Modeling International Trade and Tariffs

Elaine Liu
Mentor: George Stepaniants
Advisor: Prof. Jörn Dunkel

August 3, 2023

CONTENTS

Abstract

International trade allows for countries to expand their markets, increase the efficiency of their industries, and influence the countries with which they trade. Tariffs are a tax levied on trade imports and have become the top strategy employed by countries to protect domestic markets and exert their economic influence in response to a variety of political, economic, and natural events. Although unrestricted free trade is supported by many economic theories, targeted placements of tariffs are still used to protect new industries, regulate industry practices, and discourage political actions of other countries. The impacts of properly-placed tariffs have on the networks of international trades can be better understood with dynamical models of trade and tariff interactions. However, the trade and tariff data needed to build these models is frequently noisy and incomplete, and the size of the network is large, with over 200 countries in the world, all posing additional challenge in studying this dynamical relationship. In this work, we begin this analysis by studying how best to compress the far-from-perfect trade and tariff network data into a coordinate space more amenable for dynamical modeling. We study three classical variants of dimensionality reduction: PCA, sparse PCA, and non-negative matrix factorization, as well as present a novel method called the graphical Haar basis. We discuss the implications that these basis constructions have on the dynamical modeling of trade and tariffs.

1 Introduction

The interplay between international trade and imposed tariffs is an essential component of the global economy, with far ranging effects on economic globalization and internationalization, spreading of crises, and economic shocks [Fag16; LM20]. International trade arises from the differences in countries consumer demand and their ability to produce a set of products, leading countries to specialize in particular product trades where they have a *comparative advantage* [Ric05] over their neighbors. Trade has numerous benefits including increasing efficiency of firms and industries, expanding product varieties on the market, and driving research and product development through competition. This increase in competition between domestic and international firms however can affect a country's domestic market, particularly

in areas where it is unable to compete due to lack of resources, labor, or history of development in that field.

As a result, *free trade* between countries is often countered by *protectionist* policies [Faj+20], through import constraints such as tariffs, quotas, and product standard regulation, as well as subsidies for the production and export of domestic products [Kru87]. These policies aim to control the import of foreign products and increase the use and export of domestic products. Tariffs are one of the most important tools used to enforce trade barriers, levied as a tax on an imported product either at a fixed fee (specific tariff) or as a percentage of the item's value (ad valorem tariff).

Promoting free trade and reducing tariffs have historically been viewed to have positive long-term effects on the world economy [Ber96]. Various *preferential trade agreements* have been created to lower or eliminate tariffs between countries including the establishment of free trade areas such as the EU Economic and Monetary Union or the US NAFTA as well as agreements between developed and developing countries like the US Generalized System of Preferences. Outside of such preferential trade agreements, the World Trade Organization (WTO), comprised of 159 world countries, imposes the *most favored nation* (MFN) policy such that members must charge a fixed product tariff irrespective of the importing WTO country.

The time-dependence of trade and tariffs present a rich set of dynamical behaviors. First, tariffs can adapt depending on the exports in which a country specializes which change depending on its level of development and various global trends. Furthermore, tariffs are inherently tied to the politics of world countries. They can be used as a form of political retaliation as in the case of the EU which levied tariffs on US exports of bourbon and Harley Davidson's in response to US tariffs on EU steel and aluminum imports [HAR18]. As a more extreme example, the US-China trade war witnessed a back-and-forth increase in tariffs on US agricultural products and Chinese steel, aluminium, and various technologies [FK22]. Lastly, various global trends and events such as climate change influence the dynamics of trades and tariffs like the recent "carbon tariff" which taxes carbon-intensive imports from developing countries [CA08]. Building mechanistic data-driven models of international trade and tariff interaction are crucial for guiding the proper placement of tariffs to achieve desired economic effects [Joh60] without leading to various political and economic instabilities.

In this paper, Section 2 surveys existing models of trade-tariff interaction. Section 3 investigates the sparsity and hierarchical structure our data sets and justifies our choices of normalization methods. Section 4 investigates ways to capture trade and tariff signals with Principal Component Analysis (PCA) and Sparse Principal Component Analysis (SPCA); Section 5 likewise, with our proposed Haar basis construction. The resulting bases from PCA, SPCA, and Haar wavelets are visualized in Section 6. Moreover, Section 7 studies non-negative matrix factorization (NMF) as a remedy the lack of interpretability of some of the aforementioned techniques on tariff data. Finally, we discuss next steps on how these basis decompositions will be used to jointly model trade and tariff dynamics in Section 8 with a brief conclusion in Section 9.

2 BACKGROUND

A trade is a purchase and sale of goods from an exporting country to an importing country, for our purposes, measured in its US dollars value. International trade, in particular, is subject to the influences of political relations, which often show up in the form of tariffs. A tariff is a tax imposed by one country on the goods and services imported from another country to influence it, raise revenues, or protect competitive advantages. One notable tariff in history is the multi-decade dairy tax imposed by Canada on the United States that lasted until early 2022. The tariff charged an astounding 271% on liquid milk and 298% on butter.

2.1 Data set overview

The trade and tariff data sets we study are provided by Professor In Song Kim (MIT Political Science) in the form of CSV files. Trade data features annualized trading activities among 243 countries across 28 years, from 1989 to 2016. The product of each trade is given in a two-digit code in the Harmonized System. The Harmonized System (HS) is a standardized numerical method of classifying traded products, published by the World Customs Organization (WCO) and it is consistent across all countries. At the two-digit level, the HS code partitions products into 97 categories. The trade data set has a total of 19,112,043 entries for recorded trade transactions. Each entry specifies the importer, exporter, value (in US dollars), and year of trade. Additional information include the trade registration code and whether or not the good is re-imported, which we opt to neglect. Moreover, we choose to remove 45 countries from the data set, as well as trades that have the same importing and exporting countries. The neglected countries are typically disputed territories (e.g. Montenegro, Taiwan) or small countries with scarce data, and the neglected trade lack contextual explanation. In summary, our trade data can be represented mathematically as a tensor

$$T = \{T^p_{i,j,t} : i, j \in [N], t \in [Y], p \in [P]\} \tag{1}$$

across $N = 198$ countries, $Y = 28$ years, and $P = 97$ HS2 product categories.

Similarly, the tariff data set contains tariff information for 50 countries across 22 years, from 1996 to 2017. For each of the 42,731,540 entries, the importer, exporter, year, tariff category, product, and percentage duty are given. The product categories in the tariff data set are given at higher levels of granularity, in the form of a 6-digit, 8-digit, or 10-digit HS code, which provides finer categorization of products. The more granular product information can be coarsened to 96 HS2 codes by truncating to the first 2 digits to match the trade data set. This coarsening step creates repeated entries with different duty percentages. Section 3.2 discusses the aggregation of these repeated entries.

The tariff type of each entry belongs to one of the following categories: most favorable nation (MFN), free-trade areas, zone-zone duties, generalized system of preferences (GSP), least developed countries (LDC), other preferences not otherwise specified, and general duties. By definition, MFN tariffs applied by an importer country must be equal for all WTO countries which are exporting that good. Other types of tariffs are not subject to this requirement. We choose to only include MFN tariffs in our study, as all other tariff types are exceptional cases usually tied to free trade agreements.

To summarize, our MFN tariff data can be represented by the tensor

$$\tau = \{\tau_{i,t}^{p} : i \in [N], t \in [Y], p \in [P]\} \tag{2}$$

here defined across $N = 50$ countries, $Y = 22$ years, and $P = 96$ HS2 product categories.

We reuse the variables N, Y, and P for the number of countries, years, and products when discussing trade and tariff data but we distinguish these variables when both trade and tariff data are used in the same context. Unless it is necessary, we often remove the product superscript and write $T_{i,j,t}$ and $\tau_{i,t}$ when describing trades and tariffs for a generic product.

As a final note, the 97 HS2 product categories are often too large for us to display and analyze. Throughout this paper, we frequently aggregate these 97 products into a set of 15 aggregate product classes $C_1, \ldots, C_{15} \in [P]$ agreed upon by the WTO: https://www.foreign-trade.com/reference/hscode.htm.

2.2 Network Modeling

The trade and tariff data are defined on a directed graph $G = ([N], E)$, across $n = 198$ countries with $|E| = \binom{n}{2}$ directed trade edges.

- **Vertices** $i \in [N]$ represent countries.
- **Vertex values** τ_i represent MFN tariff imposed by the country i.
- **Directed edges** $e = (i, j) \in E$ denotes an established trade relationship from exporter country j to importer country i, consistent with notation from physics and dynamical systems theory.
- **Edge weights** $w_{i,j} = w(T_{i,j})$ represents a function of trade $T_{i,j}$ in US dollars (see Section 3.4) from exporter country j to importer country i.

Modeling trades and tariffs on graphs enables us to use graph-theoretical tools for signal decomposition. We relate these definitions to our problem with Def 1.

Definition 1. [Network] A network is a graph $G = (V, E)$, consisting of the set of countries as vertices in V and directed edges between any two countries in E.

Unless otherwise specified, all networks in this paper are assumed to be fully-connected and directed with no self-loops.

2.3 Dynamical Models in Literature

Three main families of international trade models exist in the literature: gravity model, random graphs, and stochastic difference/differential equation (SDE) models.

2.3.1 Gravity Model

The gravity model is based off of the experimentally observed phenomenon that the flow of trade between two countries decreases with their distance, much like the gravitational force between two planets. Motivated by the law of gravitation, this law takes the general form

$$T_{i,j} = G \frac{m_i^{\beta_1} m_j^{\beta_2}}{d_{i,j}^{\beta_3}} \eta_{i,j} \tag{3}$$

originally proposed by Walter Isard in 1954 [Isa54]. Here $T_{i,j}$ is the directed trade from exporter country j to importer country i, $d_{i,j}$ is the geodesic distance between the two countries, and $\eta_{i,j}$ is an error term with mean one. Importantly, the masses m_i, m_j correspond to the 'size' of the countries measured either by their GDP, population, or other factors. Here the 'gravitational constant' G and factors $\beta_1, \beta_2, \beta_3$ can be learned from data or set to default values such as one. Since it's conception in 1954, the gravity model has had numerous extensions including dynamical formulations as well as dependence on tariffs, geography, and cultural commonalities [Bha+08; GH22]. For, example a time-dependent gravity model that incorporates tariffs may take the form

$$T_{i,j,t} = G \frac{m_i^{\beta_1} m_j^{\beta_2}}{d_{i,j}^{\beta_3}(1+\tau_{i,j,t})^{\beta_4}} \eta_{i,j,t} \tag{4}$$

where t indexes the year of trade and $\tau_{i,j,t}$ is the tariff applied by importer country i on exporter country j.

2.3.2 Random Graph Models

Random graph models aim to describe the discrete year-to-year changes of trade edges in the international trade network as a random weighted directed graph. Let's denote $T_t = \{T_{i,j,t}\}_{i,j=1}^N$ as the directed trade matrix for year t. A popular model is the discrete temporal exponential random graph model (DTERGM) [HFX10; SSB22] given by

$$\mathbb{P}\Big(T_{t+1} | T_t\Big) = \frac{\exp(\langle \tau, g(T_{t+1}, T_t)\rangle)}{Z(\tau)} \tag{5}$$

where $\tau \in \mathbb{R}^p$ is a parameter vector that needs to be learned and $g : \mathbb{R}^{N\times N} \times \mathbb{R}^{N\times N} \to \mathbb{R}^p$ is a vector of network statistics. Examples of useful network statistics when $T_t \in \{0,1\}^{N\times N}$ is a binary matrix include

$$\begin{aligned}
\text{Stability: } & \frac{1}{N^2}\sum_{i,j=1}^N T_{i,j,t+1}T_{i,j,t} + (1-T_{i,j,t+1})(1-T_{i,j,t}) \\
\text{Reciprocity: } & \Big(\sum_{i,j=1}^N T_{j,i,t+1}T_{i,j,t}\Big) \Big/ \Big(\sum_{i,j=1}^N T_{i,j,t}\Big) \\
\text{Transitivity: } & \Big(\sum_{i,j=1}^N T_{k,i,t+1}T_{j,i,t}T_{k,j,t}\Big) \Big/ \Big(\sum_{i,j=1}^N T_{j,i,t}T_{k,j,t}\Big).
\end{aligned} \tag{6}$$

Here stability counts the number of trade edges which stay activated or unactivated from one year to the next, reciprocity studies whether a trade from j to i in one year, will cause a trade from i to j in the next, and transitivity studies whether a trade from i to j and from j to k in one year leads to a trade from i to k in the next. By collecting such network statistics into a vector $g(T_{t+1}, T_t)$ and learning their corresponding rates in the parameter vector τ, we can learn the dominant operations that dictate changes in the trade network from year to year.

Extensions of DTERGMs include the separable temporal exponential random graph model (STERGM) [KH14] which model formation and dissolution of trade edges as separate

processes. Yet other approaches model the generation of trade networks through preferential attachment and multiplicative processes which are consistent with the trade flow and degree distribution statistics of empirical trade networks [GMG19].

Such random graph models cannot be directly applied to predict the dynamics of trade networks but may provide useful insights into the formation, dissolution, and interaction of trade edges between countries.

2.3.3 SDE Models

Given the matrix of trades $T_i = \{T_{i,j,t}\}_{i,j=1}^{N}$ at year t, a stochastic difference equation model takes the form

$$T_{t+1} = f(t, T_t) + g(t, T_t) W_t \tag{7}$$

where $f : \mathbb{R}_+ \times \mathbb{R}^{N\times N} \to \mathbb{R}^{N\times N}$ and $g : \mathbb{R}_+ \times \mathbb{R}^{N\times N} \to \mathbb{R}$ and the entries of $W_t \in \mathbb{R}^{N\times N}$ are independent standard Gaussian for all time $t \in \mathbb{N}$. Such models have been constructed to explore the effects of shock propagation in trade networks using discrete models of disease spreading [SBS19].

An alternative to discrete time models are Itô stochastic differential equations

$$dT = f(t, T)dt + g(t, T)dW_t \tag{8}$$

where $t \in \mathbb{R}_+$ is now a continuous time variable rather than a discrete index of the year. Here $W_t \in \mathbb{R}^{N\times N}$ is now a matrix-valued Brownian motion. Prior works have fit continuous SDE models on international trade network data to model economic growth and the temporal dependence of export products within a country [CBS16; TCS18a; TCS18b].

To the best of our knowledge, most existing work on international trade networks do not study the effect of tariffs on their dynamics through SDE models. In the following sections, we discuss important preprocessing and dimensionality reduction steps for transforming trade and tariff data into a form that allow us to build dynamical models.

3 Descriptive Statistics

3.1 Trade

Trade data is represented by a 4-tensor, $T \in \mathbb{R}^{198\times 198\times 28\times 97}$. The entry (i, j, k, l) denotes the US dollar value of goods in category l, imported by country i from country j, in year $(1988 + k)$. For example, $T[1, 198, 28, 1] = x$ means Aruba imported x US dollar's worth of animal products from Zimbabwe in 2016. The set of the 198 countries featured in this data set is highlighted in Figure 1a.

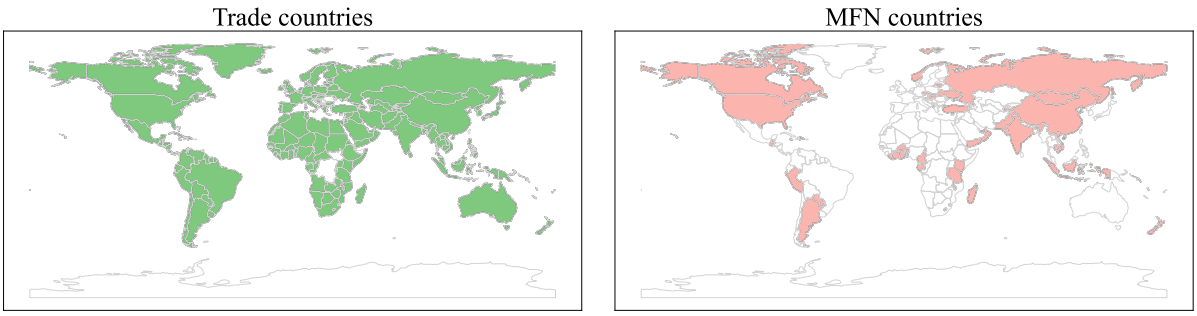
The trade data tensor T is sparse, especially in the earlier years. Across all years and products, zero values account for 84.7% of the entries of the tensor T. In year 1989, a 95.9% proportion of all trade edges were zero, whereas in year 2016, this value dropped to 81.1%. The decreasing sparsity highlights growing trade relations between countries as well as increased reporting of trades in our dataset.

3.2 Tariff

The MFN tariffs are represented by a 3-tensor, $\tau \in \mathbb{R}^{50\times 22\times 96}$. The entry (i,j,k) denotes the tariff rate imposed by country i in year $1988+j$ in product k. The set of the 50 countries featured in this data set is highlighted in Figure 1. The dis-alignment between the size of trade and tariff data is an artifact of how these datasets were collected.

MFN tariffs have the special feature that the same tariff rate applies to all countries, as enforced during the WTO GATT agreements. As such, these values can be stored at vertices of the country graph G, instead of its edges, allowing for simpler analyses of this data (e.g. spectral decomposition and NMF).

Similar to trade T, tariffs τ are sparse and zero-inflated. In particular, 26.7% entries in τ are zero. The highly-granular tariff data is coarsened by taking the non-zero mean of all entries that fall into the same HS2 category.



(a) The set of 198 countries in the trade data set. (b) The set of 50 countries in the MFN data set.

Figure 1: Countries featured in the trade and MFN tariff data set.

3.3 Hierarchical clustering

Hierarchical clustering identifies the structure of a large data set through a binary partition process that puts data points with similar features into one of the two clusters. Given the metric for similarity (e.g. Euclidean distance, Manhattan distance, inner-product) between a pair of data points, the algorithm merges the two clusters that are the most similar, iteratively, until all data points end up in the same cluster. In particular, the inter-cluster distance between clusters u and v is computed with the Ward variance minimization algorithm

$$d(u,v) = \sqrt{\frac{|v|+|s|}{|X|}\cdot d(v,s)^2 + \frac{|v|+|t|}{|X|}\cdot d(v,t)^2 - \frac{|v|}{|X|}\cdot d(s,t)^2}, \tag{9}$$

where X is the set of all data points, $u = s\cup t$, the union of two merged clusters, and $v = X\backslash u$, the unused clusters [Mül11].

3.3.1 Product Aggregation

Results from hierarchical clustering offers critical insight into the structure among the product classes. For a pair of products $p,q \in [P]$ the product-product correlation is defined by summing

the correlation in binarized trade activities in these two products over all countries $i, j \in [N]$ and years $t \in [Y]$. First define the binarized trade as

$$B^p_{i,j,t} = \mathbf{1}\{T^p_{i,j,t} > 0\}. \tag{10}$$

Then the correlation between the binarized trade networks of two products $p, q \in [P]$ can be computed as

$$\text{corr}(p,q) = \sum_{i,j=1}^{N} \sum_{t=1}^{Y} \frac{\langle B^p_{i,j,t}, B^q_{i,j,t} \rangle}{\|T^p_{i,j,t}\| \cdot \|T^q_{i,j,t}\|}. \tag{11}$$

The distance $d(p,q)$ is thus defined as $1 - \text{corr}(p,q)$.

Hierarchical clustering results is compared with the WTO's 15-class aggregation on the 97 products in Figure 2. The high value concentration along the diagonal signals that products in the same aggregated class follows similar trading patterns.

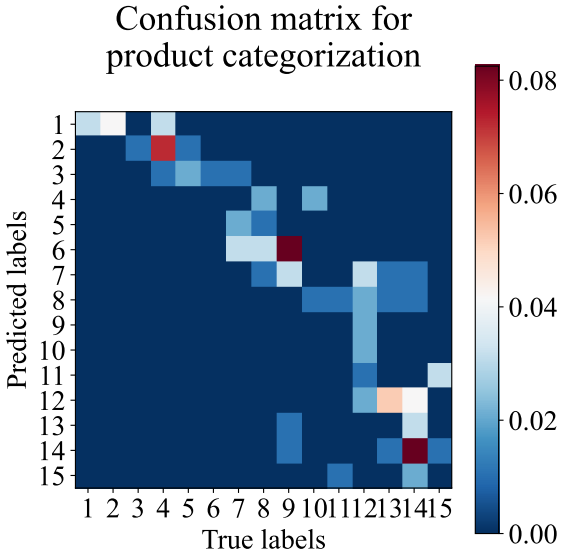


Figure 2: Confusion matrix for hierarchical clustering on the 97 products based on binarized trade patterns. An identical relationship between predicted and true product categories is observed on non-binarized (original) trade data.

3.4 Discussion on Trade Data Normalization

We construct our bases with trade data normalized in the following three ways.

1. No normalization.
2. Taking the logarithm of trade data: $\log(1 + T^p_{ij})$, where T^p_{ij} denotes the trade volume from country j to country i in product p.
3. Log-normalization with GDP correction:

$$\log(1 + T^p_{ij}) - m^p \log(1 + GDP_i \cdot GDP_j) - b^p \tag{12}$$

where m^p and b^p are product-specific correction parameters obtained with orthogonal distance regression (ODR) on $(\log(1+GDP_i \cdot GDP_j), \log(1+T_{ij}^p))$.

Contrary to linear regression that assumes the features (X) to be noisy and observations (Y) noiseless, ODR assumes that both Y and X are susceptible to error. As such, it minimizes the sum of squared perpendicular distances from the data point to its projection on the fitted line defined by m, b by minimizing

$$r_x^T M_x^{-1} r_x + r_y^T M_y^{-1} r_y, \tag{13}$$

with residuals r_x, r_y, and variance-covariance matrices M_x and M_y, respectively. The bias terms, b, aim to capture GDP-independent variation, while the slope terms, m, aim to characterize sensitivity of trade to GDP. An example of ODR is illustrated in Figure 3 on the product class HS 86 (vehicles).

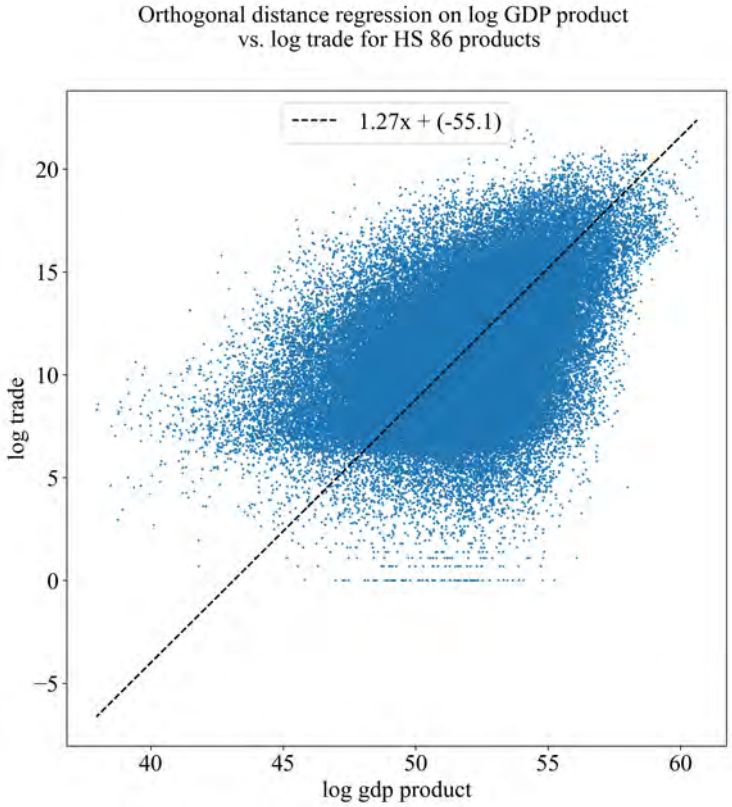


Figure 3: ODR on product class HS 86 (vehicles). The fitted line is the result of ODR, which has a slope of 1.27 and bias of -55.1.

The choice of log-normalization is inspired by literature in economics [Eck01; CG05; Sut97; BS73]. In particular, taking the logarithm allows interpretation of changes in variables as percentage changes, rather than absolute changes, which is particularly helpful for time-series data analysis [MLD08]. Another reason for log-normalization is that it is a useful preprocessing step to standardize data such as trade in US dollars that may span several orders of magnitude (e.g. trade in hundreds, thousands, millions of dollars).

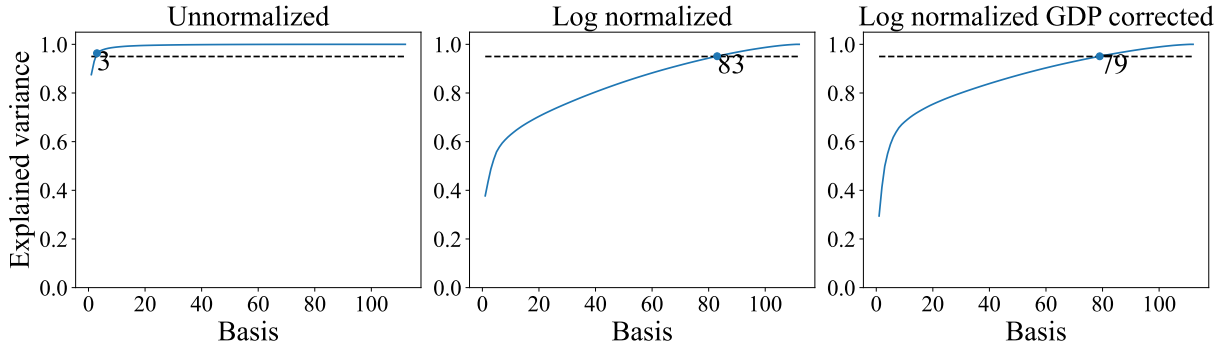


Figure 4: Recovery of trade data as a function of PCA basis size. In this example, more than 95% variance can be explained with the first 3 components on the unnormalized data, as only large trades are captured. Variance recovery rate is much slower in the log normalized and log normalized GDP corrected samples, as normalization de-emphasize outliers. The product-specific GDP correction yields minor improvement, in comparison to the log-normalization.

The rest of this paper focuses on applying the three decomposition methods (PCA, SPCA, and Haar wavelet) applied on different normalizations and comparing their performance. For the purpose of illustration and consistency, we restrict the scope of all examples to the aggregate product class of transportation for the remainder of this paper.

4 Principal Component Analysis (PCA) and Sparse PCA

Principal component analysis (and its variants) is a popular data mining technique to reduce the dimensionality of a dataset. It works to identify the direction in the feature space among which the most variance is explained. For the purpose of this work, the feature space is either country tuples (i, j) for trade data, or individual countries i for MFN tariff data. Samples are obtained across all HS2 product classes in the same aggregated product class across all years. For instance, the aggregated product class "transportation" contains four HS2 product classes: $C = \{85, 86, 87, 88\}$. As a reminder, the trade data set comprises 198 countries across 28 years, hence the sample size is 4×28 for transportation, where each sample is a vector of length 198^2 that characterizes the trade activities in one of the four HS2 products in a particular year.

In mathematical terms, given our trade tensor

$$T = \{T^p_{i,j,t} : i, j \in [N], t \in [Y], p \in [P]\} \tag{14}$$

we first subset this tensor to only include the transportation products $p \in C := \{85, 86, 87, 88\}$ which gives us

$$T^C = \{T^p_{i,j,t} : i, j \in [N], t \in [Y], p \in C\}. \tag{15}$$

In order to prepare this trade tensor for basis decomposition (using PCA, sparse PCA or Haar decompositions), we then flatten its (t, p) dimensions and (i, j) dimensions to obtain a final matrix of size $X \in \mathbb{R}^{Y|C| \times N^2}$.

The final goal of all trade data basis decompositions is to approximate the trade data as a weighted sum of basis matrices. Specifically, for each trade matrix $T_t^p = \{T_{i,j,t}^p : i, j \in [N]\}$ for a given year $t \in [Y]$ and transportation-related product $p \in C$ we approximate it using $b \in \mathbb{N}$ basis elements as

$$T_t^p \approx y_{t,1}^p U_1 + \ldots + y_{t,b}^p U_b \tag{16}$$

where $\{y_{t,l}^p \in \mathbb{R}\}_{l=1}^b$ are the weights and $\{U_l \in \mathbb{R}^{N \times N}\}_{l=1}^b$ are the basis matrices.

For MFN tariffs, our tariff tensor is given by

$$\tau = \{\tau_{i,t}^p : i \in [N], t \in [Y], p \in [P]\} \tag{17}$$

and subsetting this tensor to only include the transportation products $p \in C := \{85, 86, 87, 88\}$ gives us

$$\tau^C = \{\tau_{i,t}^p : i \in [N], t \in [Y], p \in C\}. \tag{18}$$

Again to prepare this MFN tariff tensor for basis decomposition, we flatten its (t, p) dimensions to obtain a final matrix of size $X \in \mathbb{R}^{Y|C| \times N}$.

Similarly, using basis decomposition method for tariffs (PCA, sparse PCA, or Haar) we approximate each tariff vector $\tau_t^p = \{\tau_{i,t}^p : i \in [N]\}$ for a given year $t \in [Y]$ and transportation-related product $p \in C$ as

$$\tau_t^p \approx z_{t,1}^p v_1 + \ldots + z_{t,b}^p v_b \tag{19}$$

using $b \in \mathbb{N}$ basis elements where $\{z_{t,l}^p \in \mathbb{R}\}_{l=1}^b$ are the weights and $\{v_l \in \mathbb{R}^N\}_{l=1}^b$ are the basis vectors.

4.1 PCA

Dimensionality reduction through PCA is accomplished singular value decomposition on feature-wise, mean-centered data matrix, $X \in \mathbb{R}^{n \times d}$. Here and d is the dimension of the data and n the number of data points. X is normalized to have zero mean along each column. Singular value decomposition factors the matrix $X = U\Sigma V^T$, where $U \in \mathbb{R}^{n \times b}$ a unitary matrix, $V \in \mathbb{R}^{d \times b}$ is the matrix of components, and $\Sigma \in \mathbb{R}^{b \times b}$ is a diagonal matrix of the corresponding singular values for $b = \min(n, d)$.

PCA can be equivalently expressed as a iterative optimization process, where in the i-th iteration, the basis $v_i \in \mathbb{R}^d$ that explain the most remaining variance is identified by solving the following

$$X_i = X_{i-1} - X_{i-1} v_i v_i^T, \tag{20}$$

$$v_i = \underset{||v_i||=1}{\arg\max} \{||X_i v_i||^2\} \tag{21}$$

where X_0 is the original mean-centered data.

The result of PCA is a complete, orthogonal basis, where the components are ranked by the variance they explain (formulation given in Section 4.3). Results from PCA serve as a benchmark to evaluate other decomposition methods (SPCA, NMF, and Haar wavelet).

4.2 Sparse PCA

Sparse PCA (SPCA) is a variant of PCA that extracts the set of sparse components that best reconstruct the data. By introducing an L1-penalty, SPCA suppresses small non-zero coefficients in its components when expressed as linear combinations of the original variables. In comparison to the PCA formulation in (20), the iterative optimization for SPCA uses α as a parameter to control for the level of sparsity. This work uses a default value of $\alpha = 1$.

$$X_i = X_{i-1} - X_{i-1} v_i v_i^T, \tag{22}$$

$$v_i = \underset{||v_i||=1}{\arg\min} \{ ||X_i - X_i v_i v_i^T||^2 + \alpha \|v_i\|_1 \}. \tag{23}$$

SPCA has the advantage over PCA that it can avoid noise-related overfitting and yield a more parsimonious, interpretable representation of the original data.

Unlike PCA, SPCA does not guarantee the components w_i to be orthogonal. As such, we compute the SPCA components following the algorithm in Algorithm 1, manually subtracting the explained variance in each component.

Algorithm 1 Sparse PCA

$i \leftarrow 0$
$X \leftarrow X$
while $i < b$ **do**
  $X \leftarrow X - X v v^T$
  $v \leftarrow \arg\min_{||v_i||=1} \{ ||X_i - X_i v_i v_i^T||^2 + \alpha ||v_i|| \}$
  $V[i] = v$
  $i \leftarrow i + 1$
end while

4.3 Explained Variance

For any basis set $\{v_i \in \mathbb{R}^d\}_{i=1}^b$ we can stack this basis column-wise into a matrix $V \in \mathbb{R}^{d \times b}$. Given a data matrix $X \in \mathbb{R}^{n \times d}$ consisting of n samples of dimension d, the variance explained in the data by this basis is defined as

$$\text{variance explained} = \frac{\|XV(V^TV)^+V^T\|_F^2}{\|X\|_F^2} \tag{24}$$

If all columns of V (e.g. all basis vectors v_i) are orthogonal as in the case of PCA, sparse PCA, and the Haar basis discussed below, then this formula simplifies to

$$\text{variance explained} = \frac{\|XV\|_F^2}{\|X\|_F^2}. \tag{25}$$

5 HAAR WAVELET DECOMPOSITION

5.1 Existing Wavelet Decompositions

Comparable to Fourier transformation, the wavelet transform aims to find simpler representations of signals in an alternative basis. Yet it has the advantage over Fourier transform for identifying non-periodic and sparse signals as is the case for trade and tariffs.

The classical continuous wavelet transform for $\mathbb{R}$ consists a set of integrable real-valued functions all generated by the choice of a single "mother" wavelet ψ. The basis is the span of ψ by translation and scaling with continuous parameters. As such, the wavelet basis captures signals at different locations and at different scales, and is frequently used for transient localization and multi-resolution analysis. Typical choices of ψ include the Morse wavelet, Daubechies wavelet, and the "Mexican hat" [Dau88; Mey93].

The discrete wavelet transform relaxes the continuity requirement the translation and scaling parameter of ψ. In doing so, the transform space is drastically compressed. The discrete wavelet transform can be equivalently obtained by discretely sampling the continuous wavelet transform of the same signal.

Discretizing the translation and scaling parameter of ψ makes the Haar wavelet a particularly appealing choice for basis construction, as its support is also discretized. The mother Haar wavelet $\psi(t)$, can be described as

$$\psi(t) = \begin{cases} 1 & 0 \leq t < \frac{1}{2}, \\ -1 & \frac{1}{2} \leq t < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{26}$$

Previous efforts to extend the Haar wavelets to graphs are limited to vertex-based construction [CK03; HVG11]. To the best of our knowledge, there is no systematic Haar wavelet basis construction on graph edges. This work proposes one such construction. In our definition, each basis component is built by applying graph-specific scaling factors to different intervals of the mother wavelet, hence ensuring orthogonality between all components.

5.2 Basis construction

We propose two basis constructions, one on graph edges, and one on graph vertices. The edge construction is used to decompose the trade data set, as flow information is stored on directed edges, whereas the vertex construction is used to decompose the tariff data set, as MFN tariffs are country-specific. For both the edge and vertex construction, the basis construction algorithm starts with a set of weights, and iteratively partitions the set into two roughly equal-sized sets until all sets become singlets. Basis elements are obtained from each partition by assigning different weights to elements in the two sets. To construct a basis on edges, the initial set is the set of values assigned to the edges; to construct a basis on vertices, the initial set is the set of values assigned to vertices. The algorithm is outlined in Algorithm 2.

Algorithm 2 Haar wavelet basis construction

$S \leftarrow \{v_1, v_2, \ldots, v_n\}$
$q \leftarrow [S]$
$B \leftarrow []$
function BISECTION(X)
    $m \leftarrow$ median of X
    $Left \leftarrow \arg(X < m)$
    $Right \leftarrow \arg(X > m)$
    return $Left, Right$
end function
while q **do**
    $Left, Right \leftarrow$ BISECTION$(q.pop())$
    $v \leftarrow$ all-zero array of length $|S|$
    $v[Left] = -1/|Left|$
    $v[Right] = 1/|Right|$
    $B \leftarrow$ add v
    if $Left$ has more than one element **then**
        $q \leftarrow$ add $Left$
    end if
    if $Right$ has more than one element **then**
        $q \leftarrow$ add $Right$
    end if
end while

The Haar bases for trade are defined on the edges of the graphs. For a fixed product, the weight on the directed edge between countries (i, j) is defined by the average trade volume along this edge over all years as follows

$$w_{i,j} = \frac{1}{Y} \sum_{t=1}^{Y} T_{i,j,t}. \tag{27}$$

where $Y = 28$ is the number of total years in our trade dataset. The distance between two edges (i, j) and (k, l) is the absolute difference between their corresponding average trade flows, $|w_{i,j} - w_{k,l}|$. The algorithm is initialized with all edges and their corresponding weights in the set S, as the input to Algorithm 2, then following the below process to construct the basis.

- Find the median among the set of weights, $S = \{w_1, w_2, \ldots\}$. Partition S into S_1, S_2, by the median. Add the vector v to the basis set defined as

$$v_i = \begin{cases} -\frac{1}{|S_1|} & w_i \in S_1 \\ \frac{1}{|S_2|} & w_i \in S_2 \\ 0 & \text{otherwise.} \end{cases} \tag{28}$$

- Repeat the previous step on S_1 and S_2, respectively, until there is a single element in the set.

On the MFN tariff dataset, the Haar bases decomposition is defined on nodes of the graph, where the distance between two nodes is defined as the geodesic distance between the two corresponding countries. Rather than partitioning country-country edges, the Haar bases construction for tariffs performs a binary partition over countries. In this case, the input to Algorithm 2 becomes the set of all countries.

6 Trade & Tariff Basis Visualizations

Trade and tariff decomposition result give critical insight into both the global trade and tariff signals and our normalization and decomposition choices.

As the trade basis components are on the edges, the bases are in $\mathbb{R}^{N^2}$ for the number of countries $N = 198$. Despite the large dimension, most of the variances can be explained with the significantly less components with PCA and SPCA. We choose to visualize the basis components using arrows between countries, where both the arrow and gradient (from light to dark) to denote directionality of the edge, and the thickness to denote the weight on the edge. All basis for trade are visualized in this manner as in Figure 6, Figure 7, and Figure 8, with the only exception of Haar wavelet basis. As the construction for Haar wavelet basis assigns many edges the same weights, and the number of edges in the principal components become too large to visualize, we use an alternative country-based visualization, as done in Figure 7 and Figure 8. Here, the color for country c in a component is computed as

$$k(c) = \sum_{i \neq c} w_{i,c} - \sum_{c \neq j} w_{c,j}, \tag{29}$$

where $w_{i,c}$ is the weight on the incoming edge from country i to country c, and $w_{c,j}$ is the weight on the outgoing edge from country c to country i. The color is determined by re-scaling $k(c)$ with the formula

$$\frac{1}{\tanh(5)} \cdot \tanh\left(\frac{5 \cdot k(c)}{\max_{c'}(k(c'))}\right). \tag{30}$$

As for tariffs, since tariffs are country-specific, the basis are in $\mathbb{R}^n$ for the number of countries $N = 50$. The size of the basis is dependent on the number of HS2 products in the aggregated class. Tariff bases are similarly visualized with colors re-scaled with (30).

6.1 Trade

PCA, SPCA, and Haar wavelet decomposition are applied on the transportation trade data (with three normalizations: no normalization, log normalization, and log normalization with GDP correction). Evaluating the recovery rate of the nine experiments led to following four observations.

First is the advantage of SPCA. Despite having sparse components, its variance recovery rate is comparable to that of PCA, while enforcing sparsity. PCA and SPCA identified similar edges in each components, as can be seen by comparing the first two columns in Figure 6, Figure 7 and Figure 8.

Second is the loss in sparsity with normalization. The principal components for normalized data (Figure 7, Figure 8) are much denser and weaker than that in the unnormalized data

(Figure 6). The loss of sparsity is even more apparent in Haar wavelet decomposition to the point that variance is becomes extremely hard to capture with Haar components, evidence by the slow convergence in recovery rates in Figure 5. More than 10,000 basis components are required to achieve a $> 95\%$ explained variance for both normalizations (not shown in figure). This slow convergence can be explained with the way Haar wavelets are constructed—at each partition, the binary partitions assign identical weight to edges in each subset without weighing the importance among edges in each subset. Further work is needed to find suitable normalization methods that reduce the strength of outliers without completely erasing their strength, so as to preserve sparsity in the decomposed basis.

Third is the loss of interpretability with normalization. Results given by all three decomposition methods on the unnormalized data agree with our empirical knowledge—US exports cars to Canada and Mexico; Germany exports to the UK, France, and Italy; and Japan and China exports to the USA. However, with normalization, the principal components from PCA and SPCA are no longer sparse, and identified smaller, less active countries like Uzbekistan, Liberia and the Marshall Islands. The principal Haar wavelets are so dense that our software fails to visualize the more-than-20,000 edges on the same graph. The loss of interpretability is yet another reason to search for better normalization techniques.

Fourth is the strength of certain outliers. With normalization, many bigger economies including USA and Canada disappear from the principal components with GDP normalization. However, certain trade routes the are linked to Japan and Germany persist. The robustness of these edges against normalization signals their dominating position on a global scale.

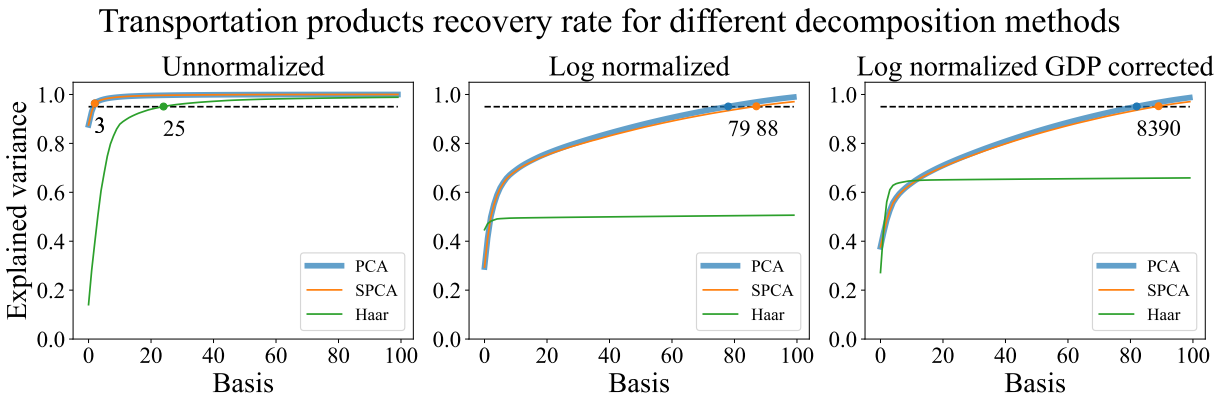
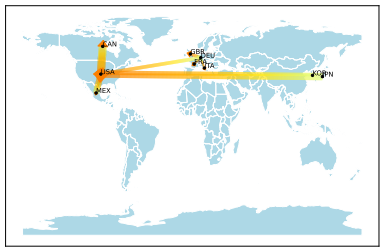
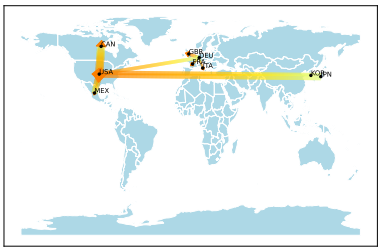


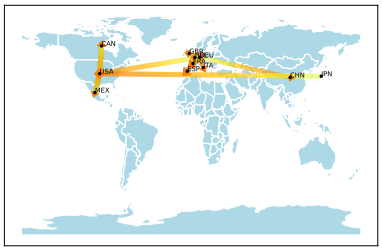
Figure 5: Scree plot for the first 100 basis components under decompositions and normalizations.



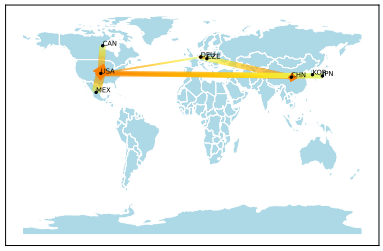
(a) PCA component 1, $\sigma = 0.88$.



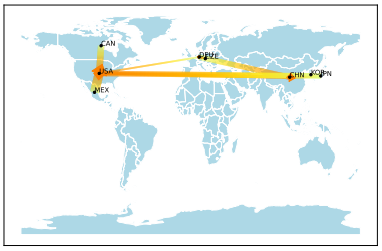
(d) SPCA component 1, $\sigma = 0.88$.



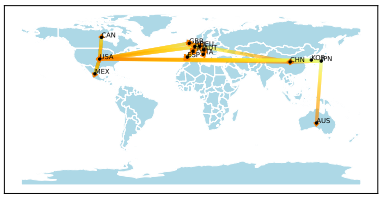
(g) Haar component 1, $\sigma = 0.14$.



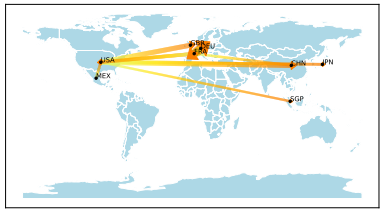
(b) PCA component 2, $\sigma = 0.93$.



(e) SPCA component 2, $\sigma = 0.93$.

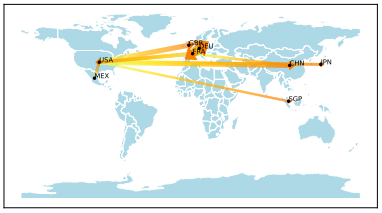


(h) Haar component 2, $\sigma = 0.28$.



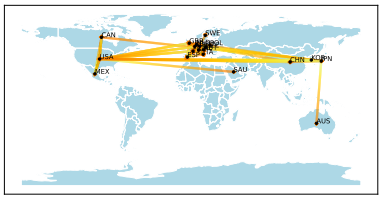
(c) PCA component 3, $\sigma = 0.96$.

First 3 PCA components.



(f) SPCA component 3, $\sigma = 0.96$.

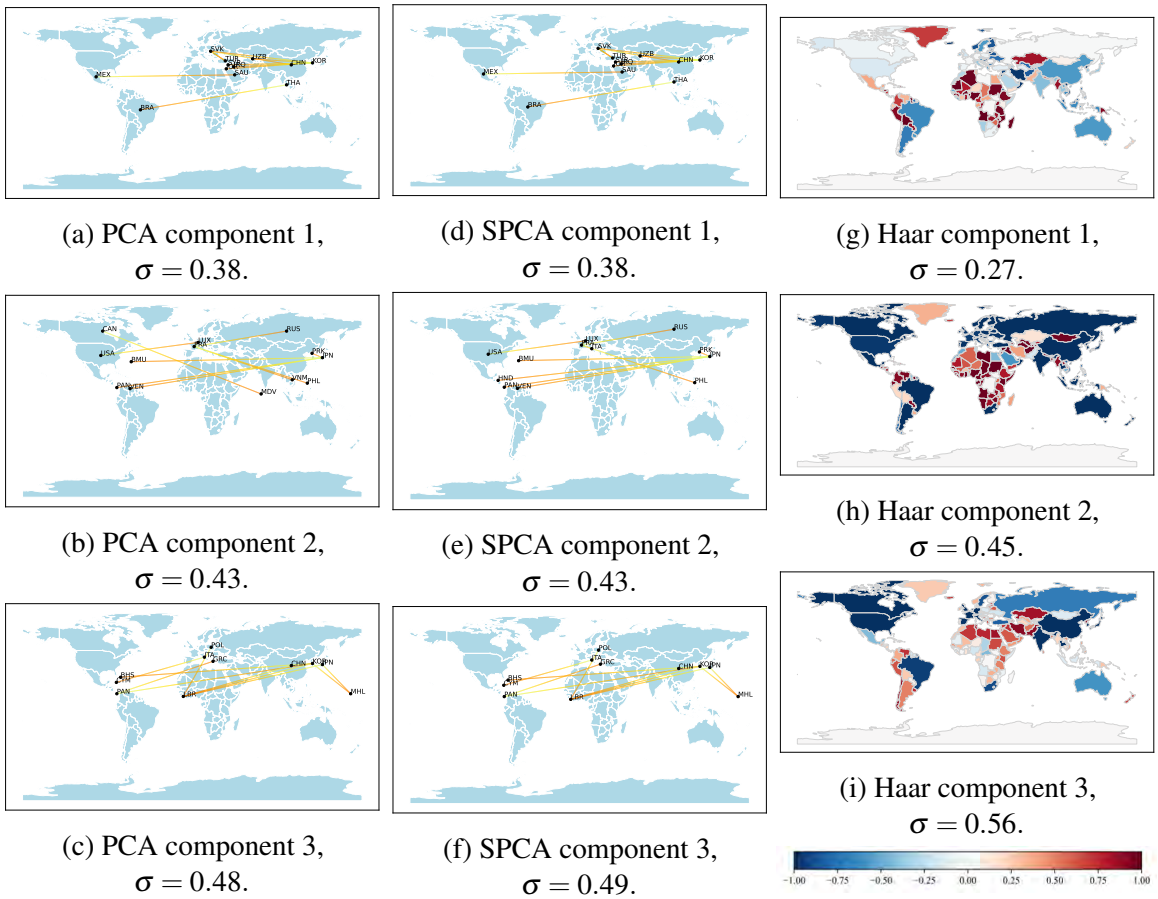
First 3 SPCA components.



(i) Haar component 3, $\sigma = 0.40$.

First 3 Haar components.

Figure 6: First 3 components for three types of basis decompositions (PCA, SPCA, and Haar), no data normalization, with explained variances σ.



(a) PCA component 1, $\sigma = 0.38$.

(b) PCA component 2, $\sigma = 0.43$.

(c) PCA component 3, $\sigma = 0.48$.

First 3 PCA components.

(d) SPCA component 1, $\sigma = 0.38$.

(e) SPCA component 2, $\sigma = 0.43$.

(f) SPCA component 3, $\sigma = 0.49$.

First 3 SPCA components.

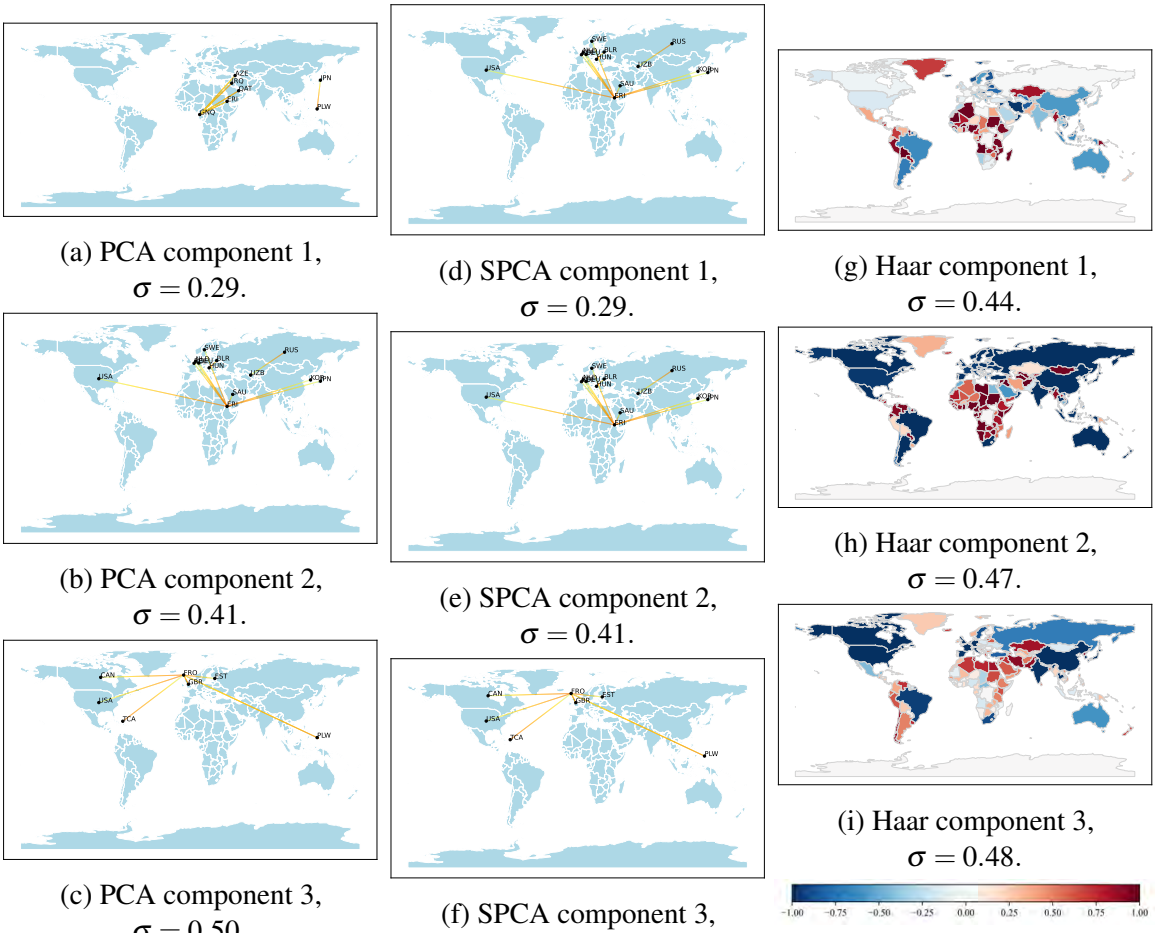
(g) Haar component 1, $\sigma = 0.27$.

(h) Haar component 2, $\sigma = 0.45$.

(i) Haar component 3, $\sigma = 0.56$.

First 3 Haar components

Figure 7: First 3 components for three types of basis decompositions (PCA, SPCA, and Haar), log-normalized trade data, with explained variances σ.



(a) PCA component 1, $\sigma = 0.29$.

(b) PCA component 2, $\sigma = 0.41$.

(c) PCA component 3, $\sigma = 0.50$.

First 3 PCA components.

(d) SPCA component 1, $\sigma = 0.29$.

(e) SPCA component 2, $\sigma = 0.41$.

(f) SPCA component 3, $\sigma = 0.50$.

First 3 SPCA components.

(g) Haar component 1, $\sigma = 0.44$.

(h) Haar component 2, $\sigma = 0.47$.

(i) Haar component 3, $\sigma = 0.48$.

−1.00 −0.75 −0.50 −0.25 0.00 0.25 0.50 0.75 1.00

First 3 Haar components.

Figure 8: First 3 components for three types of basis decompositions (PCA, SPCA, and Haar), log-normalized and GDP-corrected trade data, with explained variances σ.

6.2 Tariff

While PCA, SPCA, and Haar decompositions are effective in capturing data variation and are easy to visualize, they lack interpretability. Recovery rates with each decomposition method is given in Figure 9, with PCA and SPCA significantly out-performing Haar wavelets. Additionally, Figure 10 illustrates the first 3 principal components from each decomposition technique and variance explained up to each.

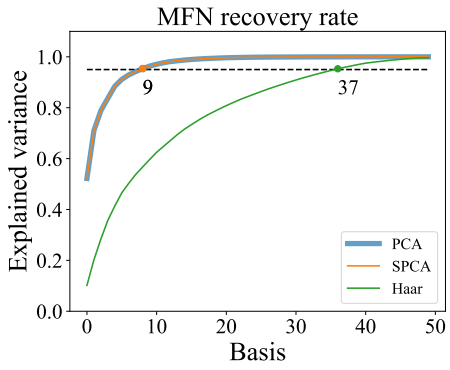
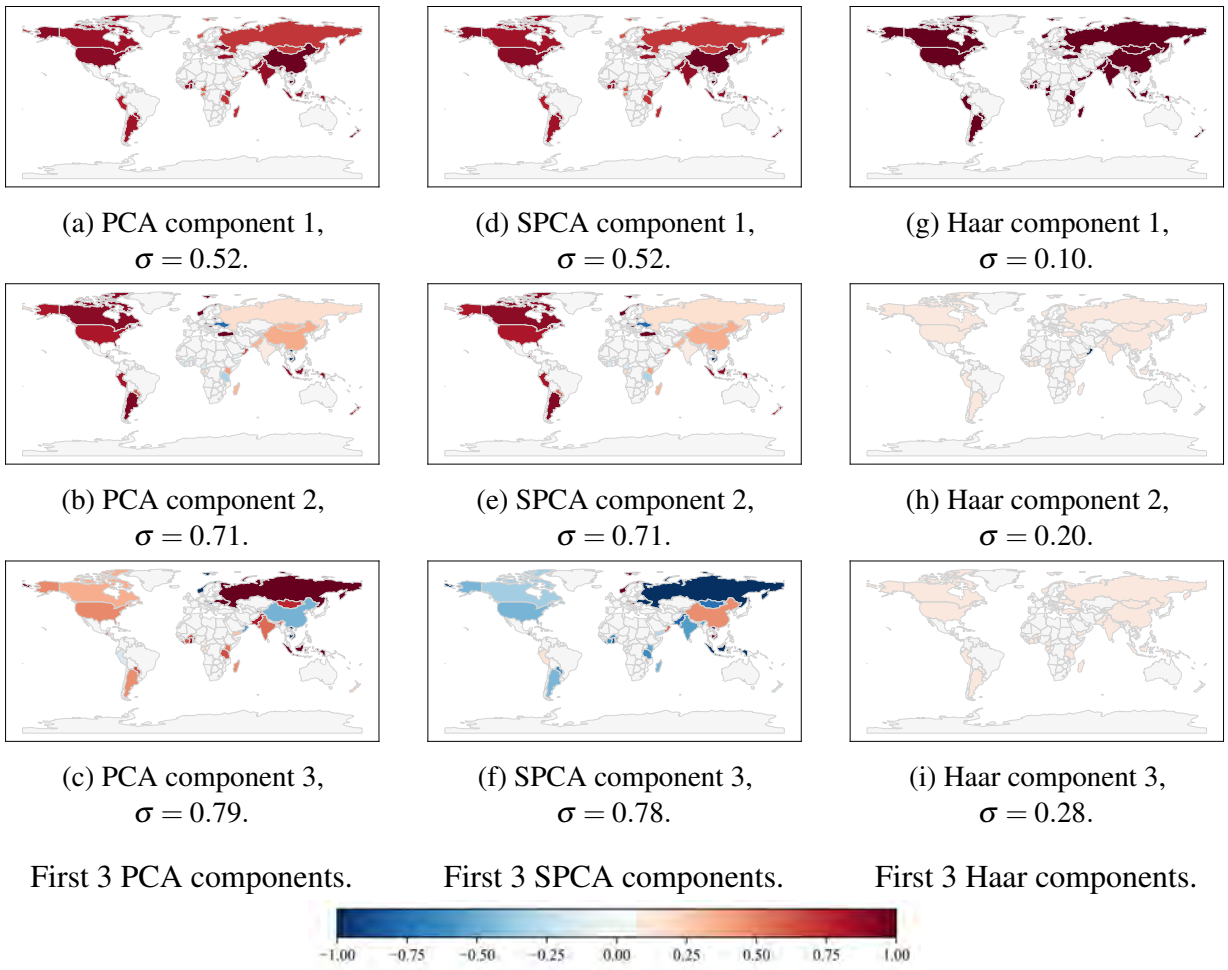


Figure 9: Recovery of MFN data as a function of basis size for various decomposition methods. In this example, PCA and SPCA decomposition yield identical recovery rates, with more than 95% variance explained by the first 9 components. Decomposition with Haar wavelets captures 95% variance with the first 37 components.



(a) PCA component 1, $\sigma = 0.52$.

(d) SPCA component 1, $\sigma = 0.52$.

(g) Haar component 1, $\sigma = 0.10$.

(b) PCA component 2, $\sigma = 0.71$.

(e) SPCA component 2, $\sigma = 0.71$.

(h) Haar component 2, $\sigma = 0.20$.

(c) PCA component 3, $\sigma = 0.79$.

(f) SPCA component 3, $\sigma = 0.78$.

(i) Haar component 3, $\sigma = 0.28$.

First 3 PCA components.

First 3 SPCA components.

First 3 Haar components.

Figure 10: First 3 components under various decompositions with explained variances σ.

7 Non-Negative Matrix Factorization

Non-negative matrix factorization (NMF) is a decomposition technique that factors a matrix of non-negative entries into the product of two non-negative matrices [PT94]. It has the advantage of being interpretable as it automatically extracts sparse and meaningful features from a set of nonnegative data vectors, and is particularly popular among the applied machine learning community [LS99; WZ13].

7.1 Non-Negative Matrix Factorization Formulation

Given a non-negative matrix $X \in \mathbb{R}^{n\times d}$, NMF finds two lower rank matrices, $W \in \mathbb{R}^{n\times b}$ and $H \in \mathbb{R}^{b\times d}$, such that WH best approximates X. Columns in H can be seen as basis elements, and rows in W, weights corresponding to the basis elements. In the context of tariffs, X is derived from the tariff tensor $\tau \in \mathbb{R}^{N\times Y\times P}$ by averaging over years

$$X_{i,p} = \frac{1}{Y}\sum_{t=1}^{Y} \tau_{i,t}^{p} \tag{31}$$

for all $i \in [N]$ and $p \in [P]$. Hence, the dimension of this data matrix is $X \in \mathbb{R}^{N\times P}$. Now the columns of H can be interpreted as tariff strategies and rows in W capture how much each strategy is adopted by each country.

7.2 Country classification based on tariff strategies

NMF is highly effective in capturing the degree of variance in MFN tariffs, reaching 95% explainability with 3 components. NMF's performance is compared against that of PCA in Figure 11. The comparability in performance between NMF and PCA as a benchmark highlights NMF as a promising choice for tariff basis construction.

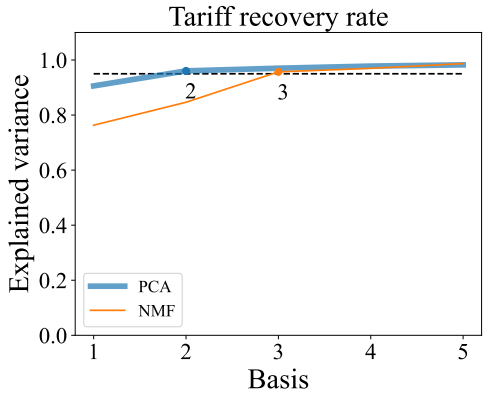


Figure 11: Recovery of MFN data as a function of basis size for PCA and NMF.

Matrices W and H lend critical insight into tariff strategies. As each row in H is a strategy profile across all product classes, summing over the rows of H reveals the level of variation in

each tariff product class. In particular, as illustrated in Figure 12, heavily weighted product classes like chemicals, textiles, and machines have greater variation among different countries.

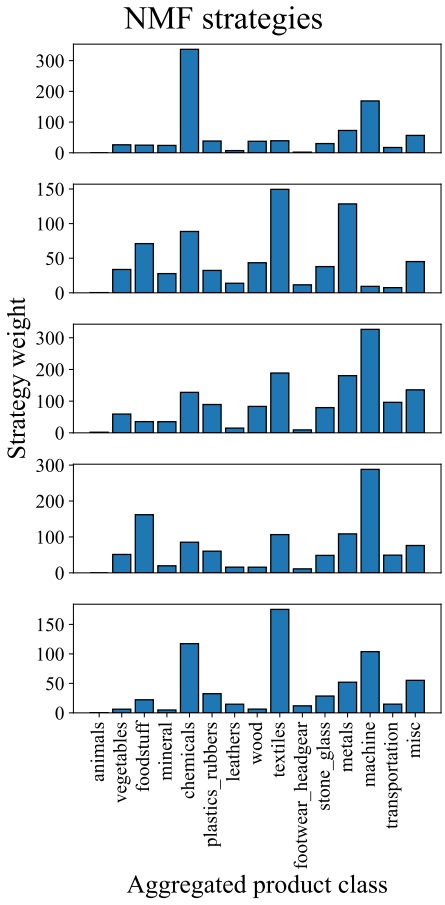


Figure 12: Top 5 strategies for applying tariffs across the 15 aggregated product classes.

Hierarchical clustering on country-wise strategy profile based on matrix W reveals structure that is consistent with international trade agreements, political interests, and the countries' income levels. As illustrated in Figure 13, at the 5-label level of granularity, European countries (in green), North American countries (in navy), equatorial countries (in grey) are grouped into the same clusters, respectively. These groupings further confirm the correlation between international politics and tariff strategies.

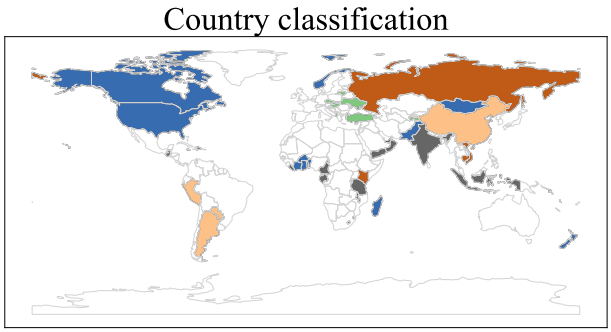


Figure 13: Hierarchical clustering on countries based on tariff strategies.

8 DYNAMICS

Here we outline next steps for building a stochastic dynamical model using the basis sets for trade and tariff data constructed above (shown on the example of products related to transportation). Recall that our trade data tensor

$$T = \{T^p_{i,j,t} : i, j \in [N], t \in [Y], p \in [P]\} \tag{32}$$

can be split into the 15 aggregate product categories $C_1, \ldots, C_{15}$ defined as

$$T^{C_k} = \{T^p_{i,j,t} : i, j \in [N], t \in [Y], p \in C_k\}, \quad k = 1, \ldots, 15. \tag{33}$$

From this, using one of the basis construction methods discussed above (PCA, sparse PCA, Haar) we can learn a separate basis set for each T^{C_k} aggregate product tensor given by

$$\mathscr{B}_k = \{U^k_l \in \mathbb{R}^{N \times N} : l \in [b_k]\}, \quad k = 1, \ldots, 15 \tag{34}$$

where $b_k \in \mathbb{N}$ is the number of basis elements constructed for aggregate product k. Finally, each trade matrix $T^p_t = \{T^p_{i,j,t} : i, j \in [N]\}$ for a given year $t \in [Y]$ and product $p \in [P]$ can be approximated in this basis as

$$T^p_t \approx y^p_{t,1} U^k_1 + \ldots + y^p_{t,b_k} U^k_{b_k} \tag{35}$$

assuming that p is in aggregate product class C_k.

Finally, we can stack the basis coefficients $y^p_t = \{y^p_{t,l}\}_{l=1}^{b_k} \in \mathbb{R}^{b_k}$ across all products $p \in [P]$ to obtain the vector

$$Y_t = (y^1_t, \ldots, y^P_t) \in \mathbb{R}^B \tag{36}$$

where $B = \sum_{k=1}^{15} b_k$. This procedure allows us to project our four-dimensional trade data tensor of size $N \times N \times Y \times P$ onto a two-dimensional matrix of size $B \times Y$ which provides significant compression since we expect that $B \ll N^2 P$.

Similarly, we can also perform a projection of our tariff data. Namely, our MFN tariff data tensor takes the form

$$\tau = \{\tau^p_{i,t} : i \in [N], t \in [Y], p \in [P]\} \tag{37}$$

which can be split into the same 15 aggregate product categories $C_1, \ldots, C_{15}$ defined as

$$\tau^{C_k} = \{\tau^p_{i,t} : i \in [N], t \in [Y], p \in C_k\}, \quad k = 1, \ldots, 15. \tag{38}$$

As before, using one of the basis construction methods (PCA, sparse PCA, Haar) we obtain a separate basis set for each τ^{C_k} aggregate product tensor given by

$$\mathscr{R}_k = \{v^k_l \in \mathbb{R}^N : l \in [r_k]\}, \quad k = 1, \ldots, 15. \tag{39}$$

where $r_k \in \mathbb{N}$ is the number of tariff basis elements constructed for aggregate product k. Finally, we approximate each tariff vector $\tau^p_t = \{\tau^p_{i,t} : i \in [N]\}$ for a given year $t \in [Y]$ and product $p \in [P]$ in this basis as

$$\tau^p_t \approx z^p_{t,1} v^k_1 + \ldots + z^p_{t,r_k} v^k_{r_k} \tag{40}$$

if p is in aggregate product class C_k.

Finally, we stack the basis coefficients $z^p_t = \{z^p_{t,l}\}_{l=1}^{r_k} \in \mathbb{R}^{r_k}$ across all products $p \in [P]$ to obtain the vector

$$Z_t = (z^1_t, \ldots, z^P_t) \in \mathbb{R}^R \tag{41}$$

where again $R = \sum_{k=1}^{15} r_k$.

Now we can combine the basis coefficients for trades and tariffs into one state vector

$$X_t = (Y_t, Z_t) \in \mathbb{R}^{B+R}. \tag{42}$$

In this projected coefficient space, we aim to learn the following stochastic difference equation (SDE)

$$X_{t+1} = f(t, X_t) + g(t, X_t) W_t \tag{43}$$

where $f : \mathbb{R}_+ \times \mathbb{R}^{B+R} \to \mathbb{R}^B$ and $g : \mathbb{R}_+ \times \mathbb{R}^{B+R} \to \mathbb{R}^{(B+R)\times m}$ are functions that need to be learned and $W_t \in \mathbb{R}^{m \times m}$ is a matrix with i.i.d. standard Gaussian entries. Such SDE models can be fit to data by specifying a parametric or nonparametric form for the unknown functions f, g and writing out the closed-form likelihood of the statistical model to find the optimal choice of f, g that maximize this likelihood through optimization.

A key challenge will be to understand the constraints we need to impose on the functional form of f and g such that they do not overfit our data consisting of $t = 1, \ldots, 21$ timepoints (the intersection of the years when both trade and tariff data were collected).

9 CONCLUSION

In this paper we have studied data of international trade and tariff dynamics across a range of products from 1989 to 2017. Our analysis studied the hierarchical clustering of products based on their network of trades, the use of nonnegative matrix factorization to extract tariff placement strategies of countries, and importantly the basis decompositions of spatiotemporal trade and tariff data using PCA, sparse PCA, and Haar wavelets. These analyses have allowed us to understand the structure of international trades and tariffs and prepare us to build stochastic dynamical models of their temporal interactions.

10 ACKNOWLEDGEMENTS

The authors thank Professor In Song Kim (MIT Political Science) for providing the trade and tariff data which inspired this project, Alasdair Hastewell for meaningful discussions, and David Jerison for constructive feedback. The authors are also grateful for the Summer Program in Undergraduate Research of the MIT Mathematics Department for financial support.

REFERENCES

[Ber96] C. F. Bergsten. "Globalizing free trade". In: *Foreign Affairs* (1996), pp. 105–120.

[Bha+08] K. Bhattacharya et al. "The international trade network: weighted network analysis and modelling". In: *Journal of Statistical Mechanics: Theory and Experiment* 2008.02 (2008), P02002.

[BS73] F. Black and M. Scholes. "The Pricing of Options and Corporate Liabilities". In: *Journal of Political Economy* 81.3 (1973), pp. 637–654.

[CA08] T. J. Courchene and J. R. Allan. "Climate change: the case for a carbon tariff/tax". In: *POLICY OPTIONS-MONTREAL-* 29.3 (2008), p. 59.

[CBS16] M. Caraglio, F. Baldovin, and A. L. Stella. "Export dynamics as an optimal growth problem in the network of global economy". In: *Scientific reports* 6.1 (2016), p. 31461.

[CG05] F. Clementi and M. Gallegati. *Pareto's Law of Income Distribution: Evidence for Grermany, the United Kingdom, and the United States*. Microeconomics. University Library of Munich, Germany, May 2005.

[CK03] M. Crovella and E. Kolaczyk. "Graph wavelets for spatial traffic analysis". In: *IEEE INFOCOM 2003. Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE Cat. No.03CH37428)*. Vol. 3. 2003, 1848–1857 vol.3.

[Dau88] I. Daubechies. "Orthonormal bases of compactly supported wavelets". In: *Communications on Pure and Applied Mathematics* 41.7 (1988), pp. 909–996.

[Eck01] M. A. Eckhard Limpert Werner A. Stahel. "Log-normal Distributions across the Sciences: Keys and Clues: On the charms of statistics, and how mechanical models resembling gambling machines offer a link to a handy way to characterize log-normal distributions, which can provide deeper insight into variability and probability—normal or log-normal: That is the question". In: (2001).

[Fag16] G. Fagiolo. "The international trade network: Empirics and modeling". In: (2016).

[Faj+20] P. D. Fajgelbaum et al. “The return to protectionism”. In: *The Quarterly Journal of Economics* 135.1 (2020), pp. 1–55.

[FK22] P. D. Fajgelbaum and A. K. Khandelwal. “The economic impacts of the US–China trade war”. In: *Annual Review of Economics* 14 (2022), pp. 205–228.

[GH22] A. Gnutzmann-Mkrtchyan and J. Hugot. “Gravity-Based Tools for Assessing the Impact of Tariff Changes”. In: *Asian Development Bank Economics Working Paper Series* 649 (2022).

[GMG19] J. García-Algarra, M. L. Mouronte-López, and J. Galeano. “A stochastic generative model of the world trade network”. In: *Scientific reports* 9.1 (2019), p. 18539.

[HAR18] R. E. N. HARTE. “US tariffs: EU response and fears of a trade war”. In: (2018).

[HFX10] S. Hanneke, W. Fu, and E. P. Xing. “Discrete temporal models of social networks”. In: (2010).

[HVG11] D. K. Hammond, P. Vandergheynst, and R. Gribonval. “Wavelets on graphs via spectral graph theory”. In: *Applied and Computational Harmonic Analysis* 30.2 (2011), pp. 129–150. ISSN: 1063-5203.

[Isa54] W. Isard. “Location theory and trade theory: short-run analysis”. In: *The Quarterly Journal of Economics* 68.2 (1954), pp. 305–320.

[Joh60] H. G. Johnson. “The cost of protection and the scientific tariff”. In: *Journal of Political Economy* 68.4 (1960), pp. 327–345.

[KH14] P. N. Krivitsky and M. S. Handcock. “A separable model for dynamic networks”. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 76.1 (2014), pp. 29–46.

[Kru87] P. R. Krugman. “Is free trade passé?” In: *Journal of economic Perspectives* 1.2 (1987), pp. 131–144.

[LM20] R. Light and J. Moody. *The Oxford handbook of social networks*. Oxford University Press, 2020.

[LS99] D. Lee and H. Seung. “Learning the parts of objects by non-negative matrix factorization”. In: *Nature* (1999).

[Mey93] Y. Meyer. *Wavelets: Algorithms & Applications*. Society for Industrial & Applied, 1993.

[MLD08] P. Mcafee, T. R. Lewis, and D. J. Dale. *Introduction to Economic Analysis, v2.1*. 2008.

[Mül11] D. Müllner. “Modern hierarchical, agglomerative clustering algorithms”. In: *ArXiv* abs/1109.2378 (2011).

[PT94] P. Paatero and U. Tapper. “Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values”. In: *Environmetrics* 5.2 (1994), pp. 111–126.

[Ric05] D. Ricardo. “From the principles of political economy and taxation”. In: *Readings in the economics of the division of labor: The classical tradition*. World Scientific, 2005, pp. 127–130.

[SBS19] M. Starnini, M. Boguñá, and M. Á. Serrano. “The interconnected wealth of nations: Shock propagation on global trade-investment multiplex networks”. In: *Scientific reports* 9.1 (2019), p. 13079.

[SSB22] A. Setayesh, Z. Sourati Hassan Zadeh, and B. Bahrak. “Analysis of the global trade network using exponential random graph models”. In: *Applied Network Science* 7.1 (2022), pp. 1–19.

[Sut97] J. Sutton. “Gibrat’s Legacy”. In: *Journal of Economic Literature* 35.1 (1997), pp. 40–59. ISSN: 00220515. (Visited on 08/01/2023).

[TCS18a] G. Teza, M. Caraglio, and A. L. Stella. “Data driven approach to the dynamics of import and export of g7 countries”. In: *Entropy* 20.10 (2018), p. 735.

[TCS18b] G. Teza, M. Caraglio, and A. L. Stella. “Growth dynamics and complexity of national economies in the global trade network”. In: *Scientific reports* 8.1 (2018), p. 15230.

[WZ13] Y.-X. Wang and Y.-J. Zhang. “Nonnegative Matrix Factorization: A Comprehensive Review”. In: *IEEE Transactions on Knowledge and Data Engineering* 25.6 (2013), pp. 1336–1353.