doi:10.20944/preprints201807.0511.v1

Article

A new and stable algorithm for economic complexity

Vito D. P. Servedio [1]*, Paolo Buttà [2], Dario Mazzilli [3], Andrea Tacchella [3], Luciano Pietronero [3]

[1] Complexity Science Hub Vienna, Josefstätter-Strasse 39, A-1080 Vienna, Austria
[2] Sapienza University of Rome, Department of Mathematics, P.le Aldo Moro 5, 00185 Roma, Italy
[3] Sapienza University of Rome, Physics Dept., P.le Aldo Moro 5, 00185 Roma, Italy
* Correspondence: v.servedio@gmail.com

Abstract: We present a non-linear non-homogeneous fitness-complexity algorithm where the presence of non homogeneous terms guarantees both convergence and stability. After a suitable rescaling of the relevant quantities, the non homogeneous terms are eventually set to zero so that this new method is parameter free. This new algorithm reproduces the findings of the original algorithm proposed by Tacchella et al. [1], and allows for an approximate analytic solution in case of actual binarized RCA matrices. This solution discloses a deep connection with the network theory of bipartite graphs. We define the new quantity of "country net-efficiency" quantifying how a country efficiently invests in capabilities able to generate innovative high quality products. Eventually, we demonstrate analytically the local convergence of the algorithm.

Keywords: Economic Complexity; Non-linear map; Bipartite networks; Network theory

1. Introduction

In the last decade a new approach to macroeconomics has been developed to better understand the growth of countries [2]. The key idea is to consider the international trade of countries as a proxy of their internal production system. By describing the international trade as a bipartite network, where countries and products are sites of the two layers, new metrics for the economy of countries and the quality of products can be constructed with a simple algorithm [1] by leveraging the network structures only. This algorithm evaluates the fitness of countries, the quality of their industrial system and the complexity of commodities, by indirectly inferring the technological requirements needed to produce them. These two new metrics have been successfully used to describe past events and to forecast the economic development of countries and commodities production [3,4].

The very same approach has been applied to different social and ecological systems presenting a bipartite network structure and a competition between the components of the system [5,6]. Thus, it is natural to interpret fitness and complexity as properties of the network underlying those systems. The revised version of the fitness-complexity algorithm that we show here, results in a clear and natural interpretation in terms of network properties and helps to better understand the different component that contribute to the fitness.

In the following, we first describe the original algorithm and its properties underlining some critical issues that we solve with the revised version. Then, we define the new algorithm step by step and study its advantages in the case of countries-products networks. Finally, we propose an approximated solution and discuss its interpretation.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 July 2018 doi:10.20944/preprints201807.0511.v1

2. Algorithm definition

2.1. The original algorithm

Object of this work is the network of countries and their exported goods. This network is of bipartite type (countries and products are mutually linked, but no link exists between countries as well as between products) and weighted (links carry a weight s_{cp}, i.e., the exported volume of product p of country c, measured in US$). Data ranging from year 1995 to year 2015 can be freely retrieved from the Web [7], though we use it after some procedure to enhance its quality [4]. Eventually, we come up with data about 161 countries and more than 4000 products, which were categorized according to the Harmonized System 2007 coding system, at 6 digits level of coarse-graining. The weighted bipartite network of countries and products can be projected onto an unweighted network described solely by the M_{cp} matrix with elements set to unity when a given country c meaningfully exports a good p and zero otherwise (See Methods).

The original algorithm is defined by the following non-linear iterative map,

$$\begin{cases} F_c^{(n)} = \sum_{p'} M_{cp'} Q_{p'}^{(n-1)} & \text{with} \quad 1 \le c \le \mathcal{C} \\ Q_p^{(n)} = \left(\sum_{c'} M_{c'p} / F_{c'}^{(n-1)} \right)^{-1} & \text{with} \quad 1 \le p \le \mathcal{P}, \end{cases} \tag{1}$$

with initial values $F_c^{(0)} = Q_p^{(0)} = 1, \forall\, c, p$. In the previous expression F_c and Q_p stand for the fitness of a country c and quality of a product p; $\mathcal{C}$ and $\mathcal{P}$ are the total number of countries and exported products respectively and from the dataset we have that $\mathcal{C} \ll \mathcal{P}$.

By multiplying all F_c and Q_p by the same numerical factor k, the map remains unaltered, so that the fixed point of the map (as $n \to \infty$) is defined up to a normalization constant. In the original algorithm this constant is chosen at each iteration n such that

$$\sum_c F_c^{(n)} = \mathcal{C} \quad \text{and} \quad \sum_p Q_p^{(n)} = \mathcal{P}. \tag{2}$$

The algorithm of Eqs. (1) and (2) successfully ranks the countries of our world according to their potential technological development and, when applied to different yearly time intervals can be used to suggest precise strategies to improve country economies. It has also been proved to give the correct ranking of importance of species in a complex ecological system [5]. Despite its success, some points can still be improved:

i. **Convergence issues:** As stated in a recent paper [8]: "If the belly of the matrix [M_{cp}] is outward, all the fitnesses and complexities converge to numbers greater than zero. If the belly is inward, some of the fitnesses will converge to zero." This means that all the products exported by the countries with zero fitness get zero quality. This is mathematically acceptable but heavily underestimates the quality of such products: even natural resources need the right know-how to be extracted so that their quality would be better represented by a positive quantity. To cure this issue one has to introduce the notion of "rank convergence" rather than absolute convergence, i.e., the fixed point is considered achieved when the ranking of countries stays unaltered step by step.

ii. **Zero exports:** The countries that do not export any good do have zero fitness independently from their finite capabilities.

iii. **Specialized world:** In an hypothetical world where each country would export only one product, different from all other products exported by other countries, the algorithm would assign a unity fitness and quality to all countries and products. Though mathematically acceptable, this solution does not take into account the intrinsic complexity of products.

iv. **Equation symmetry:** This is rather an aesthetic point, in that Eq. (1) are not cast in a symmetric form.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 July 2018 doi:10.20944/preprints201807.0511.v1

2.2. The new algorithm

First, we reshape Eq. (1) in a symmetric form by introducing the variable $P_p = Q_p^{-1}$, i.e.,

$$\begin{cases} F_c^{(n)} = \sum_{p'} M_{cp'} / P_{p'}^{(n-1)} & \text{with} \quad 1 \le c \le \mathcal{C} \\ P_p^{(n)} = \sum_{c'} M_{c'p} / F_{c'}^{(n-1)} & \text{with} \quad 1 \le p \le \mathcal{P}. \end{cases} \tag{3}$$

Now the quality of products are given by the quantities P_p^{-1} and the algorithm is trivially equivalent to the original one provided one uses the normalization conditions $\sum_c F_c^{(n)} = \mathcal{C}$ and $\sum_p (P_p^{(n)})^{-1} = \mathcal{P}$.

Next, we introduce two set of quantities $\delta_c > 0$ and $\delta_p > 0$ and consider the inhomogeneous non-linear map defined as

$$\begin{cases} F_c^{(n)} = \delta_c + \sum_{p'} M_{cp'} / P_{p'}^{(n-1)} & \text{with} \quad 1 \le c \le \mathcal{C} \\ P_p^{(n)} = \delta_p + \sum_{c'} M_{c'p} / F_{c'}^{(n-1)} & \text{with} \quad 1 \le p \le \mathcal{P}. \end{cases} \tag{4}$$

Since the map is no more defined up to a multiplicative constant, the normalization condition is not required anymore, while the initial condition can be set as in the original algorithm $F_c^{(0)} = P_p^{(0)} = 1, \forall\, c, p$. The fixed point of the transformation is now trivially characterized by the conditions

$$F_c \ge \delta_c, \quad P_p \ge \delta_p, \quad F_c P_p > M_{cp}. \tag{5}$$

The parameters δ_c and δ_p can be interpreted as follows. The parameter δ_c represents the intrinsic fitness of a country. In fact, for a country k that does not export any good we have $M_{kp} = 0\; \forall p$ so that its fitness is simply equal to δ_k. Irrespective of its exports any country has a set of capabilities that characterize it.

The parameter δ_p is more intriguing. If no country exports it (probably because no country produces it), the product q has not been invented yet and its quality lies at its maximum value δ_q^{-1} since $M_{cq} = 0\; \forall c$. Therefore, the inverse of δ_q may be interpreted as a sort of innovation threshold: the smaller the parameter is, the higher is the quality of the product in his outset and more sophisticated capabilities are necessary to produce it. On the other hand, products like natural resources may be associated with a larger value of the parameter since require less complex capabilities for their extraction.

In order to keep the algorithm simple and parameter free as the original one, we first set a common value $\delta_c = \delta_q = \delta$, then we study the dependence of the algorithm on δ and finally we set $\delta = 0$.

3. Results

3.1. Dependence on the non-homogeneous parameter

We consider $\delta_c = \delta_q = \delta$ and address the dependence of the fixed point upon δ. To outline the dependence of F_c and P_p from the parameter δ, we can use the relations defined in Eq. (4) and introduce the rescaled quantities $\tilde{P}_p = P_p/\delta$ and $\tilde{F}_c = F_c \delta$. After some trivial algebra we get from Eq. (4),

$$\begin{cases} \tilde{F}_c^{(n)} = \delta^2 + \sum_{p'} M_{cp'} / \tilde{P}_{p'}^{(n-1)} & \text{with} \quad 1 \le c \le \mathcal{C} \\ \tilde{P}_p^{(n)} = 1 + \sum_{c'} M_{c'p} / \tilde{F}_{c'}^{(n-1)} & \text{with} \quad 1 \le p \le \mathcal{P}, \end{cases} \tag{6}$$

from which we deduce that, as soon as the parameter δ^2 is much smaller than the typical value of M_{cp} matrix elements, i.e., much smaller than unity, the fixed point in terms of $\tilde{F}_c$ and $\tilde{P}_p$ almost does not depend on δ (see Fig. 1). It is worth noting that the values of fitness F_c and quality $Q_p = P_p^{-1}$ of the original map defined by Eqs. (1) and (2) cannot be obtained from this new algorithm when the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 July 2018 doi:10.20944/preprints201807.0511.v1

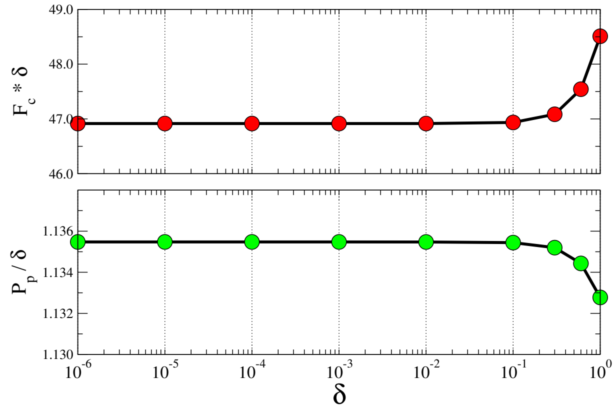

Figure 1. Dependence on the non-homogeneous parameter: Dependence of fitness and quality at the fixed point on the parameter δ. One country (Afghanistan) and one product (live horses) were chosen arbitrarily from the sample of year 2014.

parameter δ tends to zero. In terms of $\tilde{F}_c$ and $\tilde{P}_p$ the fitness and quality obtained from the original algorithm can be expressed as $F_c = \tilde{F}_c\,\delta^{-1}$ and $Q_p = \tilde{P}_p^{-1}\,\delta^{-1}$. Since the new algorithm provides finite non vanishing values of $\tilde{F}_c$ and $\tilde{P}_p$, by taking the limit $\delta \to 0$ would deliver infinite values of F_c and Q_p. We might think that the normalization procedure necessary in the old algorithm in order to fix the arbitrary constant would get rid of the common factor δ^{-1} and deliver the same values of the new method. Unfortunately, this is not the case since the new method does not rely on a normalization procedure. Therefore, since a self-consistent procedure of normalization, i.e., a projection on the double simplex defined by Eq. (2), is missing in the new algorithm, the results cannot coincide. Since the quantities $\tilde{F}_c$ and $\tilde{P}_p$ are well defined in the limit $\delta \to 0$, we shall focus on them only, in the following. We remind that the complexities of products delivered by the original method are connected to the set of P_p^{-1} and thus to the $\tilde{P}_p^{-1}$. In particular, the second of Eq. (6) can be interpreted at the fixed point as $\tilde{P}_p = 1 + \tilde{Q}_p^{-1}$ with the $\tilde{Q}_p$ expressed as in the second of Eq. (1), but with the tilde quantities calculated with the new algorithm. Therefore, we shall assign to $\tilde{Q}_p = (\tilde{P}_p - 1)^{-1}$ the meaning of complexity of products in our new algorithm. The differences between the old and new algorithm are depicted in Fig. 2, while the evolution of the fitnesses in time is shown in Fig. 4.

3.2. Analytic approximate solution

Despite their symmetric shape, Eq. (4) are not symmetric at all since in case of actual countries and products, the matrix M_{cp} is rectangular with the number of its rows $\mathcal{C}$ being much less than the number of its columns $\mathcal{P}$. To estimate the effect of this asymmetry, we first consider Eq. (4) in a mean field fashion, where each element of M_{cp} is set to the average value $\langle M\rangle = \sum_{c,p} M_{cp}/\mathcal{C}\mathcal{P}$, and write, at the fixed point,

$$\begin{cases} \tilde{f} = \delta^2 + \mathcal{P}\langle M\rangle\,\tilde{p}^{-1} \\ \tilde{p} = 1 + \mathcal{C}\langle M\rangle\,\tilde{f}^{-1}, \end{cases} \tag{7}$$

with now all $\tilde{F}_c$ and $\tilde{P}_p$ set to be equal to their mean field value $\tilde{f}$ and $\tilde{p}$ respectively. By setting $\delta = 0$, we find $\tilde{p} = 1/(1 - \frac{\mathcal{C}}{\mathcal{P}}) \approx 1 + \frac{\mathcal{C}}{\mathcal{P}}$ and $\tilde{f} = \mathcal{P} - \mathcal{C}$.

Indeed, an approximate expression for the fixed point of Eq. (6) in the regime $\delta \ll 1$ and $\mathcal{C} \ll \mathcal{P}$ can be derived also beyond the mean field approximation. To this end, we set again $\delta = 0$ and consider the corresponding fixed point equation associated to Eq. (6), i.e.,

$$\begin{cases} \tilde{F}_c = \sum_{p'} M_{cp'}/\tilde{P}_{p'} & \text{with} \quad 1 \le c \le \mathcal{C} \\ \tilde{P}_p = 1 + \sum_{c'} M_{c'p}/\tilde{F}_{c'} & \text{with} \quad 1 \le p \le \mathcal{P}. \end{cases} \tag{8}$$

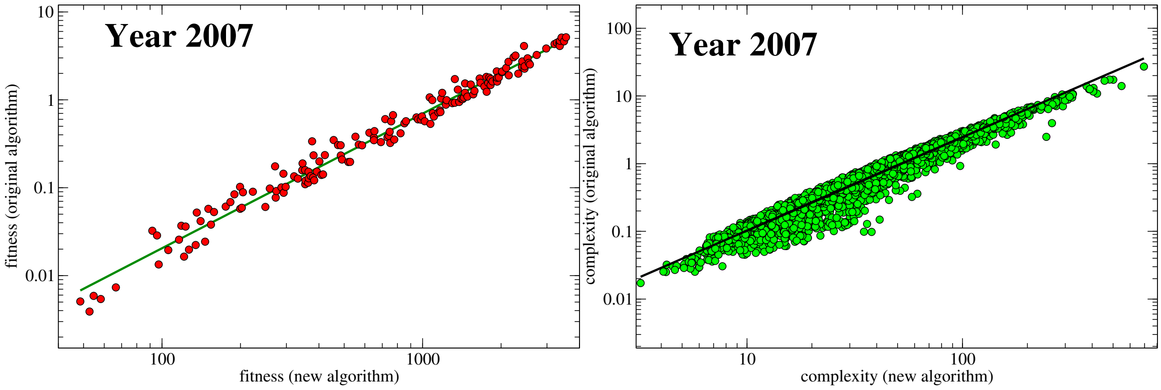

Figure 2. Comparison between the original and the revised method: Differences in country fitness (left panel) and product complexity (right panel) calculated with the original method of Ref. [1] (vertical axes) and new method (horizontal axes) as referred to year 2007. The green line in the left panel is the best least square approximation of power-law type (correlation coefficient 0.989) with exponent ca. 1.53. The dark line in the right panel is the best power-law approximation (correlation coefficient 0.971) resulting with an exponent of ca. 1.38.

We observe that the quantity $D_c = \sum_p M_{c,p}$, representing the diversification of country c, i.e., the number of different products exported, is of the order of $\mathcal{P}$ (at least for the majority of countries). Therefore, setting $\tilde{P}^* = \max_p \tilde{P}_p$ and $\tilde{F}_* = \min_c \tilde{F}_c$, Eq. (8) implies,

$$\begin{cases} \tilde{F}_c \geq D_c/\tilde{P}^* \approx \text{const}\, \mathcal{P}/\tilde{P}^* \quad \text{with} \quad 1 \leq c \leq \mathcal{C} \\ \tilde{P}^* \leq 1 + \mathcal{C}/\tilde{F}_*. \end{cases}$$

From the first estimate, $\tilde{F}_* \geq \text{const}\, \mathcal{P}/\tilde{P}^*$, and therefore, by the second estimate, $\tilde{P}^* \leq 1 + \text{const}\, \frac{\mathcal{C}}{\mathcal{P}} \tilde{P}^*$. As $P_p \geq 1$, we conclude that $\tilde{P}_p = 1 + W_p$ with W_p in the order of magnitude of $\mathcal{C}/\mathcal{P}$, and, as a consequence, $\tilde{F}_c$ is of the order of magnitude of $\mathcal{P}$.

We next compute explicitly the values of $\tilde{F}_c$ and $\tilde{P}_p$ at the first order in this approximation. The calculation of second order terms can be found in Appendix A. By using the first order approximation $(1+a)^{-1} \approx 1 - a$ twice, from Eq. (8) we have,

$$W_p \approx \sum_{c'} \frac{M_{c'p}}{D_{c'}} \left(1 + \frac{1}{D_{c'}} \sum_{p'} M_{c'p'} W_{p'} \right).$$

Let now $\mathbf{H}$ be the square matrix of elements $H_{pp'} = \sum_{c'} M^T_{pc'} D^{-2}_{c'} M_{c'p'}$. Letting D^{-1} be the column vector with components $1/D_c$, the last displayed formula reads,

$$(\mathbf{1} - \mathbf{H})W \approx \mathbf{M}^T D^{-1}.$$

We now observe that $H_{pp'} \leq \sum_{c'} 1/D^2_{c'} \leq \text{const}\, \mathcal{C}/\mathcal{P}^2$. Therefore, the matrix $(\mathbf{1} - \mathbf{H})$ is close to the identity (the correction is of order $\mathcal{C}/\mathcal{P}^2$) and hence invertible (with also the inverse close to the identity). In this approximation, $W = \mathbf{M}^T D^{-1}$, so that the rescaled (reciprocals of the) qualities of products are given by

$$\tilde{P} = 1 + \mathbf{M}^T D^{-1}. \tag{9}$$

In the same approximation, we obtain the rescaled fitnesses $\tilde{F}_c$; since

$$\tilde{F}_c = \sum_{p'} \frac{M_{cp'}}{1 + W_{p'}} \approx \sum_{p'} M_{cp'}(1 - W_{p'}),$$

 doi:10.20944/preprints201807.0511.v1

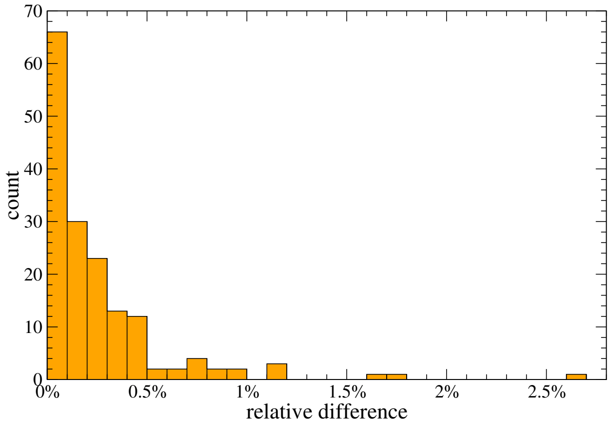

Figure 3. Numerical vs Analytic relative error: The histogram of the relative difference $(\tilde{F}_c^{(\text{fixed point})} - \tilde{F}_c^{(\text{approximated})})/\tilde{F}_c^{(\text{fixed point})}$ is plotted with the number of countries on the vertical axis. The approximated values are calculated using Eq. (10).

we have

$$\tilde{F} = D - \mathbf{K}D^{-1}, \tag{10}$$

having introduced the *co-production* matrix $\mathbf{K} = \mathbf{M}\mathbf{M}^T$ with elements $K_{cc'} = \sum_{p'} M_{cp'} M_{p'c'}$, representing the number of the same products exported by the two countries c and c'.

Interesting to note how, up to the first order approximation, the values of the fitness of countries are depending on the co-production matrix only. The goodness of the approximations above can be appreciated in Fig. 3 that shows how the relative difference between the numerical values at the fixed point and the approximate solution of Eq. (10) is below 0.5% for more than 85% of the countries.

3.3. Country inefficiency and net-efficiency

From Eq. (10) we deduce that the leading part of fitness $\tilde{F}_c$ is given by the diversification D_c. The diversification of a country is indeed an important quantity, for the calculation of which we do not need any complicated algorithm. On the other hand, what the non-linear algorithm proposed does, is to quantify how a country manages to successfully differentiate its products, and indirectly offers an estimate of the capabilities of a nation. In fact, a country exporting mainly raw materials would be less efficient with respect to a country exporting high technological goods, when they have the same diversification value. For this reason, we introduce the new quantity $I_c = D_c - \tilde{F}_c$, inefficiency of country c: the smaller the value I_c the more efficient is the diversification it chooses. From the approximate solution displayed in Eq. (10), we get that $I_c \approx \sum_{c'} K_{cc'}/D_{c'}$, so that the inefficiency of a country is a weighted average of its co-production matrix elements. The dependence of the country inefficiency on the diversification is displayed in Fig. 5, while a visual representation of it is displayed in Fig. 8. It is interesting to notice how a clear power-law dependence exists between the inefficiency and the diversification of a country. By indicating with $I_c = qD_c^m$ the least square best fit of yearly data we find that over the range 1995-2014, $m = 0.751 \pm 0.0029$ and $q = 0.318 \pm 0.015$.

The structure of the $\mathbf{M}$ matrix is such that those countries with high diversification also export low quality goods in average. Therefore to a large diversification would statistically correspond a large inefficiency, though the found power-law is not trivial and depends on the structure of the $\mathbf{M}$. A similar power-law behaviour is found between the fitness calculated with the traditional method and the diversification, but with a different exponent (from the left panel of Fig. 2 we deduce that there is a power-law relation between the fitnesses calculated with the original method and this new method, and the exponent is around 1.53; since the fitness F_c calculated with the new method goes as D_c at the first order, then the old fitnesses also go as $D_c^{1.53}$). In order to better appreciate the production strategies of countries, we subtracted the common power-law trend of the dependency

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 July 2018 doi:10.20944/preprints201807.0511.v1

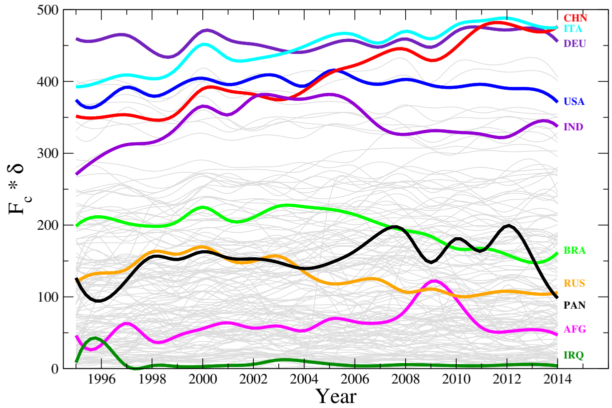

Figure 4. Country fitness evolution: Country fitness as calculated by the new algorithm. Curves were artificially smoothed by a cubic spline for a better visual representation.

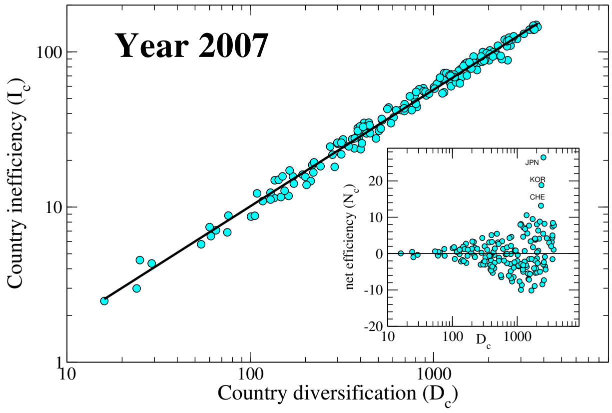

Figure 5. Role of diversification: The country inefficiency ($I_c = D_c - \tilde{F}_c$ vs the diversification D_c with the black line representing the power-law relation $I_c \approx D_c^{0.75}$ (linear regression with correlation coefficient 0.994). In the inset the net efficiency N_c, defined as the difference between the black line and the inefficiency of the main graph, is shown. Data pertain to year 2007.

of the inefficiency on the diversification for each year, changed its sign and plotted the result in Fig. 6, which thus shows the time evolution of a quantity that we call country *net-efficiency* N_c (*net* in the sense opposed to *gross*) over the years 1995-2014. It interesting to note how countries behave differently over the time lapse considered. Some countries display a decreasing net-efficiency, others an increasing or a constant one. What many of these curves have in common is the decreasing behaviour after year 2007, i.e., the year considered the beginning of the last large financial crisis.

3.4. Local convergence

From the simulations it is clear that the fixed point obtained by iterating Eq. (4) is locally stable. We can also prove it by resorting to the Jacobian of the transformation, in the case of countries and products. First we recall that the sum over the indexes c and p of Eq. (4) run from 1 to $\mathcal{C}$ and $\mathcal{P}$ respectively, with usually $\mathcal{C} \ll \mathcal{P}$. In the case of countries and products $\mathcal{C}/\mathcal{P} \approx 10^{-1}$. We also fix $\delta_c = \delta_q = \delta \ll 1$, so that the fitnesses and the (reciprocals of the) qualities at the fixed point are approximately given by $F_c = \tilde{F}_c/\delta$ and $P_p = \delta\tilde{P}_p$ with $\tilde{F}_c$ and $\tilde{P}_p$ the components of the vectors $\tilde{F}$ and $\tilde{P}$ given in Eq. (10) and Eq. (9) respectively.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 July 2018 doi:10.20944/preprints201807.0511.v1

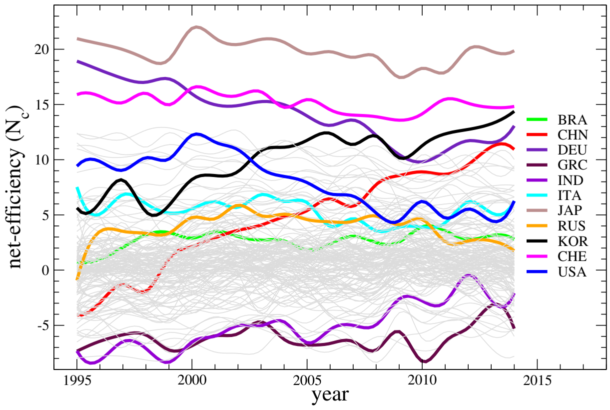

Figure 6. Yearly evolution of net efficiency: yearly time evolution of country net efficiency. The net efficiency is a detrended version of the inefficiency defined in the text and already displayed in the inset of Fig. 5 in the year 2007. Curves were artificially smoothed by a cubic spline for a better visual representation.

Next, we calculate the Jacobian of the transformation at the fixed point which can be simply expressed as the block anti-diagonal matrix

$$\mathbf{J} = \begin{pmatrix} \mathbf{0} & -\mathbf{M}^T\mathbf{F}^{-2} \\ -\mathbf{M}\mathbf{P}^{-2} & \mathbf{0} \end{pmatrix}, \tag{11}$$

155 having introduced the diagonal matrices $\mathbf{F} = \text{diag}(F_1, F_2, \dots, F_c)$ and $\mathbf{P} = \text{diag}(P_1, P_2, \dots, P_p)$
156 respectively.

We claim that the spectral radius $\rho(\mathbf{J})$ of the square matrix $\mathbf{J}$ is strictly smaller than one. Denoting by $\sigma(\mathbf{J})$ the spectrum of $\mathbf{J}$, this means that $\rho(\mathbf{J}) := \max\{|\lambda| \colon \lambda \in \sigma(\mathbf{J})\} < 1$. From this it follows [9] that the fixed point is asymptotically stable and the convergence exponentially fast. To prove the claim we consider the square of the Jacobian that can be written as a block diagonal matrix,

$$\mathbf{J}^2 = \begin{pmatrix} \mathbf{M}^T\mathbf{F}^{-2}\mathbf{M}\mathbf{P}^{-2} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}\mathbf{P}^{-2}\mathbf{M}^T\mathbf{F}^{-2} \end{pmatrix}, \tag{12}$$

and note that the traces of the two matrices on the diagonal is the same by applying a cyclic permutation. Noticing that $F_c P_p = \tilde{F}_c \tilde{P}_p$ and using the approximate solutions in Eq. (10) and Eq. (9), we find with simple algebra that

$$\text{Tr}(\mathbf{J}^2) = 2\sum_{c,p} \frac{M_{c,p}^2}{F_c^2 P_p^2} \approx 2\sum_{c,p} \frac{M_{c,p}^2}{D_c^2} = 2\sum_c \frac{1}{D_c} \approx \frac{\mathcal{C}}{\mathcal{P}} < 1. \tag{13}$$

Moreover, we can write the two non trivial matrices composing $\mathbf{J}^2$ as

$$\mathbf{M}^T\mathbf{F}^{-2}\mathbf{M}\mathbf{P}^{-2} = \mathbf{P}(\mathbf{P}^{-1}\mathbf{M}^T\mathbf{F}^{-1})(\mathbf{F}^{-1}\mathbf{M}\mathbf{P}^{-1})\mathbf{P}^{-1} = \mathbf{P}\mathbf{A}^T\mathbf{A}\mathbf{P}^{-1} \tag{14}$$

and

$$\mathbf{M}\mathbf{P}^{-2}\mathbf{M}^T\mathbf{F}^{-2} = \mathbf{F}(\mathbf{F}^{-1}\mathbf{M}\mathbf{P}^{-1})(\mathbf{P}^{-1}\mathbf{M}^T\mathbf{F}^{-1})\mathbf{F}^{-1} = \mathbf{F}\mathbf{A}\mathbf{A}^T\mathbf{F}^{-1}, \tag{15}$$

with $\mathbf{A} = \mathbf{F}^{-1}\mathbf{M}\mathbf{P}^{-1}$. The matrices $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$ are symmetric and positive-semidefinite so that their eigenvalues are real and non negative, and the matrices $\mathbf{F}\mathbf{A}\mathbf{A}^T\mathbf{F}^{-1}$ and $\mathbf{P}\mathbf{A}^T\mathbf{A}\mathbf{P}^{-1}$ have the same

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 July 2018 doi:10.20944/preprints201807.0511.v1

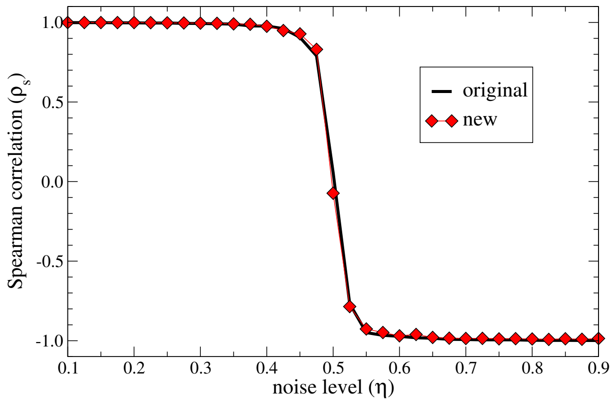

Figure 7. Noise robustness: Spearman correlation between the ranking of countries based on fitness at zero noise and at different noise levels η (see Sec. 3.5 in the main text). The performance of the two algorithms is practically indistinguishable. Note that at $\eta = 1$ all elements are flipped so that the perturbed system is perfectly anti-correlated with the original one.

eigenvalues. Therefore, the eigenvalues of $\mathbf{J}^2$ are real and non negative and we can write according to Eq. (13)

$$\mathrm{Tr}(\mathbf{J}^2) = \sum_i \lambda_i^2 < 1, \tag{16}$$

with λ_i eigenvalues of $\mathbf{J}$. Finally, from the preceding equation we have $\max \lambda_i^2 < \max |\lambda_i| < 1$ so that at the fixed point $\rho(\mathbf{J}) < 1$.

3.5. Robustness to noise

Fitness and complexity (and quality) values depend on the structure of the matrix M_{cp}. Noise can affect its elements by flipping their value. Thus, we test the robustness of the algorithm to noise as described in [10]. The idea is to introduce random noise by flipping each single bit of the matrix with probability η, which then is a parameter tuning the noise level. The rank of country fitnesses in presence of noise R_c^η is then compared with the rank obtained without noise R_c^0. The Spearman correlation ρ_s is then evaluated between these two sets and shown in Fig. 7 as a function of η for both the original and the new algorithm: the new algorithm shows a perfect stability to random noise as the original one with an unavoidable transition around $\eta \approx 0.5$, where noise is so strong to alter significantly the structure of the matrix M_{cp}.

4. Discussion

The proposed new inhomogeneous algorithm of economic complexity defined in Eq. (4) and in Eq. (6) carries many advantages with respect to the original one. The fitnesses and complexities coming out from these two methods are not identical, but highly correlated to each other, as witnessed by the plots in Fig. 2. This high correlation between the two methods ensures that all the studies carried on with the original method so far, can be obtained by applying this new method as well.

Besides the stability of the algorithm and its robustness, one advantage of this method is that the fitness is well defined also for those countries that have low exportation volumes and that in the original method had their fitness tending to zero. For those countries it is now possible to undertake a comparative study based on hypothetical investments (changing the elements of the $\mathbf{M}$ matrix) so to make predictions on their economic impact.

By first symmetrising the original equations, by adding an inhomogeneous parameter and by rescaling the quantities, one obtains Eq. (6), where the parameter can be safely set to zero. This ensures that this new algorithm is parameter free as the original one. As a side effect, the fixed point of the map can be well approximated analytically, with an error with respect to the iterative fixed point of

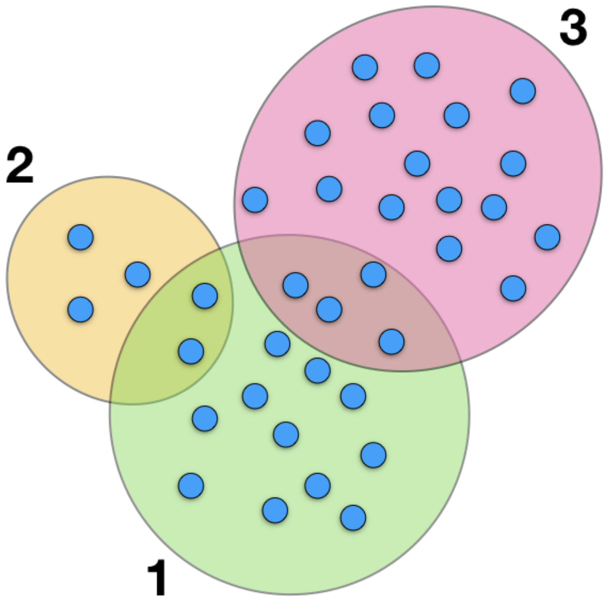

Figure 8. Inefficiency cartoon: Large ovals represent three countries, while small circles represent products. In this simple example, the inefficiency I_1 of country 1 is $I_1 = K_{12}/D_2 + K_{13}/D_3$. From the figure we get $K_{12} = 2$ and $K_{13} = 4$, i.e. the number of products exported by both countries (the cardinality of the intersection sets), and the diversifications $D_1 = 17$, $D_2 = 5$, $D_3 = 20$. Thus, $I_1 = 2/5 + 4/20 = 0.6$ and the approximated fitness $\tilde{F}_1 \approx 16.4$.

less than 3% (see Fig. 3). The result is represented by Eq. (9) and Eq. (10) at the first order (Eq. (19) and Eq. (20) at the second order), which allow for a simple intuitive explanation of the complexity of products and fitness of countries. Let us discuss Eq. (10) first. The result suggests that the fitness of a country is trivially related, at the first order, to its diversification: the more products a country exports, the larger is its fitness, i.e., the more developed its capabilities. This simple explicit dependence of the fitness on the diversification is also an advantage with respect to the original method, where the dependence was not explicitly clear. The second term of Eq. (10), which we call *inefficiency*, is also very interesting. If a country is the only one to export a given product, the contribution of this product to its fitness is a full one, or in other words, the contribution to the inefficiency is zero. This situation mimics a condition of monopoly on that product and it is logical that the exporting country has the full benefit of it. When a product is exported by multiple nations then it is critical to assess whether those countries export few or many other products (see Fig. 8). If a product is exported by a country c' with low diversification (low capabilities), then that product is not supposed to be of high complexity. The result is that the ratio $K_{cc'}/D_{c'}$ can be close to one ($c = 1, c' = 2$ in the figure) and the inefficiency associated to the common products is high, resulting in a small contribution to the fitness of c. The inefficiency can be interpreted in terms of the bipartite network of countries and products: the $K_{cc'}$ counts the number of links that connect countries c and c' to the same products, while the differentiation D_c is the node degree of country c. In other words, for a country c the inefficiency counts the links to common products of all other countries and weights them according to the degree of those. To our knowledge, this kind of measure has never been considered in complex networks so far. Since, statistically, countries with an high diversification also export many less complex products, the inefficiency is an increasing function of the diversification (Fig.5, main graph). If we subtract the general trend, which stems from the structure of the matrix M_{cp}, we can appreciate the net effect of choosing the goods to export. We call this new de-trended quantity *net-efficiency*. In this way we somehow remove the negative effect of less valuable products and highlight the contribution of more sophisticated goods. In the inset of Fig. 5 we show the net-efficiency as a function of diversification

 doi:10.20944/preprints201807.0511.v1

and underline the three nations (Japan, Korea and Switzerland) that stand out among the others. The complexity of products is estimated by Eq. (9) as the reciprocal of the second term of the sum. Since the diversification of a country D_c is a direct measure of its capabilities, we expect to find a simple relation between it and the complexities of products $\tilde{Q}_p$. Indeed, if we indicate with c_i those countries exporting the product p, for which obviously we have $M_{c_i p} = 1$, and with $m = \sum_c M_{cp}$, we can write

$$\tilde{Q}_p \approx \left(\frac{1}{D_{c_1}} + \frac{1}{D_{c_2}} + \ldots + \frac{1}{D_{c_m}} \right)^{-1}$$

from which we corroborate the main idea that the complexities of products are driven by the countries with low diversification (capabilities) that export it. Just for amusement, we observe how the complexity of products can be considered as the equivalent resistor of a parallel of resistors each one with resistance D_c. Somehow, a high D_c represents an effective resistance to the creation of a product and its export, so that if a country exists with a low diversification exporting it, the effort (resistance) of producing that product is also low.

5. Materials and Methods

*5.1. Construction of the **M** matrix*

Given the export volumes s_{cp} of a country c in a product p one can evaluate the Revealed Comparative Advantage (RCA) indicator [11] defined as the ratio

$$\mathrm{RCA}_{cp} = \frac{s_{cp}}{\sum_{c'} s_{c'p}} \Bigg/ \frac{\sum_{p'} s_{cp'}}{\sum_{c'p'} s_{c'p'}} \tag{17}$$

in this way one can filter out size effects. As described in the Supplementary information of [4], from the time series of the RCA we can evaluate the productive competitiveness of each country in each product by assigning to it a productivity state from 1 to 4. State 1 means that the country does not produce (or is very uncompetitive in producing) a product, state 4 means that it is one of the main producer in the world. We can then project this states onto the binarized matrix M_{cp} by simply setting its elements to unity whenever a state larger than 2 is encountered, and set them to null otherwise.

Supplementary Materials: None.

Acknowledgments: We acknowledge A. Zaccaria and E. Pugliese for interesting discussions. We acknowledge the EU FP7 Grant 611272 project GROWTHCOM and CNR PNR Project “CRISIS Lab” and the Austrian Research Promotion Agency FFG under grant #857136, for financial support. FFG is also covering the costs to publish in open access.

Author Contributions: Formal analysis, Paolo Butta; Investigation, Vito D. P. Servedio, Dario Mazzilli and Andrea Tacchella; Methodology, Vito D. P. Servedio and Luciano Pietronero; Project administration, Luciano Pietronero; Supervision, Luciano Pietronero; Writing – original draft, Vito D. P. Servedio, Paolo Butta, Dario Mazzilli, Andrea Tacchella and Luciano Pietronero.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Second order expansion of fitness and qualities

In this section we compute explicitly the values of $\tilde{F}_c$ and $\tilde{P}_p$ for $\mathcal{C} \ll \mathcal{P}$ up to the second order of magnitude of $\mathcal{C}/\mathcal{P}$. Letting $\varepsilon = \mathcal{C}/\mathcal{P}$, we expand $W_p = \varepsilon W_p^{(1)} + \varepsilon^2 W_p^{(2)} + O(\varepsilon^3)$. By assuming D_c of the order of $\mathcal{P}$ and by using the second order approximation $(1+a)^{-1} \approx 1 - a + a^2$ twice, Eq. (8) implies that

$$\tilde{F}_c = D_c \left(1 - \varepsilon \sum_{p'} \frac{M_{cp'}}{D_c} W_{p'}^{(1)} - \varepsilon^2 \sum_{p'} \frac{M_{cp'}}{D_c} [W_{p'}^{(2)} - (W_{p'}^{(1)})^2] + O(\varepsilon^3) \right), \qquad \text{with } 1 \le c \le \mathcal{C}, \tag{18}$$

and

$$\varepsilon W_p^{(1)} + \varepsilon^2 W_p^{(2)} = \sum_{c'} \frac{M_{c'p}}{D_{c'}} + \varepsilon \sum_{c',p'} \frac{M_{c'p}}{D_{c'}} \frac{M_{c'p'}}{D_{c'}} W_{p'}^{(1)} + \varepsilon^2 \sum_{c',p'} \frac{M_{c'p}}{D_{c'}} \frac{M_{c'p'}}{D_{c'}} [W_{p'}^{(2)} - (W_{p'}^{(1)})^2]$$
$$+ \varepsilon^2 \sum_{c',p',p''} \frac{M_{c'p}}{D_{c'}} \frac{M_{c'p'}}{D_{c'}} \frac{M_{c'p''}}{D_{c'}} W_{p'}^{(1)} W_{p'}^{(1)} + O(\varepsilon^3), \qquad \text{with} \;\; 1 \le p \le \mathcal{P}.$$

By the assumption on the magnitude of D_c, the first sum in the right-hand side is of the order of ε, the second one is of the order of ε^2, while the last two sums are of the order ε^3. Therefore,

$$\varepsilon W_p^{(1)} = \sum_{c'} \frac{M_{c'p}}{D_{c'}}, \qquad \varepsilon^2 W_p^{(2)} = \sum_{c',p'} \frac{M_{c'p}}{D_{c'}} \frac{M_{c'p'}}{D_{c'}} \varepsilon W_{p'}^{(1)}.$$

Recalling $\mathbf{H}$ denotes the square matrix of elements $H_{pp'} = \sum_{c'} M_{pc'}^T D_{c'}^{-2} M_{c'p'}$ (hence $H_{pp'} \approx \varepsilon/\mathcal{P}$) and D^{-1} the column vector with components $1/D_c$, we have just showed that $W = \mathbf{M}^T D^{-1} + \mathbf{H}\mathbf{M}^T D^{-1} + O(\varepsilon^3)$. Therefore, in the second order approximation, the rescaled (reciprocals of the) qualities of products are given by

$$\tilde{P} = 1 + \mathbf{M}^T D^{-1} + \mathbf{H}\mathbf{M}^T D^{-1}. \tag{19}$$

In the same approximation, from Eq. (18) we finally calculate the rescaled fitnesses $\tilde{F}_c$. Denoting by $(\mathbf{M}^T D^{-1})^2$ the column vector with components $(\mathbf{M}^T D^{-1})_p^2$ we get

$$\tilde{F} = D - \mathbf{K}D^{-1} + \mathbf{M}(\mathbf{M}^T D^{-1})^2 - \mathbf{M}\mathbf{H}\mathbf{M}^T D^{-1}, \tag{20}$$

where the *co-production* matrix $\mathbf{K} = \mathbf{M}\mathbf{M}^T$ has been introduced just below Eq. (10).

Bibliography

1. Tacchella, A.; Cristelli, M.; Caldarelli, G.; Gabrielli, A.; Pietronero, L. A New Metrics for Countries' Fitness and Products' Complexity. *Scientific Reports* **2012**, *2*, 723 EP –.
2. Hidalgo, C.A.; Hausmann, R. The building blocks of economic complexity. *Proceedings of the National Academy of Sciences* **2009**, *106*, 10570–10575.
3. Saracco, F.; Di Clemente, R.; Gabrielli, A.; Squartini, T. Detecting early signs of the 2007-2008 crisis in the world trade. *Scientific Reports* **2016**, *6*, 30286.
4. Tacchella, A.; Mazzilli, D.; Pietronero, L. A Dynamical Systems approach to GDP Forecasting. *To appear in Nature Physics* **2018**.
5. Domínguez-García, V.; Muñoz, M.A. Ranking species in mutualistic networks. *Scientific Reports* **2015**, *5*, 8182. Article.
6. Cimini, G.; Gabrielli, A.; Sylos Labini, F. The Scientific Competitiveness of Nations. *PLOS ONE* **2014**, *9*, 1–11.
7. http://comtrade.un.org/. Last accessed on Sep. 22nd 2016.
8. Pugliese, E.; Zaccaria, A.; Pietronero, L. On the convergence of the Fitness-Complexity algorithm. *The European Physical Journal Special Topics* **2016**, *225*, 1893–1911.
9. Hasselblatt, B.; Katok, A. *A first course in dynamics. With a panorama of recent developments.*; Cambridge University Press: New York, 2003. Proposition 2.2.19.
10. Battiston, F.; Cristelli, M.; Tacchella, A.; Pietronero, L. How metrics for economic complexity are affected by noise. *Complexity Economics* **2014**, *3*, 1–22.
11. Balassa, B. Trade Liberalisation and "Revealed" Comparative Advantage. *The Manchester School* **1965**, *33*, 99–123.