Politecnico di Torino

DIPARTIMENTO DI SCIENZA APPLICATA E TECNOLOGIA
Corso di Laurea in Physics of Complex Systems

Tesi di Laurea Magistrale

Economic Complexity Algorithms in Complex Networks: Applications to Economics and Ecology

Laureando:
Emanuele Calò
Matricola 288272

Relatore:
Prof. Alfredo Braunstein

Correlatore:
Dr. Vito D.P. Servedio

Anno Accademico 2021-2022

ABSTRACT

Economic Complexity (EC) algorithms estimate the fitness and complexity of the nodes of bipartite networks, such as the network of countries and their exported products. In their application to country-product networks, EC algorithms try to shed light on the hidden capabilities of countries. Capabilities represent the intangible assets driving the development and wealth of countries, such as infrastructures and educational systems. We begin by analyzing the linear Economic Complexity Index (ECI) method, for which a country is fit if it exports complex products. Analogously, a product is complex if fit countries export it. We then study the non-linear Economic Fitness Complexity in its non-homogeneous version (NH-EFC). In this case, countries with high fitness export various products, from very simple to very complex. If a non-fit country exports a product, this product has low complexity. The primary outcome of both algorithms is to deliver the list of countries ranked according to their fitness. These algorithms cease to work when the network is not bipartite, even if there is only one "weak" link between two nodes of the same class (e.g., country-country or product-product). Our task has been to generalize them to deal with non-bipartite networks. We show that the NH-EFC is more stable after introducing small non-bipartite perturbations (random uncorrelated noise), i.e., the perturbation leaves the ranking of countries almost unchanged. Eventually, we apply the NH-EFC to study the complexity of the prey-predator ecosystem in Florida Bay and disclose information on the hidden capabilities of the organisms in the system.

Da queste mura, una risata invaderà il mondo, infettando di coraggio il laborioso peone
dell'antichità chino sul suo lavoro ∼ Jack Kerouac

Contents

Introduction

Recently, new tools have been introduced to help assess the health of countries' economic growth. These tools complement those already used in macroeconomics and are generally considered to belong to the Economic Complexity (EC) field. In this new view, the international trade among countries can be described as a bipartite network composed of countries on the one side and thousands of their exported goods on the other. Countries and products are mutually linked, but no link is considered between countries nor between products. By exploiting the network structure, new metrics for evaluating the economy of countries and the quality of products can be therefore built. EC algorithms aim to gain information on the capabilities of the countries, that is, all the intangible assets driving their development and wealth, such as infrastructures and educational systems. The first work in this new direction we will introduce is the Economic Complexity Index (ECI) algorithm [1]. The key idea behind the linear map defining the algorithm is that a country is fit if it exports complex products. Analogously, a product is complex if fit countries export it. The second EC algorithm we will introduce is the non-linear Economic Fitness Complexity (EFC) method [2]. It is based on the fact that countries with high Fitness export a large variety of products, from very simple to very complex, while countries with low Fitness export only the products exported by most countries. The third and final algorithm we will see is the non-homogeneous version of the latter (NH-EFC), which solves the convergence issues affecting the original algorithm [3]. All these algorithms cease to work as soon as the network is not bipartite anymore, even if there is only one "weak" link between two nodes of the same class (country-country or product-product). The aim of this thesis is thus to generalize EC algorithms to apply them also to non-bipartite networks.

The structure of this work is the following. In Chapter 1, we will present an overview of the EC literature, first by presenting and explaining the concept of capability, then introducing the above three algorithms and the equations defining them. Then, in Chapter 2, we will express the ECI algorithm in matrix representation and study its spectral properties. Moreover, we will see that by recasting the formalism in terms of the adjacency matrix of the network, we manage to use EC algorithms even in the case of non-bipartite networks. Then, we will add random uncorrelated noise (i.e., linking countries with countries and products with products randomly). Focusing firstly on the ECI algorithm, we will compare the results obtained by directly applying the iterative method with those obtained using the components of the eigenvectors. Secondly, we will analyze the stability of both the ECI and the NH-EFC algorithms with respect to this perturbation. Finally, in Chapter 3, we will apply the NH-EFC method to the Florida Bay prey-predator network to assess the hidden capabilities of the system. Moreover, we will investigate the correlation between

Fitness and Complexity and use two different methods to classify approximately the nodes into two classes, prey and predator.

Background & Notation

Here we introduce the background definitions and results we shall use throughout the thesis. First of all, we rigorously introduce the concepts of graph and bipartite/tripartite graph.

Definition 0.1 (Digraph, Graph). A *directed graph* (digraph) G is an ordered pair $G = (V, E)$ where V is called the set of vertices (or nodes) and $E \subseteq V \times V$ is called the set of edges. Self-edges (i, i) will not be allowed, that is $(i, i) \notin E$. An *undirected graph* (sometimes called simply a graph) is an ordered pair $G = (V, E)$ where V is called the set of vertices (or nodes) while $E \subset [V \times V]$ is called the set of edges and it is a subset of pairs of vertices in which opposite pairs $[(v, w)] = [(w, v)]$ are identified in a single equivalence class. The equivalence class symbol $[\cdot]$ is normally omitted.

Definition 0.2 (Bipartite Graph). A graph is said to be *bipartite* if $G = (V, E)$ with $V = A \cup B$, $\emptyset = A \cap B$ and $E \subset A \times B$.

Definition 0.3 (Tripartite Graph). A graph is said to be *tripartite* if $G = (V, E)$ with $V = A \cup B \cup C$, $\emptyset = A \cap B = B \cap C = C \cap A$ and $E \subset A \times B \vee B \times C \vee C \times A$.

We now define the important concept of stochastic matrix.

Definition 0.4 (Stochastic Matrix). A square matrix is said to be *stochastic* if each of its entries is a nonnegative real number and each row (or column) sums to 1.

In Chapter 2 we will use an important result concerning (also) stochastic matrices, that is the Perron Frobenius theorem.

Theorem 0.5 (Perron Frobenius Theorem). *Let* $\mathbf{T}$ *be an* $N \times N$ *stochastic matrix such that for a certain power* $\mathbf{T}^m$ *all entries are positive. Then* $\mathbf{T}$ *has one (up to a scalar) eigenvector* $\underline{\rho}_{\tau_{max}}$ *with positive components and no other eigenvectors with nonnegative components. The eigenvalue* τ_{max} *corresponding to* $\underline{\rho}_{\tau_{max}}$ *is simple, equal to* 1 *and greater than the absolute values of all other eigenvalues.*

In terms of notation, a bold letter will denote a matrix, for instance $\mathbf{X}$, while X_{ij} will indicate its entries, and $\mathbf{X}^T$ its transpose.

1. Economic Complexity Literature

One of the main objectives of economic complexity (EC) algorithms is to estimate the capabilities of countries from the products they export. Capabilities represent the intangible assets that drive the development and wealth of countries, i.e., the socio-economic substrate on which the national economic system is built, such as infrastructures and educational systems. Nevertheless, there is no universal standard measure for characterizing them. Each product requires specific and necessary capabilities that a country must own to produce and then export that product. Thus the basket of exported products of a country contains hidden information about its fundamental capabilities. We can represent this framework as a tripartite network country (C) - capability (K) - product (P) in which capabilities are the intermediate layer between countries and products. The non-observability of capabilities means that we can only access the "contraction" of this tripartite network into the bipartite country-product network. In this sense, the export of countries can be informative about capabilities. From now on, we will focus on the bipartite graph C - P.

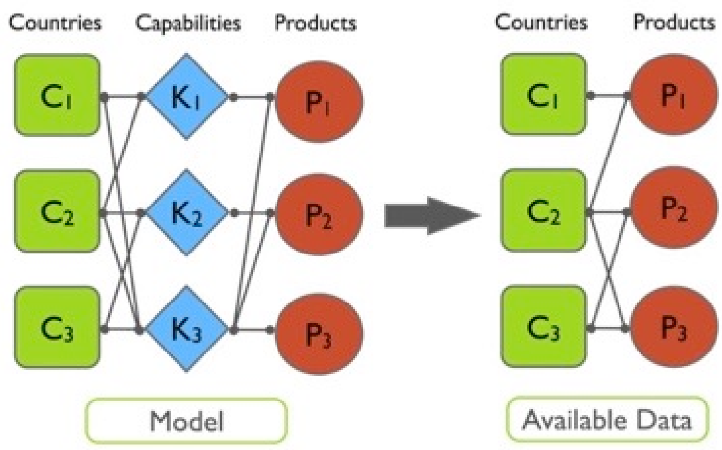

Figure 1.1: Tripartite network C - K - P projection into C - P bipartite network. Image taken from [4].

1.1 Economic Complexity Index

In 2009, Hidalgo and Hausmann (from now on, HH) published [1], introducing the so-called Method of Reflections. This algorithm allows to rank countries and products in the international market by using only the information contained in the country-product binary matrix $\mathbf{M}$, whose entries M_{cp} are 1 if the country c exports the product p and 0 otherwise. For a rigorous derivation of the matrix $\mathbf{M}$ and more information on the dataset that we will use, see appendix A.

Let N_c and N_p be respectively the number of countries and the number of products. The total number of exported products by each country is thus given by

$$k_c = \sum_{p=1}^{N_p} M_{cp}, \tag{1.1}$$

while the number of countries exporting a certain product is given by

$$k_p = \sum_{c=1}^{N_c} M_{cp}. \tag{1.2}$$

These two quantities respectively represent the *diversification* of c and the *ubiquity* of p. The iterative map defining the algorithm is the following

$$\begin{cases} k_c^{(n+1)} = \frac{1}{k_c} \sum_{p=1}^{N_p} M_{cp} k_p^{(n)} \\ k_p^{(n+1)} = \frac{1}{k_p} \sum_{c=1}^{N_c} M_{cp} k_c^{(n)}, \end{cases} \tag{1.3}$$

with the initial conditions $k_c^{(0)} \equiv k_c$ and $k_p^{(0)} \equiv k_p$. That is, the diversification k_c represents the zero$^{\text{th}}$ order measure of the *quality* of country c (the more products a country exports, the best its position in the market). While the ubiquity k_p represents the zero$^{\text{th}}$ order measure of the *dis-value* of product p in the global competition (the more countries produce a product, the least is its value on the market). $k_c^{(n)}$ and $k_p^{(n)}$ are respectively called Economic Complexity Index (ECI) and Product Complexity Index (PCI). The countries, sorted by their ECI, will form a ranking, and analogously the products sorted by their PCI. We will refer to this algorithm as ECI.

In the standard view of Ricardian theory [5], an optimal situation occurs when the national economies have a high degree of specialization so that it would be possible to rearrange rows and columns of **M** to make it almost block-diagonal. Conversely, by reordering the rows and columns by k_c and k_p, **M** approximately takes a triangular shape, as shown in Fig. 1.2. Thus evolved countries become more complex by acquiring a higher degree of diversification instead of specializing.

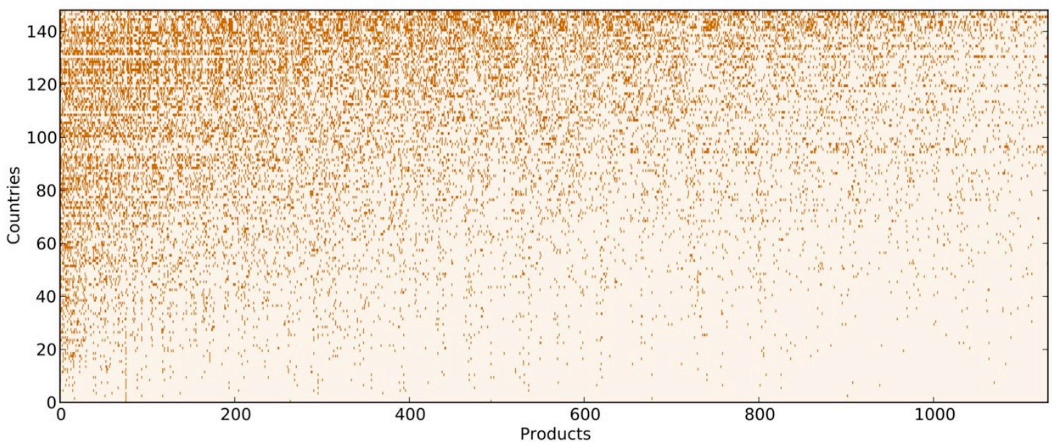

Figure 1.2: Matrix **M** for the year 2010 with rows and columns reordered by respectively decreasing k_c and k_p. Image taken from [4].

1.2 Economic Fitness Complexity

The approximate triangular structure of $\mathbf{M}$ implies that the information that a diversified country exports a product conveys little information about the complexity of the product itself. Indeed, these countries export almost all products. Conversely, the products exported by less competitive countries are supposed to have low complexity. To account for this, Tacchella et al. [6] introduced a non-linear algorithm, i.e., the Economic Fitness Complexity (EFC). The non-linear iterative map defining the algorithm is the following

$$\begin{cases} \tilde{F}_c^{(n+1)} = \sum_p M_{cp} Q_p^{(n)} \\ \tilde{Q}_p^{(n+1)} = \dfrac{1}{\sum_c M_{cp} \dfrac{1}{F_c^{(n)}}} \end{cases} \rightarrow \begin{cases} F_c^{(n)} = \dfrac{\tilde{F}_c^{(n)}}{\langle \tilde{F}_c^{(n)} \rangle_c} \\ Q_p^{(n)} = \dfrac{\tilde{Q}_p^{(n)}}{\langle \tilde{Q}_p^{(n)} \rangle_p}, \end{cases} \tag{1.4}$$

with initial conditions $\tilde{F}_c^{(0)} = 1 \ \forall c$ and $\tilde{Q}_p^{(0)} = 1 \ \forall p$. $\tilde{F}_c^{(n+1)}$ and $\tilde{Q}_p^{(n+1)}$ are the intermediate values of $F_c^{(n+1)}$ and $Q_p^{(n+1)}$. Indeed, at each step the intermediate values are normalized by their algebraic means. The normalization is required for the stabilization of the non-linear map. The two new quantities introduced in Eq. 1.4 are the Complexity of a product (Q_p) and the Fitness of a country (F_c). Fitness is obtained as the sum of the Complexity of the exported products, while Complexity non-linearly depends on the Fitness of the exporting countries. Indeed, as explained above, countries with high Fitness should have less weight in determining the Complexity of a product, while countries with low Fitness should contribute strongly. That is, if a country with high Fitness exports a product, this does not affect much its Complexity, while if a country with low Fitness exports a product, then this product has low Complexity (i.e., it is simple).

Comparing (1.3) and Eq. (1.4) we notice a fundamental difference. Indeed, for the first order of the ECI, the successfulness of a country is given by the average Complexity of its products. While, for EFC, Fitness is equal to the sum of the Complexity of its products. Moreover, as shown in [4], ECI rapidly loses correlation with the capabilities of the countries when iterated. On the other hand, Fitness preserves the information on the diversification of the export baskets. For instance, let us consider two countries, one developed and one not. The first will have a more diversified export basket, and on average, the Complexity of its products will be higher than the one of the second. Therefore the next iteration should increase the gap between the two compared to the starting point. This reasoning applies to all the orders of the iterative method.

1.3 Non-Homogeneous Economic Fitness Complexity

Despite the success of EFC, which has been used for state-of-the-art forecasting for economic growth [2, 7], it suffers from a couple of issues. The first one is a convergence issue [8]. In the original work, they have addressed it by introducing the notion of "rank convergence" instead of dealing with absolute convergence. That is, they consider the fixed point achieved when the ranking of countries stays unaltered step by step. The second one is that despite their finite capabilities, the countries that export no good have zero Fitness. Servedio et al. [3] solved these issues by introducing a non-linear non-homogeneous map (NH-EFC). Introducing the new variable $P_p = Q_p^{-1}$ and the parameters ϕ_c and π_p, the set of iterated variables thus becomes

$$\begin{cases} F_c^{(n)} = \phi_c + \sum_{p'} M_{cp'} / P_{p'}^{(n-1)} \\ P_p^{(n)} = \pi_p + \sum_{c'} M_{c'p} / F_{c'}^{(n-1)}, \end{cases} \tag{1.5}$$

with $1 \leq c \leq N_c$ and $1 \leq p \leq N_p$, and initial conditions $F_c^{(0)} = 1 \; \forall c$ and $P_p^{(0)} = 1 \; \forall p$. In contrast to the original map, the non-homogeneous one is not defined up to a multiplicative constant. The normalization procedure is thus not necessary. The parameter ϕ_c represents the intrinsic Fitness of a country. Indeed, each country has a set of capabilities characterizing it regardless of its exports. The inverse of the parameter π_p can be considered an innovation threshold. If a product has not been invented yet, then no country exports it, and its quality lies at its maximum value. The smaller the parameter is, the higher the Complexity of the product and more sophisticated capabilities are necessary to produce it. Let now $\phi_c = \pi_p = \delta \; \forall c, p$. Introducing the new variables $\tilde{P}_p = P_p/\delta$ and $\tilde{F}_c = F_c \delta$, Eq. (1.5) can be recast in the form

$$\begin{cases} \tilde{F}_c^{(n)} = \delta^2 + \sum_{p'} M_{cp'} / \tilde{P}_{p'}^{(n-1)} \\ \tilde{P}_p^{(n)} = 1 + \sum_{c'} M_{c'p} / \tilde{F}_{c'}^{(n-1)}. \end{cases} \tag{1.6}$$

In particular, as soon as the parameter δ^2 is much smaller than the typical value of M_{cp} (i.e., $\delta^2 \ll 1$), the fixed point does not depend on δ. In this new metric, $\tilde{Q}_p = (\tilde{P}_p - 1)^{-1}$ represents the Complexity of a product.

2. Introducing Non-Bipartite Perturbations

2.1 ECI

Let us now focus on the ECI algorithm. To homogenize the notation, in analogy with the EFC algorithm, let $F_c^{(n)}$ and $Q_p^{(n)}$ identify the quantities previously labeled as $k_c^{(n)}$ and $k_p^{(n)}$ defined in Eq. (1.3). We thus have

$$\begin{cases} F_c^{(n+1)} = \frac{1}{k_c} \sum_p M_{cp} Q_p^{(n)} \\ Q_p^{(n+1)} = \frac{1}{k_p} \sum_c M_{cp} F_c^{(n)}, \end{cases} \tag{2.1}$$

with $c = 1, \ldots, N_c$ and $p = 1, \ldots, N_p$. Inserting the second into the first and vice versa, we get

$$\begin{cases} F_c^{(n)} = \frac{1}{k_c} \sum_{c'p} \frac{1}{k_p} M_{cp} M_{c'p} F_c^{(n-2)} \\ Q_p^{(n)} = \frac{1}{k_p} \sum_{cp'} \frac{1}{k_c} M_{cp} M_{cp'} Q_p^{(n-2)}. \end{cases} \tag{2.2}$$

2.1.1 Matrix Representation and Spectral Analysis

We now want to recast the ECI algorithm in a vectorial form to study it through its spectral properties. We begin by introducing the matrices $\mathbf{D}^{-1}$ and $\mathbf{U}^{-1}$, defined as follows. Let $\vec{d}$ and $\vec{u}$ be respectively the N_c dimensional vector of the diversification of the countries and the N_p dimensional vector of the ubiquity of the products. That is, let $\vec{\mathbb{1}}_{N_c}$ and $\vec{\mathbb{1}}_{N_p}$ be respectively the N_c and N_p dimensional vectors of 1s. We have

$$\vec{d} = \mathbf{M}\vec{\mathbb{1}}_{N_c}, \qquad \vec{u} = \mathbf{M}^T \vec{\mathbb{1}}_{N_p}. \tag{2.3}$$

We thus construct the diagonal matrices $\mathbf{D}$ and $\mathbf{U}$ whose diagonal entries are the elements of the vectors $\vec{d}$ and $\vec{u}$. We can finally define $\mathbf{D}^{-1}$ and $\mathbf{U}^{-1}$ as

$$\begin{aligned} \mathbf{D}^{-1} &\doteq \begin{pmatrix} \frac{1}{d_1} & & \\ & \ddots & \\ & & \frac{1}{d_{N_c}} \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{k_{c=1}} & & \\ & \ddots & \\ & & \frac{1}{k_{c=N_c}} \end{pmatrix} \end{aligned} \tag{2.4}$$

and

$$
\begin{aligned}
\mathbf{U}^{-1} &\doteq \begin{pmatrix} \frac{1}{u_1} & & \\ & \ddots & \\ & & \frac{1}{u_{N_p}} \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{k_{p=1}} & & \\ & \ddots & \\ & & \frac{1}{k_{p=N_p}} \end{pmatrix}.
\end{aligned}
\tag{2.5}
$$

That is, $\mathbf{D}^{-1}$ is the $N_c \times N_c$ diagonal matrix with (diagonal) entries the inverse of the diversification of the countries, while $\mathbf{U}^{-1}$ is the $N_p \times N_p$ diagonal matrix with (diagonal) entries the inverse of the ubiquity of the products. Therefore, Eq. (2.2) is equivalent to

$$
\begin{cases} \vec{F}^{(n)} = \mathbf{D}^{-1}\mathbf{M}\mathbf{U}^{-1}\mathbf{M}^T\vec{F}^{(n-2)} \\ \vec{Q}^{(n)} = \mathbf{U}^{-1}\mathbf{M}^T\mathbf{D}^{-1}\mathbf{M}\vec{Q}^{(n-2)}, \end{cases}
\tag{2.6}
$$

where $\vec{F}^{(n)}$ and $\vec{Q}^{(n)}$ are respectively N_c and N_p dimensional vectors $\forall n$. We then define the matrices $\mathbf{S}_1$ and $\mathbf{S}_2$ as $\mathbf{S}_1 = \mathbf{M}\mathbf{U}^{-1}\mathbf{M}^T$ and $\mathbf{S}_2 = \mathbf{M}^T\mathbf{D}^{-1}\mathbf{M}$. We notice that these matrices are symmetric, indeed

$$
\begin{cases} \mathbf{S}_1^T &= \left(\mathbf{M}\mathbf{U}^{-1}\mathbf{M}^T\right)^T = \mathbf{M}\mathbf{U}^{-1}\mathbf{M}^T = \mathbf{S}_1 \\ \mathbf{S}_2^T &= \left(\mathbf{M}^T\mathbf{D}^{-1}\mathbf{M}\right)^T = \mathbf{M}^T\mathbf{D}^{-1}\mathbf{M} = \mathbf{S}_2, \end{cases}
\tag{2.7}
$$

since $\mathbf{U}^{-1}$ and $\mathbf{D}^{-1}$ are diagonal. Finally, defining the matrix $\mathbf{N}_1$ as $\mathbf{N}_1 = \mathbf{D}^{-1}\mathbf{S}_1$ and the matrix $\mathbf{N}_2$ as $\mathbf{N}_2 = \mathbf{U}^{-1}\mathbf{S}_2$, we can express Eq. (2.6) as

$$
\begin{cases} \vec{F}^{(n)} = \mathbf{N}_1\vec{F}^{(n-2)} \\ \vec{Q}^{(n)} = \mathbf{N}_2\vec{Q}^{(n-2)}, \end{cases}
\tag{2.8}
$$

or, in terms of the initial conditions, we have

$$
\begin{cases} \vec{F}^{(2n)} = \mathbf{N}_1^n\vec{F}^{(0)} \\ \vec{Q}^{(2n)} = \mathbf{N}_2^n\vec{Q}^{(0)}. \end{cases}
\tag{2.9}
$$

Eq. (2.8) is a linear equation, finding the ECI by the Method of Reflections can thus be reformulated as a fix-point problem and solved using the spectral properties of the (linear) system. In particular, since $\mathbf{N}_1$ and $\mathbf{N}_2$ are stochastic ergodic matrices[1], for the Perron Frobenius theorem (see Theorem 0.5), their spectrum of eigenvalues is bounded in absolute value by their unique upper eigenvalue $\lambda_1^{1,2} = 1$, where superscripts 1,2 refer to $\mathbf{N}_1$ and $\mathbf{N}_2$, respectively [9]. We now want to find an expression that allows us to compute the eigenvector associated with the second eigenvalue (for reasons that will soon become clear) without solving the full eigenvalue problem. Let us now focus on $\vec{F}^{(n)}$, that is on countries (analogous results hold for the products). We start by noticing that the eigenvectors of $\mathbf{N}_1$ are not orthogonal since $\mathbf{N}_1$ is not symmetric, we thus define the symmetric matrix $\mathbf{H}$

$$
\mathbf{H} = \mathbf{D}^{-\frac{1}{2}}\mathbf{S}_1\mathbf{D}^{-\frac{1}{2}} = \mathbf{D}^{+\frac{1}{2}}\mathbf{N}_1\mathbf{D}^{-\frac{1}{2}}.
\tag{2.10}
$$

[1]The fact that they are stochastic matrieces follows from their definition. Indeed, multiplying $\mathbf{M}$ and $\mathbf{M}^T$ respectively by $\mathbf{D}^{-1}$ and $\mathbf{U}^{-1}$ is equivalent to normalizing their rows making them sum to one.

$\mathbf{H}$ is symmetric, thus its eigenvectors are orthogonal. Moreover, it has the same eigenvalues of $\mathbf{N}_1$. Indeed, the eigenvalue problem for $\mathbf{H}$ can be expressed as

$$\mathbf{H}\vec{\psi}_i = \lambda_i \vec{\psi}_i, \tag{2.11}$$

which, using Eq. (2.10), becomes

$$\mathbf{D}^{\frac{1}{2}} \mathbf{N}_1 \mathbf{D}^{-\frac{1}{2}} \vec{\psi}_i = \lambda_i \vec{\psi}_i. \tag{2.12}$$

Thus

$$\mathbf{N}_1 \big(\mathbf{D}^{-\frac{1}{2}} \vec{\psi}_i\big) = \lambda_i \big(\mathbf{D}^{-\frac{1}{2}} \vec{\psi}_i\big) \qquad \rightarrow \qquad \mathbf{N}_1 \vec{\varphi}_i = \lambda_i \vec{\varphi}_i, \tag{2.13}$$

where we defined $\vec{\varphi}_i = \mathbf{D}^{-\frac{1}{2}} \vec{\psi}_i$. We can express $\vec{\varphi}_i^{(n)}$ as

$$\vec{\varphi}_i^{(n)} = \mathbf{N}_1 \vec{\varphi}_i^{(n-1)} = \cdots = \mathbf{N}_1^n \vec{\varphi}_i^{(0)}. \tag{2.14}$$

We can thus compute λ_i with the following formula.

$$\lambda_i = \lim_{n\to\infty} \frac{\left|\mathbf{N}_1 \vec{\varphi}_i^{(n)}\right|}{\left|\vec{\varphi}_i^{(n)}\right|}. \tag{2.15}$$

For the first eigenvector, since $\lambda_1 = 1$, we have

$$\vec{\varphi}_1^{(n)} = \mathbf{N}_1 \vec{\varphi}_1^{(n-1)}. \tag{2.16}$$

Exploiting the orthogonality of $\vec{\psi}_i$, we have

$$\delta_{ij} = \big(\vec{\psi}_i\big)^T \vec{\psi}_j = \big(\vec{\varphi}_i\big)^T \mathbf{D}^{\frac{1}{2}} \mathbf{D}^{\frac{1}{2}} \vec{\varphi}_j = \big(\vec{\varphi}_i\big)^T \mathbf{D} \vec{\varphi}_j, \tag{2.17}$$

implying

$$\big(\vec{\varphi}_i\big)^T \mathbf{D} \vec{\varphi}_j = \delta_{ij}. \tag{2.18}$$

Thus $\vec{\varphi}_i$ can also be orthogonalized if appropriately multiplied by $\mathbf{D}$. We can therefore use Gram–Schmidt process to compute the second eigenvector, that is

$$\vec{\varphi}_2^{(n)} = \vec{\varphi}_1^{(n)} - \left(\frac{\vec{\varphi}_1^{(n)} \cdot \vec{d}}{\big|\vec{d}\big|^2}\right) \vec{d}. \tag{2.19}$$

The eigenvector corresponding to $\lambda_1 = 1$ is simply a uniform vector with identical components. This is the reason why Hidalgo and Hausmann prescribe to stop their algorithm after a finite number of iterations, otherwise all countries would have the same asymptotic value (ECI). By ordering the eigenvalues λ_i in decreasing order $\lambda_1 > |\lambda_2| \geq \cdots \geq |\lambda_{N_c}|$ and expanding $\vec{F}^{(0)}$ as

$$\vec{F}^{(0)} = a_1 \vec{\varphi}_1 + a_2 \lambda_2 \vec{\varphi}_2 + \cdots + a_{N_c} \lambda_{N_c} \vec{\varphi}_{N_c}, \tag{2.20}$$

we can express the first equation of Eq. (2.9) as

$$\begin{aligned} \vec{F}^{(2n)} &= a_1 \vec{\varphi}_1 + a_2 \lambda_2^n \vec{\varphi}_2 + \cdots + a_{N_c} \lambda_{N_c}^n \vec{\varphi}_{N_c} \\ &= a_1 \vec{\varphi}_1 + a_2 \lambda_2^n \vec{\varphi}_2 + O((\lambda_3/\lambda_2)^n). \end{aligned} \tag{2.21}$$

Therefore, at sufficiently large n, the ranking of the countries is determined by the components of $\vec{\varphi}_2$. Thus, computing the components of the eigenvector associated to the second eigenvalue means determining the ranking of countries. With Eq. (2.19), we can do this without solving the full eigenvalue problem. We finally notice that the ranking is independent of the initial conditions $\vec{F}^{(0)}$.

2.1.2 Adjacency Matrix Representation

Up until now, we have represented the bipartite network of countries and products in terms of its matrix $\mathbf{M}$. We now recast this formalism in terms of the adjacency matrix $\mathbf{A}$ given by the $(N_c + N_p) \times (N_c + N_p)$ square matrix

$$\mathbf{A} = \begin{pmatrix} 0 & \mathbf{M} \\ \mathbf{M}^T & 0 \end{pmatrix}. \tag{2.22}$$

The diagonal blocks are blocks of zeros (the first of dimensions $N_c \times N_c$ and the second $N_p \times N_p$) since the graph is bipartite. There are therefore no links among countries, nor among products. We notice that $\mathbf{A}$ is a symmetric matrix. Let $\vec{k}$ be the $N_c + N_p$ dimensional vector whose first N_c components are the k_c associated to the countries (their diversification) and whose last N_p are the k_p associated to the products (their ubiquity). Thus, in terms of the adjacency matrix, we have

$$\begin{cases} F_i^{(n+1)} = \frac{1}{k_i} \sum_j A_{ij} Q_j^{(n)} \\ Q_i^{(n+1)} = \frac{1}{k_i} \sum_j A_{ij} F_j^{(n)}, \end{cases} \tag{2.23}$$

with $i, j = 1, \ldots, N_c + N_p$. Now, $\vec{F}^{(n)}$ and $\vec{Q}^{(n)}$ are $N_c + N_p$ dimensional vectors $\forall n$. The first N_c components of $\vec{F}^{(n)}$ represent the Fitness of countries[2], while the last N_p components of $\vec{Q}^{(n)}$ represent the Complexity of products. Although expressed in a different form, Eq. (2.23) is equivalent to Eq. (2.1). Indeed, let us consider for instance $F_c^{(n+1)}$, we have

$$\begin{aligned} F_c^{(n+1)} &= \frac{1}{k_c} \sum_j A_{cj} Q_j^{(n)} \\ &= \frac{1}{k_c} \left(\sum_{j=1}^{N_c} A_{cj} Q_j^{(n)} + \sum_{j=N_c+1}^{N_c+N_p} A_{cj} Q_j^{(n)} \right) \\ &= \frac{1}{k_c} \sum_{j=N_c+1}^{N_c+N_p} A_{cj} Q_j^{(n)} \\ &= \frac{1}{k_c} \sum_p M_{cp} Q_p^{(n)}, \end{aligned} \tag{2.24}$$

where the first sum in the second line is zero since $A_{cj} = 0 \ \forall j = 1, \ldots, N_c$. Moreover, the last two expressions are equivalent, they only differ for the notation ($j = N_c + 1, \ldots, N_c + N_p \to p = 1, \ldots, N_p$). We notice that by using the adjacency matrix representation, we can define two new quantities, the Complexity of countries (represented by the first N_c components of $\vec{Q}^{(n)}$) and the Fitness of products (represented by the last N_p components of $\vec{F}^{(n)}$). In the case of the ECI algorithm, since it is symmetric, given the same initial conditions, the Complexity of countries is equivalent to their Fitness, and analogously the Fitness of products is equivalent to their Complexity, as we can see expressing these new quantities in terms of the matrix $\mathbf{M}$

$$\begin{cases} F_p^{(n+1)} = \frac{1}{k_p} \sum_c M_{cp} Q_c^{(n)} \\ Q_c^{(n+1)} = \frac{1}{k_c} \sum_p M_{cp} F_p^{(n)}. \end{cases} \tag{2.25}$$

[2]By the term Fitness, we now refer to the quantity defined in the first equation of Eq. (2.23), even though we are dealing with the ECI algorithm. We made this choice because we will soon introduce the concept of Fitness of products and the Complexity of countries. Therefore, to continue calling the quantity F by the name Economic Complexity Index would have been misleading.

The latter is equivalent to Eq. (2.1) except for the notation ($F_c \to Q_c$ and $Q_p \to F_p$), while in the EFC algorithm this is no longer the case. We now want to find the matrices $\mathbf{D}_A^{-1}$ and $\mathbf{U}_A^{-1}$ that are analogous to $\mathbf{D}^{-1}$ and $\mathbf{U}^{-1}$. In analogy with what was done before, we can define the vectors $\vec{d}_A$ and $\vec{u}_A$ as

$$\vec{d}_A = \vec{u}_A = \mathbf{A}\vec{\mathbb{1}}_{N_c+N_p} \tag{2.26}$$

where $\vec{\mathbb{1}}_{N_c+N_p}$ is the $N_c + N_p$ dimensional vector of 1s. We thus have

$$\begin{aligned}\mathbf{U}_A^{-1} = \mathbf{D}_A^{-1} &\doteq \begin{pmatrix} \frac{1}{d_1} & & & & & \\ & \ddots & & & & \\ & & \frac{1}{d_{N_c}} & & & \\ & & & \frac{1}{d_{N_c+1}} & & \\ & & & & \ddots & \\ & & & & & \frac{1}{d_{N_c+N_p}} \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{k_1} & & & & & \\ & \ddots & & & & \\ & & \frac{1}{k_{N_c}} & & & \\ & & & \frac{1}{k_{N_c+1}} & & \\ & & & & \ddots & \\ & & & & & \frac{1}{k_{N_c+N_p}} \end{pmatrix}.\end{aligned} \tag{2.27}$$

That is, $\mathbf{D}_A^{-1}$ and $\mathbf{U}_A^{-1}$ are the diagonal matrices whose (diagonal) entries are the inverse of the elements of the vector $\vec{k}$ (that is the diversification of the countries and the ubiquity of the products). We notice that now $\mathbf{D}_A^{-1} = \mathbf{U}_A^{-1}$. Eq. (2.6) can thus be recast with the adjacency matrix representation as

$$\begin{cases} \vec{F}^{(n)} = \mathbf{D}_A^{-1}\mathbf{A}\mathbf{U}_A^{-1}\mathbf{A}^T\vec{F}^{(n-2)} = \mathbf{N}_A\vec{F}^{(n-2)} \\ \vec{Q}^{(n)} = \mathbf{U}_A^{-1}\mathbf{A}^T\mathbf{D}_A^{-1}\mathbf{A}\vec{Q}^{(n-2)} = \mathbf{N}_A\vec{Q}^{(n-2)}. \end{cases} \tag{2.28}$$

In the latter we have introduced the matrix $\mathbf{N}_A \doteq \mathbf{D}_A^{-1}\mathbf{A}\mathbf{U}_A^{-1}\mathbf{A}^T = \mathbf{U}_A^{-1}\mathbf{A}^T\mathbf{D}_A^{-1}\mathbf{A}$ which is analogous to the matrices $\mathbf{N}_1$ and $\mathbf{N}_2$ in the previous formalism. The fact that the evolution of $\vec{F}^{(n)}$ and of $\vec{Q}^{(n)}$ is characterized by the same matrix $\mathbf{N}_A$ is in accordance with what was said before about the Complexity of countries being equivalent to their Fitness, and analogously for the products. We now express $\mathbf{N}_A$ in terms of the matrix $\mathbf{M}$ in order to study its eigenvalues and eigenvectors. We have

$$\begin{aligned}\mathbf{N_A} &= \mathbf{D}_A^{-1}\mathbf{A}\mathbf{U}_A^{-1}\mathbf{A}^T \\ &= (\mathbf{D}_A^{-1}\mathbf{A})^2.\end{aligned} \tag{2.29}$$

The product $\mathbf{D}_A^{-1}\mathbf{A}$ can be expressed as

$$\begin{aligned}\mathbf{D}_A^{-1}\mathbf{A} &= \begin{pmatrix} \mathbf{D}^{-1} & 0 \\ 0 & \mathbf{U}^{-1} \end{pmatrix}\begin{pmatrix} 0 & \mathbf{M} \\ \mathbf{M}^T & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & \mathbf{D}^{-1}\mathbf{M} \\ \mathbf{U}^{-1}\mathbf{M}^T & 0 \end{pmatrix}.\end{aligned} \tag{2.30}$$

Thus

$$
\begin{aligned}
\mathbf{N_A} &= \begin{pmatrix} 0 & \mathbf{D}^{-1}\mathbf{M} \\ \mathbf{U}^{-1}\mathbf{M}^T & 0 \end{pmatrix} \begin{pmatrix} 0 & \mathbf{D}^{-1}\mathbf{M} \\ \mathbf{U}^{-1}\mathbf{M}^T & 0 \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{D}^{-1}\mathbf{M}\mathbf{U}^{-1}\mathbf{M}^T & 0 \\ 0 & \mathbf{U}^{-1}\mathbf{M}^T\mathbf{D}^{-1}\mathbf{M} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{N}_1 & 0 \\ 0 & \mathbf{N}_2 \end{pmatrix}.
\end{aligned} \tag{2.31}
$$

In terms of the initial conditions, we have

$$
\begin{cases} \vec{F}^{(2n)} = (\mathbf{N}_A)^n \vec{F}^{(0)} \\ \vec{Q}^{(2n)} = (\mathbf{N}_A)^n \vec{Q}^{(0)}. \end{cases} \tag{2.32}
$$

In particular, $\mathbf{N}_1$ and $\mathbf{N}_2$ are the products of the same matrices but in different order. Thus all the N_c eigenvalues of $\mathbf{N}_1$ are also eigenvalues of $\mathbf{N}_2$ (but since it is $N_p \times N_p$ dimensional it also has other eigenvalues). Moreover, since the spectrum of a block diagonal matrix is the union of the spectra of the blocks, $\mathbf{N}_A$ has as upper eigenvalue $\lambda = 1$ with degeneracy 2. Therefore, as seen before, at sufficiently large n the eigenvector $\vec{\varphi}_2$ of $\mathbf{N}_1$ associated with its second eigenvalue will completely determine the ranking of countries. It can be expressed in terms of eigenvectors of $\mathbf{N}_A$ as

$$
v_{N_1} = \begin{pmatrix} (\varphi_2)_{c=1} \\ \vdots \\ (\varphi_2)_{c=N_c} \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{2.33}
$$

where the last N_p entries are zeroes. Analogously, let $\vec{\phi}_2$ be the eigenvector associated with the second eigenvalue of $\mathbf{N}_2$. At sufficiently large n, $\vec{\phi}_2$ will completely determine the ranking of products. It can be expressed in terms of eigenvectors of $\mathbf{N}_A$ as

$$
v_{N_2} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ (\phi_2)_{p=1} \\ \vdots \\ (\phi_2)_{p=N_p} \end{pmatrix}, \tag{2.34}
$$

where the first N_c entries are zeroes.

2.1.3 Non-Bipartite Perturbations

We now want to study what happens if random perturbations are added to the system, i.e., if we consider random links between countries and between products, resulting in the adjacency matrix $\mathbf{A}$ having non-zero elements in the diagonal blocks. That is, we are now considering a graph that is no longer bipartite. We begin by considering an adjacency matrix of the form

$$
\mathbf{A} = \mathbf{A}_0 + \varepsilon \mathbf{A}_1, \tag{2.35}
$$

where $\mathbf{A}_0$ is the adjacency matrix of the bipartite case, ε is a small parameter (whose values we will vary in the next steps) representing the strength of the perturbation, and $\mathbf{A}_1$ is a random matrix of the following form

$$\mathbf{A}_1 = \begin{pmatrix} \mathbf{A}_{1a} & 0 \\ 0 & \mathbf{A}_{1b} \end{pmatrix}. \tag{2.36}$$

In particular, $\mathbf{A}_{1a}$ and $\mathbf{A}_{1b}$ are binary random matrices with the constraint of having the diagonal elements set to zero to avoid self loops ($\mathbf{A}_{1a}$ is a $N_c \times N_c$ dimensional matrix while $\mathbf{A}_{1b}$ is a $N_p \times N_p$ dimensional matrix). We sample from two different probability distributions. In the first one the probability that the entries of $\mathbf{A}_{1a}$ and $\mathbf{A}_{1b}$ are 1 is 50%, and in the second it is 1% (henceforth, we will refer to these probability distributions as the first and the second, respectively).

We now study how varying ε the difference between the first two eigenvalues of $\mathbf{N}_A$ changes. Indeed, the new links added to the system act as a perturbation that removes the degeneracy of its eigenvalues. Since $\mathbf{N}_A$ is still a stochastic matrix, the largest eigenvalue will be 1, independently of the perturbation. On the other hand, the second largest eigenvalue, which we call λ, will no longer be equal to 1 once we add a perturbation. It means that if we apply the ECI algorithm, as we explained before, it converges to the second eigenvector of $\mathbf{N}_A$. However, it is no longer the eigenvector that provides the ranking, since it corresponds to the almost constant eigenvector with almost identical components. It means that the ECI algorithm, used in its iterative formulation, will lead to improper results. In Fig. 2.1, for the first and the second probability distribution, we compare the original (unperturbed) ranking of countries with each perturbed one (that is, one for each value of ε). We make the comparison by counting the minimum number of swaps necessary to make the two rankings equal[3]. For the first probability distribution, even for small values of the

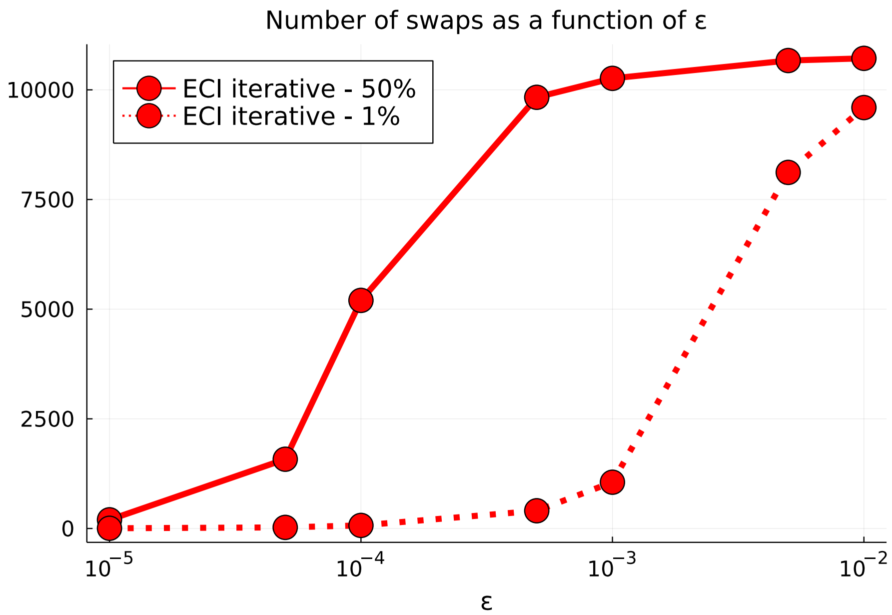

Figure 2.1: Number of swaps in the ranking of countries between unperturbed and perturbed case for ECI iterative algorithm for the two probability distributions.

[3]For this and all the following plots, error bars will not be present. We have run the code several times, obtaining approximately the same results (i.e., variation in the number of swaps on the order of units). However, for values of the parameter of almost one, the computation time significantly increases, not allowing us to repeat the computation a large number of times.

parameter as $\varepsilon = 5 \cdot 10^{-5}$, there are more than one thousand swaps, meaning that the iterative application of the ECI algorithm is strongly affected by the perturbation, giving a very different ranking as we expected. For the second probability distribution, we reach more than one thousand swaps for $\varepsilon = 10^{-3}$. We will see soon that we can obtain better results looking at the eigenvector components, despite performing this kind of analysis is more tricky in the perturbed case. In Fig. 2.2, it is shown the behavior of the quantity $\Delta = 1 - \lambda$ as a function of ε for the two probability distributions. We thus see that with the increase of the strength of the perturbation, the difference

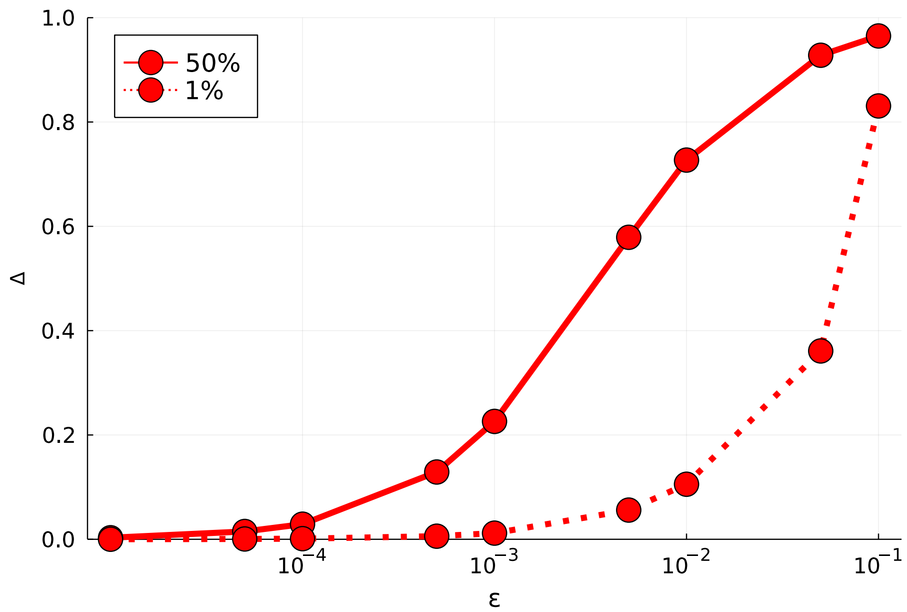

Figure 2.2: Δ dependence on ε for the two probability distributions.

between the first two eigenvalues significantly increases, growing from the magnitude order of one-thousandth (for the first probability distribution, while for the second one it is even smaller) for $\varepsilon = 10^{-5}$ to almost one for $\varepsilon = 10^{-1}$. We notice that for each value of the parameter ε, both the difference between the first two eigenvalues and the number of swaps in the ECI perturbed ranking are smaller for the second probability distribution with respect to the first one. Moreover, analogous results will hold for the next computations. The reason is that when we sample from the second probability distribution, we add fewer links to the adjacency matrix, which is thus more similar to the unperturbed one.

We now focus on the second couple of eigenvalues of $\mathbf{N}_A$, that is, the third and the fourth ones, which we denote λ_3 and λ_4. In the bipartite case, the eigenvectors associated with these two eigenvalues are v_{N_1} and v_{N_2}, that we defined in Eq. (2.33) and Eq. (2.34). They determine the ranking of countries and the ranking of products, respectively. Fig. 2.3 shows the behavior of λ_3 and λ_4 as a function of ε for the two probability distributions. We notice again that the perturbation removes the degeneracy for both the eigenvalues. A significant consequence of the perturbation is that the eigenvectors determining the ranking of countries and products are now mixed. Therefore, it is not trivial to understand which of the two eigenvectors determines the ranking of countries and which determines the ranking of products. Thus, we analyze both, focusing on the ranking of countries. In Fig. 2.4 and Fig. 2.5, for the first and the second probability distribution, respectively, we compare the original (unperturbed) ranking with each perturbed one. The behavior is similar for the two probability distributions. The main difference is that, for the first one, the number of

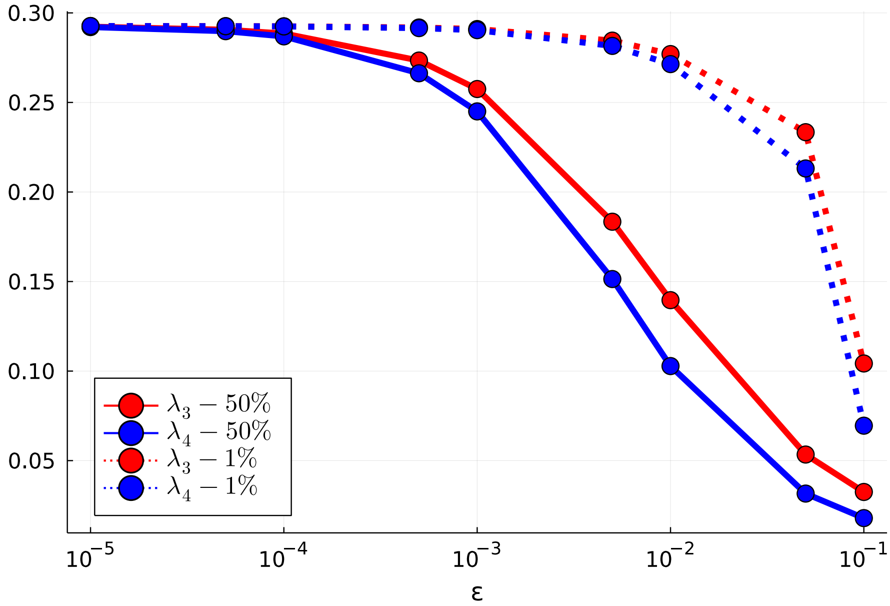

Figure 2.3: λ_3 and λ_4 dependence on ε for the two probability distributions.

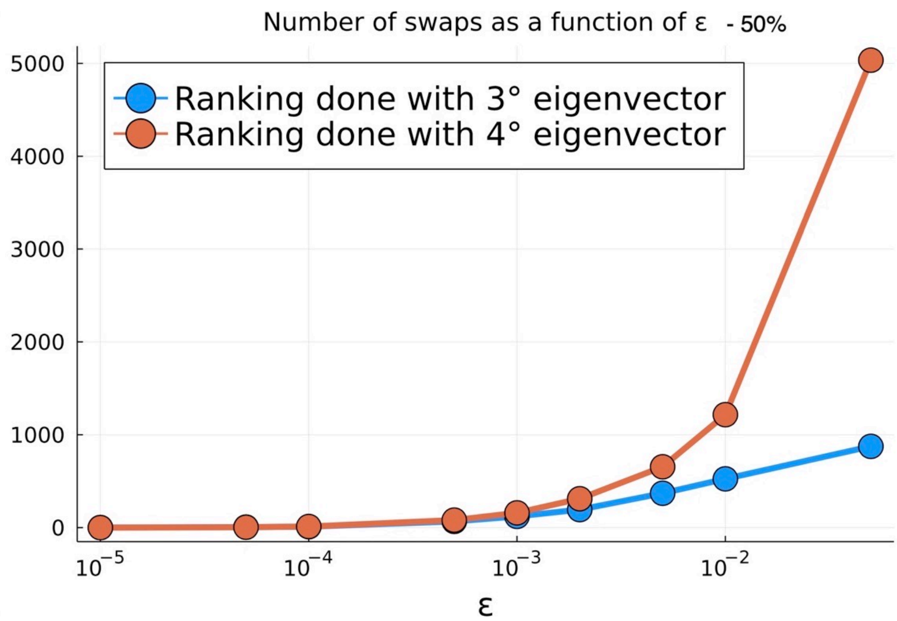

Figure 2.4: Number of swaps in the ranking of countries between the unperturbed case and the ranking done with third and fourth eigenvectors as a function of ε for the probability distribution of 50%.

swaps between the unperturbed ranking and the one done with the fourth eigenvector explodes before ($\varepsilon = 0.05$) compared to the second one ($\varepsilon = 0.5$). The ranking done with the components of the third eigenvector is more similar to the original one. We thus conclude that this eigenvector determines the ranking of countries.

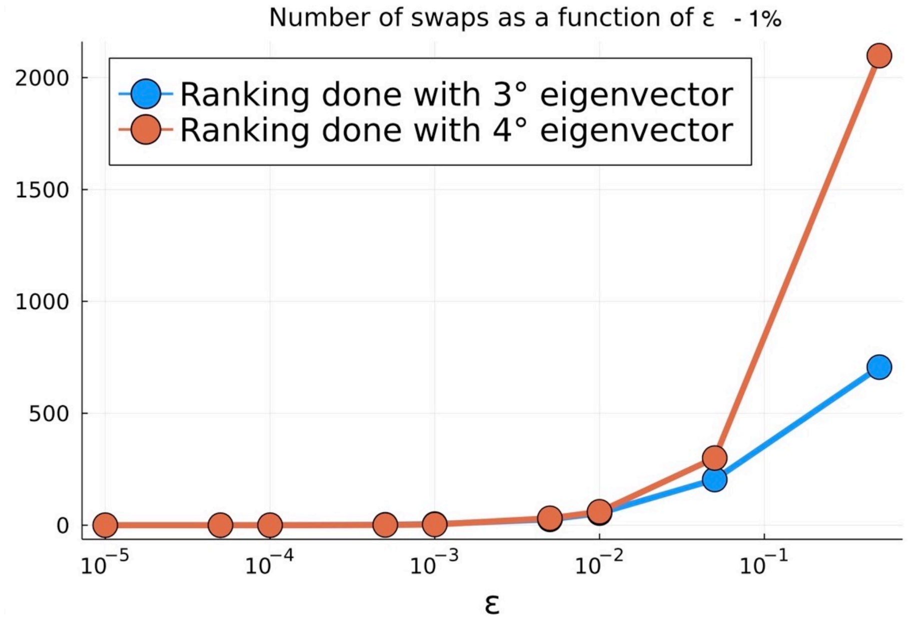

Figure 2.5: Number of swaps in the ranking of countries between the unperturbed case and the ranking done with third and fourth eigenvectors as a function of ε for the probability distribution of 1%.

2.2 NH-EFC

In this section, we will conduct a similar study on NH-EFC and compare the results with those obtained with ECI. We can rewrite the NH-EFC algorithm in the adjacency matrix formalism as

$$\begin{cases} \tilde{F}_i^{(n)} = \delta^2 + \sum_j A_{ij}/\tilde{P}_j^{(n-1)} \\ \tilde{P}_i^{(n)} = 1 + \sum_j A_{ji}/\tilde{F}_j^{(n-1)}. \end{cases} \tag{2.1}$$

Since this algorithm is not linear, we can no longer study Fitness and Complexity through the spectral properties of a matrix. Thus, we have to apply the algorithm in its iterative form.

Again, we compare the ranking without perturbation with those obtained for different values of the parameter ε by counting the number of swaps. Fig. 2.6 shows the results for the first probability distribution of both NH-EFC and ECI (computed with the third eigenvector)[4]. The picture inside Fig. 2.6 is a photo with uncorrelated white noise. Despite the presence of the noise, we are still able to see and recognize two people and several colored balloons. Analogously, spurious links in the country-product network do not affect the NH-EFC algorithm. Although the perturbation, it manages to reproduce the unperturbed ranking almost perfectly, with swaps on the order of units up to $\varepsilon = 0.05$. It then reaches values of a few hundred for $\varepsilon = 0.5$. On the other hand, for ECI, the number of swaps is hundreds from the value $\varepsilon = 10^{-3}$. In Fig. 2.7, the results for the second probability distribution are shown. In this case, the number of swaps for NH-EFC is almost constant (between 15 and 20 swaps for all the values of the parameter). Regarding ECI, for small values of ε there are fewer swaps than NH-EFC, while for $\varepsilon = 5 \cdot 10^{-3}$ they are more, and finally it explodes for $\varepsilon \approx 0.1$. We conclude that the NH-EFC algorithm is much more stable than the ECI algorithm with respect to non-bipartite perturbations.

[4]The number of swaps of ECI is computed with respect to the unperturbed ECI ranking and the number of swaps of NH-EFC is computed with respect to the unperturbed NH-EFC ranking.

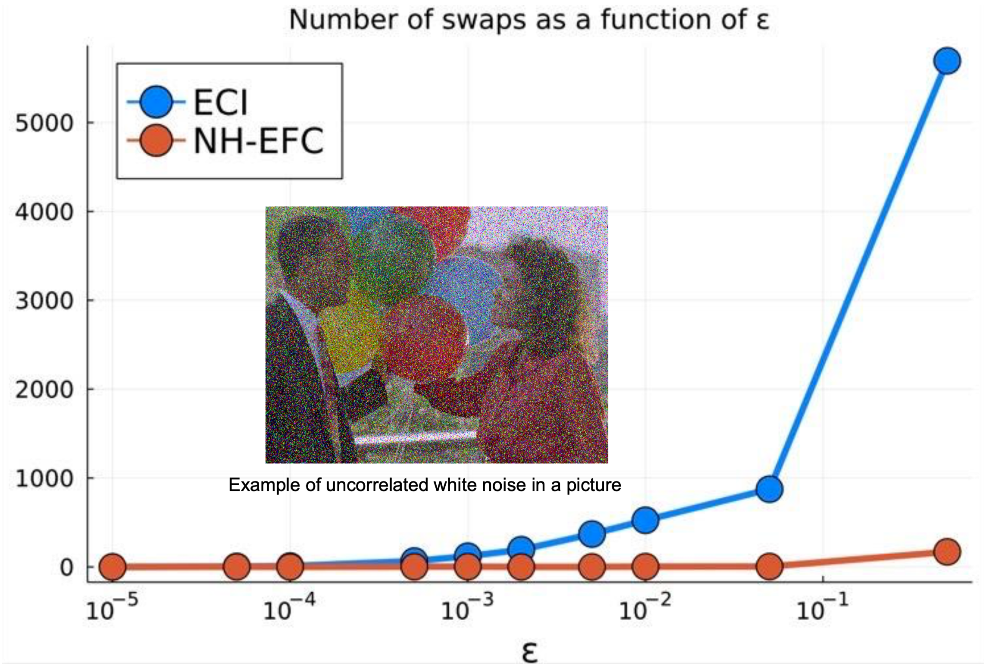

Figure 2.6: Number of swaps in the ranking of countries between unperturbed and perturbed case for both ECI (third eigenvector) and NH-EFC with probability distribution of 50%.

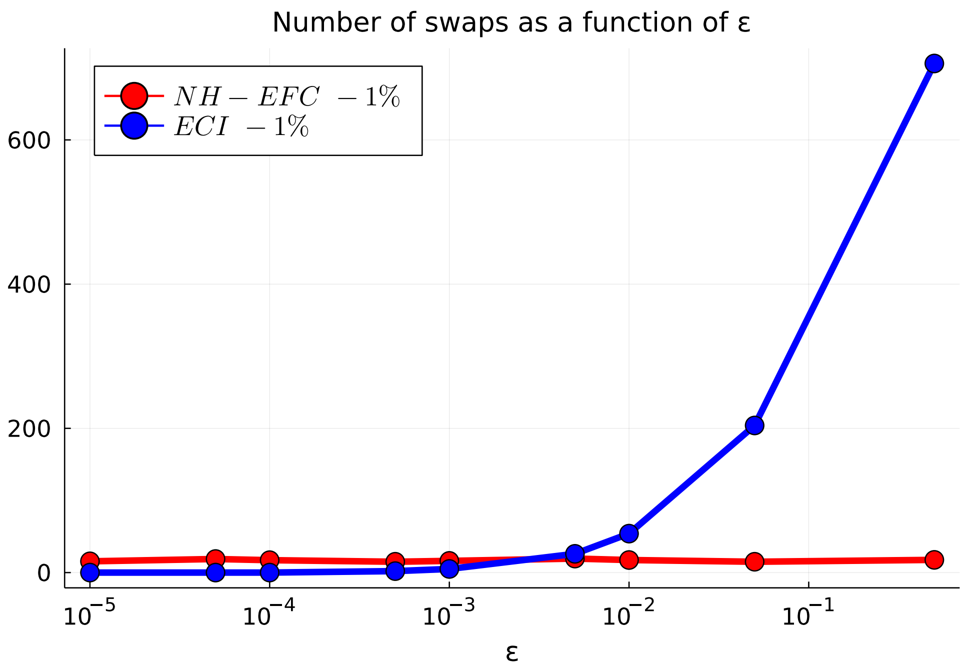

Figure 2.7: Number of swaps in the ranking of countries between unperturbed and perturbed case for both ECI (third eigenvector) and NH-EFC with probability distribution of 1%.

3. Prey-Predator Network

In this chapter, we will use the results of Section 2.2 to study an example of the prey-predator network. Indeed, the formalism introduced in the previous chapter allows us to deal also with directed graphs. The prey-predator network is a relevant example of a directed non-bipartite network since in the food chain, organisms that are neither top predators nor primary producers both eat and are eaten. In the specific example we will consider, the nodes are organisms while the edges represent directed carbon exchange in Florida Bay[1]. Therefore there is an edge from i to j if organism i eats species j[2]. In the network, 128 nodes are present, but we removed the last six since they do not represent organisms (they indicate Water POC, Benthic POC, DOC, Input, Output, and Respiration). Our network thus consists of 122 nodes and 1767 edges, as shown in Fig. 3.1. The nodes are colored using the HITS algorithm [10]. This algorithm assigns two scores for each node: a hub value and an authority value. The more nodes it points to, the higher the hub value, while the more nodes point towards it, the higher the authority value. Specifically, lighter nodes have a low hub value, while darker nodes have a high hub value.

In [11], the EFC algorithm is applied to ecological networks. Specifically, they consider the bipartite networks such as plant-pollinator, seed-disperser, and anemone-fish networks. In each of these, the first element represents a passive agent and corresponds to the products of the original Economic Complexity algorithm. On the other hand, the second element is active and corresponds to the countries. Moreover, rather than Fitness and Complexity, they use the terms *importance* and *vulnerability*. The importance of active species is determined by the number of its mutualistic passive partners, each weighted with its vulnerability. That is, the more partners and the more vulnerable they are, the more important an active element is. On the other hand, the vulnerability of a passive element is bounded by the less important species it interacts with, analogously with the country-product network. However, in our case, it is no longer possible to distinguish between passive and active agents since most elements are both.

We have applied the NH-EFC and computed the Fitness and the Complexity of each organism, as shown in Fig. 3.2. We highlighted four organisms, i.e., phytoplankton, spotted seatrout, green turtle, and barracuda. The first has both low Fitness and low Complexity, the second has intermediate Fitness and intermediate Complexity, the third has high Fitness and low Complexity, and the fourth has high Fitness and high Complexity. In particular, predators with high Fitness eat a large variety of prey, from very simple to very complex. Predators with low Fitness eat only the organisms eaten by most predators. On the other hand, if a predator with low Fitness eats one prey, then the Complexity of the prey is low.

We now try to interpret these results in terms of the hidden capabilities of the system. Since Fitness is determined by the number of organisms eaten (each weighted with its Complexity), it means that if a predator has high Fitness, then it has more capabilities than others (i.e., more clever, more versatile, or faster), as in the case of the barracuda. Moreover, because the organisms

[1] Available at `https://snap.stanford.edu/data/Florida-bay.html`

[2] In the original edge list, the nodes are switched, meaning that organism j eats species i.

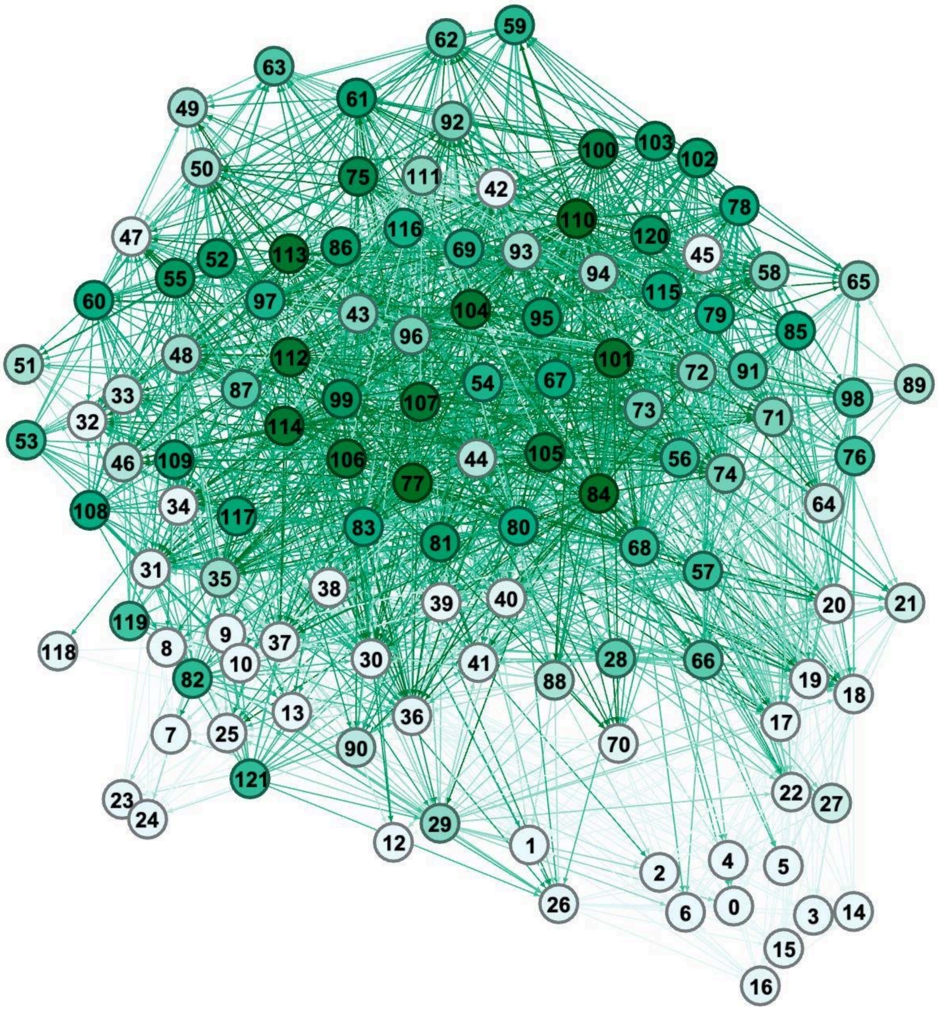

Figure 3.1: Florida Bay prey-predator network. Nodes are colored according to their hub value computed with the HITS algorithm. The higher it is, the darker they are.

with high Complexity weigh more in determining the Fitness of a predator, it is thus more difficult to eat them, like the green turtle.

Moreover, we notice that in the Fitness-Complexity plane, we can identify two regions. The first one is close to the diagonal line going from low Fitness and low Complexity to high Fitness and high Complexity. The second one is the one in the top left with low Fitness and high Complexity. Therefore, Fitness and Complexity are correlated. In particular, we see that no organisms with low Complexity have high Fitness.

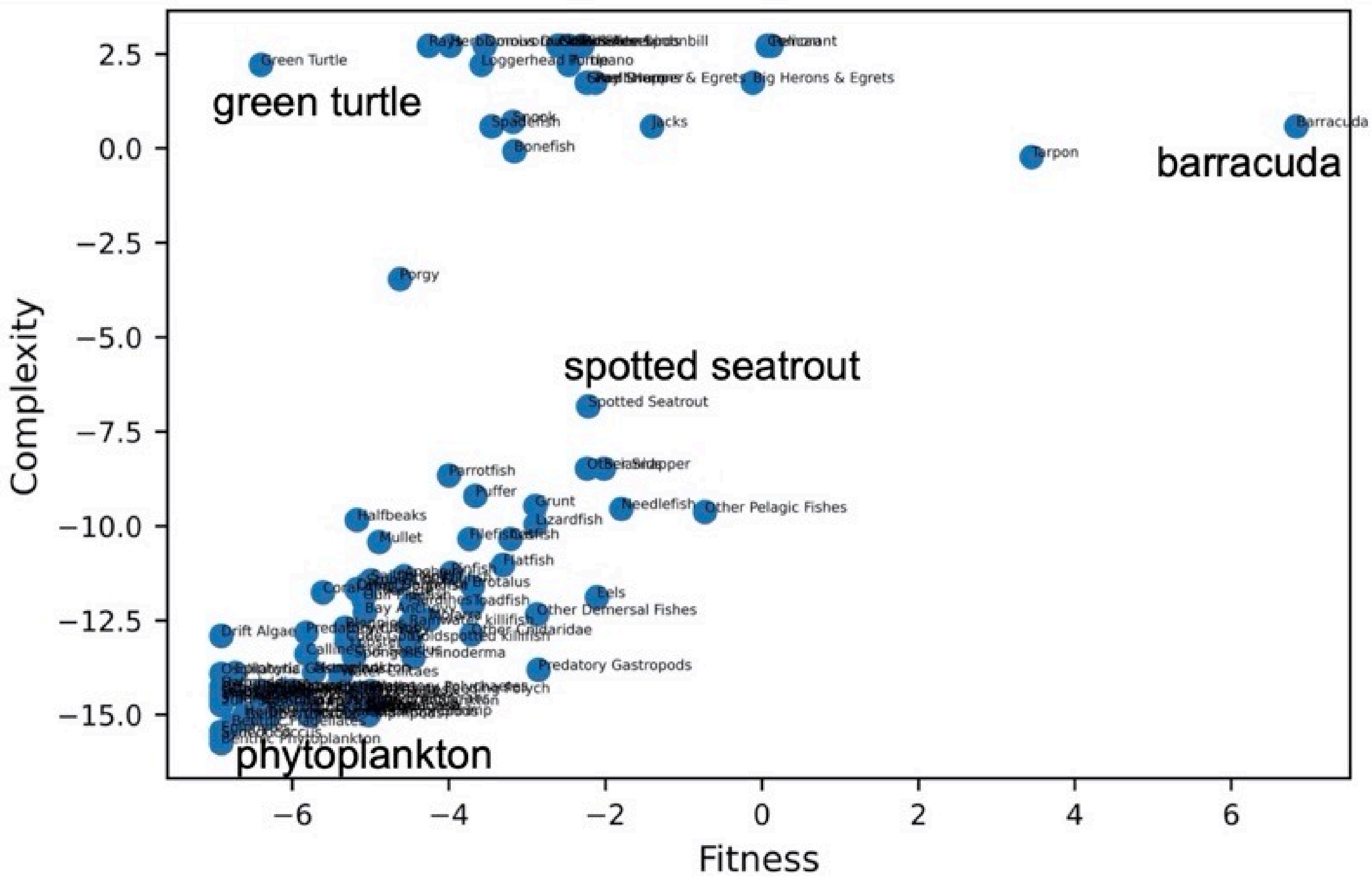

Figure 3.2: Fitness VS Complexity for each organism of the Florida Bay prey-predator network.

Finally, we said that in our case, it is no longer possible to separate passive and active agents since most elements are both. However, we can still look for the best division of the nodes into two classes: prey and predator. One way to proceed is by using Newman's modularity [12]. Modularity is a measure of the structure of networks used in community detection. Indeed, maximizing the modularity corresponds to favoring connections between nodes of the same class (or module) and disfavoring links between nodes of different modules. On the contrary, minimizing the modularity corresponds to maximizing the bipartivity of the network [13]. That is, favoring links between nodes of different classes and disfavoring connections between nodes of the same module. We then obtain that the prey class consists of the first fifteen elements (the ones that do not eat other organisms, corresponding to plankton species and seagrass) plus meroplankton and raptors. Another way to do so is by using the HITS algorithm we used to color the nodes. We (arbitrarily) put in the prey class the elements with the two lowest hub values, i.e., the first fifteen organisms plus four zooplankton species. We notice that we obtain different results with the two methods (and by changing the arbitrary choice of the second one) because since the network is not bipartite, the division is not unique. That is, different algorithms bring different results.

Conclusions

In this thesis, we have expressed both the ECI and the (NH-)EFC algorithm in terms of the adjacency matrix of the network, which in the case of a bipartite network is a block matrix whose diagonal blocks are zeroes. In this way, we managed to study the non-bipartite case simply by considering a network adjacency matrix whose diagonal blocks were nonzero. We have thus applied the ECI method to a network with random links between nodes of the same class. In the bipartite case, the eigenvectors associated with the second eigenvalues of the matrices $\mathbf{N}_1$ and $\mathbf{N}_2$ determine the countries' ranking and products' ranking, respectively. In the perturbed case, the eigenvector associated with the second eigenvalue of $\mathbf{N}_A$ has nearly identical components, so it contains no ranking information. Indeed, the eigenvectors determining the ranking of countries and products are now mixed. For the countries' one, we saw that the third gives better results, meaning that for high values of the parameter defining the strength of the links, it has fewer swaps (while the fourth explodes) compared to the original one, deducing that the third eigenvector determines the ranking of countries. Moreover, by using eigenvectors components, we get better results than by directly applying the iterative procedure. Because the new eigenvalues are different, thus the number of steps required for the iterative method is also different. Then, we compared the stability of ECI with the ranking done with the third eigenvector and NH-EFC and showed that the NH-EFC is more stable after introducing small non-bipartite random perturbations. Remarkably for NH-EFC, the number of swaps is very few, even when the parameter reaches values of almost one. Finally, we applied the NH-EFC algorithm to study the prey-predator ecosystem in Florida Bay, interpreting our results in terms of the hidden capabilities of the system. In particular, if an organism has high Fitness, it means that it has more capabilities than others (i.e., more clever, more versatile, or faster), as in the case of the barracuda. Moreover, if an organism has high Complexity, it is thus harder to eat it, as the green turtle. Furthermore, we saw that in the Fitness-Complexity plane, we can identify two regions and that no organisms with low Complexity have high Fitness. Lastly, we used two approaches to divide the nodes into the two classes of prey and predator. The first was minimizing Newman's modularity, while the second was using the HITS algorithm and arbitrarily classifying the prey as the elements with the two lowest hub values. We obtained similar results, i.e., in both cases, the prey class was composed of the first fifteen organisms that do not eat other organisms and some other species of plankton. However, the two results were not equal since, with the first method, also raptors appear to be in the prey class. Because the network is not bipartite, the division is not unique, and different algorithms thus produce different results.

The interesting methodology that we developed can be further improved. For instance, we could consider different weights for different types of swaps. In the sense that, e.g., one country changing its ranking position from the last to the first position should be considered differently than an equal number of swaps between neighbors (i.e., each country changes ranking position by a few places). Moreover, we could compare Fitness and Complexity in the context of ecological networks with existing ecological indices to study if there are correlations among them and to better validate and interpret our results. A significant additional development would be using Fitness and Complexity to detect endangered species. Lastly, we could apply the EC algorithms in the version we developed to any other non-bipartite network, which has never been done because it was not possible before.

A. Methods

A.1 Dataset

We used data extracted from the BACI dataset [14]. We focused on the trading data of the year 1996. We removed the countries exporting nothing and the products exported by no one, obtaining a total of 161 countries and 5036 products classified according to a four-digit code (six-digit code coarse-grained by considering only the first four digits) [15].

A.2 Country-Product Binary Matrix

Starting from the flows q_{cp} of US dollars representing the export in dollars of the product p by the country c, we can then construct the binary matrix $\mathbf{M}$ through the so-called Revealed Comparative Advantage (RCA) criterion. That is, the ratio between the export of the product p by country c and the global export of p done by all countries is divided by the ratio between the total export of c and the whole world export, i.e.

$$RCA_{cp} = \frac{\frac{q_{cp}}{\sum_{c'} q_{c'p}}}{\frac{\sum_{p'} q_{cp'}}{\sum_{c'p'} q_{c'p'}}}. \tag{A.1}$$

In order to build $\mathbf{M}$ from the RCA matrix, we consider $M_{cp} = 1$ if $RCA_{cp} \geq 1$ and zero otherwise.

Bibliography

[1] C. A. Hidalgo and R. Hausmann *The building blocks of economic complexity*, PNAS **106**, 10570–10575, (2009), available at `https://www.pnas.org/doi/full/10.1073/pnas.0900943106`

[2] M. Cristelli, A. Tacchella, M. Cader, K. Roster and L. Pietronero *On the Predictability of Growth*, Policy Research Working Paper 8117; The World Bank: Washington, DC, USA (2017), available at `https://openknowledge.worldbank.org/handle/10986/27620`

[3] V. D. P. Servedio, P. Buttà, D. Mazzilli A. Tacchella and L. Pietronero *A New and Stable Estimation Method of Country Economic Fitness and Product Complexity*, Entropy 20(10):783 (2018), available at `https://doi.org/10.3390/e20100783`

[4] M. Cristelli, A. Gabrielli A. Tacchella, G. Caldarelli, and L. Pietronero *Measuring the Intangibles: A Metrics for the Economic Complexity of Countries and Products*, PLoS ONE 8(8): e70726 (2013), available at `https://doi.org/10.1371/journal.pone.0070726`

[5] D. Ricardo *On the Principles of Political Economy and Taxation*, John Murray, (1817)

[6] A. Tacchella, M. Cristelli, G. Caldarelli, A. Gabrielli and L. Pietronero *A New Metrics for Countries' Fitness and Products' Complexity*, Sci. Rep. **2**, 723 (2012), available at `https://doi.org/10.1038/srep00723`

[7] A. Tacchella, D. Mazzilli and L. Pietronero *A dynamical systems approach to gross domestic product forecasting*, Nat. Phys. 14, 861 (2018), available at `https://www.nature.com/articles/s41567-018-0204-y`

[8] E. Pugliese, A. Zaccaria and L. Pietronero *On the convergence of the Fitness-Complexity algorithm*, Eur. Phys. J. Spec. Top. 225, 1893–1911 (2016), available at `https://link.springer.com/article/10.1140/epjst/e2015-50118-1`

[9] G. Caldarelli, M. Cristelli, A. Gabrielli, L. Pietronero, A. Scala and A. Tacchella *A Network Analysis of Countries' Export Flows: Firm Grounds for the Building Blocks of the Economy*, PLoS ONE 7(10): e47278 (2012), available at `https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0047278`

[10] J. M. Kleinberg *Authoritative Sources in a Hyperlinked Environment*, Journal of the ACM 46 (5), 604–632, (1999).

[11] V. Domínguez-García and M. A. Munoz *Ranking species in mutualistic networks*, Scientific reports 5(1), 1-7, (2015), available at `https://www.nature.com/articles/srep08182`

[12] M. E. J. Newman *Modularity and community structure in networks*, PNAS **103**, 8577-8582, (2006), available at https://www.pnas.org/doi/10.1073/pnas.0601602103

[13] M. E. J. Newman *Finding community structure in networks using the eigenvectors of matrices*, Physical review E 74(3), 036104, (2006), available at https://journals.aps.org/pre/abstract/10.1103/PhysRevE.74.036104

[14] G. Gaulier and S. Zignago, *BACI: International Trade Database at the Product-level.*, Centre d'Etudes Prospectives et d'Informations Internationales, (2010), available at http://www.cepii.fr/anglaisgraph/workpap/pdf/2010/wp2010-23.pdf

[15] https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Harmonized_commodity_description_and_coding_system

Acknowledgements

Sono numerose le persone che devo ringraziare, comincerò dalla fine.

Vito e Giordano, grazie per il vostro aiuto, la vostra disponibilità e la vostra pazienza, senza di voi questa tesi non sarebbe mai nata.

Ringrazio il Complexity Science Hub Vienna per l'ospitalità durante il tirocinio della tesi. Al CSH ho avuto la possibilità di interagire con persone fantastiche, sia professionalmente che socialmente, e dimostrare le mie abilità di bartender e giocatore di calcio balilla.

Andando un po' più indietro, grazie a tutti i miei compagni PCS, originali e acquisiti. Il semestre a Parigi è stato davvero divertente (credo vi siate accorti di qualche mia risata).

Dulcis in fundo, ringrazio con tutto il cuore la mia famiglia, per essere sempre presenti e per sostenermi in ogni momento della mia vita. Un grazie speciale a mamma, e a nonna, anche se non ci sei più, vi voglio bene.