} \cdots \frac{\widehat{\mathbf{\Sigma}}^{(N)}\widehat{\mathbf{\Sigma}'}^{(N)}}{Tr(\widehat{\mathbf{\Sigma}}^{(N)}\widehat{\mathbf{\Sigma}'}^{(N)})}$

It is important to notice that $\widehat{\mathbf{\Sigma}}^{(n)}\widehat{\mathbf{\Sigma}'}^{(n)}$ are not identifiable because by multiplying one of the covariance matrices for a scalar w and another covariance matrix for the inverse value w^{-1}, the optimization logarithm reaches the same value. For this reason, and to have covariance matrices which are comparable in magnitude, we estimate the covariance matrices normalized by their trace, that is the sum of the elements of the main diagonal. In this paper, we do not assume

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77977-1_31

any structure of the error term and compute the covariance matrices only as post-modeling diagnostic to check if any residual information is present in the error term, i.e. strongly non-diagonal covariance matrices, and to check if the autocovariance matrix (the fourth mode covariance matrix) exhibits any sort of dynamics that can be exploited in a forecasting setting. Yet, the latter analysis is beyond the scope of the present paper since our main focus is on the factor loadings matrices. From each factor loading matrix, we compute a distance and a similarity matrix which are then used as inputs in the DBHT algorithm. For the distance matrix, we use the Euclidean distance[7], i.e.:

$$D_{a,b}^{(i)} = \left\| \mathbf{\Lambda}_a^{(i)} - \mathbf{\Lambda}_b^{(i)} \right\| \tag{5}$$

where $\mathbf{\Lambda}_j^{(i)}$ is the j-th element (country or product) of the i-th factor loadings matrix. The distance synthesizes the dissimilarity between two items' factor loadings. For the similarity matrix, we use the Gaussian kernel [22,30], i.e:

$$S_{a,b}^{(i)} = e^{\frac{-\left\| \mathbf{\Lambda}_a^{(i)} - \mathbf{\Lambda}_b^{(i)} \right\|^2}{2\sigma^2}}, \tag{6}$$

where σ^2 is the variance of the set of distances in D. This matrix weights more pairwise distances $D_{a,b}^{(i)}$ near to 0 and less values with higher distances. These two matrices suffice for the DBHT to cluster the data as the algorithm first extracts the PMFG and then uses its three-cliques elements to hierarchically cluster the nodes. In the next section, we present the results of our analysis of the FAO international trade network using the HHSC method.

3.3 Application of HHSC to FAO data

The number of factors introduced in Section 3.2 used in the Higher-order Tucker decomposition can be chosen in two ways: either by using a theoretically or economically motivated number of factors or by using some data-driven methods [29] which heuristically choose the best model compared to its complexity. However, standard information criteria cannot be used in this context because of the strong imbalance between the huge multidimensional data and the number of parameters. To select the number of components, we fit various models with increasing number of factors, starting from the [1 1 1] specification (one-factor model) to the [50 50 50] specification and search for the elbow in a scree type of plot [8] in which we compare the log of the Explained Sum of Squares (ESS) and the number of parameters in the model. Figure 1 shows that the right combination of the explained variance and the number of parameters is [28 28 17], which corresponds to the elbow of the curve. Using this rank specification, we obtain three factor loading matrices, i.e. the import matrix $\mathbf{\Lambda}^{(1)} \in \mathbb{R}^{128\times 28}$, the export matrix $\mathbf{\Lambda}^{(2)} \in \mathbb{R}^{128\times 28}$, and the product matrix $\mathbf{\Lambda}^{(3)} \in \mathbb{R}^{137\times 17}$. These matrices represent the information related to each mode of the tensor filtered by the effect

[7] We row normalize the factor loadings before computing the distance and similarity measures in order to harmonize the different items, i.e. countries or products.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77977-1_31

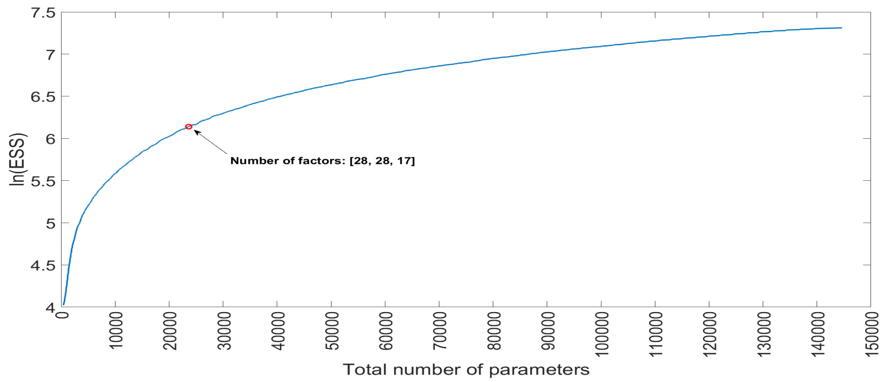

Fig. 1: Logarithm of the Explained Sum of Squares vs the number of model parameters. The circle corresponds to the specification at the elbow of the plot.

of the other modes. They can be used both as outputs to directly analyze or as inputs for a clustering algorithm. In fact, for each row of the matrix (country or product) we identify a set of factor loadings and these provide the information on which factor (hub) the country or product is more related to. Countries or products with similar factor loadings in terms of distribution and magnitude are expected to have a strong similarity and to be allocated to the same cluster. However, by the use of a clustering algorithm, the analysis is made more robust as the full distribution of factor loadings for each country or product will be used instead of taking only the maximum value. By applying the DBHT algorithm on the set of latent factor matrices, we obtain 12 clusters for the import mode, 13 clusters for the export mode, and 9 clusters for the products mode. Results are graphically reported in Figures 2-4.[8] From Figure 2, we can observe that clusters are mainly explained by geographical proximity and economic growth. Indeed, we detect the European clusters highlighted in blue, cyan, yellow, and orange on the center left of the plot. However, even if geographically close to each others, they have different trading patterns for some products which make them to fall in different clusters. There is a second block of European countries on the opposite side of the plot, highlighted in magenta, violet and light blue. It is important to observe that these represent Eastern European and ex URSS countries. The Asian countries are shown in purple on the top of the plot and a mixture of high growth countries of Asia and America are highlighted in green emerald on the top of the plot. Then, there is a cluster of mostly African countries highlighted light green on the bottom left. The two remaining clusters (red and green) are more convoluted. Indeed, they both share slow growth and fast growth countries in Africa and South America.

Regarding the export countries clusters, also in this case the main drivers can be attributed to geographical proximity and economic activity. We can easily identify the European cluster and a few ex URSS countries in magenta, purple, violet, blue, light blue, and light green on the top of the plot. Moreover, we can observe that there are dissimilarities in how different European groups cluster together. We can observe that Eastern European and ex URSS countries cluster

[8] A set of additional figures is reported in in Appendix C.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77977-1_31

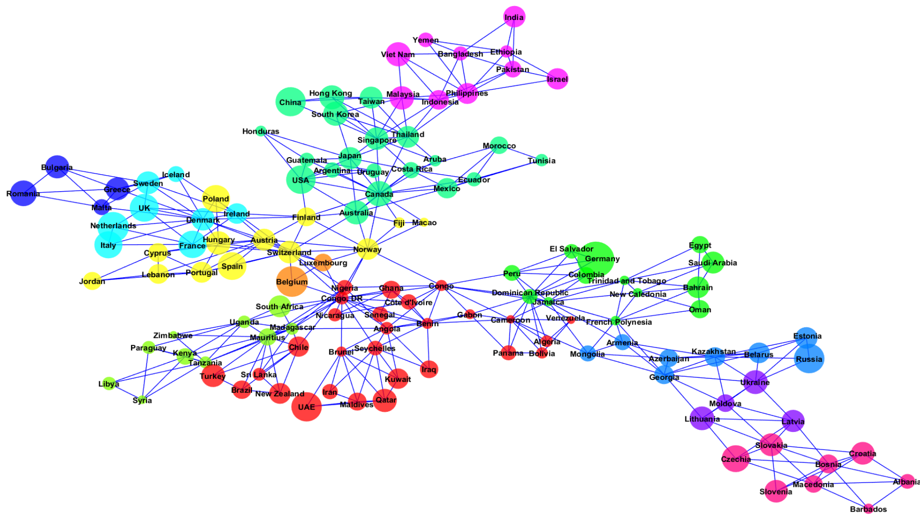

Fig. 2: HHSC clustering for imports. The size of the node is proportional to the total rate of growth of products the country imported between 1986 and 2018.

together as well as ex Yugoslavia countries. There are then the Mediterranean countries clustering together and the North Eastern countries. In the European macro-cluster, it is possible to notice a cluster composed by the Francophone countries, i.e. France, Belgium, and Luxembourg. On the bottom of the plot, it is possible to observe the Asian counties in turquoise and emerald green, and the Arabic countries highlighted in cyan. At the center right of the plot, the South American countries are shown in orange. There is a small cluster highlighted in green composed by Chile, South Africa, and Zimbabwe. The first two have enhanced trading agreements while South Africa is the leading exporter and importer for Zimbabwe. The other two clusters (in red and yellow) are mixed and contain the leading importers and exporters, which do not necessarily follow geographical proximity. Indeed, it is important to mention that even though most of the clusters' countries can be explained by geographical proximity, import/export size, or growth rates, there are some of them, especially the world leading importers/exporters, which do not always follow this pattern. This is because their interrelations with other countries exhibit deeper connections, which go beyond their geographical positions. With respect to the products cluster, Figure 4 shows that the DBHT has good performance also in this case. Indeed, the fruits and vegetable clusters are on the center left part of the plot highlighted in turquoise and red, while on the bottom of the plot there are two meat-related clusters highlighted in magenta and violet, with the second one more related to pig meat type of meat. We can then observe the beverages cluster in light blue and an hyper cluster, highlighted in orange, which connects the other clusters. Indeed, this hyper cluster has different products, but those products are similar to the nearest clusters. Finally, there are other two mixed clusters which correspond to high growth/high amount exported products in green and light green. Finally, in Figure 5 we report the modes' covariance matrices estimated by using the flip-flop Algorithm 1. As it is possible to see, the data is heteroskedastic, as

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77977-1_31

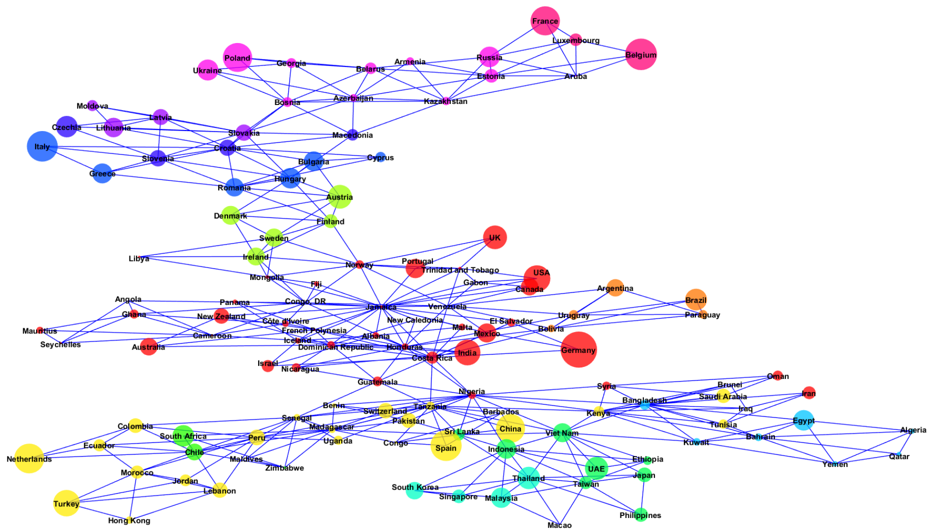

Fig. 3: HHSC clustering for exports. The size of the nodes is proportional to the total rate of growth of products the country exported between 1986 and 2018.

the diagonal elements of the covariance matrices have different magnitudes. It is important to highlight that the static modes have diagonal covariance matrices, meaning that the model correctly factorizes the data while the time mode has a negative first-order autocovariance, in line with an autoregressive specification.

4 Conclusions

In this paper, we have proposed a new methodology, the Higher-order Hierarchical Spectral clustering (HHSC), to cluster multidimensional data by means of a 2-steps approach. In the first step, we decompose the multidimensional data via the Tucker decomposition, while in the second step, we use the DBHT algorithm on the factor loading matrices. We can appreciate that the clusters retrieved by this methodology can be easily explained by economic and geographical factors. Therefore, the tensor factor model in combination with hierarchical clustering is a promising tool to extract clusters and analyse the bilateral food network. Moreover, to better understand and predict the specific factors which drive the formation of clusters, an econometric model based on the Multinomial Logit can also be implemented to link the clusters to economic variables, in particular to understand cases where geographical proximity and rate of growth are not enough to explain the clustering results, e.g. Germany. Finally, the model can be extended to perform a forecasting analysis. This can be done by exploiting the dynamics of the core factor tensor by assuming a Tensor Autoregressive model [6,5]. A further extension would consist in adopting a fully dynamic setting in the spirit of data assimilation through a Kalman filter approach [18]. The HHSC algorithm proposed in this paper is general enough to be used to exploit information contained in a variety of empirical networks which evolve over time with multidimensional interactions, e.g. ecological networks, financial networks.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77977-1_31

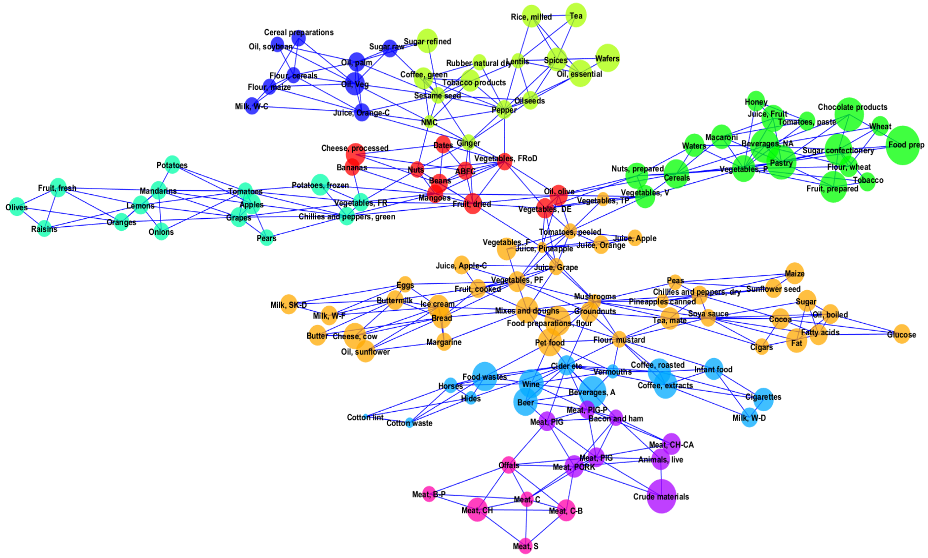

Fig. 4: HHSC clustering for products. The size of the nodes is proportional to the total rate of growth of each product all the countries imported and exported between 1986 and 2018.

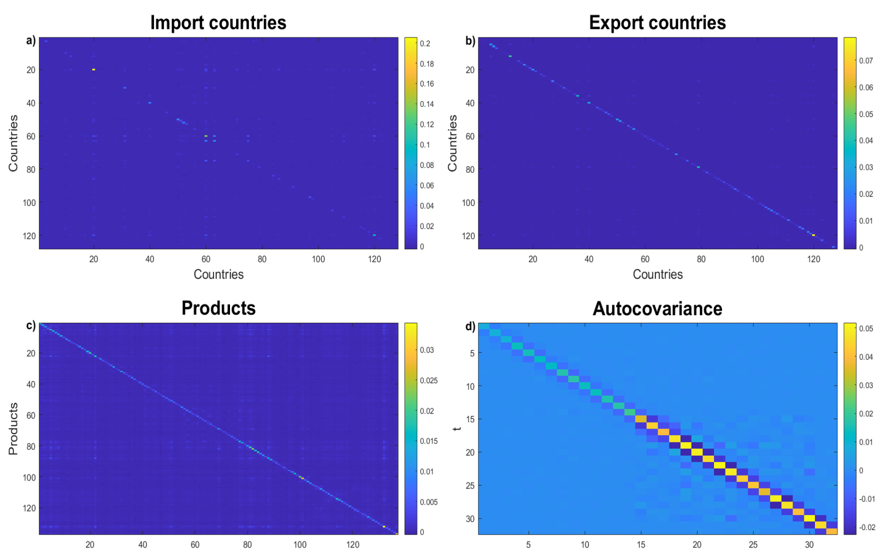

Fig. 5: Covariance of the error tensors modes estimated via the flip-flop algorithm: a) imports mode, b) exports mode, c) products mode, d) Autocovariance matrix.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77977-1_31

References

[1] Angelini, O., Matteo, T.D.: Complexity of products: The effect of data regularisation. Entropy **20**(11), 814 (Oct 2018)
[2] Bader, B.W., Kolda, T.G., Harshman, R.A.: Temporal analysis of social networks using three-way dedicom. Tech. rep., Sandia National Laboratory (2006)
[3] Bai, J., Ng, S.: Determining the number of factors in approximate factor models. Econometrica **70**(1), 191–221 (2002)
[4] Brandi, G.: Decompose et Impera: tensor methods in high-dimensional data. Ph.D. thesis, LUISS Guido Carli (2018)
[5] Brandi, G., Di Matteo, T.: A new multilayer network construction via tensor learning. In: Krzhizhanovskaya, V.V., Závodszky, G., Lees, M.H., Dongarra, J.J., Sloot, P.M.A., Brissos, S., Teixeira, J. (eds.) Computational Science – ICCS 2020. pp. 148–154. Springer International Publishing, Cham (2020)
[6] Brandi, G., Di Matteo, T.: Predicting multidimensional data via tensor learning. Accepted in Journal of Computational Science (2021)
[7] Brandi, G., Gramatica, R., Di Matteo, T.: Unveil stock correlation via a new tensor-based decomposition method. Journal of Computational Science (2020)
[8] Cattell, R.B.: The scree test for the number of factors. Multivariate Behavioral Research **1**(2), 245–276 (Apr 1966)
[9] De Lathauwer, L., De Moor, B., Vandewalle, J.: A multilinear singular value decomposition. SIAM journal on Matrix Analysis and Applications **21**(4), 1253–1278 (2000)
[10] Ercsey-Ravasz, M., Toroczkai, Z., Lakner, Z., Baranyi, J.: Complexity of the international agro-food trade network and its impact on food safety. PLoS ONE **7**(5), e37810 (May 2012)
[11] Escaith, H., Gaudin, H.: Clustering value-added trade: Structural and policy dimensions. World Trade Organization Economic Research and Statistics Division Staff Working Paper No. ERSD-2014-08 (2014)
[12] Garlaschelli, D., Matteo, T.D., Aste, T., Caldarelli, G., Loffredo, M.I.: Interplay between topology and dynamics in the world trade web. The European Physical Journal B **57**(2), 159–164 (May 2007)
[13] Harshman, R.A.: Foundations of the parafac procedure: models and conditions for an" explanatory" multimodal factor analysis. UCLA Working Papers in Phonetics (1970)
[14] Hoff, P.D.: Separable covariance arrays via the tucker product, with applications to multivariate relational data. Bayesian Analysis **6**(2), 179–196 (2011)
[15] Kolda, T.G., Bader, B.W.: Tensor decompositions and applications. SIAM review **51**(3), 455–500 (2009)
[16] Kroonenberg, P.M., De Leeuw, J.: Principal component analysis of three-mode data by means of alternating least squares algorithms. Psychometrika **45**(1), 69–97 (1980)

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77977-1_31

[17] Lin, X., Dang, Q., Konar, M.: A network analysis of food flows within the united states of america. Environmental Science & Technology **48**(10), 5439–5447 (May 2014)
[18] Nadler, P., Arcucci, R., Guo, Y.K.: Data assimilation for parameter estimation in economic modelling. In: 2019 15th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS). IEEE (Nov 2019)
[19] Phan, A., Cichocki, A., Tichavský, P.: On fast algorithms for orthogonal tucker decomposition. In: 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 6766–6770 (2014)
[20] Puma, M.J., Bose, S., Chon, S.Y., Cook, B.I.: Assessing the evolving fragility of the global food system. Environmental Research Letters **10**(2), 024007 (Feb 2015)
[21] Shutters, S.T., Muneepeerakul, R.: Agricultural trade networks and patterns of economic development. PLoS ONE **7**(7), e39756 (Jul 2012)
[22] Song, W.M., Matteo, T.D., Aste, T.: Hierarchical information clustering by means of topologically embedded graphs. PLoS ONE **7**(3), e31929 (Mar 2012)
[23] Spearman, C.: "general intelligence, " objectively determined and measured. The American Journal of Psychology **15**(2), 201 (Apr 1904)
[24] Stock, J.H., Watson, M.W.: Forecasting using principal components from a large number of predictors. Journal of the American statistical association **97**(460), 1167–1179 (2002)
[25] Thurstone, L.L.: Multiple-factor analysis; a development and expansion of The Vectors of Mind. University of Chicago Press (1947)
[26] Torreggiani, S., Mangioni, G., Puma, M.J., Fagiolo, G.: Identifying the community structure of the food-trade international multi-network. Environmental Research Letters **13**(5), 054026 (May 2018)
[27] Tucker, L.R.: Some mathematical notes on three-mode factor analysis. Psychometrika **31**(3), 279–311 (1966)
[28] Verma, A., Angelini, O., Matteo, T.D.: A new set of cluster driven composite development indicators. EPJ Data Science **9**(1) (Apr 2020)
[29] Verma, A., Vivo, P., Matteo, T.D.: A memory-based method to select the number of relevant components in principal component analysis. Journal of Statistical Mechanics: Theory and Experiment **2019**(9), 093408 (sep 2019)
[30] Von Luxburg, U.: A tutorial on spectral clustering. Statistics and computing **17**(4), 395–416 (2007)
[31] Wu, F., Guclu, H.: Global maize trade and food security: Implications from a social network model. Risk Analysis **33**(12), 2168–2178 (May 2013)

A Time complexity

The HHSC algorithm is composed by two components, i.e. Tucker decomposition and the DBHT clustering algorithm. Assuming a third order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ with $I_1 = I_2 = I_3 = I$, the time complexity of the DBHT algorithm is of the order $\mathcal{O}(I^3)$ [22] for each mode of the tensor. Regarding the Tucker

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77977-1_31

decomposition with Tucker rank $\in \mathbb{R}^{R_1 \times R_2 \times R_3}$ with $R_1 = R_2 = R_3 = R$, the time complexity is in the order of $\mathcal{O}(I^3 R + IR^4 + R^6)$ [19]. Being the HHSC the combination of the two algorithms, also the time complexity follows. However, being R of much smaller dimension of I, the latter dominates the algorithm's running time.

B Data filtering

The complete dataset corresponds to 255 countries, 425 products, and 33 years. We first filter the countries that were inactive for more than 10 years and the products for which there were no transactions for more than 10 years. We then filter the dataset with respect to sparseness. In particular, we filter out countries and products for which the density is less than 1%. This resulted in a final dataset of 128 countries, 137 products, and 33 years.

C Clustering with growth related size of the nodes

In this Appendix we report the same clustering plots reported in Figures 2-4 in which the dimension of the nodes is proportional to the total amount (in dollars) exchanged during the period analysed.

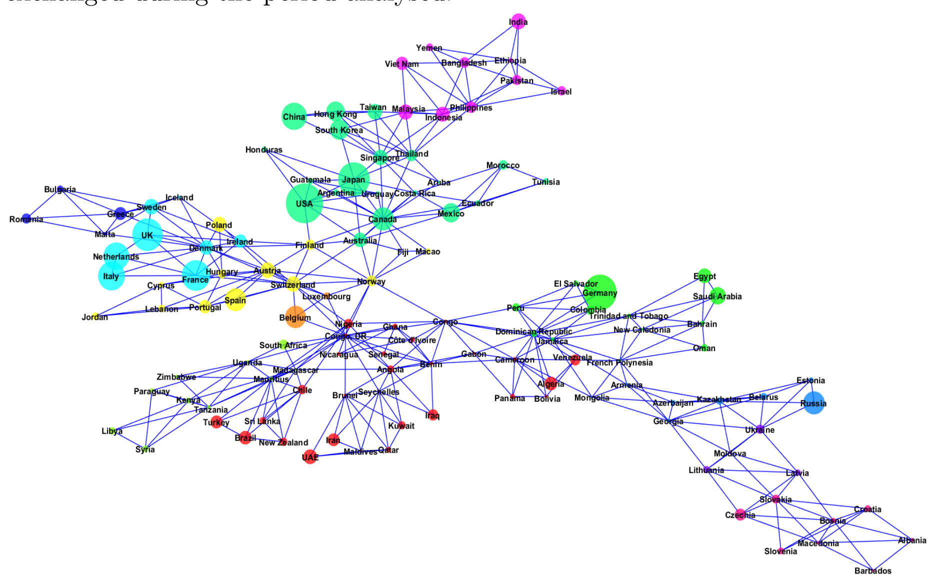

Fig. 6: HHSC clustering for the imports. The size of the nodes is proportional to the amount of products (in $) the country import between 1986 and 2018.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77977-1_31

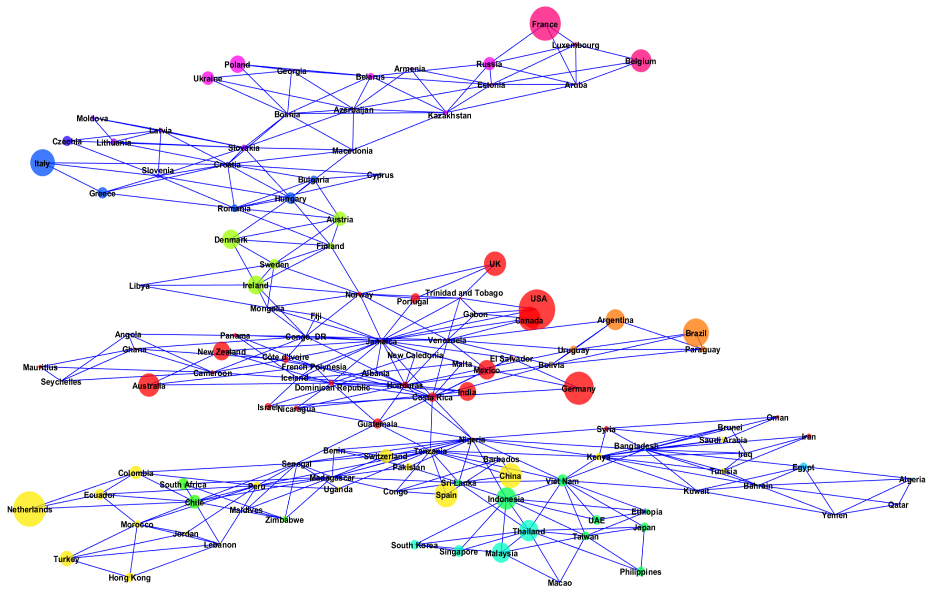

Fig. 7: HHSC clustering for exports. The size of the nodes is proportional to the amount of products (in $) the country exported between 1986 and 2018.

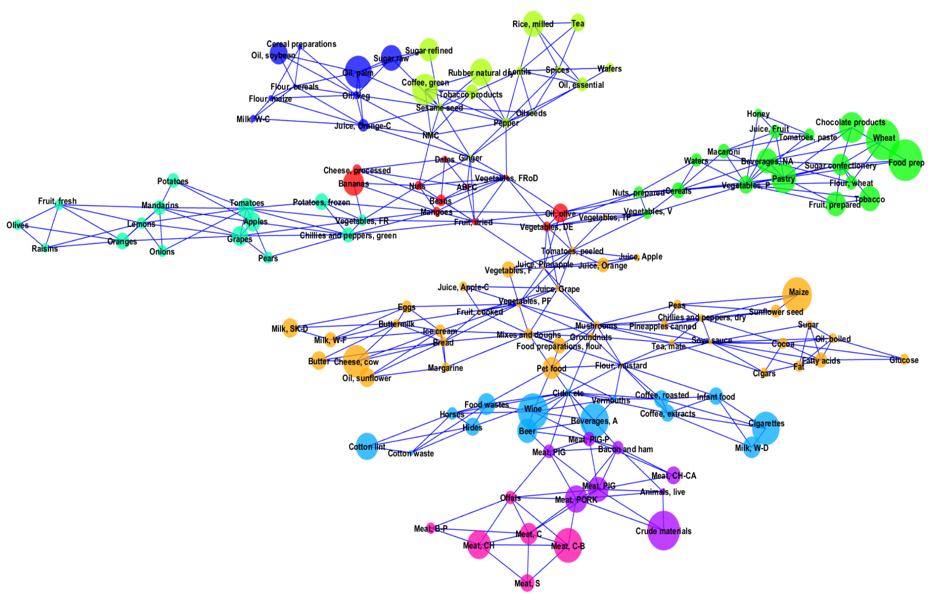

Fig. 8: HHSC clustering for products. The size of the nodes is proportional to the amount of the each product (in $) all the countries imported and exported between 1986 and 2018.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77977-1_31