The complex dynamics of products and its asymptotic properties

Orazio Angelini[1]*, **Matthieu Cristelli**[1], **Andrea Zaccaria**[1], **Luciano Pietronero**[1,2]

1 ISC-CNR, Institute for Complex Systems, Rome, Italy, **2** Physics Department, Sapienza University of Rome, Rome, Italy

* angelini.orazio@gmail.com

Abstract

We analyse global export data within the Economic Complexity framework. We couple the new economic dimension Complexity, which captures how sophisticated products are, with an index called logPRODY, a measure of the income of the respective exporters. Products' aggregate motion is treated as a 2-dimensional dynamical system in the Complexity-logPRODY plane. We find that this motion can be explained by a quantitative model involving the competition on the markets, that can be mapped as a scalar field on the Complexity-logPRODY plane and acts in a way akin to a potential. This explains the movement of products towards areas of the plane in which the competition is higher. We analyse market composition in more detail, finding that for most products it tends, over time, to a characteristic configuration, which depends on the Complexity of the products. This market configuration, which we called asymptotic, is characterized by higher levels of competition.

Introduction

In this paper we report the results of an analysis that has been triggered by recent developments in the field of Economic Complexity. The proposal of the Fitness and Complexity measures [1], which capture respectively the advancement of the productive system of a country and the level of sophistication of a given kind of product found in the export market, allows for the extraction of useful information from countries' export data. This enabled a novel approach that compared monetary metrics, such as Gross Domestic Product per capita (GDPpc or Y from now on), with non-monetary metrics, such as Fitness and Complexity [2]. The interplay between these quantities reveals a wealth of information about economic phenomena. With this method, we explore the dynamics of the Complexity (Q from now on), comparing it to logPRODY, a monetary metric defined as the weighted average of the GDPpc's of a product's exporters [3]. This lets us gather insight into the dynamics of the global export markets for different products, and find a number of relevant regularities in their behavior.

The Complexity measure stems from a promising body of research. The hypothesis underlying the definition of this non-monetary metrics is that counties' competitiveness is connected with their *capabilities* [4–7], which are non-exportable and hard-to-measure features that allow the production of sophisticated goods [8]. For example, property rights, regulation,

---

educational system, infrastructure, the specific know-how of an industry, etc., could all be considered to be capabilities. In principle, though, there can be any number of capabilities with any relative importance, and it is impossible to "a priori", or normatively, determine them. This makes it practically impossible to measure them directly. A recent body of research focused on leveraging the economic output of a country as a proxy for its capabilities, by looking at which products it is capable of exporting [1, 4, 8–10]. The fact that a country is capable of exporting a given product p signals that its production system is competitive enough to stand out in the global market for p. This, in turn, implies that it has the necessary capabilities to produce p. In other words, products encode information on the (up to now) elusive capabilities. How to infer the presence of capabilities from export data? World trade flow data is representable as a bipartite network, where country c exporting product p is represented as a link between c and p. This allows a bottom-up approach enabling us to exploit the network structure in order to measure properties related to the capabilities from data regarding economic output. Conceptually this body of literature is doing something similar to PageRank [11, 12], since it leverages topological properties of the network in order to measure properties of the nodes. The first attempt is due to Hidalgo and Hausmann [4], followed by the aforementioned Fitness and Complexity measures. The adjacency matrix of the export bipartite network is nested [1]. This means that there are some countries that export almost all the products, and some products that are exported only by the few most diversified countries, that also happen to usually be the richest ones from a monetary point of view. On the other hand, the few products exported by poorly diversified countries are also exported by almost everyone else. A finding in itself, that drives the subsequent analysis, is that the evolution of an economy in time is characterized by increasing diversification, rather than specialization [13]. The situation is analogous to what is found in some biological systems [14], namely that the network is *nested*, and this offers a natural way to sort products in terms of increasing *Complexity* and countries in terms of *Fitness* [9] (indeed a similar algorithm has actually been applied to biological networks [15]). The definition and calculation techniques for Fitness and Complexity are discussed at length in the Methods section.

Comparing non-monetary metrics to monetary ones promises to be a particularly insightful approach. As an example, a fruitful way to look at the Fitness non-monetary metric is to compare it with the monetary Gross Domestic Product per capita. By representing countries with their position on the Fitness-GDPpc plane, it has been possible to apply techniques of dynamical systems in order to model their coupled motion [2]. In some cases, the motion is sufficiently regular to enable predictions on the future change in Fitness and GDPpc.

In light of this promising approach in literature, we set out to analyze Complexity, first by identifying a suitable monetary metric to compare it with, and then looking at its dynamics with a method similar to the one used for Fitness. We considered various possibilities for the monetary index, the most relevant being product prices, value added, and two metrics called PRODY [3] and Sophistication [16].

Prices are very impractical to use, since it is difficult to collect reliable data about them, and they fluctuate in response to a large number of variables that might have little to do with their Complexity. We tried extrapolating prices from the BACI dataset (see Methods for information on the datasets), which contains, for each product and couple of countries, the unit quantity exported and the price paid for it in dollars. Unfortunately the data is partly unreliable because of non-standardized customs' procedures, and also because a given class of products allows very different pricing in it. Additionally, given how heterogeneous products are, there is no universal definition for the quantity of goods being exported, making it difficult to compare across different kinds of goods. Because of this, the only relevant information that can be extracted for the data is the relative change in time of prices for highly aggregated categories.

This results in extremely noisy data, with which we are not able to reliably explore any hypothesis. Value added, being defined as the difference between production cost and selling price, retains all the problem prices have, adding to them the problem of reliably measuring the production cost. Since it is an extensively studied quantity, though, it is still possible to find reliable data, but it is very aggregated. The main datasets are the OECD Trade in Value Added [17] data and the World Input Output Database [18]. Both contain a handful of broad categories, with which we couldn't perform the kind of analysis presented in this paper.

Finally, PRODY and Sophistication appear to be good candidates for our study. The quantities used to calculate them are widely available and reliable. They have already been used to classify products and make predictions [3, 16, 19–21], and they depend on the GDPpc's of exporting countries, so they look like a natural counterpart to Complexity as GDPpc is to Fitness. We slightly modifed PRODY by defining *logPRODY*, which looks to us like a more natural measurement. What *logPRODY* represents is a weighted average GDPpc of the exporters of a certain product (the exact definition of PRODY and logPRODY is reported in the Methods section). The biggest the share of a certain product *p* in the total export of a country *c*, the more *c* is weighted in the average that defines $logPRODY_p$. logPRODY can be interpreted as a proxy for the composition of the market for a given product, signaling what is the mean income of countries export it. A change in logPRODY is linked to a change in the underlying composition in the set of countries that export a given product and, as we will see, these changes can be modeled in analogy with dynamical systems. This is essential, because high correlation is expected (and observed) between logPRODY and Complexity: high-*Q* products are exported by high-Fitness countries, which are usually the richest. By just using traditional parametric regression techniques, we would not be able to gather much more information than this simple fact. Instead, the technique we used allows us to check what happens, for example, to products that don' follow, or stop following, the trivial correlation trend.

What we actually observe in this analysis is the motion of the products on the ranked Complexity *r*(*Q*) vs. ranked logPRODY *r*(*logPRODY*) plane (called RCLP from now on). Note that the ranking procedure is simply an assignment of a 0 score to the product with the lowest yearly value and a 1 score to the highest, while the other products are equally spaced in this interval; see Methods on why the ranking is our choice in analysis. Analysing time displacements allows us to detect non-trivial interactions between the two quantities, that would be hard to detect with other methods. We also assign a value of competition on the export market, as measured by the Herfindahl index (explained further on), to each area of the plane, as it seems competition has some explanatory power for the motion we observed.

Results

Here we will illustrate the findings of our analysis. This section is divided in paragraphs illustrating respectively the motion we observe, a model we suggest to explain its cause, the concept of asymptotic market that we will introduce further on, and finally a minor discrepancy between the model and the observation, that is explained outside of the model. The analysis is conducted on two different datasets, which we call BACI and Feenstra respectively, and are detailed in the Methods section. We want to stress that in all cases the two datasets give results that are consistent with each other.

Kinematics. To gather information on the dynamics of the products' Complexity, we can look at the ranked Complexity (*r*(*Q*)) vs. ranked logPRODY (*r*(*logPRODY*)) plane. We find that, as expected, there is a strong relation between *r*(*Q*) and *r*(*logPRODY*): most of the points tend to gather on the diagonal of the plane (for the dataset provided by Feenstra et al. in [22], see Methods for details on the data used), with a dramatic change in the density of points. The
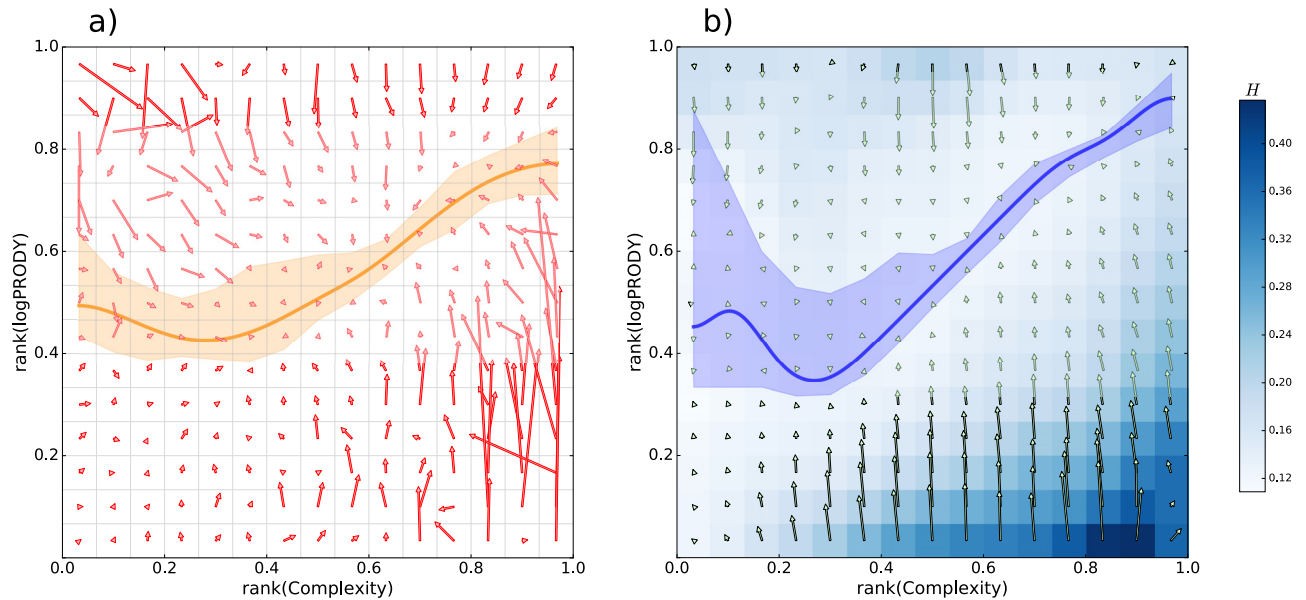
Fig 1. Panel a). In red, the average velocity field of the products on the RCLP, $\vec{v}$. Average velocities always point toward the asymptotic zone, marked by the orange line, which is calculated by finding the per-column minima of the velocity field after applying a kernel regression on the field. The result of a bootstrap, in orange, indicates the 5% confidence interval on the values. $\vec{v}$ indicates that points that stray away from the asymptotic zone tend to go back to it. Low velocities do not mean that products do not move, but rather that they move away from their box in all directions. The motion develops mainly across the vertical direction, with points randomly diffusing away from the asymptotic zone, and then moving back to it with regularity. **Panel b**). Average Herfindahl per box, i.e. the H field (in blue color), with its gradient calculated according to Eq 3 (green arrows). Higher values of H signal that there is less competition, on average, for products in that box. The blue line is a kernel regression of the minima of the Herfindahl field per column, with the blue colored area showing the 5% confidence interval on the value, calculated with a bootstrap. Notice its position is compatible with that of the velocity field minima. The H field acts in a way akin to a potential, as the products move towards its minima. This strongly suggests that changes in the vertical position of products is due to changes in the underlying markets. All the vectors in both panels are to 1:1 scale with each other and with the plot's axes.

Spearman correlation coefficient between $r(Q)$ and $r(logPRODY)$ is 0.72. From the most populated to the least populated area of the plane, density changes by about 3 orders of magnitude (A plot of the datapoints is shown in the Supplementary Information addendum). Next, we define the vector field $\vec{v}$ representing the average velocities on said plane, see Fig 1. To calculate $\vec{v}$, we divide the plane into a grid of boxes. Each arrow in Fig 1, panel a), represents the average of all 1-year products' displacements whose starting point is found in the corresponding grid of the box; refer to Methods section for further information on this regression technique. The motion resembles a laminar flow: points on the plane tend to diffuse out of the grid boxes with a certain average velocity, which is a function of the position of the box. The average velocities always point toward a line that is slightly skewed from the diagonal, which we will call the *asymptotic zone*, highlighted in orange in Fig 1. The velocities are higher the farther away one gets from the asymptotic zone. Lower velocities on the asymptotic zone, however, do not mean that products do not move; the low average velocity here means instead that points generally move away from the asymptotic zone in all different directions. In general, there is an ordered movement towards the asymptotic zone, shown in Fig 1 panel a), and a random movement away from the asymptotic zone, which cannot be pictured via the average velocity field $\vec{v}$, since the random component averages to zero. The shape of $\vec{v}$ holds at different time scales, i.e. if we look at the displacements over time intervals ranging from 1 to 10 years, and if we decrease or increase the number of boxes in the grid up to the point where there is at least a number of order 5 points per box (see Supplementary Information). The areas away from the asymptotic

zone, though, do not get emptied over time. What seems to happen is that points tend to diffuse randomly away from the asymptotic zone, but to fall back to it with regularity, their typical return trajectories being captured by $\vec{v}$. We attribute the movement away from the asymptotic zone to contingencies that change the export market for a product and move its position away in a random direction. The sum of the orderly movement towards the asymptotic zone and the random movement away form it returns a picture that is self-consistently stationary on the RCLP, meaning that it keeps its characteristics unchanged over time. Further calculations in support of this stationarity are shown in the Supplementary Information. The velocity field can appear non-stationary because it shows where a product will go, on average, after it is found in a given box, but it does not describe how products enter the boxes. Calculating this "inward" velocity (shown in Supporting Information), a a clear pattern of diffusion out of the equilibrium diagonal emerges. Changes in density of points per box over time (also found in the Supporting Information) are negligible. This stationary behavior is consistent across all datasets examined. These findings suggest that points in the highly populated asymptotic zone are in a sort of equilibrium situation (on which we will elaborate further on), where the average logPRODY is at the "required" value with respect to the Complexity of a product. Points occasionally diffuse away from it, but whenever a point is away from the asymptotic zone, meaning that logPRODY is too high or too low with respect to Complexity, it tends to move towards it. Products farther away from the asymptotic zone also tend to move considerably faster than those lying close to it. This is the observation that led us to the conclusion that there is an asymptotic value of logPRODY for a product, which depends on its Complexity.

What drives the motion. What is the process driving the motion on this plane, i.e. what causes the products to diffuse away from the asymptotic zone, and then go back to it? The answer can be sought by looking at what happens to the export market for a given product in time, and how does this affect the motion and the position of the product on the plane.

We characterize the market by using the Herfindahl index [23, 24]:

$$H_p = \sum_c \left(s_{cp} \right)^2; \qquad s_{cp} = \frac{E_{cp}}{\sum_c E_{cp}} \tag{1}$$

where s_{cp} is the share of country c in the global export market for product p. A sum of the squared market shares of each country, the Herfindahl index equals 1 when the market is a monopoly, and decreases as the competition increases. Calculating the average Herfindahl index per box on the $r(Q) - r(logPRODY)$ plane reveals another pattern. Since such a calculation returns a scalar field sampled at discrete intervals, we call the result the H (Herfindahl) field. The scalar field obtained is quite irregular in the high frequencies, therefore we looked at the lower frequencies of the field by smoothing it with a Gaussian kernel convolution. Superimposing the velocity field $\vec{v}$ shows not only that products drift toward areas of lower Herfindahl index (higher competition), but that the minima of the H field lie in the asymptotic zone, where the velocity field's modulus is minimum as well, as shown in Fig 1. This suggests that H is acting in a way similar to a potential [25]. The motion of producs on the RCLP is the result of actions of their producers and consumers. A product moving from low-competition to a high-competition is, usually, the result of previously less relevant producers increasing their market share in a low-competition market, thus making it, by definition, more competitive. So what we observe can be interpreted as producers actually making the obvious choice: ceteris paribus, it's easier to enter a "blue ocean" [26] than a "red ocean". The interesting finding is that when competition on a product increases, this is associated on the RLCP with movement towards the "asymptotic zone". The opposite—products moving or staying into low-competition areas when their market competition increases—does not seem to happen. In other

words, there seems a tendency for low-competition products to have a logPRODY different from what their Complexity would suggest (or vice-versa).

Note that the derivatives of the Herfindahl field are proportional to the average velocities directly, and not to accelerations. We can explain this result as the outcome of our averaging process, if we interpret the derivatives of H as forces $\vec{F}$ acting on the products in the boxes defined by the grid on the plane. Each product spends an interval Δt in a given box, and exits the box at the end of the interval with a velocity $\vec{v}$. This velocity can be decomposed into two components. One, $\vec{v}_H$, is proportional to $\vec{F}\Delta t$, and is due to the action of H. The other, $\vec{v}_O$ depends on the other degrees of freedom of the system. When we average all the outgoing velocities in a box, we obtain:

$$\langle\vec{v}\rangle = \frac{1}{N}\sum_i^N \left(\vec{v}_{i,H} + \vec{v}_{i,O}\right) = \langle\vec{v}_{i,H}\rangle + \langle\vec{v}_{i,O}\rangle. \tag{2}$$

where $\langle\vec{v}_{i,H}\rangle \propto \vec{F}\Delta t$. If $\langle\vec{v}_{i,O}\rangle$ is reasonably small, this line of reasoning explains our finding that $\langle\vec{v}\rangle$ is correlated to $\vec{F}$.

We calculate the gradient of H, see Fig 1 panel b, and find that the following relation holds well:

$$\vec{v} \simeq -k_x\frac{\partial H}{\partial x}\vec{x} - k_y\frac{\partial H}{\partial y}\vec{y} \equiv -\vec{\nabla}_k H \tag{3}$$

where, for readability reasons, we refer to the Complexity direction on the plane as $\vec{x}$ and the logPRODY direction as $\vec{y}$. To check the validity of equations we evaluate a linear regression between the vertical and horizontal components of the fields. The k_x, k_y coefficients in this formula are the first-order coefficients of this regression; zero-order coefficients are set to zero by hypothesis. The gradient of H has to be scaled differently along the two components to correctly reproduce the velocity field $\vec{v}$, an operation we defined with the symbol $\vec{\nabla}_k$. The concordance between Eq 3 and the data is further discussed in the Supplementary Information addendum. The lower value for the x axis is caused mainly by big outliers in the lower right corner of the plane, where $\vec{v}$ has very high velocities compared to H's gradient. Another feature is the difference between $\vec{v}$ and $-\vec{\nabla}_k H$ in the upper left quadrant of the plane: here Complexity tends to increase slightly, but the model does not predict this phenomenon, and we will discuss the issue further on. Eq 3 holds at different time scales, from 1 to 10 years time interval, and at different resolutions of the grid (See Supplementary Information). It suggests very strongly that the process underlying the motion on the plane is driven by a change in the market for the represented products. The number of exporters of a given product, when it is in an area of the plane with low average H, tends to regularly increase, and this seems to be the driving force changing *logPRODY* and Q. Another clue is that velocity is the lowest where H is at the minimum for each column.

There is a further argument towards the hypothesis that motion on the plane is driven by a change in the set of exporters of a product. The motion of products on the plane from one year to another is too fast for it to be caused just by inter-year change in the GDPpc and Fitness values of exporters. These quantities change too slowly (a few percent at most, yearly) with respect to the speed of products, that can jump most of the ranking in a single time interval. A change in the set of exporters of a product is thus necessary to cause the observed trajectories.

The finding of Eq 3 suggests a scheme where most of the products are at their asymptotic market, with the logPRODY reflecting their Complexity. Their moving away from the asymptotic market is unpredictable and accompanied, on average, by a decrease of the competition

on the market for a given product. Their return to asymptotic market is instead much more regular and evident (see Fig 1 panel a) and corresponds to an increase in competition. This mechanism seems to be reliable enough that we can describe the motion on the plane with an equation that connects velocities to competition, and we can confidently say that vertical motion on the plane corresponds to shifts in the market composition.

Definition of asymptotic market. We showed that the motion on the plane of products is related to the state of their markets. But what are the actual changes in the markets? To answer this question, we consider our interpretation of logPRODY as a synthetic indicator of the market's composition, as it represents a weighted average of GDPpc's. By analyzing the $\tilde{R}$ weights, as defined in Eq 7, we can have a more detailed description of what the market looks like. We remind that the $\tilde{R}_{cp}$'s are the weighs in the logPRODY. They amount to the RCA normalized to one, and they are therefore proportional to the ratio of p's export of country c to all of c's export. The definition is: $\tilde{R}_{cp} = \frac{R_{cp}}{\sum_c R_{cp}}$. In order analyse the $\tilde{R}$ weights, we produced Fig 2. For each box we calculated a histogram that shows the average $\tilde{R}$ values of countries exporting the products therein contained. Each bar of the histograms shows the average RCA of countries with a Fitness value between two consecutive deciles of the Fitness distribution. So the first bar represents the average RCA of countries which are exporting the products in the box and have Fitness lower than the first decile; the second bar shows the RCA of countries with Fitness between the first and the second decile, and so on. $\tilde{R}_{cp}$ is proportional to the comparative advantage country c has in making product p. In short, we are representing the state of the market for products in a certain box in terms of the comparative advantage countries have in making them.

The result is shown in Fig 2, with the red highlighting the minima of the H field for each column. One can clearly see that the distributions on the minima of H go from flat for the lowest Complexity level to markedly peaked on high Fitness for the highest Complexity. These distributions tell us something about the shape of the market for products found in a given box: the $\tilde{R}$ values, as discussed above, tell us the relative importance of a certain export for the corresponding country. While the high complexity products show a peak of RCA on the countries with the highest Fitness, the lowest complexity products exhibit more variation. Most of the low complexity products are found between the lower edge of the plot and the line drawn by the minima of the H field. In this area the market shape shows either a gentle peak around the lower end of the Fitness spectrum or a flat line. This denotes two different market regimes for low-complexity products. One is held by the majority of products, where the comparative advantage is slightly higher for low-Fitness or mid-Fitness countries (but without the sharp peak seen in high-Complexity products), and one at the minimum of H, in which all countries tend to develop about the same comparative advantage, causing the minimum in H. The velocity in this area is very low and points generally from the mildly peaked distributions towards flatter ones.

All these observations point to the fact that there seems to be a characteristic configuration of the export market for a product, that depends on its Complexity levels. We call this the "asymptotic market". High-Complexity products show a sharp peak of comparative advantage on high-Fitness countries, and very low comparative advantage on low-fitness countries. This is expected, as Complexity is bounded by the lowest Fitness found among exporting countries. Low-Complexity products, on the other hand, seem to have two possible configurations: either a mild peak, or a flat distribution. There is a mild, but consistent, tendency to go from slightly peaked towards flat. High competition is associated, on the RCLP, with having a market in the asymptotic state, and deviations from it are associated with lower levels of competition.
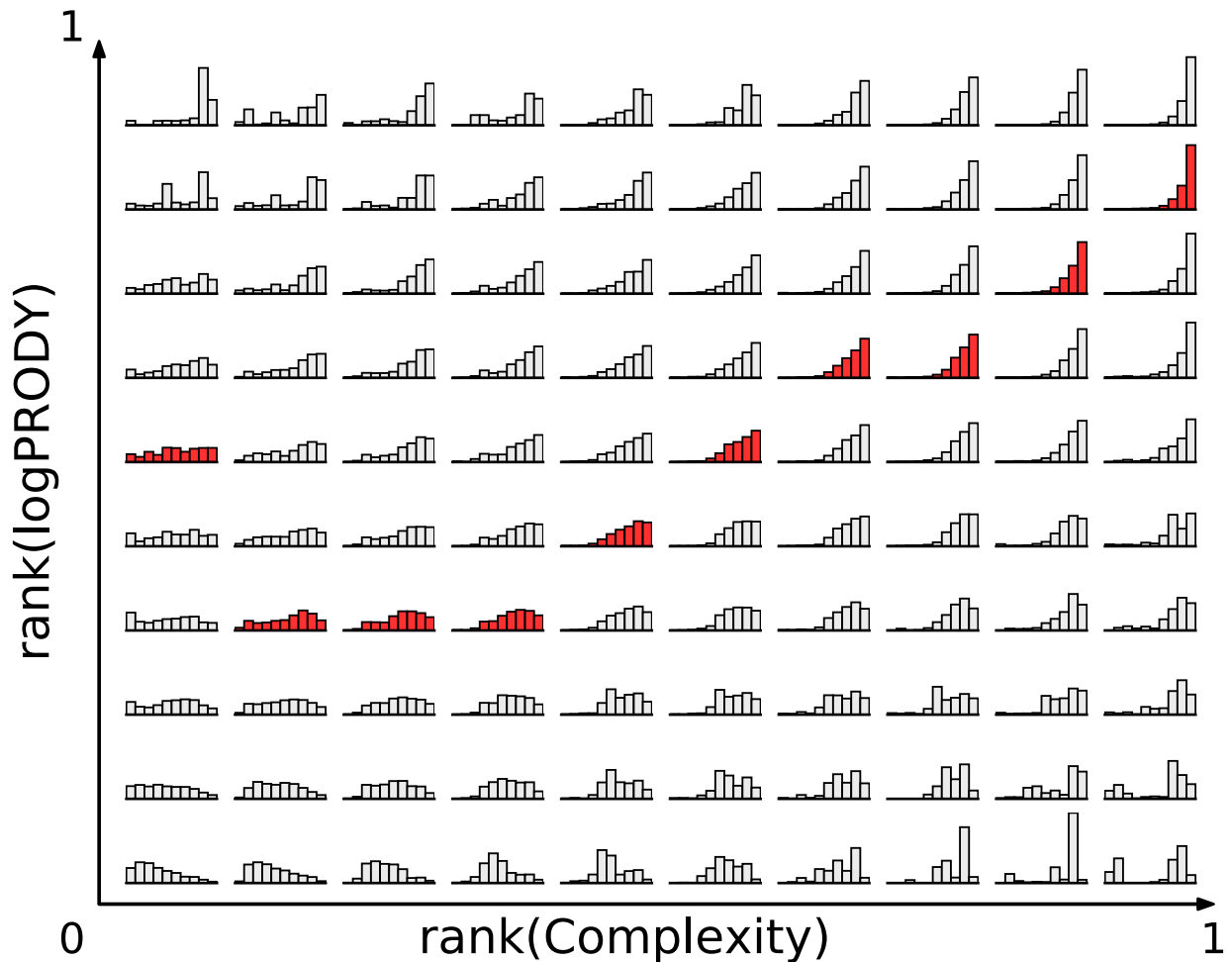
Fig 2. Average RCA weights per box. For each box *b* we plot a histogram showing the average $\tilde{R}$ values of countries exporting the products contained in *b*. Each bar of the histograms shows the average RCA of countries with a Fitness value between two consecutive deciles of the Fitness distribution. We remind that $\tilde{R}$ represents the share of a product in a country's total exports. The distributions on the minima of *H*, which correspond to the asymptotic market shape of the market, go from flat for the lowest Complexity level to markedly peaked on high Fitness for the highest Complexity, and are highlighted in red. Most of the high-Complexity products are found between the asymptotic zone and the top of the plane, and they show a distribution of comparative advantage that is markedly peaked on high-Fitness countries. More interesting is the case of low-Complexity products, that are mostly found between the asymptotic zone and the bottom of the plane. Here the asymptotic market market shape is either rather flat or gently peaked on low-Fitness countries.

https://doi.org/10.1371/journal.pone.0177360.g002

Beyond the Herfindahl approximation. Finally, a word on the discrepancy between Eq 3's predictions and the observed positive velocities along the Complexity axis in the top left part of the plane. This situation can be well understood in terms of the following reasoning. Products in this area have high GDPpc exporters, but both high and low Fitness exporters. When they move towards the diagonal, their exporters' GDPpc and Fitness tend to become more correlated to each other (through changes in the set of exporters). For some of them, this means losing low-Fitness exporters, and this causes an increase in Complexity, generating positive Complexity velocities in this area. logPRODY adjusts accordingly to the new Complexity value, but our model does not predict positive velocities because a high *H* value only signals that the market shape is away from asymptotic market, while a change in Complexity implies a change in the asymptotic market value of a product. On the contrary, products in the lower

right part of the plane tend to have high-Fitness exporters only, since the Complexity of the products found in this area is bound by their lowest Fitness exporter. It is extremely improbable for low-Fitness countries to start exporting high-Complexity products, and this cancels negative velocities on the Complexity axis in the area. To confirm this, we look at the exporter set of products. For each product p, we take its top 30% exporters, as measured with RCA. We then run a regression of these countries' rank(GDPpc) vs. rank(Fitness), and measure the Spearman correlation coefficient R^2. A plot of the pattern formed by this indicator on the RCLP is shown in Fig 3. It is clearly visible that products in the top left corners of the plane have exporters with high GDPpc, but both high and low Fitness, since correlation in these areas is lower. These results explain another aspect of a products' asymptotic market: for a
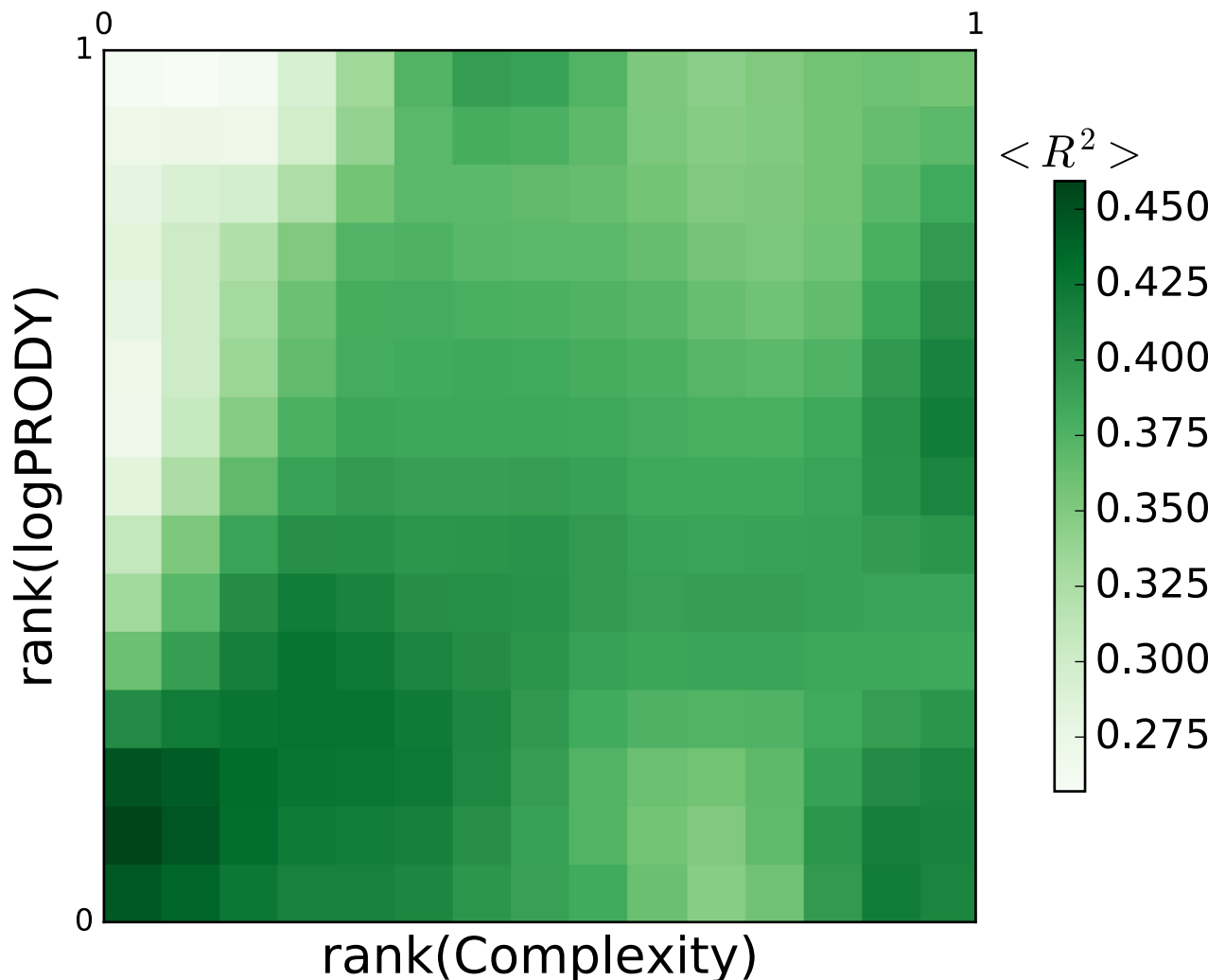

Fig 3. Beyond Herfindahl: Fitness and GDP correlation. We try to explain why the model does not predict positive velocities along the Complexity axis in the top left part of the plane (see Fig 1). We look at the exporter set of products. For each product p, we take its top 30% exporters. We then run a regression of these countries' r(GDPpc) vs. r(Fitness), and measure the Spearman correlation coefficient R^2. We finally average the coefficients per box. This figure shows the pattern formed by the indicator on the RCLP. It is clearly visible that products in the top left corners of the plane have exporters with high GDPpc, but both high and low Fitness. Losing low-Fitness exporters causes them to increase their Complexity, and this generates positive velocities along the horizontal axis. Products in the bottom right corner are much more homogeneous, having only high-Complexity exporters, and undergo less change in Complexity. This figure refers to the Feenstra dataset.

low-Complexity product, around the asymptotic market, the Fitness level of its exporters is correlated to their GDPpc level. In order to capture this aspect of the dynamics on the RCLP, one needs to go beyond the information supplied by just H and use additional data.

In conclusion, the position and movement of the products on the plane is related to the configuration of their respective export markets. In general, market tend to the highest competition possible given the Complexity level of the product, and this high level of competition is associated with a characteristic position on the RCLP plane and a characteristic shape of the market, which we called "asymptotic", both depending on the Complexity level. Markets seem to randomly move away from the asymptotic state, and this is associated with movement away from the asymptotic zoneon the RCLP. The velocity field shows that there is a strong tendency to move back again to the asymptotic configuration of the market.

Discussion

To summarise, the first finding is that there is a value of logPRODY products tend to move towards, which depends on their Complexity value. The value of logPRODY a product tends towards is related to its Complexity in a predictable way; this suggests that the Complexity of a product determines the way its global export market is shaped. Further analysis seems to confirm that there is an asymptotic configuration for a product's export market that we called asymptotic market, which we were able to characterize. High-complexity products have the simplest asymptotic market, as they see countries with the highest Fitness develop the highest comparative advantage in making them. A more interesting finding is that, at asymptotic market, comparative advantage for lower Complexity products is not peaked on low Fitness countries, but seems instead to be, in general, more evenly distributed among all countries. On the plane there is an area that we called the asymptotic zone, which is also characterized by the highest competition values, and the average motion of products tends towards it. Products on the asymptotic zone have on average a particular export market configuration, that we called asymptotic market, whose shape depends on their Complexity. Once products are in the asymptotic zone, they can move away from it in a random fashion, because of a random noise effect. We attribute this effect to contingencies that change the shape of the market into something different from its asymptotic market and, on average, decrease the competition. So in order to understand the motion on the RCLP, both this smooth average motion towards the asymptotic zone and the random motion away from it need to be taken into consideration. As a result of the two, the full picture is self-consistently stationary, with products regularly going towards their asymptotic market, and randomly exiting it.

The comparison of monetary and non-monetary metrics proposed in the Economic Complexity framework, and the use of complexity tools such as dynamical system methods, allows us to observe regularities in the export markets' behavior that are hard to notice with more conventional mathematical tools. By looking at the trajectories of products on the RCLP, we are able to infer that there seems to be an asymptotic value of logPRODY which depends on the level of Complexity, and whenever a product is away from its asymptotic logPRODY it tends to move towards it. Study of the underlying markets showed that vertical movement in the RCLP is associated with moving from low-competition areas of the plane towards zones with higher average competition, and this happens with such regularity that we can use Eq 3 to describe the average motion. Therefore the change in logPRODY is caused by a shift in the set of countries that export a product at a given time: logPRODY can be seen as a synthetic index describing the export market for a product, whereas it cannot convey all the information given by a market representation such as those of Fig 2. The fact that the asymptotic logPRODY for a product is determined for the most part by the corresponding value of Complexity suggests

therefore that the shape of the export market is determined by the complexity of a product. Complexity, in turn, expresses how difficult it is to make a product, and hints to some bounds on which countries can make certain products.

Study of the coupled time evolution of the two indices is telling us that the logPRODY of a product is usually aligned with its Complexity level, meaning that the players in the product's export market are distributed in a certain typical way, that we called asymptotic market. Whenever this does not happen, logPRODY tends to change to remove this misalignment. This change reflects a change in the market for the product, that tends to move towards an asymptotic marketsituation. The Complexity level of a product determines how the asymptotic market market is shaped, and in turn the way the market is shaped informs us on the Complexity level of a product. Higher Complexity products tend to have markets in which only high-Fitness countries develop a significant comparative advantage. On the other hand, lower-Complexity products do not show a peak on the lowest-Fitness countries: a "flat" distribution of comparative advantage is observed among all countries at asymptotic market (even though some of these products exhibit a distribution with a slight peak, it is nowhere near the sharpness seen for complex products). The observed phenomena could be said to be a process of return to equilibrium on a short time scale, as there certainly are well behaved regularities in the aggregated behavior of products, consistent across all the data examined. On longer time scales, there is evidence that low-Complexity products tend to be replaced by high-Complexity products as countries develop [27].

We speculate that the observed asymptotic market market shapes is a consequence of the triangular shape of the export matrix [1], which in turn follows from the fundamental empirical finding [4] that development of a country grows together with diversification, from low-Complexity export to more sophisticated products. We remind that the export matrix is the adjacency matrix of the export network, connecting a country c to all the products $\{p\}$ that it exports. High-Complexity products are exported only by a few extremely diversified countries, that also tend to have very high Fitness, and this could be the cause of the narrow peak of comparative advantages observed in Fig 2. Products with lower Complexity, instead, are exported by almost all countries, and not just by low-Fitness ones; this suggests an explanation for why, at asymptotic market, we generally see flat or gently peaked market shapes for this kind of products. Because of the shape of the adjacency matrix, and the definition of Complexity, there is an approximate maximum level of competition for a product. This maximum decreases with the Complexity of the product, because there are few countries that can export high-Complexity products. Our finding is the observation that, when at this maximum of competition, the market is shaped in a characteristic way, that also depends on the Complexity, and that we called asymptotic market.

Furthermore the dynamics out of the asymptotic regime raises a natural question concerning the creation of new opportunities or movements driven by the attempt of reducing the competition or the creation of commercial niche: roughly speaking, whether schemes like red ocean/blue ocean [26] can be somehow mapped into this kind of analysis and, in general, whether they would applicable to bilateral trade networks. However, moving forward in this line of reasoning would require reliable and granular product value added data and not simply raw measures of product value. As previously mentioned, there is a general scarcity of value added data. We basically find value added data only for US internal production at a level of aggregation comparable to HS2007 4 digits. Using only this data would require the clearly poor assumption that all countries share the same structure of product value added. On the other hand, we have other sources of data which cover approximately 50–60 countries and approximately one decade (mainly OECD countries plus few large developed non OECD countries) but unfortunately they are defined at a much more aggregated level (comparable

with 1 or 2 digit level of HS2007). Including value added data represents one of the challenging next steps of our analyses and the latter data-sets are clearly the most promising candidates but they require additional modeling in order to break down aggregate value added at a more granular level.

Finally, a finding in itself is that results obtained with the Herfindahl field H suggest that models inspired by statistical physics and complexity science can be used effectively to explore and understand the behavior of economic phenomena. Just regressing the value of H versus the value of logPRODY cannot take into account the dependency on Complexity and changes in time, resulting in a correlation coefficient near to zero. Measuring the average value per box of an observable on the RCLP, and representing it as a scalar field, is a useful extension of the Economic Complexity toolbox, that could be applied in the future in to shed light on other processes being examined, where more traditional approaches might fail. We also believe that the great consistency of results across different datasets speaks to the coherence and well-groundedness of our methods, and the Economic Complexity framework as a whole.

Methods

Here we add some information on the data and methodologies needed for the reproduction of our results. This section is organized into sub-sections. In this order, we will illustrate the data-sets used and their structure, clarify the meaning of the non-monetary Fitness and Complexity metrics, explain the role of the ranking function in our analysis, and finally make some remarks on the regression techniques we used.

Datasets

In this work we use two different datasets, that contain all the information of the E_{cp} matrix (from which all the metrics considered in this paper can be calculated, except for GDP). We will call the first dataset BACI, and it is described in [28]. It consists in data obtained by the UN-COMTRADE, and elaborated by CEPII, from which it has been purchased. For this reason, it cannot be made publicly available, although a free version without data cleaning is available on the BACI section of the organization's website [29]. We use the data for 148 countries, and spanning the 20 years from 1995 to 2014. In this dataset, products are classified according to the Harmonized System 2007 [30] which denotes them with a set of 6-digit codes organized in a hierarchical fashion. The code is divided into three 2-digit parts, each specifying one level of the hierarchy: so the first part indicates the broadest categories, such as "live animals and animal products" (01xxxx) or "plastics and articles thereof" (39xxxx). The second two digits specify further distinctions in each category, such as "live swine" (0103xx), and "live bovines" (0102xx). The last two digit are even more specific. For the analysis mentioned in the paper, we look at data for products aggregated at 4-digit level (1131 products). For the analysis on prices, for which we do not mention any results, we use a similar dataset, again released by BACI, containing all the transactions between countries in the 2008–2014, at 6-digit level. The second dataset for which we show results in the paper is available on the NBER-UN world trade data website [31], extensively documented by Feenstra et al. in [22], therefore we call it the Feenstra dataset in this work. After data cleaning procedures, we retain the data for 158 countries and 538 products, classified according to the SITC rev.2 [32]. Data cleaning procedures have been performed with the same methodologies on both datasets for which we present results. The procedures consist in the elimination of extremely small countries and countries with fragmented data; the aggregation of some product categories that are closely related, and a regularization of the M_{cp} matrices. The dataset spans 36 years, from 1963 to

2000. GDPpc data has been downloaded from the World Bank Open Data website [33]. All results shown in this paper are consistent across all the datasets examined.

Non-monetary metrics

As already discussed, Fitness and Complexity measures stem from an attempt to improve current theories about economic growth. It has been shown that developed countries show a high diversity of export, while poorly diversified economies tend to be competitive in just those products that are exported by almost all other countries [4]. A model to explain this phenomenon is given in [3]: the bipartite products-countries network is actually the only measurable part of a tripartite products-capabilities-countries network. The capabilities are unobservable, and are linked to both the countries that possess them and the products that need them in order to be exported. A link is added between country c and product p if c is linked to all capabilities k needed to export p. This will result in a nested products-countries network, as some capabilities are common to almost all countries, and others are rarer, and associated to developed countries only. Since diversification appears to be crucial to development of an economy, Fitness and Complexity are designed to efficiently extract nestedness information from the bipartite network of exporting countries and exported products. This is obtained via an algorithm iterating a highly non-linear map onto the adjacency matrix of the export network. The map is designed according to the following two observations. First, a product being made by a very diversified country is uninformative, while if it is made by at least one underdeveloped country we have grounds to think that it is a low-Complexity one. Vice-versa, if a product is made only by high-Fitness countries, it is reasonable to expect that it has high Complexity. The result is one of the simplest algorithms capable to be coherent and compatible with the kind of information being manipulated, taking advantage of the highly nested nature of the export network. An important feature of the algorithm, which is important to our analysis, is that the Complexity of a product has an upper bound, namely the lowest Fitness value found among its exporters. We define M as the adjacency matrix of such a network: M_{cp} is equal to 1 if country c is an exporter of product p (defined as $R_{cp} \geq 1$), and 0 if it is not ($R_{cp} < 1$). Fitness and Complexity values are defined as the fixed point of this non-linear coupled map:

$$F_c^{(0)} = 1 \quad \forall c, \quad Q_p^{(0)} = 1 \quad \forall p. \tag{4}$$

$$\tilde{F}_c^{(n)} = \sum_p M_{cp} Q_p^{(n-1)}, \tilde{Q}_p^{(n)} = \frac{1}{\sum_c M_{cp} \frac{1}{F_c^{(n-1)}}} \tag{5}$$

$$F_c^{(n)} = \frac{\tilde{F}_c^{(n)}}{\left\langle \tilde{F}_c^{(n)} \right\rangle_c}, \quad Q_p^{(n)} = \frac{\tilde{Q}_p^{(n)}}{\left\langle \tilde{Q}_p^{(n)} \right\rangle_p}. \tag{6}$$

This defines an iterative process coupling the F_c to the total complexities Q_p of products exported by c. The equation for Q is non-linear, and bounds the complexity of a product to be smaller than the lowest Fitness value found among its exporters. The iteration is run until the values of F and Q reach a fixed point. The fixed point can be defined numerically in various ways. Here we used the definition proposed by Pugliese et al. [34], based on the stability of the ranking, which has to stop changing for a high number of successive iterations. This algorithm is capable of capturing the characteristic nestedness of the bipartite export networks, which is relevant to examining the dynamics of industrialization found in developing countries, but also to analyze in detail the production structure of developed countries [35–37]. Further

details can be found in [1, 9], the convergence problem is studied in [34, 38], and a study of the stability of these metrics with respect to noise in the data has been performed in [39, 40].

Monetary metrics

Both aforementioned Sophistication and PRODY for a product p are defined as a suitably weighted average of the GDPpc's of p's exporters. In Sophistication, the weights are the export market shares of countries for product p, while PRODY uses the RCA values, defined in the following. We found that results are essentially similar whether one uses PRODY or Sophistication, as some literature seems to already suggest [41]. We identify logPRODY as the monetary metric of choice, which is a modification of the PRODY index proposed by Hausmann [3], who employed it to investigate the relationship between exports and growth of a country. logPRODY is defined, for a product p, as follows:

$$logPRODY_p \equiv \sum_c \frac{R_{cp} \log_{10}(Y_c)}{\sum_c R_{cp}} = \sum_c \tilde{R}_{cp} \log_{10}(Y_c) \qquad (7)$$

where R_{cp} is the so-called Revealed Comparative Advantage (RCA), or Balassa index [42], and we defined the weighs $\tilde{R}_{cp} = R_{cp}/\sum_c R_{cp}$. If we define the value in dollars of product p exported by country c as E_{cp}, then the R_{cp} is defined as:

$$R_{cp} = \frac{\dfrac{E_{cp}}{\sum_j E_{cj}}}{\dfrac{\sum_i E_{ip}}{\sum_{kl} E_{kl}}}. \qquad (8)$$

The original PRODY is defined the same way, except that $\log_{10}(Y_c)$ is replaced by Y_c in the sum. The change to logarithms has been chosen because GDPpc's of countries span about four orders of magnitude, and the geometric mean is better suited to represent such a numeric distribution of values. Normalization of the weight $\tilde{R}_{cp}$ takes away the dependency of RCA on the denominator in its definition. Therefore $\tilde{R}_{cp}$ is proportional to E_{cp} divided by the total export of country $\sum_j E_{cj}$, i.e. the share of p in all of c's exports. This means that the biggest the share of a certain product p in the total export of a country c, the more c is weighted in the average that defines $logPRODY_p$. As with all economic measures, logPRODY has its own pitfalls. It is less reliable when used to characterize products whose production is location-specific, such as raw materials, some kinds of vegetables. An interesting example is live swine, which is very common in rich western countries, and almost absent throughout the islamic world. As a result, it is one of the products with the highest logPRODY. We ran all our calculations both with and without these peculiar products, and didn't notice any significant change in the signal, therefore we believe this shortcoming of the metric does not impact negatively on our findings. In this respect, we stress that the non linearity of the Complexity measure solves this kind of issues, because it is enough to have only one low fitness country that export a raw material to obtain a low complexity score for that product. These intrisic differences make the comparison of the two quantities particularly meaningful.

Ranking

The coordinates of choice for exploring the dynamics of products are not directly the value of Complexity and logPRODY, but their yearly tied ranking (normalized to 1, with zero being the lowest ranking). This is because of a known phenomenon observed during the

convergence of the map in Eq 6 defining Fitness and Complexity, namely that some values tend asymptotically to zero [34]. At the numerically estimated fixed point of the map, Complexity values span about 300 orders of magnitude, from 10^1 to 10^{-300}, this number being limited only by computer precision. Some of the products reach complexities so low that they have to be approximated to 0. Additionally, the yearly change depends on the order of magnitude of the Complexity, i.e. a complexity of about 10^{-100} has typically changes of about 10^{-100}, while a complexity of order 1 moves of about order 1. This makes it hard to compare velocities across the plane, even when it is represented with logarithmic proportions. Ranking has the nice property of assigning equal spacing between one value and the next, solving all of these problems. It also introduces potential distortions in the signal, though, as a product's rank is directly dependent on all other products' Complexity values. In order to evaluate this potential distortion, we study a model in which we rank a number of values that evolve in time with a random walk. We find that the distortions introduced by the ranking are progressively higher as one moves towards the edges of the ranking, and depend on the process underlying the random walk. In the case of our dataset's particular kind of motion, we find that the distortions introduced by the ranking are negligible in size.

Regression technique

The main regression technique used throughout this work consists in gathering information, in the form of a field, from a series of discrete trajectories on a plane. To measure velocity, for example, one can consider all the positions associated to the trajectories as points on the plane. To each point a velocity is associated by considering how the point moved along the trajectory it belongs to. One can then divide the plane into boxes with a square grid, and average all the velocities of points in the same box, obtaining a discrete vector field. The same can be done for other observables, i.e. for the Herfindahl index one can associate to each point (representing the state of a product at a certain time) the corresponding value of H, and then make an average per box obtaining a scalar vector field. It is fundamentally a non-parametric regression technique, based on Lorenz's method of analogues [43]. We checked that results presented in this paper hold using two more regression techniques, namely a Nadaraya-Watson kernel regression [44] and the *bdynsys* R package by Ranganathan et al. [45], which is essentially a utility for finding the best function to fit on the data among all the possible polynomials up to a certain degree. Both produce fields that are very similar to those obtained with the average box velocity estimation, and comparable results in all cases.

Supporting information

S1 Fig. Comparison between the 1-year average velocity field, $\vec{v}$, and the field obtained from the gradient of H. Calculations done on the Feenstra dataset. *Top*: comparison between the horizontal and vertical components of the field. In the left panel, showing the regression for velocities along the logPRODY axis, the points removed by the RANSAC regression are shown in red. The *Bottom*: Comparison between orientations and moduli of the field. (PDF)

S2 Fig. Bootstrapping the regression of $\vec{v}_y$ versus $-\vec{\nabla}_k H$ at 1 year displacement time and 10×10 grid. Each iteration of the bootstrap randomly removed 10% of the samples, for 50k iterations. Two peaks are clearly visible: one at about $R^2 = 4$, and one at about .7. The bimodal distribution confirms our hypothesis that the points follow a very clear linear trend, with a few very big outliers. The peak at .7 is caused by the bootstrap randomly removing the outliers from the regression. The peak at .4 comes from the bootstrap removing points following the

linear trend, and leaving the outliers in. This clear bimodal distribution allows us to use RAN-SAC, which is an algorithm apt to take out a few big outliers from a regression.
(PDF)

S3 Fig. A plot of the R^2 values obtained by regressing $\vec{v}_y$ versus $-\vec{\nabla}_k H$ at various resolution and displacement time values, Feenstra dataset. On the horizontal axis there is the resolution of the grid, with the number of boxes per side. In blue to red color, we show the time interval used to calculate the products' displacements, which are then averaged into $\vec{v}$. The yellow line indicates the minimum number of points per box at a given resolution. The accuracy of our model's prediction peaks at a resolution between 10 and 15. Less resolution is probably too little detail to capture the features of the system. Accuracy starts to drop, especially for longer time displacements, as soon as the number of points per box is 1; this is probably caused by an increase in noise in the less populated areas of the plane. All the regressions are linear, calculated with RANSAC, and with the zero-order coefficient set to zero by hypothesis.
(PDF)

S4 Fig. $\vec{v}$ and $-\vec{\nabla}_k H$ fields for the BACI dataset. The results are extremely similar to those shown in the paper for the Feenstra dataset, with great consistency across datasets. **Panel a**). In red, the average velocity field of the products on the *RCLP* plane, $\vec{v}$. The asymptotic zone is marked by the orange line, together with the bootstrap result for the 5% confidence interval. **Panel b**). Average Herfindahl per box, the H field (in blue), with the derivatives (green arrows). The line is the kernel regression of the minima of the Herfindahl field per column, with 5% confidence interval on the value superimposed. All the arrows in both panels are to 1:1 scale with each other and with the plot's axes.
(PDF)

S5 Fig. Average RCA weights per box, BACI dataset. For each box b we plot a histogram showing the average $\tilde{R}$ values of countries exporting the products contained in b. Each bar of the histograms shows the average RCA of countries with a fitness value between two consecutive deciles of the fitness distribution. We remind that $\tilde{R}$ represents the share of a product in a country's total exports. Here we see the same patterns found in the Feenstra dataset: the distributions on the minima of H going from flat for the lowest Complexity level to markedly peaked on high fitness for the highest Complexity. Again, the results show consistency.
(PDF)

S6 Fig. Comparison between the 1-year average velocity field, $\vec{v}$, and the field obtained from the gradient of H defined in the paper. Calculations done on the BACI dataset. *Top*: comparison between the horizontal and vertical components of the field, the latter regressed with RANSAC. *Bottom*: Comparison between orientations and moduli of the field.
(PDF)

S7 Fig. All individual data points of the BACI dataset. One can clearly see higher density on the diagonal, and a decreasing number of points in the upper left and lower right quadrants of the RCLP plane.
(TIF)

S8 Fig. Panel a) Maximum value of the H field. Given that the less points one has per box, the higher the standard deviation for H in that box is, one expects to find higher values of H where the density of points is lowest on the *RCLP* plane. We check whether the null model can produce a $\tilde{H}$ field with the same highest value found in H, which we will call max H. Running the

model 1M times produced few samples of comparable highest value max $\tilde{H}$. In the histogram we show the frequency of max $\tilde{H}$; the vertical line shows max H for the Feenstra dataset. A similar result (not shown) is found in the BACI dataset. **Panel b**) Check of the potential-like equation's validity versus the null model. We ran the null model 1M times, and for each $\tilde{H}$ field obtained we calculated the correlation between $\tilde{H}$'s and $\vec{v}$'s vertical and horizontal components with a linear regression. The sum of the two Pearson's R^2 coefficients obtained from this process, R^2_{tot} (which is a number between 0 and 2), is used here as a measure of significance for the $\tilde{H}$. In this histogram we show the frequency of the R^2_{tot} values; the vertical line represents R^2_{tot} obtained from the H field. The null model produces gradients that correlate both positively or negatively with each component of $\vec{v}$: in some cases, one gets a high R^2_{tot}, but the field being reproduced is the inverse of $\vec{v}$. In light blue, we show the fraction of the $\tilde{H}$'s that correctly reproduce the direction of $\vec{v}$ on both components, while dark blue represents the fraction that reproduces at least one of the components of $-\vec{v}$. The result is that, among all generated $\tilde{H}$ fields (both with right and wrong direction), those with a significance level higher than that of H are 1.5%.
(PDF)

S9 Fig. The field obtained by calculating the gradient of the H^R field.
(PDF)

S10 Fig. Comparison between the 1-year average velocity field, $\vec{v}$, and the field obtained from the gradient of H^R. Calculations done on the Feenstra dataset. *Top*: comparison between the horizontal and vertical components of the field *Bottom*: Comparison between orientations and moduli of the field.
(PDF)

S11 Fig. *Left*: Relation between values of the S and H field in the BACI dataset. Left, in black, a linear fit, which is good enough for the fields, that aggregate many individual points. The actual relationship between entropy and Herfindahl index of a distribution is not linear, as we show in the next section. *Right*: The field obtained by using S.
(PDF)

S12 Fig. Comparison between the 1-year average velocity field, $\vec{v}$, and the field obtained from the gradient of S. Calculations done on the Feenstra dataset. *Top*: comparison between the horizontal and vertical components of the field *Bottom*: Comparison between orientations and moduli of the field.
(PDF)

S13 Fig. In green, the field obtained by regressing the average value of U (Ubiquity) per box in the *RCLP* plane. It is different from the field obtained by regressing H, because the average Ubiquity of products decreases with increasing Complexity; therefore the left part of the *RCLP* plane has higher values than the right. The plot shows the regression for the Feenstra dataset.
(PDF)

S14 Fig. "inward" velocity field $\vec{w}$. To calculate this field, considered all products that entered a given box, and averaged their displacements. The vectors are not to scale with those shown in the rest of the paper. To allow comparison, we present a depiction of both $\vec{w}$ and $\vec{v}$ fields in the next figure. The figure refers to the BACI dataset.
(PDF)

S15 Fig. Both $\vec{w}$ and $\vec{v}$ fields, to scale with each other. The figure refers to the BACI dataset.
(PDF)

S16 Fig. Occupation number timeseries for each box onb the RCLP plane. Each of the plots in this figure shows the evolution in the number of products contained in each box. The horizontal axis of the plots represents time in years, and the vertical axis the number of points. The plots are to scale relative to each other. The figure refers to the BACI dataset.
(PDF)

S17 Fig. Inward (blue) and outward (red) fluxes for each box on the RCLP plane. Each of the plots in this figure shows the yearly change in the number of products contained in each box. The horizontal axis of the plots represents time in years, and the vertical axis the difference in number of points (for the outward flux the difference is negative). The plots are to scale relative to each other. The figure refers to the BACI dataset.
(PDF)

S18 Fig. A scatter-plot of the Herfindahl index vs. the entropy value for individual products' market share distributions, from the BACI dataset.
(PDF)

Acknowledgments

Author Contributions

Conceptualization: OA MC AZ LP.

Data curation: OA MC AZ LP.

Formal analysis: OA MC AZ LP.

Funding acquisition: OA MC AZ LP.

Investigation: OA MC AZ LP.

Methodology: OA MC AZ LP.

Project administration: OA MC AZ LP.

Resources: OA MC AZ LP.

Software: OA MC AZ LP.

Supervision: OA MC AZ LP.

Validation: OA MC AZ LP.

Visualization: OA MC AZ LP.

Writing – original draft: OA MC AZ LP.

Writing – review & editing: OA MC AZ LP.

References

1. Tacchella A, Cristelli M, Caldarelli G, Gabrielli A, Pietronero L. A New Metrics for Countries' Fitness and Products' Complexity. Scientific Reports. 2012; 2. https://doi.org/10.1038/srep00723 PMID: 23056915

2. Cristelli M, Tacchella A, Pietronero L. The Heterogeneous Dynamics of Economic Complexity. PLOS ONE. 2015; 10(2):e0117174. https://doi.org/10.1371/journal.pone.0117174 PMID: 25671312

3. Hausmann R, Hwang J, Rodrik D. What you export matters. Journal of Economic Growth. 2007; 12(1): 1–25. https://doi.org/10.1007/s10887-006-9009-4

4. Hidalgo CA, Hausmann R. The building blocks of economic complexity. Proceedings of the National Academy of Sciences. 2009; 106(26):10570–10575. https://doi.org/10.1073/pnas.0900943106

5. Dosi G. Sources, Procedures, and Microeconomic Effects of Innovation. Journal of economic literature. 1988; 26(3):1120–1171.

6. Lall S. Technological capabilities and industrialization. World Development. 1992; 20(2):165–186. https://doi.org/10.1016/0305-750X(92)90097-F

7. Teece DJ, Rumelt R, Dosi G, Winter S. Understanding corporate coherence. Theory and evidence. Journal of Economic Behavior and Organization. 1994; 23(1):1–30. https://doi.org/10.1016/0167-2681(94)90094-9

8. Hidalgo CA, Klinger B, Barabasi AL, Hausmann R. The Product Space Conditions the Development of Nations. Science. 2007; 317(5837):482–487. https://doi.org/10.1126/science.1144581 PMID: 17656717

9. Cristelli M, Gabrielli A, Tacchella A, Caldarelli G, Pietronero L. Measuring the Intangibles: A Metrics for the Economic Complexity of Countries and Products. PLoS ONE. 2013; 8(8). https://doi.org/10.1371/journal.pone.0070726 PMID: 23940633

10. Zaccaria A, Cristelli M, Tacchella A, Pietronero L. How the Taxonomy of Products Drives the Economic Development of Countries. PLoS ONE. 2014; 9(12):e113770. https://doi.org/10.1371/journal.pone.0113770 PMID: 25486526

11. Page L, Brin S, Motwani R, Winograd T. The PageRank Citation Ranking: Bringing Order to the Web. World Wide Web Internet And Web Information Systems. 1998; 54(1999–66):1–17.

12. Caldarelli G, Chessa A, Pammolli F, Gabrielli A, Puliga M. Reconstructing a credit network. Nature Physics. 2013; 9(3):125–126. https://doi.org/10.1038/nphys2580

13. Cadot O, Carrère C, Strauss-Kahn V. Export Diversification: What's behind the Hump? Review of Economics and Statistics. 2011; 93(2):590–605. https://doi.org/10.1162/REST_a_00078

14. Bascompte J, Jordano P, Meliãn CJ, Olesen JM. The nested assembly of plant-animal mutualistic networks. Proceedings of the National Academy of Sciences of the United States of America. 2003; 100(16):9383–9387. https://doi.org/10.1073/pnas.1633576100 PMID: 12881488

15. Domínguez-García V, Muñoz MA, Dunne J, Williams R, Martinez N, Dunne J, et al. Ranking species in mutualistic networks. Scientific Reports. 2015; 5:8182. https://doi.org/10.1038/srep08182 PMID: 25640575

16. Lall S, Weiss J, Zhang J. The "sophistication" of exports: A new trade measure. World Development. 2006; 34(2):222–237. https://doi.org/10.1016/j.worlddev.2005.09.002

17. OECD. OECD Trade in Value Added Database;. Available from: http://www.oecd.org/sti/ind/measuringtradeinvalue-addedanoecd-wtojointinitiative.htm.

18. WIOD (World input output Database) tables;. Available from: http://www.wiod.org/new_site/database/wiots.htm.

19. Jarreau J, Poncet S. Export sophistication and economic growth: Evidence from China. Journal of Development Economics. 2012; 97(2):281–292. https://doi.org/10.1016/j.jdeveco.2011.04.001

20. Ghani E, Goswami AG, Kharas H. Can services be the next growth escalator?; 2011. Available from: http://voxeu.org/article/can-services-be-next-growth-escalator?quicktabs_tabbed_recent_articles_block=0.

21. Jarreau J, Poncet S. Sophistication of China's exports and foreign spillovers. Journal of Economic Surveys. 2009;(7):149–161.

22. Feenstra R, Lipsey R, Deng H, Ma A, Mo H. World Trade Flows: 1962–2000. Cambridge, MA: National Bureau of Economic Research; 2005. Available from: http://www.nber.org/papers/w11040.

23. Rhoades SA. Herfindahl-Hirschman Index, The. Federal Reserve Bulletin. 1993;(79):188.

24. Kelly WA. A Generalized Interpretation of the Herfindahl Index. Southern Economic Journal. 1981; 48(1):50. https://doi.org/10.2307/1058595

25. Goldstein H, Poole C, Safko J. Classical Mechanics; 2007.

26. Kim WC, Mauborgne R. Blue ocean strategy: from theory to practice. California Management Review. 2005; 47(3):105–121. https://doi.org/10.2307/41166308

27. Klimek P, Hausmann R, Thurner S. Empirical confirmation of creative destruction from world trade data. PloS one. 2012; 7(6):e38924. https://doi.org/10.1371/journal.pone.0038924 PMID: 22719989

28. Gaulier G, Zignago S. BACI: International Trade Database at the Product-Level (the 1994–2007 Version). SSRN Electronic Journal. 2010; https://doi.org/10.2139/ssrn.1994500

29. CEPII BACI dataset;. Available from: http://www.cepii.fr/CEPII/en/welcome.asp.

30. World Customs Organization;. Available from: http://www.wcoomd.org/.

31. National Bureau of Economic Research dataset;. Available from: http://www.nber.org/data/.

32. United Nations Statistics Division—SITC rev.2 Classifications Registry;. Available from: http://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=8.

33. The World Bank Open Data;. Available at: http://data.worldbank.org/.

34. Pugliese E, Zaccaria A, Pietronero L. On the convergence of the Fitness-Complexity algorithm. The European Physical Journal Special Topics. 2016; 225(10):1893–1911. https://doi.org/10.1140/epjst/e2015-50118-1

35. Zaccaria A, Cristelli M, Kupers R, Tacchella A, Pietronero L. A case study for a new metrics for economic complexity: The Netherlands. Journal of Economic Interaction and Coordination. 2016; 11(1): 151–169. https://doi.org/10.1007/s11403-015-0145-9

36. Pugliese E, Chiarotti GL, Zaccaria A, Pietronero L. Complex economies have a lateral escape from the poverty trap. 2015; PloS one 12 (1), e0168540 https://doi.org/10.1371/journal.pone.0168540

37. Cristelli M, Tacchella A, Zaccaria A, Pietronero L. Growth scenarios for sub-Saharan countries in the framework of economic complexity. 2014; submitted. Available at https://mpra.ub.uni-muenchen.de/71594/

38. Wu RJ, Shi GY, Zhang YC, Mariani MS. The mathematics of non-linear metrics for nested networks. Physica A: Statistical Mechanics and its Applications. 2016; 460:254–269. https://doi.org/10.1016/j.physa.2016.05.023

39. Battiston F, Cristelli M, Tacchella A, Pietronero L. How metrics for economic complexity are affected by noise. Complexity Economics. 2014; 3(Number 1 / 2014):1–22.

40. Mariani MS, Vidmer A, Medo M, Zhang YC. Measuring economic complexity of countries and products: which metric to use? The European Physical Journal B. 2015; 88(11):293. https://doi.org/10.1140/epjb/e2015-60298-7

41. Minondo A. Technology and sophistication: a tale of two indexes. The Empirical Economics Letters;(7): 331–339.

42. Balassa B. Trade Liberalisation and "Revealed" Comparative Advantage. The Manchester School. 1965; 33(2):99–123. https://doi.org/10.1111/j.1467-9957.1965.tb00050.x

43. Lorenz EN. Atmospheric Predictability as Revealed by Naturally Occurring Analogues. Journal of the Atmospheric Sciences. 1969; 26(4):636–646. https://doi.org/10.1175/1520-0469(1969)26%3C636:APARBN%3E2.0.CO;2

44. Nadaraya EA. On Estimating Regression. Theory of Probability & Its Applications. 1964; 9(1):141–142. https://doi.org/10.1137/1109020

45. Ranganathan S, Spaiser V, Mann RP, Sumpter DJT. Bayesian Dynamical Systems Modelling in the Social Sciences. PLoS ONE. 2014; 9(1):e86468. https://doi.org/10.1371/journal.pone.0086468 PMID: 24466110