UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia "Galileo Galilei"

Corso di Laurea in Fisica

Tesi di Laurea

Misure di complessità economica nella rete commerciale globale

Economic complexity measures in the world trade web

Relatore:
Prof. Fulvio Baldovin
Co-Relatori:
Prof. Attilio Stella
Dr. Gianluca Teza

Laureando:
Giacomo Veglio
Matricola:
2085010

Anno Accademico 2024/2025

ABSTRACT

ITA Negli ultimi anni, diversi econofisici hanno lavorato allo sviluppo di metriche, elaborando dati della rete commerciale globale, per quantificare la complessità economica di paesi e prodotti. Queste metriche riflettono meglio la competitività di un paese nella rete commerciale globale rispetto alle metriche economiche tradizionali, poiché sono costruite sfruttando una quantità significativamente maggiore di informazioni commerciali anziché semplici valori monetari. In questa tesi, riesamineremo lo stato dell'arte dei principali metodi sviluppati negli ultimi anni e, in particolare, ci concentreremo su misure entropiche sviluppate utilizzando il concetto di entropia di Shannon preso in prestito dalla teoria dell'informazione classica. Rapporteremo quindi queste metriche per presentare un'immagine la più completa possibile della rete commerciale globale.

ENG In recent years, several econophysicists have worked on developing metrics, by elaborating world trade web data, to quantify the economic complexity of countries and products. These metrics better reflect the competitiveness of countries in the world trade web compared to traditional economic metrics, as they are built upon a significantly higher amount of trade information rather than mere monetary value. In this thesis, we will review the state of the art of the main methods developed in recent years and, in particular, focus on entropic measures developed using the concept of Shannon entropy borrowed from classical information theory. We will then relate these metrics to present as complete a picture as possible of the world trade web.

What's in a name?
In the case of Shannon's measure the naming was not accidental. In 1961 one of us (Tribus) asked Shannon what he had thought about when he had finally confirmed his famous measure. Shannon replied: " My greatest concern was what to call it. I thought of calling it 'information', but the word was overly used, so I decided to call it 'uncertainty'. When I discussed it with John von Neumann, he had a better idea. Von Neumann told me, 'You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, no one knows what entropy really is, so in a debate you will always have the advantage.' "
[23]

Contents

1 Introduction

1.1 In short

In this thesis work we will dive into the world of complexity economics and econophysics. In particular we will focus on the development of metrics to measure the complexity of countries and products in the world trade web. We will argue on the necessity for the introduction of these metrics and we will dive in a review of recent literature pertaining to the topic by analyzing prominent economic complexity metrics, such as the Economic Complexity Index constructed via the Method of Reflections[**10**][**8**] and the Fitness metrics constructed via the Fitness-Complexity algorithm[**21**][**20**], and their relative issues to then close the review with an econophysical metric, which arguably constitutes the current state of the art in the matter of economic complexity metrics, constructed upon a rather common tool provided by classical information theory: Shannon Entropy[**18**][**22**].

In **Section 2** we will introduce some relevant mathematical tools in our analysis and the crucial concepts of Product Space and World Trade Web.

In **Section 3** we will present the Economic Complexity Index constructed via the iterative algorithm called the Method of Reflection. This metric pioneered the research on economic complexity and is based upon an elegant iterative algorithm. Unfortunately it is affected by some pathologies that we will discuss about in **Subsection 3.3**.

In **Section 4** we will present the Fitness and Complexity indexes that are constructed via the Fitness complexity algorithm. This method first appeared as an alternative to the previous one by solving the mathematical issues affecting it. This is done relying on a non-linear iterative algorithm whilst, at the same time, more or less maintaining the underlying economic theory. Unfortunately, this method is affected by some issue related to so called hidden divergences that derive from how data are aggregated and/or elaborated. This will be discussed in the dedicated **Subsection 4.3**.

In **Section 5** we will present a novel method that leverages Shannon Entropy to reconcile the ideas already present in the literature in a robust mathematical framework which yields reasonable results and makes one naturally wonder with regard to the possible physical nature of economics.

In **Section 6**, the conclusions, we will wrap up the discussion by presenting a final comparison between the various metrics and some future prospects in the field.

1.2 Data

In the literature different academics utilize various data classification related to the world trade web. Among the most commonly used are the Standard International Trade Classification (SITC) [**25**], the COMTRADE Harmonized System (HS), and the North American Industry Classification System (NAICS) [**24**]. For simplicity and consistency, we employ data from the CEPII's BACI dataset [**6**], which provides polished HS data. The HS is a classification of products that is structured as a multilayered nomenclature, where the leftmost digits are shared by products belonging to a more general macro-category[**22**]. This hierarchical structure makes it possible to aggregate data to achieve different levels of granularity in the analysis. Some studies focus solely on the coarse-grained 4-digits macro-categories P whilst the HS actually provides a 6-digit classification for each product p for a batch of approximately 5000 products depending on year and version of the HS. These studies often assume that algorithms would perform equally well at different levels of granularity, without considering the intra-sectorial complexity that may become significant when moving from a 4-digit to a 6-digit classification.

Specifically, we use the 1992 HS standard, the oldest available, as it offers the most extensive dataset. This is because it is possible to retrofit new products into older classifications, but not the other way around. Regardless of that, yearly variations in the datasets may happen due to a variety of factors such as policy changes or products disappearing from the trade web. Timeseries based on rough data hence lack of reliable information. Therefore, we will focus on a specific year for the analysis in this work.

1.3 The Role of Physics in Economic Complexity

Before starting off with the main discussion in this work it is appropriate to discuss the relationship between economics and physics from which the subject of our studies stems. To begin such an endeavor we must point out the use of Shannon entropy in the following pages since it is renowned for its versatily in physical application and especially in information theory. We recall here the definition and properties of the Shannon function [**18**].

Def. 1: Shannon Entropy

Given a complete set of non null probabilities $\{p_1, ..., p_N\}$ such that $\sum_{i=1}^{N} p_i = 1$ we can define the Shannon Entropy of such set as

$$H := -\sum_{i=1}^{N} p_i \log(p_i). \tag{1}$$

Considering p_i as relative occupation of N states, the Shannon function in **Eq.**(**1**) can be taken as a measure of the diversity of the distribution. H is bounded, for $N > 1$, in the region $[0, \log N]$, its maximum is reached in case of equipartition $p_i = 1/N \; \forall i$ and its minimum when a single state is occupied[**22**].

The physical nature of information in its thermodynamical sense has been well known and commonly accepted in the scientific community in agreement with Landauer's Principle [**13**].

Thm. 1: Landauer's Principle

Any logically irreversible manipulation of information, such as the erasure of a bit, must be accompanied by a minimum amount of heat dissipation.

A practical example of this principle that is often offered in literature and in the classrooms is the Szilard's engine. Let's suppose to have a chamber at equilibrium at temperature T with a thermal reservoir. Let's place in this container a particle that is only subject to elastic collisions with the chamber's walls. We may now reversibly introduce an adiabatic barrier in the chamber and split it into two separate chambers (A and B). This fundamentally allows us to consider the chamber as a single bit register. By doing this we isolated the particle in one half of the chamber registering a bit of information. Already when the barrier is introduced, a counterweight can be linked to it as illustrated in **Figure 1**. By hitting the barrier the particle can do work on the counterweight. Supposing that the piston is fully frictionless would apparently allow us to extract infinite work by moving it and the counterweight in a convenient way. This situation is in clear violation to the principles of thermodynamics and it is exactly by taking into account Landauer's Principle that one can fix it.

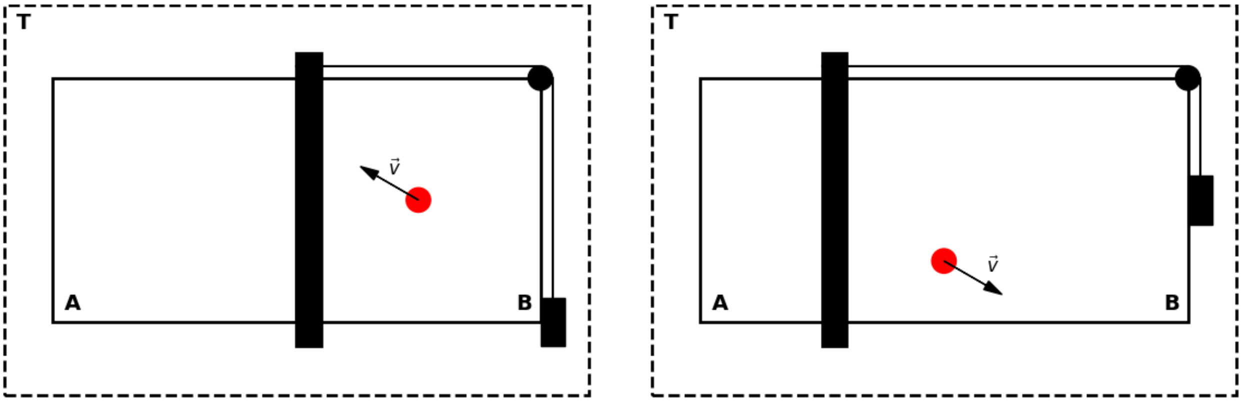

Figure 1: Schematic illustration of Szilard's engine.

As previously mentioned, when the barrier is splitting the chamber in 2 we fundamentally have a 1 bit register characterized by an initial entropy of $S_i = k_B \log 2$ and when we lift the barrier to readjust it we are deleting the bit and reducing the chamber as a 0 bit register with final entropy of $S_f = 0$ hence yielding an entropy variation in the process of $\Delta S_{chamber} = -k_B \log 2$. By the second principle we have $\Delta S_{tot} = \Delta S_{reservoir} + \Delta S_{chamber} = \frac{Q_{reservoir}}{T} - k_B \log 2 \geq 0$, where we imposed Clausius formula for an infinite thermal reservoir, implying $Q_{reservoir} \geq k_B T \log 2$. By the first principle the heat exchanged should equal the work performed, hence the deleting of information requires, as Landauer's principle imposes, the consumption of energy of at least $\Delta E = k_B T \log 2$.

We have just presented a textbook argument on the physical nature of information and one may now argue about the relation of economics to information on a merely philosophical sphere. When two players in the market agree with a handshake to assign a price[**14**] and value[**3**] to a product or a service they are indeed creating or modifying information related to that product. The most obvious example is the price of shares, constantly oscillating depending on the market players' offers and demands. In a broad sense we could state that economics is but the study of the specific information game that takes life in the markets. And as a game of information those who own more of it find themselves in an advantageous position. It is hence reasonable to assume that we might find insights via scientific domains that at first appear unrelated to economics but may regardless prove to be useful.

2 On The Necessity

In [9] Hidalgo et al. first questioned theories regarding the structure of the so called product space deriving from classical economic theory [19][16]. They presented the product space as a space containing all existing goods into which a subject, such as a company or a nation, can move by unlocking new productive capabilities, or loosing them. The products may arbitrarily be positioned and connected in this space in a plethora of ways. For instance one may even hypothesize the space to be completely uniform to then figure out a way to test such theory. A more intriguing theory is the quality ladder's one [7] whereby it is theorized that countries can arbitrarily improve their goods production based on what they already produce. This would imply a strong degree of relatedness among similar products and a low one among very different products almost suggesting a tree like structure for a graph representation of the product space. However, the authors provide simple yet convincing evidence against this assumptions and show that the product space is neither boringly homogeneous nor necessarily well structured and is instead rather discombobulated.

2.1 Revealed Comparative Advantage and Proximity

To begin our discussion we must introduce some numerical quantity and crucial mathematical results in our discussion. The first of these is the Revealed Comparative Advantage (RCA) which acts as a measure of whether or not we can consider a country as a significant exporter of a given good and is defined as follows.

Def. 2: Revealed Comparative Advantage (RCA)

According to Balassa's definition [1], the Revealed Comparative Advantage,

$$\mathrm{RCA}_{cp} := \frac{M_{cp}}{\sum_{c'=1}^{N_c} M_{c'p}} \Bigg/ \frac{\sum_{p'=1}^{N_p} M_{cp'}}{\sum_{c'=1}^{N_c}\sum_{p'=1}^{N_p} M_{c'p'}} \tag{2}$$

measures whether a country c exports more of a product p, as a share of total exports, than the average country [9](484).

Hence Balassa's definition fundamentally states that a country has Revealed Comparative Advantage in a product if it exports more than its "fair" share, that is, a share that is equal to the share of total world trade that the product represents [8](25).

Having introduced the RCA we can focus on the Product Space by considering what the authors of [9] call product proximity which, according to them, formalizes the intuitive idea that the ability of a country to produce a product depends on its ability to produce other products.

Def. 3: Proximity

The proximity ϕ between products p and p' is the minimum of the pairwise conditional probabilities of a country c exporting a good given that it exports another estimated through RCA, in mathematical terms

$$\phi_{pp'} := \min\left\{P(\mathrm{RCA}_{cp} \mid \mathrm{RCA}_{cp'}), P(\mathrm{RCA}_{cp'} \mid \mathrm{RCA}_{cp})\right\}. \tag{3}$$

Given the definition of proximity we may now compute, with data from the BACI dataset, the proximity matrix which, given its size, calls for a graphical representation in the form of a bi-dimensional heatmap such as the one presented in **Figure 2**. From it we can infer some information regarding the nature of the product space. In the first place we can clearly see that the space is not homogeneous for it was the plot would be fundamentally uniform in color. However, we can also appreciate the fact that we have plenty of non null entries far from the diagonal which is a trivial evidence against the theory of the product quality ladder [7] according to which we would expect considerably higher proximity values only close to the diagonal. In **Figure 2** we can also appreciate the sparseness of the proximity matrix via the distribution of its values which is concentrated towards lower values.

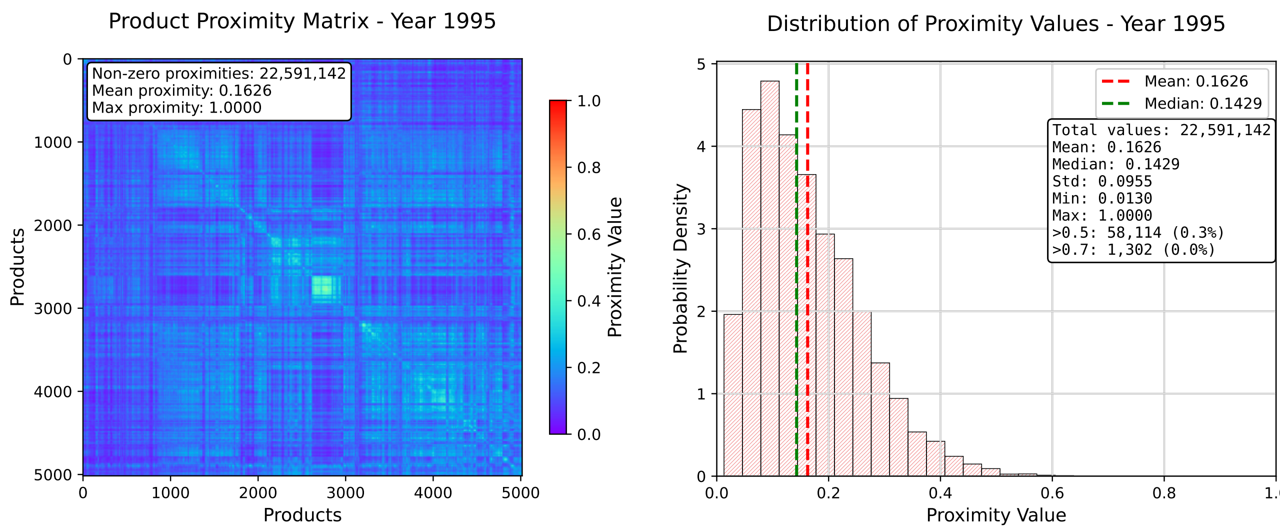

Figure 2: (Left) Proximity matrix HS standard. (Right) Distribution of proximity values in the matrix

2.2 The Product Space and the World Trade Web

We may now represent the product space as a network as shown in **Figure 3**. In this representation all available products in the BACI dataset were plotted in a hierarchical serie of concentrical circles and, considering the low number of higher proximity values in the proximity matrix, only the edges corresponding to the highest proximity value were plotted for each product in order to avoid excessive visual cluttering. Already from this simplistic graph we can clearly see that products are connected to each others in a non trivial way and in particular we see that often the most relevant connection for each product is not with a neighbor belonging to the same HS macro-category but with a completely different type of good.

Having now illustrated the idea of product space we may wonder how the producers relate to it. The producers are the countries listed in the HS and we have already utilized information

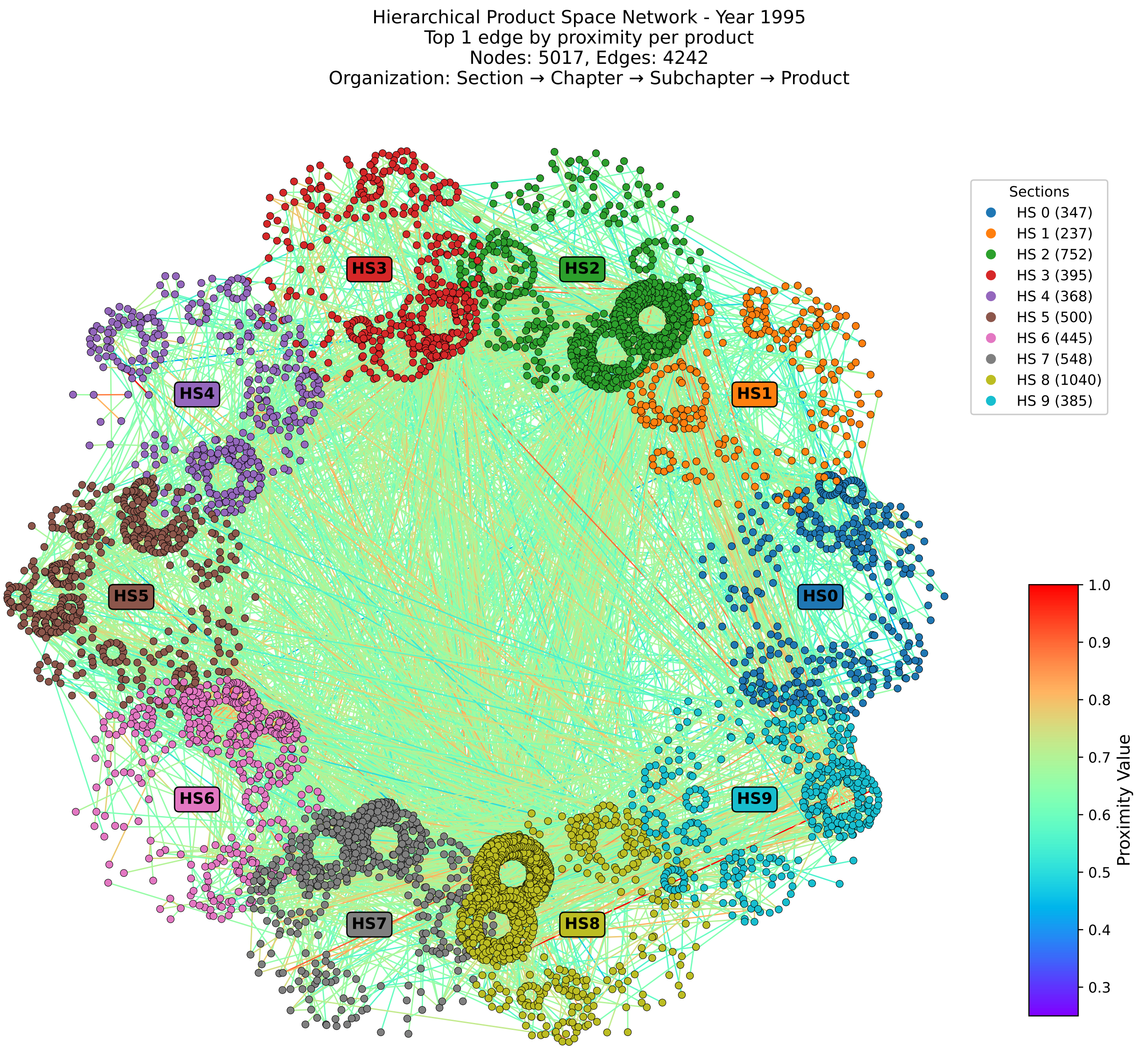

Figure 3: Illustrative Network Representation of the Product Space

relating to them to compute the RCA_{cp} bi-adjacency matrix according to **Def. 2**. A simpler yet relevant bi-adjacency matrix can also be constructed by direct data aggregation from the BACI dataset. We will refer to such matrix as the bi-adjacency matrix M_{cp} which has a number of rows equivalent to the number of countries and a number of columns equivalent to the number of products available in the BACI dataset with entries representing the total amount of product p exported by a country c towards the rest of the world in a given year. The use of either one of the two matrixes depends on the aims of a method or of an analysis for one may be considered intrinsic and the other extrinsic with respect to monetary data in US$. This RCA_{cp} matrix is however not profusely adoperated in literature for both in the Method of Reflections and in the Fitness-Complexity algorithm is used a binarized bi-adjacency matrix m_{cp} constructed by imposing the condition

$$m_{cp} = \begin{cases} 1 & \text{if } \mathrm{RCA}_{cp} \geq \alpha \\ 0 & \text{otherwise} \end{cases} \qquad \alpha \in \mathbb{R}_0^+. \tag{4}$$

In particular, the main choice in literature for α is 1 even though any positive number is allowed. This will prove to be pathological as shown in **Subsection 4.3**. All three of these matrixes are shown in **Figure 4**.

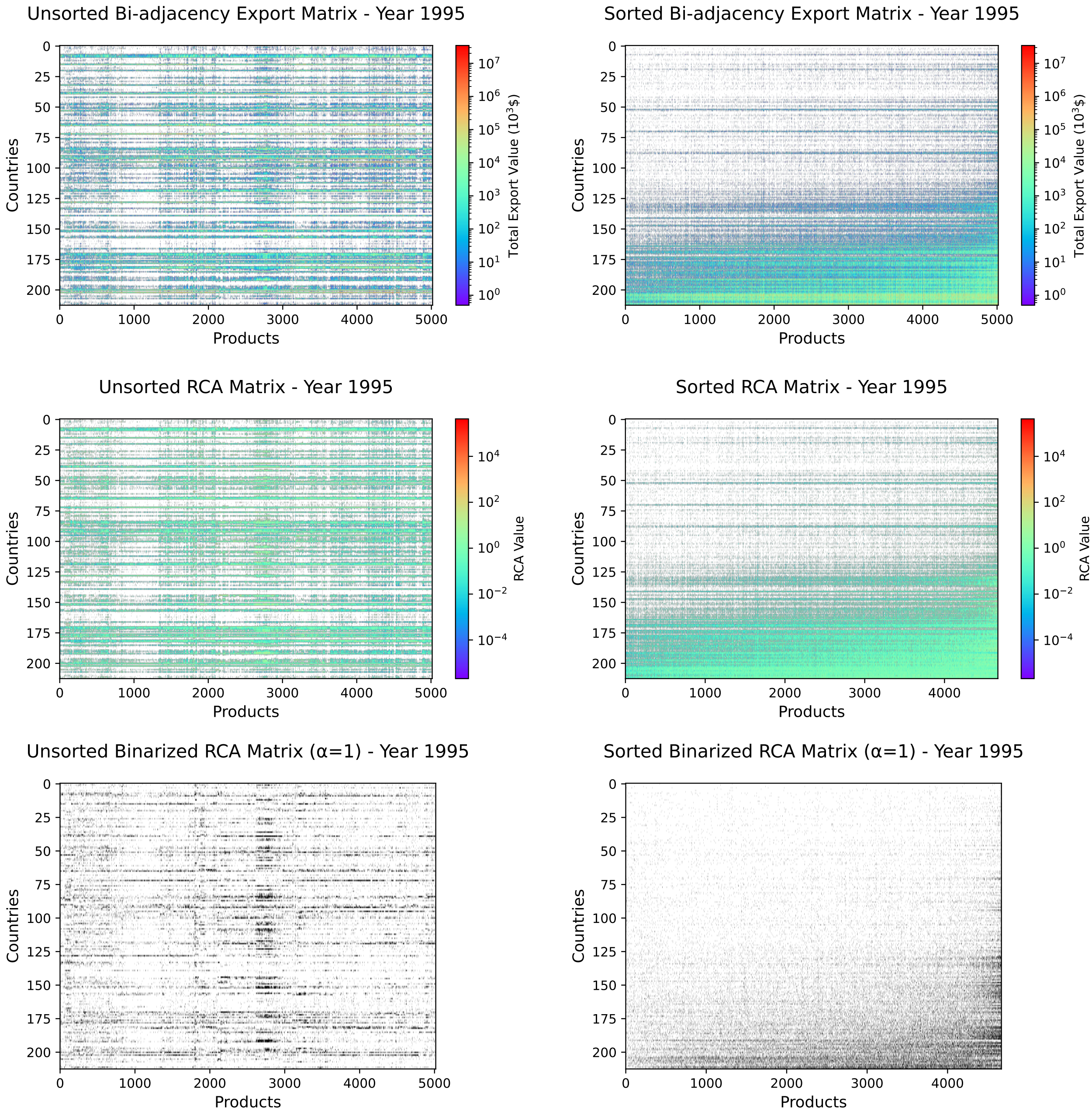

Figure 4: 6-digit HS classification. Bi-adjacency matrix M_{cp}: (Top left) not sorted; (Top right) sorted according to bare TCS diversification and ubiquity. RCA_{cp} matrix: (Middle left) unsorted; (Middle right) sorted according to bare TCS diversification and ubiquity. Binarized bi-adjacency matrix m_{cp} with threshold RCA ≥ 1: (Bottom left) not sorted; (Bottom right) sorted according to bare TCS diversification and ubiquity.

From the bi-adjacency matrix, the RCA one in this case, we may offer a simple network representation of what we can unequivocally refer to as the World Trade Web. In **Figure 5** we have an illustration of the web with each country connected with its top 3 products by RCA. It is evident that countries connect to products in a 2 layer country-product network and since we debunked the idea based on the quality ladder that products should be connected primarily to their neighbors we may ask ourselves what role do countries play in determining the connections of products.

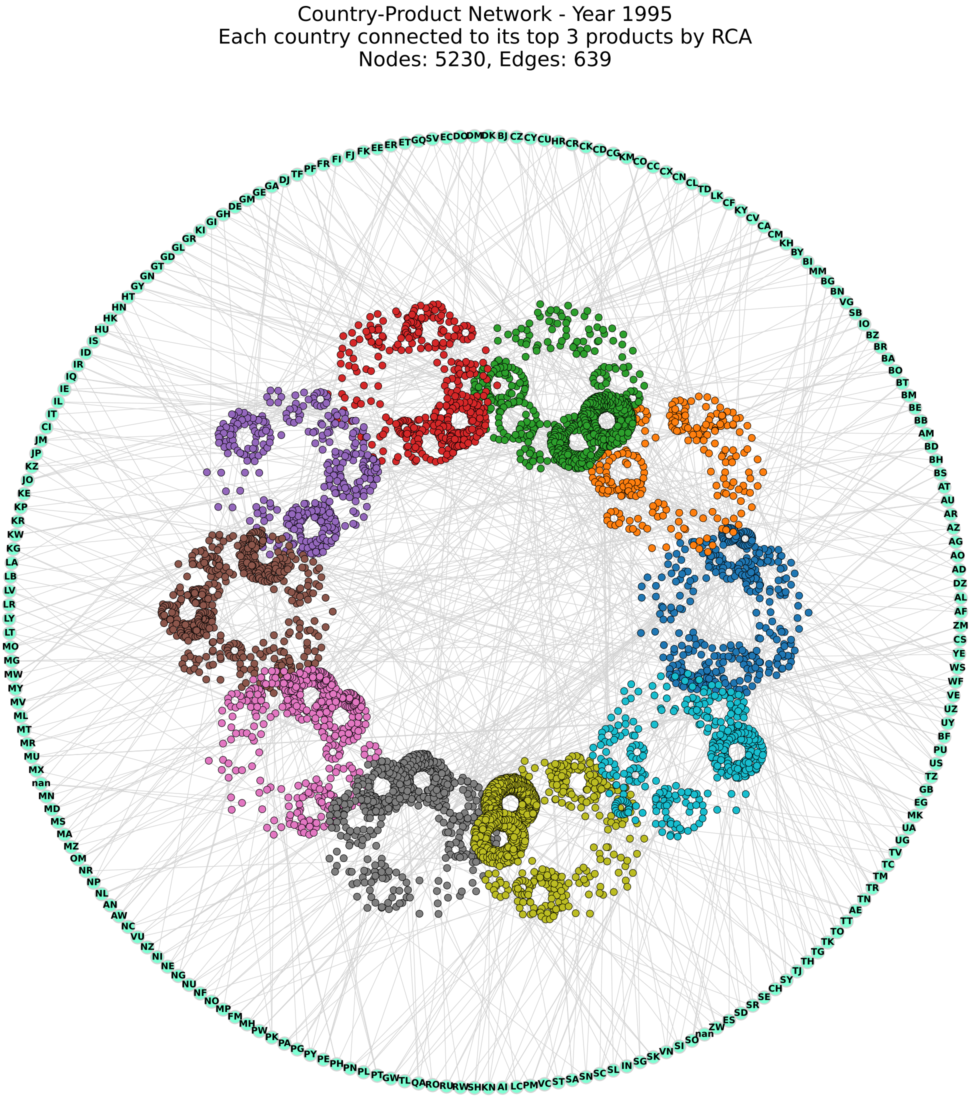

Figure 5: Illustrative Network Representation of the World Trade Web. Each country is connected with its top 3 exports by RCA value.

How do countries decide what products to produce and what not? In **[10]** César A. Hidalgo and Ricardo Hausmann propose the idea of a hidden layer between countries and products. Such layer ought to represent what the authors of **[20]** refer to as the intangibles. The intangibles are the hidden capabilities that countries posses which are fundamentally impossible to estimate numerically. Work ethics, creativity or risk appetite of citizens, the detailed structure of the educational system or the repercussions of specific governmental investments are all intangible features that can influence the productive capabilities of a country and influence the development of its export basket. We may hence state that the main reason why research began on the development of economic complexity metrics is precisely the attempt at summarizing and quantifying this hidden capabilities of countries in their complexity assuming its proportionality to their performance in the world trade web. These metrics could actually turn out to have a practical impact on our society, as an obvious example they could be included by rating agencies in their evaluation of countries which may for instance directly affect the rates of bonds offered by a country and hence the investors market.

3 Economic Complexity

3.1 The Method of Reflections (MR)

One of the first and now most diffused metrics to estimate the complexity of a country's export basket in the world trade web is by far the Economic Complexity Index, mainly developed by César A. Hidalgo and Ricardo Hausmann [**10**][**8**].

In their works [**10**][**8**], the authors introduce the Method of Reflections (MR) to characterize countries and products by trying to capture the information relative to the structure of the bipartite network defined by the binarized bi-adjacency matrix. The method consists in the iterative construction of 2 vectors containing respectively 2 symmetrical sets of variables that characterize the nodes of the network. The initial variable for such iterative process are the "bare" HH diversification of a country and the "bare" HH ubiquity of a product[1].

Def. 4: "Bare" HH diversification

We define as diversification the measure, in the context of the world trade network, of the number of products from $p = 1$ to $p = N_p$ with RCA ≥ 1 exported by a specific country c, which HH propose to be

$$k_c^{(0)} := \sum_{p=1}^{N_p} m_{cp}. \tag{5}$$

Def. 5: "Bare" HH ubiquity

We define as ubiquity the measure, in the context of the world trade network, of the number of countries from $c = 1$ to $c = N_c$ with RCA ≥ 1 exporting a given product p, which HH propose to be

$$k_p^{(0)} := \sum_{c=1}^{N_c} m_{cp}. \tag{6}$$

The Method of Reflections hence consists in the iterative calculation of the average value of the previous level properties of a node's neighbors and is defined, for $N \geq 1$ as

$$k_c^{(N)} = \frac{1}{k_c^{(0)}} \sum_{p=1}^{N_p} m_{cp} k_p^{(N-1)} = \left\langle k_p^{(N-1)} \right\rangle_c, \tag{7}$$

$$k_p^{(N)} = \frac{1}{k_p^{(0)}} \sum_{c=1}^{N_c} m_{cp} k_c^{(N-1)} = \left\langle k_c^{(N-1)} \right\rangle_p. \tag{8}$$

Which yields the characteristic country and product vectors

$$\vec{k}_c = \left(k_c^{(0)}, k_c^{(1)}, k_c^{(2)}, ..., k_c^{(N)}\right), \tag{9}$$

$$\vec{k}_p = \left(k_p^{(0)}, k_p^{(1)}, k_p^{(2)}, ..., k_p^{(N)}\right). \tag{10}$$

[1] We put HH in front of the term diversification and ubiquity to highlight the paternity of the definition in order to avoid confusion with similar objects introduced by Teza, Caraglio, and Stella (TCS) that are going to be presented in **Section 5**.

For the countries' vectors, we may consider the even-positioned variables as generalized measures of HH diversification of country c and the odd ones as a generalized measure of the HH ubiquity of its exports. Vice versa for the products' vectors.

3.2 Economic Complexity Index (ECI)

Starting from the Method of Reflections, in the Atlas of Economic Complexity [8](**24**), HH provided a formal definition of Economic Complexity Index (ECI). The authors compose **Eq.**(**7**) and **Eq.**(**8**) obtaining

$$\begin{aligned} k_c^{(N)} &= \frac{1}{k_c^{(0)}} \sum_{p=1}^{N_p} m_{cp} \frac{1}{k_p^{(0)}} \sum_{c'=1}^{N_c} m_{c'p} k_{c'}^{(N-2)} \\ &= \sum_{c'=1}^{N_c} k_{c'}^{(N-2)} \sum_{p=1}^{N_p} \frac{m_{cp} m_{c'p}}{k_c^{(0)} k_p^{(0)}} \\ &= \sum_{c'=1}^{N_c} \tilde{m}_{cc'} k_{c'}^{(N-2)}, \end{aligned} \tag{11}$$

where we defined

$$\tilde{m}_{cc'} := \sum_{p=1}^{N_p} \frac{m_{cp} m_{c'p}}{k_c^{(0)} k_p^{(0)}}, \tag{12}$$

which fundamentally is the transposed of an ergodic Markov transition operator [4](**16**) characterizing the iterative algorithm.

HH then point out that **Eq.**(**11**) is satisfied when $k_c^{(N)} = k_c^{(N-2)} = 1$, which is respected for the eigenvector of $\tilde{m}_{cc'}$ associated to its highest eigenvalue. This being a uniform vector of ones, it is fully deprived of any useful information. They hence propose to focus on the eigenvector $\vec{K}$ associated to the second highest eigenvalue of $\tilde{m}_{cc'}$, which captures the largest amount of variance in the system, and they take it as a measure of economic complexity by defining the Economic Complexity Index.

Def. 6: Economic Complexity Index

$$\mathrm{ECI} := \frac{\vec{K} - \left\langle \vec{K} \right\rangle}{\mathrm{StDev}\left(\vec{K}\right)}. \tag{13}$$

Due to the symmetry of the problem, it is also possible to define the mathematically analogous Product Complexity Index by considering the eigenvector $\vec{Q}$ associated to the second highest eigenvalue of

$$\tilde{m}_{pp'} := \sum_{c=1}^{N_c} \frac{m_{cp} m_{cp'}}{k_c^{(0)} k_p^{(0)}}. \tag{14}$$

Def. 7: Product Complexity Index

$$\mathrm{PCI} := \frac{\vec{Q} - \left\langle \vec{Q} \right\rangle}{\mathrm{StDev}\left(\vec{Q}\right)}. \tag{15}$$

3.3 Criticisms to the MR

Most likely due to its intuitive and linear nature, the method of reflections and the ECI gained a lot of traction as econonomic metrics. Albeit there being some logical fallacies and mathematical issues that have being brought up by several academics in the years soon following the publications by HH, the ECI still remains predominant in rankings and trade visualization platforms. This raises some concerns as a physical or juridical person may refer to such rankings, potentially ending up in ill-guided decision-making.

In **[20](1685)** and [4]**(15)**, Cristelli et al. present some arguments regarding the main conceptual and mathematical flaws of the Method of Reflections.

The authors report the first issue being the change in economic meaning of variables in the characteristic country and product vectors at each iteration. In particular, they highlight the linearity in the relation between $k_c^{(2N)}$ and $k_c^{(2N-2)}$ that **Eq.(11)** grants and the conceptual weirdness they perceive in the fact that the iterations preserve the "dimension" (diversification or ubiquity) of the object only at even or odd steps. They further argue that the economical and statistical interpretation is quickly lost with the increase of the iteration order N. They point out that a valid alternative would be an iterative algorithm to solve a self-consistent fixed point equation such as the one that the same authors introduce and that is presented in **Section 4**.

The second problem is that the information is shrinking when the method heads towards convergence, yielding a uniform and de facto information-less vector. For instance, one can appreciate the fact that in the Method of Reflections the $2N+1$ order complexity of a product $k_p^{(2N+1)}$ is given by the average of the $k_c^{(2N)}$ of its producers, and vice versa, which means that the information relative to the most complex countries associated with the product of interest is as relevant as the information from the least complex ones. In short, at each iteration, information is progressively lost through averaging of complexity values for products and countries, rendering the tails of the **Eq.(9)** and **Eq.(10)** vectors uniform and information-less. This furthermore implies that the final element of the country vector $k_c^{(N)}$ ends up not being correlated with countries' capabilities, which is an antagonistic result to the conceptual underlining economic model.

The third issue that is pointed out by the authors also lies within the averaging that is carried out in the iterative steps. The starting point of the algorithm, the "Bare" HH diversification in **Eq.(5)** and "Bare" HH ubiquity in **Eq.(6)**, directly constructed from the binarized bi-adjacency matrix, is extensive in the number of products and countries. However, the following orders turn out to be intensive with respect to the same quantities due to the averages that are considered. The non-extensiveness with respect to the exported products of the competitiveness of a country implies that its diversification does not necessarily constitute an advantage, which, again, goes against the conceptual framework that HH themselves introduced.

At last, the authors point out that the algorithm is conceptually wrong. Even though we would expect the differences among countries to be generally magnified, they are instead reduced at each iteration. According to the conceptual model, which is well sound, we would expect a magnification in ranking distances. This would allow us to compare, for instance, a scarcely diversified country that only produces ubiquitous products from a well-diversified country. The ranking position of the latter must increase when the diversification of products is taken into consideration, for the ubiquitous products are going to be less worthy than the others.

3 Economic Complexity

In **[12]** Eric Kemp-Benedict presented similar critics and made an attempt in giving an alternative interpretation to the Method of Reflections specifically discussing the orthogonality of diversification and economic complexity as they are defined in the MR.

4 Fitness

4.1 A Non-Linear Approach to Complexity Measures

Considering the previously mentioned criticisms regarding the Method of Reflections, Tacchella et al. proposed in **[21][20] [4]** a non-linear iterative algorithm and an economic complexity metric called Fitness alternative to the one proposed by Hidalgo and Haussmann. The different approach brought up by the authors aims at resolving the mathematical inconsistencies of the results of the Method of Reflections within the conceptual framework in which it was conceived.

The authors observe that the fact that a product is produced in a country considerable as developed does not provide much information relative to the actual complexity of the product for, according to the data, said country tends to export a wide variety of products. However, when we appreciate the fact that an underdeveloped country exports a certain product it is likely that the production of said good requires a low level of sophistication. Starting from this consideration the authors propose a correction to the economic theory standing behind the Method of Reflections. They consider reasonable the calculation of countries' features, such as their competitiveness, based on the properties, such as complexity, of the products they export but unreasonable the assumption that the vice versa must also be true. They hence postulate that the complexity of a product cannot be defined as the average of the fitnesses of the countries producing it. The authors hence propose a highly non-linear iterative algorithm that, unlike the linear one of the Method of Reflections, penalizes the weight of products based on whether or not they are present in export baskets of undeveloped economies.

4.2 The Fitness-Complexity Algorithm (FC)

In **[21]** the authors introduce the definitions of fitness of a country and complexity of a product which are two related quantities that can be iteratively computed starting from the initial condition

$$\tilde{F}_c^{(0)} = 1 \;\; \forall c, \tag{16}$$

$$\tilde{Q}_p^{(0)} = 1 \;\; \forall p. \tag{17}$$

The algorithm is characterized by two steps per iteration, the first of which is

$$\tilde{F}_c^{(N)} := \sum_{p=1}^{N_p} m_{cp} Q_p^{(N-1)}, \tag{18}$$

$$\tilde{Q}_p^{(N)} := \left[\sum_{c=1}^{N_c} \frac{m_{cp}}{F_c^{(N-1)}} \right]^{-1}. \tag{19}$$

From which the second step which entails the definition of Fitness of countries and Complexity of products.

Def. 8: Fitness of a country

$$F_c^{(N)} := \frac{\tilde{F}_c^{(N)}}{\left\langle \tilde{F}_c^{(N)} \right\rangle_c} = \frac{\sum_{p=1}^{N_p} m_{cp} Q_p^{(N-1)}}{\left\langle \sum_{p=1}^{N_p} m_{cp} Q_p^{(N-1)} \right\rangle_c}. \tag{20}$$

Def. 9: Complexity of a product

$$Q_p^{(N)} := \frac{\tilde{Q}_p^{(N)}}{\left\langle \tilde{Q}_p^{(N)} \right\rangle_p} = \frac{\left[\sum_{c=1}^{N_c} \frac{m_{cp}}{F_c^{(N-1)}} \right]^{-1}}{\left\langle \left[\sum_{c=1}^{N_c} \frac{m_{cp}}{F_c^{(N-1)}} \right]^{-1} \right\rangle_p}. \tag{21}$$

As one can easily appreciate the algorithm proposed in **Eq.(20)** and **Eq.(21)** is evidently highly non-linear unlike the one proposed in **Eq.(7)** and **Eq.(8)**. In and of itself, the formula in **Eq.(20)** for the Fitness of a country is not much dissimilar from the one in **Eq.(7)** for the HH Diversification. The main difference among the two iterative algorithms lies in how the complexity of products is estimated through **Eq.(21)**. The authors proposed such equation as it is the simplest way to implement the desired behavior. The sum at the denominator is dominated by the producers with the lowest fitness and increases as the total number of exporter increases, thus considering the ubiquity of the product. Finally inverting the sum makes the metric Q coherent with its positive meaning of complexity **[21](2)**.

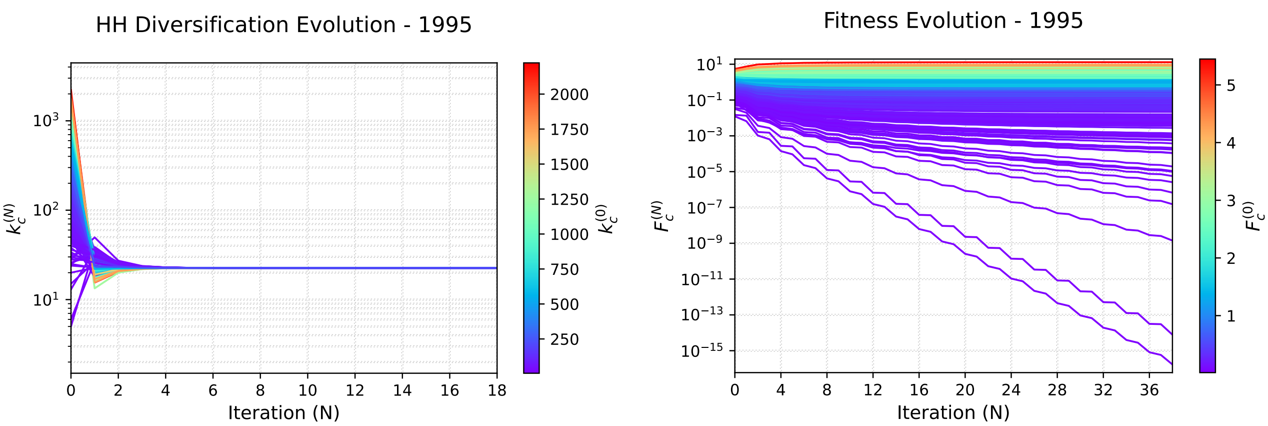

Figure 6: Comparison between countries' HH Diversification metric k_c convergence at each iteration (left) and Fitness F convergence at each iteration (right) in log scale.

4.3 Criticisms to the FC

In **[15]** Morrison et al. argue that the FC Algorithm proposed in **Subsection 4.2** does not yet constitute the state of the art of the measures of complexity in the world trade web due intrinsic instability in the definition of the algorithm itself. They highlight evidences of theoretically predicted instabilities and show that complex products tend to be those that countries rarely produce rather than those that are intrinsically more difficult to produce, hence defying the framework alterations proposed in **Subsection 4.1**.

A first main criticism that the authors bring forward is that the definition of Product Complexity seems rather far fetched and not properly justified. They argue that this function was chosen as the simplest monotonically decreasing function that maintained the dimensionality of product complexity but that it is not necessarily the right choice. This clearly leaves a layer of vagueness in a definition that should be the backbone of the entire algorithm and that should have been better argued. However, the original authors of the FC algorithm and other academics did address this issue in a later article [**17**].

The authors than claim that the FC method is inherently unstable to minimal perturbations in the world trade web. This is especially relevant in sparse bipartite networks. The use of the coarse-grained 4-digits cataloging system instead of the 6-digits one may veil this issue in that it basically eliminates the possibility of having a significant country-product combination in the binarized bi-adjacency matrix whereby that specific country is the sole exporter in the whole network of that specific product. If that was the case than the FC algorithm would have the fitness of most countries to converge to zero due to the unproper classification of the complexity of the lonely product[**15**](**5**). The authors played with the sparsity of the matrix m_{cp} with variations to the requirements for its binarization by using different RCA threshold (RCA $\geq \alpha$, $\alpha \in \mathbb{N}_0$) to offer in **Figure 7** [**15**](**7**) a graphical comparison that makes the issue evident.

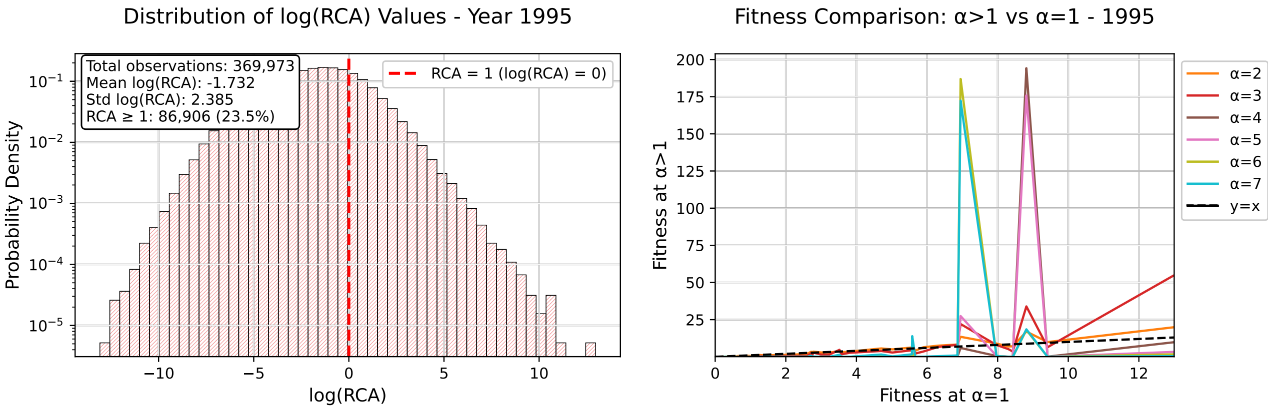

Figure 7: (Left) RCA distribution in year 1995 BACI (Right) Fitness at different thresholds VS Fitness at threshold = 1.

5 The Entropic Approach

5.1 Entropic Metrics

In [22] Gianluca Teza, Michele Caraglio and Attilio L. Stella propose a novel method to estimate complexity of countries and products in the world trade web that leverages the use of entropy functions. On one end, unlike the FC algorithm, they revert to a 'symmetric' iterative algorithm where the diversification of a country and the ubiquity of a product are calculated in much the same way, in a similar fashion to the one proposed in the MR algorithm. On the other end they maintain the use of non-linear functions, as proposed in the FC algorithm, to avoid convergence towards an information-less variable and to achieve instead convergence towards a fixed point. In particular the authors point out that in various scientific fields Shannon entropy is often appreciated as a diversity indicator for a system and wonder how come it never came to anyone's mind, albeit having a lot of time elapsed since the first works by HH were published, to adoperate it as an indicator for economic complexity [22](1).

Another novelty proposed by the authors is to drop the binarization of the M_{cp} matrix via RCA threshold, used both in the MR and in the FC algorithms, for it just reduces the available information regarding the economic volumes in terms of monetary value and may alter the results of the analysis based on the set threshold by even hiding potential divergences as seen in **Subsection 4.3** without offering any real relevant advantages, neither logical nor computational.

The analysis proposed in [22] hence begins with the computation of the shares of products in the basket of a country and analogously the shares of countries exporting a given product. This is fundamentally done by a row-wise or column-wise normalization to 1 that renders the rows or columns entries a complete set of probabilities with which we can than compute a Shannon Entropy. In mathematical terms such shares are defined as follows.

Def. 10: Shares of a product p in the basket of a country c

$$\xi_{cp}^{(0)} := \frac{M_{cp}}{\sum_{p'=1}^{N_p} M_{cp'}}. \tag{22}$$

Def. 11: Shares of countries c that export a given product p

$$\zeta_{cp}^{(0)} := \frac{M_{cp}}{\sum_{c'=1}^{N_c} M_{c'p}}. \tag{23}$$

In **Figure 8** we can appreciate the relation between the total exports of countries towards the rest of the world and the median shares of each country's basket. The 'Poverty Trap Line' in log-log scale highlights the dramatic monotonic drop in exports for countries with approximately less than 1000 products in their export basket. In the proximity of such line we find the so called underdeveloped economies such as Djibouti, Cayman Islands, Benin, Rwanda, Tuvalu, etc. At the very top we have the developed economies with important export volumes and with almost the totality of the HS classified products in their basket such as the USA, Italy, Japan, the Netherlands, etc. In the central part we have the semi-developed economies with discrete

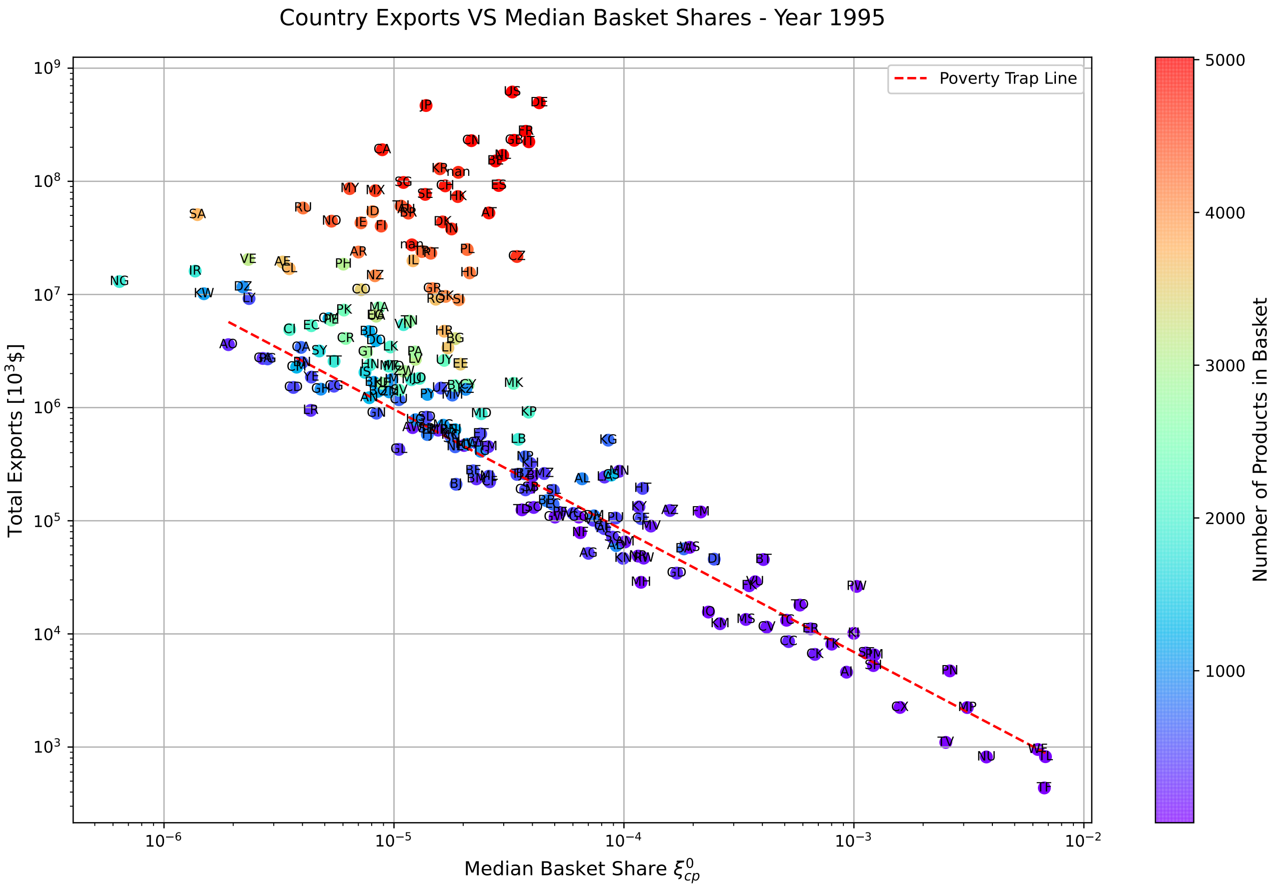

Figure 8: Scatter plot of the total exports of countries towards the rest of the world in 10^3\$ vs median ξ^0_{cp} shares of each country. The dashed red line denominated the 'Poverty Trap Line' represent a linear fit in log-log scale of countries with less than 1000 items in their export basket.

export volumes and number of products in their baskets such as Portugal, Romania, Mexico, Indonesia, etc. On the leftmost side we can identify a group of countries that we could name as oil-economies such as Iraq, Kuwait, Venezuela, Angola, etc. Regardless of the fact that this countries are characterized by discrete export values they fall dangerously close to the poverty trap line. This is in part imputable to their scarce economic diversification which makes them fragile to global market fluctuations and may impede proper wealth distribution.

5.2 Entropic Algorithm

The algorithm proposed in **[22](3)** begins with the definitions of bare entropic TCS diversification and bare entropic TCS ubiquity.

Def. 12: "Bare" TCS diversification

According to TCS we take the bare diversification of the export basket of a given country c as the Shannon entropy of the probability set given by the N_p non-zero $\xi^{(0)}_{cp}$ shares in the row of the M_{cp} matrix associated with such country. Formally:

$$H^{(0)}_c := -\sum_{p=1}^{N_p} \xi^{(0)}_{cp} \log(\xi^{(0)}_{cp}). \qquad \textbf{(24)}$$

Def. 13: "Bare" TCS ubiquity

According to TCS we take the bare ubiquity in the world trade web of a given product p as the Shannon entropy of the probability set given by the N_c non-zero $\zeta_{cp}^{(0)}$ shares in the column of the M_{cp} matrix associated with such product. Formally:

$$H_p^{(0)} := -\sum_{c=1}^{N_c} \zeta_{cp}^{(0)} \log(\zeta_{cp}^{(0)}). \tag{25}$$

The bare TCS diversification and ubiquity values are presented in **Figure 9** where they are highlighted by the color gradient. The sorted ranking for the bare entropic measure are the ones used to sort the bi-adjacency matrix M_{cp} in **Figure 4**.

TCS's entropic algorithm is based on the idea proposed by HH according to which we would want important products to weigh more in a country wealth indicator whilst, at the same time, having such importance be determined by how many developed countries are exporting said products. The algorithm proposed in **[22](4)** is hence 'symmetric' like the MR and starting from the initial conditions given by **Eq.**(**24**) and **Eq.**(**25**) we can update the values of the shares at the (N)-th step as

$$\xi_{cp}^{(N)} = \frac{M_{cp} f\left(H_p^{(N)}\right)}{\sum\limits_{p'=1}^{N_p} M_{cp'} f\left(H_{p'}^{(N)}\right)}, \tag{26}$$

$$\zeta_{cp}^{(N)} = \frac{M_{cp} g\left(H_c^{(N)}\right)}{\sum\limits_{c'=1}^{N_c} M_{c'p} g\left(H_{c'}^{(N)}\right)}. \tag{27}$$

where $f\left(H_p\right)$ and $g\left(H_c\right)$ are functions the form of which is not necessarily unique and is dictated by the aims of the analysis in order to achieve the desired qualitative behavior. In **[22]** they are defined as

$$f(H_p) := \log(N_c) - H_p, \tag{28}$$

$$g(H_c) := \log(N_p) - H_c. \tag{29}$$

At the $(N+1)$-th iteration of the algorithm the TCS diversification and ubiquity then become

$$H_c^{(N+1)} = -\sum_{p=1}^{N_p} \xi_{cp}^{(N)} \log\left(\xi_{cp}^{(N)}\right), \tag{30}$$

$$H_p^{(N+1)} = -\sum_{c=1}^{N_c} \zeta_{cp}^{(N)} \log\left(\zeta_{cp}^{(N)}\right). \tag{31}$$

The choice for the weight functions in **Eq.**(**28**) and **Eq.**(**29**) needs to reflect the fact that the more a product is ubiquitous the less it should contribute to the diversity of a basket of a country and that the more a country is wealthy and developed, the less it should count in determining the ubiquity of a product. The two functions are hence designed to invert the concept of diversity and ubiquity respectively.

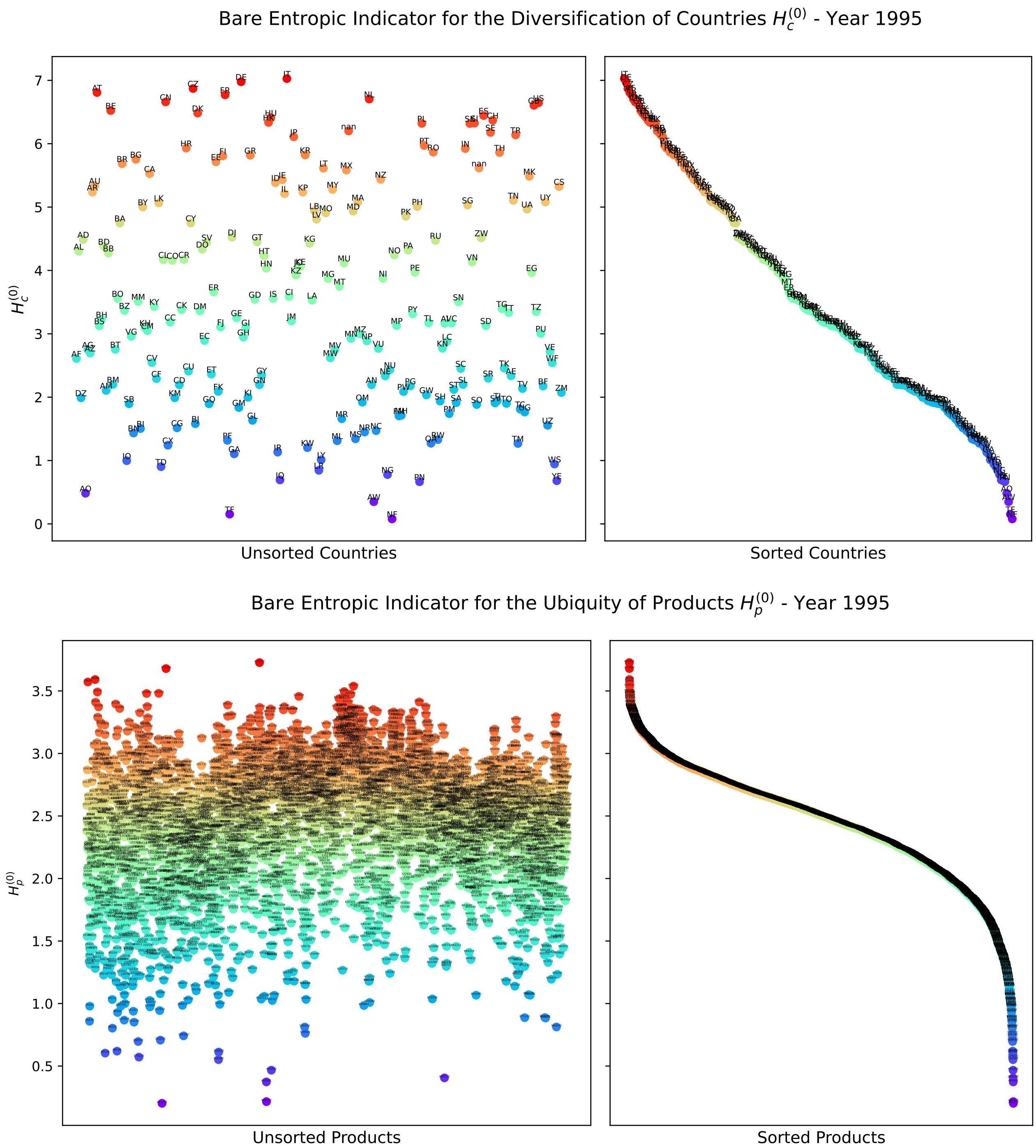

Figure 9: (Top left) bare TCS diversification; (Top right) sorted bare TCS diversification; (Bottom left) bare TCS ubiquity; (Bottom right) sorted bare TCS ubiquity.

Due to the properties of Shannon entropies, the iterative scheme can be seen as the continuous map

$$\varphi : [0, \log(N_p)]^{N_c} \otimes [0, \log(N_c)]^{N_p} \to [0, \log(N_p)]^{N_c} \otimes [0, \log(N_c)]^{N_p} \quad (32)$$

of a compact set in itself and, given its continuity, the existence of a fixed point to which the algorithm can stably converge is granted by Brower's fixed point theorem [11].

Thm. 2: Brower's fixed point theorem

Let $D \neq \emptyset$ be a compact convex set in $\mathbb{R}^n$ and let $\varphi : D \to D$ be continuous. Then the existence of a fixed point x such that $x = \varphi(x)$ is guaranteed.

The euclidean distance between iterative steps is

$$d^{(N)} = \sqrt{\sum_{c=1}^{N_c} \left(H_c^{(N+1)} - H_c^{(N)}\right)^2 + \sum_{p=1}^{N_p} \left(H_p^{(N+1)} - H_p^{(N)}\right)^2}. \tag{33}$$

By iterating **Eq.(30)** and **Eq.(31)** we can appreciate the exponential convergence to the fixed point as seen in **Figure 10**

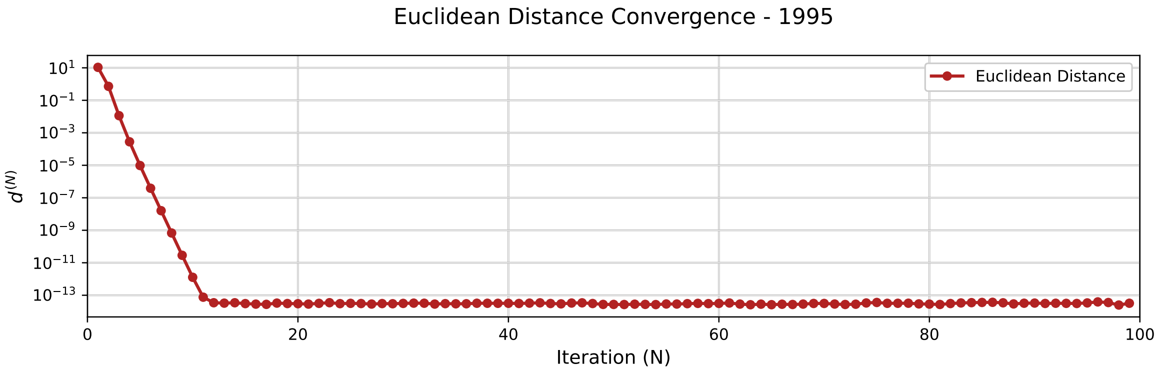

Figure 10: Euclidean distance $d^{(N)}$ evaluated between consecutive steps of the iterative entropic algorithm.

In **Figure 11** we can appreciate the step-by-step evolution of the TCS diversification and the TCS ubiquity for 100 iterations and the quick convergence of the algorithm to the final values. In **Figure 12** we then have the scatter plots of total exports in 10^3\$ against the final values for countries' TCS diversification and ubiquity.

5.3 Intra and Inter Sectorial Entropies

The Shannon entropy function offers us another advantage as its behavior at different level of data aggregation (HS4 or HS6) allows us to observe, via the study of intra and intersectorial entropies, the emergence of economic complexity without the issues present in other methods. As mentioned in **Subsection 1.2** we can aggregate products p with a 6 digits categorization in 4 digit macro-categories P obtaining coarse grained shares of the form

$$\xi_{cP} = \sum_{p \in P} \xi_{cp} \tag{34}$$

with which we can compute a coarse grained Shannon entropy

$$H_c^{CG} = -\sum_{P=1}^{N_P} \xi_{cP} \log(\xi_{cP}) \tag{35}$$

that serves as a coarse grained measure of diversification of a country. In particular, due to the definition of Shannon Entropy we have that

$$H_c \geq H_c^{CG} \tag{36}$$

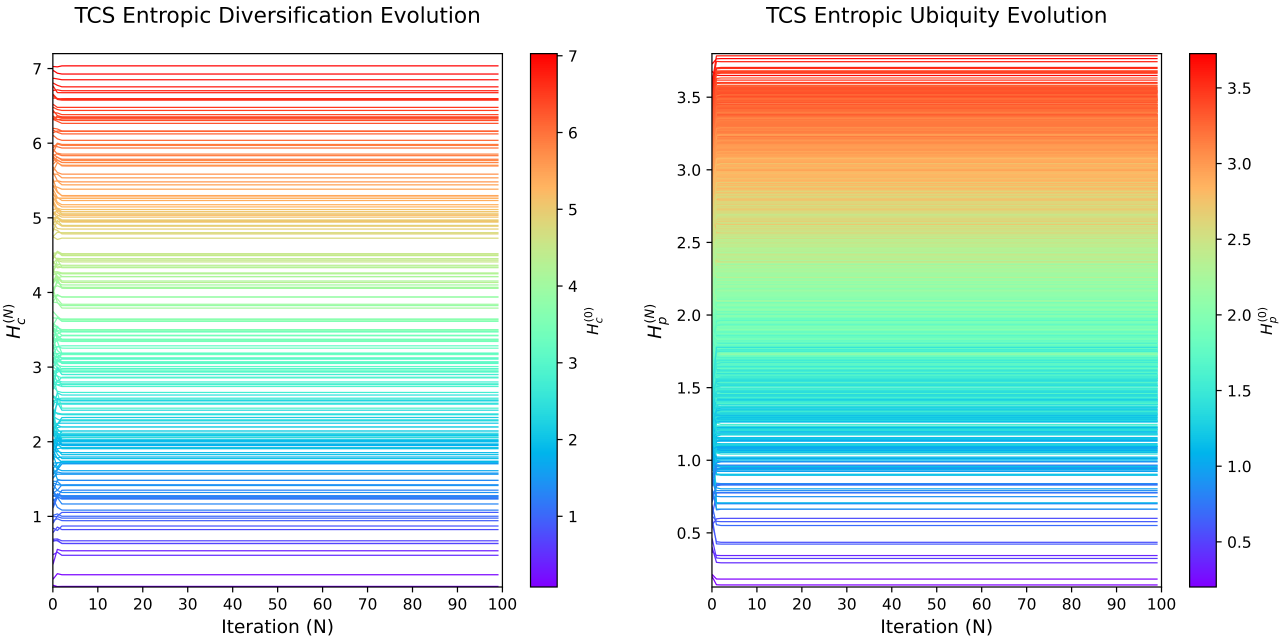

Figure 11: (Left) Iterative evolution of TCS diversification H_c; (Right) iterative evolution of TCS ubiquity H_p.

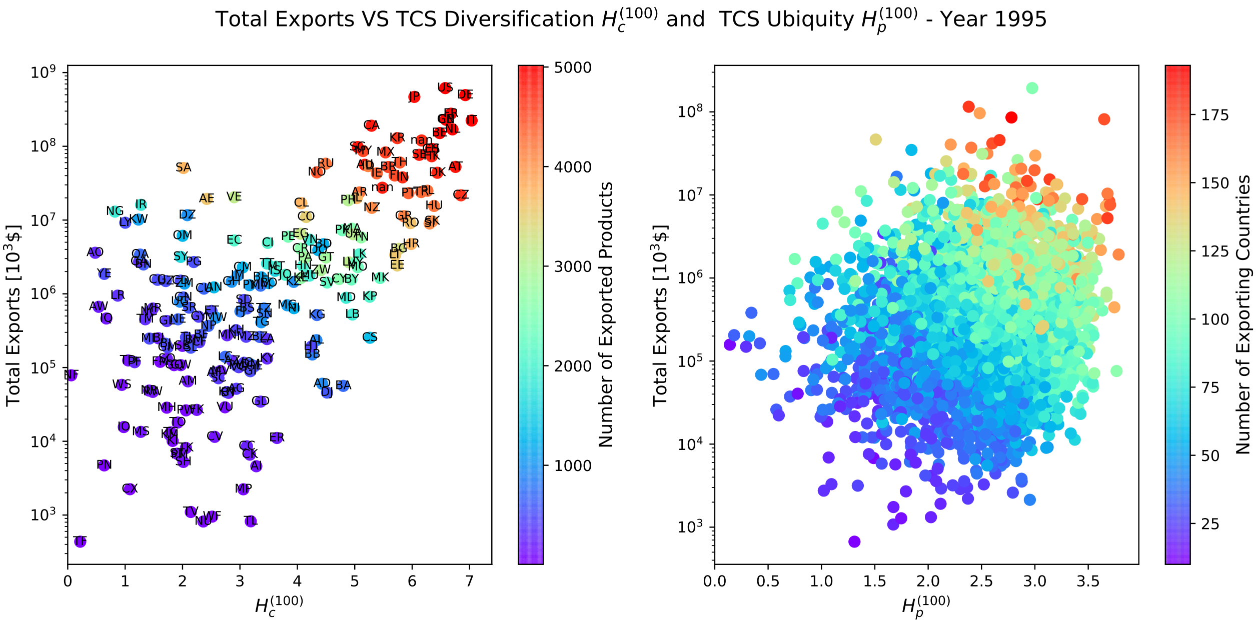

Figure 12: Scatter plots of total exports in 10^3\$ against (Left) TCS diversification H_c and (Right) TCS ubiquity H_p

where the equality holds if and only if every macro-category P contains only one product p. We can take H_c^{CG} as defined in **Eq.(35)** as a measure of inter-sectorial diversification. In the leftmost graph of **Figure 13** we can appreciate how for more diversified countries the dots in the scatter plot of the TCS diversification H_c against the TCS coarse diversification H_c^{CG} tend to distance themselves from the bisector of the plane quadrant unlike the ones from less diverse countries. This tells us that more underdeveloped countries have a inter-sectorial entropy similar to their diversification. Indeed from their difference we have

$$\Delta H_c = H_c - H_c^{CG} = \sum_{P=1}^{N_P} \xi_{cP} H_{cP}, \tag{37}$$

where

$$H_{cP} = -\sum_{p \in P} \frac{\xi_{cp}}{\xi_{cP}} \log\left(\frac{\xi_{cp}}{\xi_{cP}}\right), \tag{38}$$

which we can take as a measure of intra-sectorial entropy.

In the rightmost graph of **Figure 13** we can see how the the more diversified economies also have a much higher intra-sectorial diversification ΔH_c highlighting the level of sophistication in their productive capabilities and the emergence of complexity in the world trade web.

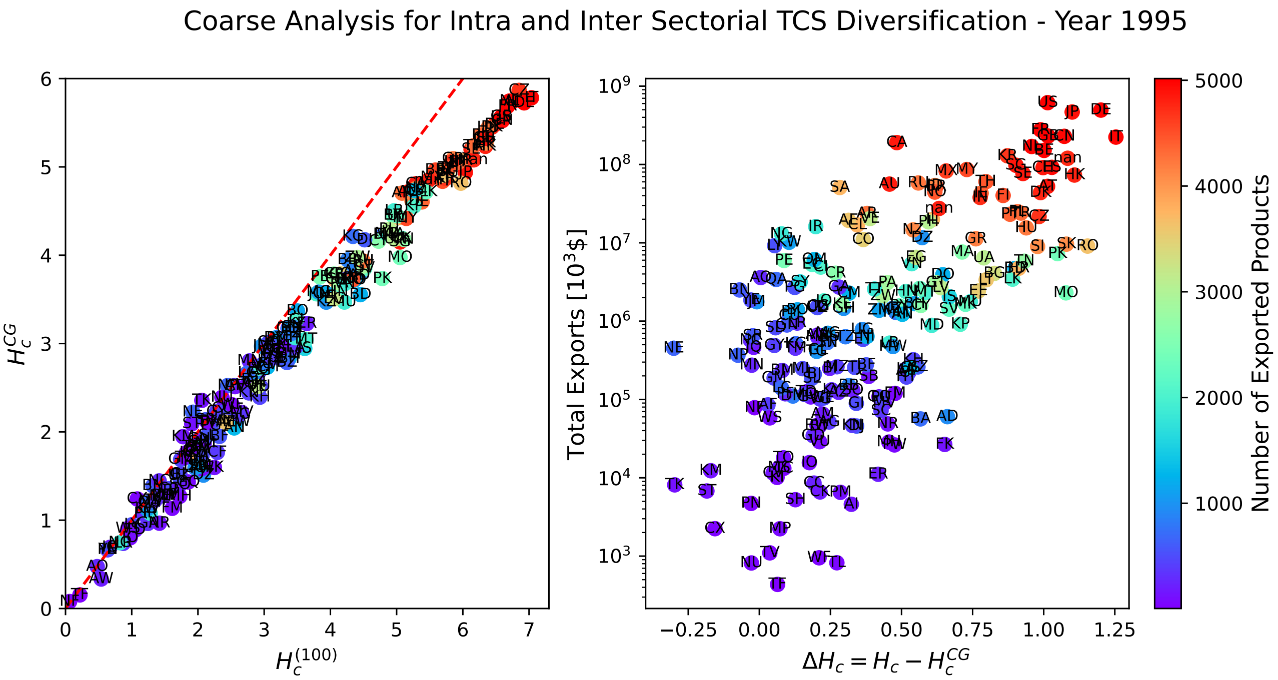

Figure 13: (Left) Coarse grained diversification H_c^{CG} against TCS diversification; (Right) Total exports in 10^3\$ against intra-sectorial diversification ΔH_c.

6 Conclusions

In this thesis work we walked through more than a decade worth of research in complexity metrics in the world trade web illustrating the trial and error process that incurred before reaching what we may now consider as the state of the art in economic complexity metrics. In **Figure 14** is proposed an illustrative comparison between the bare HH diversification ranking, the Fitness ranking and the TCS entropic diversification ranking for all countries in the BACI dataset. Whilst the first one doesn't hold much information, the differences between the second and the third rankings are evidently relevant and this is clearly related, as we discussed and beside the nature of the iterative algorithms, to the difference in information content in the data used for the analysis due to the use of the binarized RCA_{cp} bi-adjacency matrix or the simple bi-adjacency matrix M_{cp}.

The results obtained with the method presented in [**22**] by Gianluca Teza, Michele Caraglio and Attilio L. Stella can furthermore prove to be useful fields other than economics. With the right adjustments they might for instance be applied in ecology to quantify the complexity of an ecosystem described as a bipartite network of predators and prays or plants and herbivores. The complexity of a given ecosystem could arguably be taken as a measure of its biodiversity and consequent wellness. This highlights the interdisciplinarity that characterizes the physics of complexity.

Regardless of the already relevant achievement of econophysics, such as the one presented in this thesis, the antagonism and skepticism of some academics towards the so called 'heterodox' economic theory persists [**2**][**5**]. The most relevant issue may be the perceived difference, which seems hard to even out, in the approach of classically trained economists and physicists to the subject as the two hardly find occasions to properly debate their ideas on the same playing field. It is only auspicable that the two groups may work closer together in the future to further develop the field of complexity economics and econophysics in general.

6 Conclusions

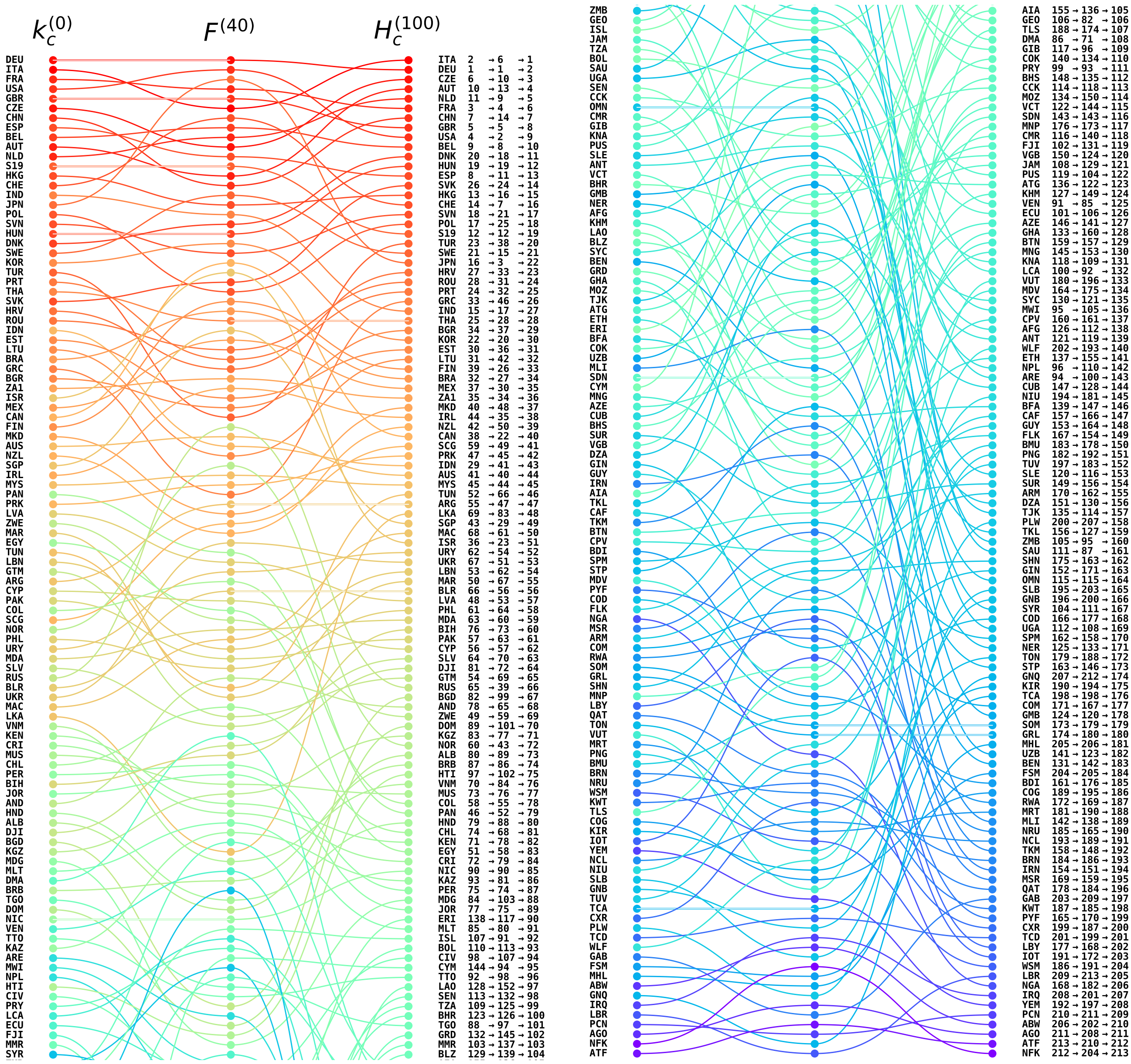

Figure 14: $k_c^{(0)}$ HH Diversification VS $F_c^{(40)}$ Fitness VS $H_c^{(100)}$ TCS Diversification ordered rankings

References

1 Bela Balassa. “Comparative Advantage in Manufactured Goods: A Reappraisal”. In: *The Review of Economics and Statistics* 68.2 (1986), pp. 315–319.

2 Philip Ball. “Culture Crash”. In: *Nature* (2006).

3 William J. Baumol. *utility and value*. Encyclopedia Britannica. Feb. 2024. (Visited on 08/25/2025).

4 Matthieu Cristelli et al. “Measuring the Intangibles: A Metrics for the Economic Complexity of Countries and Products”. In: *PLoS ONE* 8.8 (2013), e70726. DOI: 10.1371/journal.pone.0070726. URL: https://doi.org/10.1371/journal.pone.0070726.

5 Mauro Gallegati et al. “Worrying trends in econophysics”. In: *Physica A: Statistical Mechanics and its Applications* (May 2006). Available online 12 May 2006.

6 Guillaume Gaulier and Soledad Zignago. *BACI: International Trade Database at the Product-Level. The 1994-2007 Version.* Working Papers 2010-23. CEPII, 2010. URL: https://www.cepii.fr/CEPII/fr/publications/wp/abstract.asp?NoDoc=2726.

7 Gene M. Grossman and Elhanan Helpman. “Quality Ladders in the Theory of Growth”. In: *The Review of Economic Studies* 58.1 (Jan. 1991), pp. 43–61. ISSN: 0034-6527. URL: https://www.jstor.org/stable/2298044.

8 Ricardo Hausmann et al. *The Atlas of Economic Complexity: Mapping Paths to Prosperity.* Accessed on October 12, 2023. MIT Press, 2011. ISBN: 9780615546629. URL: https://atlas.cid.harvard.edu/.

9 C. A. Hidalgo et al. “The Product Space Conditions the Development of Nations”. In: *Science* 317.5837 (2007), pp. 482–487.

10 César A. Hidalgo and Ricardo Hausmann. “The Building Blocks of Economic Complexity”. In: *Proceedings of the National Academy of Sciences* 106.26 (2009), pp. 10570–10575. DOI: 10.1073/pnas.0900943106. URL: https://www.pnas.org/doi/10.1073/pnas.0900943106.

11 R. B. Kellogg, T. Y. Li, and J. Yorke. “A Constructive Proof of the Brouwer Fixed-Point Theorem and Computational Results”. In: *SIAM Journal on Numerical Analysis* 13.4 (1976), pp. 473–483.

12 Eric Kemp-Benedict. “An Interpretation and Critique of the Method of Reflections”. In: *MPRA Paper* 60705 (Dec. 2014). Accessed on 2023-10-17. URL: https://mpra.ub.uni-muenchen.de/60705/.

13 Rolf Landauer. “Irreversibility and heat generation in the computing process”. In: *IBM Journal of Research and Development* 5.3 (1961), pp. 183–191. DOI: 10.1147/rd.53.0183.

14 Karl Montevirgen. *price*. Encyclopedia Britannica. Aug. 2025. (Visited on 08/25/2025).

15 Greg Morrison et al. “On Economic Complexity and the Fitness of Nations”. In: *Scientific Reports* 7 (2017), p. 15332. DOI: 10.1038/s41598-017-14603-6. URL: https://doi.org/10.1038/s41598-017-14603-6.

16 David Ricardo. *On the Principles of Political Economy and Taxation.* London: John Murray, 1817.

17 Vito D. P. Servedio et al. “A new and stable estimation method of country economic fitness and product complexity”. Version 2. In: *arXiv preprint* (Oct. 2018). eprint: 1807.10276. URL: https://arxiv.org/abs/1807.10276.

18 Claude E. Shannon. “A Mathematical Theory of Communication”. In: *Bell System Technical Journal* 27.3 (July 1948), pp. 379–423. DOI: 10.1002/j.1538-7305.1948.tb01338.x.

19 Adam Smith. *An Inquiry into the Nature and Causes of the Wealth of Nations*. London: W. Strahan and T. Cadell, 1776.

20 A. Tacchella et al. “Economic complexity: Conceptual grounding of a new metrics for global competitiveness”. In: *Journal of Economic Dynamics and Control* 37 (2013), pp. 1895–1919. DOI: https://doi.org/10.1016/j.jedc.2013.03.005.

21 Andrea Tacchella et al. “A New Metrics for Countries’ Fitness and Products’ Complexity”. In: *Scientific Reports* 2 (2012), p. 723. DOI: 10.1038/srep00723. URL: https://doi.org/10.1038/srep00723.

22 Gianluca Teza, Michele Caraglio, and Attilio L. Stella. “Entropic Measure Unveils Country Competitiveness and Product Specialization in the World Trade Web”. In: *Scientific Reports* 11.10189 (2021). DOI: 10.1038/s41598-021-89519-3. URL: https://doi.org/10.1038/s41598-021-89519-3.

23 Myron Tribus and Edward C. McIrvine. “Energy and Information”. In: *Scientific American* 225.3 (1971), pp. 179–190. URL: http://www.jstor.org/stable/24923125.

24 U.S. Census Bureau. *North American Industry Classification System (NAICS) Manual*. U.S. Census Bureau. 2022. URL: https://www.census.gov/naics/reference_files_tools/2022_NAICS_Manual.pdf.

25 United Nations Statistics Division. *Standard International Trade Classification, Revision 4*. Sales No. E.06.XVII.10. New York: United Nations, 2006. ISBN: 92-1-161493-7.