scientific reports

OPEN

Entropic measure unveils country competitiveness and product specialization in the World trade web

Gianluca Teza [1], Michele Caraglio [2] & Attilio L. Stella [3,4]✉

We show how the Shannon entropy function can be used as a basis to set up complexity measures weighting the economic efficiency of countries and the specialization of products beyond bare diversification. This entropy function guarantees the existence of a fixed point which is rapidly reached by an iterative scheme converging to our self-consistent measures. Our approach naturally allows to decompose into inter-sectorial and intra-sectorial contributions the country competitivity measure if products are partitioned into larger categories. Besides outlining the technical features and advantages of the method, we describe a wide range of results arising from the analysis of the obtained rankings and we benchmark these observations against those established with other economical parameters. These comparisons allow to partition countries and products into various main typologies, with well-revealed characterizing features. Our methods have wide applicability to general problems of ranking in bipartite networks.

One of the aims of the economic complexity approach[1–6] is that of establishing consistent rankings of countries in terms of productive efficiency and of products in terms of sophistication[7,8], with the final goal of estimating both measurable and hidden growth potentials of the economies. These rankings should result from an analysis of bipartite networks which connect the countries to their products. The yearly amounts of the various exports of each country offer the possibility to construct such networks and, on their basis, various algorithms have been proposed so far to obtain measures on which to base the rankings[1–3].

While the general ideas inspiring these methods appear quite convincing and promising, the status of the art regarding their interpretation and implementation is not completely satisfactory, in spite of the time elapsed since the first proposals. The mathematical structure of the approach of Ref.[1] poses problems of interpretation[9], while the methods of Refs.[2] have been shown to suffer poor convergence[10] requiring substantial reformulations[11]. In addition, most implementations so far make use of binarizations of the bipartite network linking countries to products. This binarization, although based on RCA[12] criteria, makes only distinction between countries which export and countries which do not export a given product, hence causing a substantial loss of the information provided by the available data concerning exports. Therefore it would be advisable to construct and use bipartite networks that make full use of the information contained in the original data.

On a more general level, one should note that, with a single exception[13], all approaches to economic complexity proposed so far emphasize the importance of diversity as an essential ingredient of the measures[1,2,5]. In various contexts, ranging from ecology[14] to economics itself[15,16], the entropy function first introduced by Boltzmann in statistical physics and later by Shannon in communication theory[17,18] is universally used and appreciated as an indicator of diversity[19]. It is thus somehow surprising that the use of the Shannon entropy function for the construction of economic complexity measures was proposed only in one case so far[5].

In this report, following lines already drawn in Refs.[5,20], we show that diversity, as recorded through the Shannon entropy function, can be taken as a basis for establishing meaningful rankings which are fully consistent with the aims of the economic complexity approach, and exempt from difficulties of both interpretation and implementation. To initially qualify the bare diversity of export baskets or the ubiquity of productions no distinction is made among products and among producing countries: only monetary amounts are counted,

[1]Department of Physics of Complex Systems, Weizmann Institute of Science, 7610001 Rehovot, Israel. [2]Institut für Theoretische Physik, Universität Innsbruck, Technikerstraße 21A, 6020 Innsbruck, Austria. [3]Dipartimento di Fisica e Astronomia, Università di Padova, Via Marzolo 8, 35131 Padua, Italy. [4]Sezione INFN Università di Padova, Via Marzolo 8, 35131 Padua, Italy. ✉email: stella@pd.infn.it
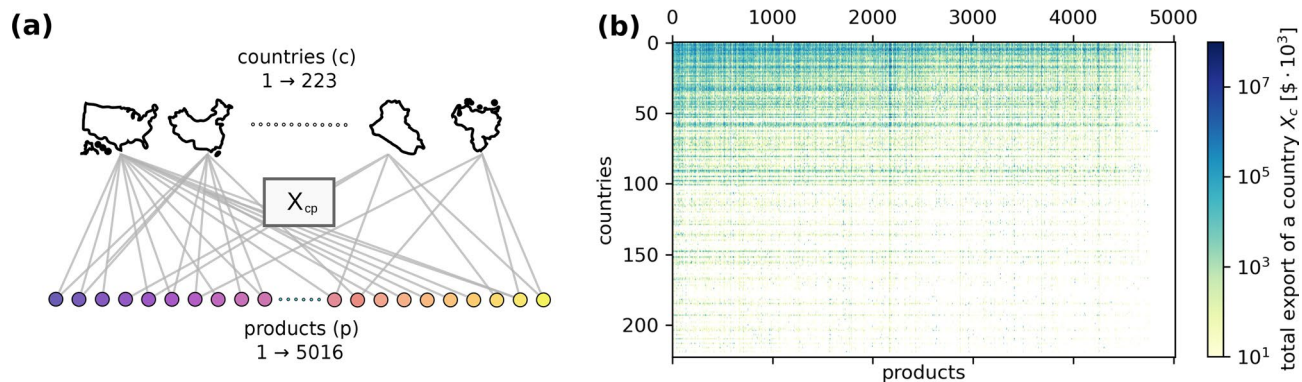
Figure 1. (**a**) Sketch of the bipartite network associated to the weighted bi-adjacency matrix X_{cp}. Every link represents the overall amount (in thousands of current US dollars) of product p exported by a country c towards the rest of the World in a given year. (**b**) X_{cp} matrix. Countries (rows) are sorted from top to bottom with decreasing diversity H_c, while products (columns) are sorted from left to right with decreasing ubiquity H_p. Sorting according to these criteria highlights a nested structure of the bipartite network.

irrespective of the specific export or exporter they refer to. This corresponds to the standard use of the entropy function in the above mentioned fields of application. Distinction among productions and countries is then obtained here by reweighting self-consistently, in terms of the searched measures, the amounts corresponding to products and countries when evaluating the entropy functions. Such self-consistency can be achieved by running an iterative procedure in which bare entropies enter as starting inputs. The final measures searched are then obtained as fixed points of the iterative scheme. While iteration to reach self-consistency is common to all methods proposed so far to obtain complexity measures, in our case its realization within a scheme dictated by the Shannon function guarantees the existence of a fixed point. This feature, and the fact that the iterations reveal the character of contractions rapidly converging to the fixed point, represents an advantage with respect to previous alternative approaches.

The fact that our measure of country competitiveness is expressed as a Shannon entropy function of the weighted percentual shares of the exports in the basket offers also the unique possibility to decompose the same measure into intra- and inter-sectorial contributions, once the shares are summed to represent different sectors of production. This is allowed by a key, characterizing property of the function[17] and opens the possibility to revisit in the spirit of the economic complexity approach some methods of analysis already in use in development economics[16].

Another virtue of our approach is the wide potential applicability, mainly due to the fact that it is based on the universally used Shannon function. It can indeed be used in any problem involving (weighted or not weighted) bipartite networks[21,22], when the searched rankings depend in a simple way on the number and strength of the links connecting the two network's layers. Possible contexts of application range from ecology[23], e.g. the case of plant-pollinator[24] and predator-pray[25] networks, to collaboration networks[26] as well as gene sharing networks in microbial communities[27].

This article is organized as follows. In "Data and bare entropic measures" section we describe how the data on the exports are used for setting up our bipartite network and provide empirical evidence in favor of the use of the Shannon function. "Iterative scheme" section is devoted to the exposition of the iterative algorithm, while "Results" section presents the main results. "Coarse-grained analysis, and intra- vs inter-sectorial contributions" section illustrates an example of intra- and inter-sectorial analysis. Conclusions are in "Conclusion" section.

Data and bare entropic measures

The data we use for the analysis are extracted from the BACI database[28], which is a refined version of the freely accessible COMTRADE database[29]. This contains export data on a country-to-country level covering 21 years between 1995 and 2015. The total number of countries contained in the database is $N_c = 223$ while products are classified in 5016 categories according to the Harmonized System 2007 (HS07)[30], which consists in a 6 digit code of hierarchical nomenclature. Every digit represents a category by which the good is classified, and this category becomes more specific as the number of digits increases. This hierarchical structure consents to naturally coarse-grain the data at different levels, aggregating products sharing the most significant digits. Being not limited by problems of convergence, our analysis will make use of the full information contained in the original dataset by considering all the 6 digits that characterize the HS07 classification with a total of $N_p = 5016$ different product categories. It is precisely the aggregation represented by the four-digit classification with respect to the six-digit one that we will consider below in order to illustrate the decomposition into inter- and intra-sectorial contributions.

A preliminary aggregation of the data is carried over the importing countries, in order to obtain the bi-adjacency matrix $X_{cp}(n)$, representing the total amount of product p exported by a country c in a given year n (Fig. 1). Since the analysis does not mix data from different years, from now on we will drop the year index n and specify the year considered in the text.
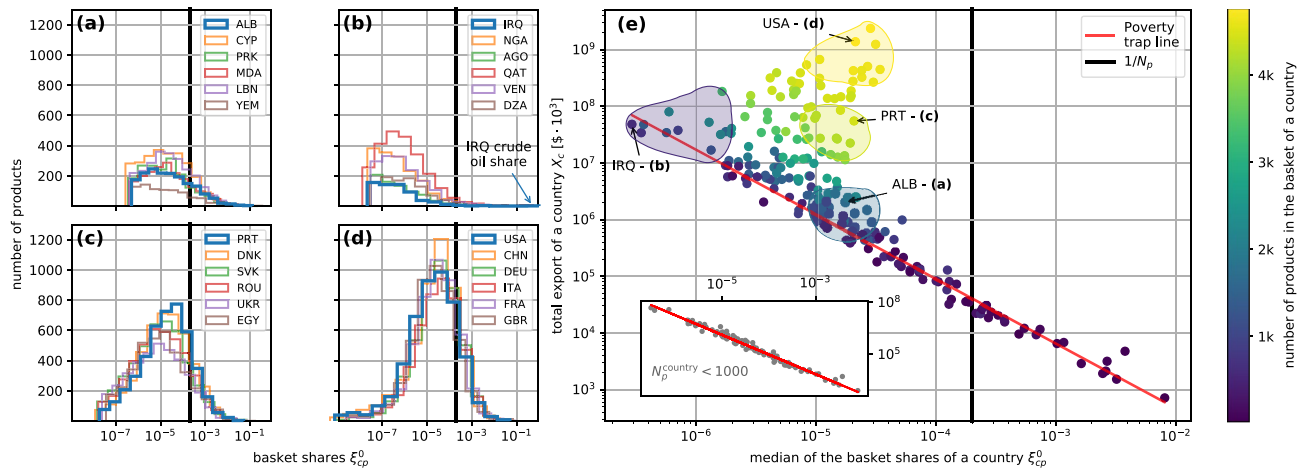
Figure 2. (**a–d**) Examples of histograms of the 2015 basket shares ξ_{cp}^0 for different countries grouped among underdeveloped (**a**), oil dependent (**b**), semi-developed (**c**) and fully developed (**d**). Albania (ALB), Iraq (IRQ), Portugal (PRT) and the USA will serve as representative examples of these categories throughout the text (country codes follow ISO-3166 standard[31]). (**e**) Median of the distribution of the shares vs. total export of a country, colored with respect to the number of exported products. The colored areas show the locations of the four groups of countries whose histograms are reported in (**a–d**). In the inset we highlight the countries exporting less than 1000 products defining a "poverty trap" area. The red line is a linear regression performed over such countries in log-log scale.

The starting point for the construction of our measure is a bare Shannon entropy of the nodes of each layer. Given a set of probabilities $\{p_i\}_{i=1\dots N}$, $\sum_i p_i = 1$, the Shannon function is defined as

$$H = -\sum_{i=0}^{N} p_i \log(p_i) \tag{1}$$

If we think the p_i's as relative occupations of a collection of N states, Eq. (1) expresses the diversity of the corresponding distribution. Indeed, H increases with both the total number of states or entries, N, and with the evenness of the distribution of the p_i-s. For any $N > 1$, H is bounded in the region $[0, \log(N)]$, with the maximum value being reached in the case of exact equipartition $p_i = 1/N$ and the minimum when the occupation is concentrated in one single state.

In the context of the countries-products bipartite network the shares of products in the basket of a given country c can be interpreted as a set of probabilities which altogether provide a measure of the diversification of a country's exports[5]. Something similar holds also for the products: defining for each country a share of an exported product with respect to the worldwide export of that same product p one can represent somehow the ubiquity of that export[5,20]. Similar uses of the Shannon function to estimate diversification are not new in fields like development economics[15,16].

So, the shares $\xi_{cp}^0 = X_{cp}/\sum_{p'} X_{cp'}$ of the products p in the basket of country c can be plugged in Eq. (1), providing us with a "bare" entropic indicator of productive diversification:

$$H_c^0 = -\sum_{p=1}^{N_p} \xi_{cp}^0 \log\left(\xi_{cp}^0\right) \tag{2}$$

In this sum there are of course terms which are zero corresponding to products which do not appear in the basket of the considered country. It is the number of nonzero terms which corresponds to the N-dependence we were discussing in connection with Eq. (1). In Fig. 2a-d we present histograms of the basket shares ξ_{cp}^0 of 2015 in the case of countries included in the four different regions highlighted in Fig. 2e. Countries belonging to the same area exhibit qualitatively similar distributions, regardless of the different number of products exported (histograms are not normalized). Remarkably, we find that such a qualitative classification through the basket shares distributions allows to characterize countries in categories in line with the economic narrative: Fig. 2a reports histograms of underdeveloped countries, including Albania (ALB); Fig. 2b of oil dependent economies, including Iraq (IRQ); Fig. 2c of moderately developed economies, including Portugal (PRT), and Fig. 2d of fully developed nations, including USA. The different total numbers of products exported by the various countries are given by the areas under the corresponding histograms. Up to differences in these numbers the shapes of the histograms are rather similar within each group. Therefore, throughout the rest of the paper, USA, IRQ, PRT and ALB will serve as representative examples for countries characterized by these differently structured economies. We also notice that fully and moderately developed countries like USA and PRT present a relatively narrow peak rather close to the equipartition value of the full range of products $1/N_p$, while both IRQ and ALB show a broader peak, with IRQ's peak being much further away from this equipartition value. Moreover, while for example PRT

and IRQ share a very similar total export $X_c = \sum_p X_{cp} \sim 5 \times 10^{10}\$$, we can clearly see how the distribution are very different: not only the number of exported products (area of the histogram) is much larger for PRT, but also the distribution of basket shares in the case of IRQ is extremely uneven. This is due to the overwhelming dominance of the few oil related exports (highlighted with a blue arrow in Fig. 2b), with total shares holding 98%.

One parameter that also captures and partly quantifies the above considerations is the median of the distributions of the basket shares. With the due precautions connected to the fact that the median of the distribution is also depending on the number of products exported by the country and thus cannot be naively compared to the world equipartition value $1/N_p$, this parameter provides an insight complementary to that of the total export. In Fig. 2e we show how such quantity relates with the total export X_c for each one of the 223 countries of the data-set, always in the year 2015. The color scheme used in Fig. 2e reflects the number of products exported by each country. It is remarkable how all the countries exporting less than 1000 products are found to lie on a line, exhibiting a very high correlation between the overall export of a country and the median of its basket shares. Such behavior defines a sort of "poverty trap" on which are lying the countries whose export depends only on a blind exploitation of the natural resources at disposal. Note that, in spite of its relatively high total export, also Iraq is lying on this line because of its complete dependence on oil exports. Countries that managed to improve their overall economic wealth had to increase both the number of products and the diversification of the total production. This strategy allowed them to detach from the "poverty trap" line in a direction which brought them closer to the ideal equipartition value $1/N_p$.

The above considerations point to basket shares distributions as a key ingredient for the construction of a measure for the productive efficiency of the countries. At the same time, the Shannon entropy function introduced in Eq. (2) qualifies as a natural candidate to represent a single value quantifier of all the "information" contained in such distributions.

Similarly, one can think of an analogous argument also for the layer of the products[5]. For any given product p we can define an ubiquity measure that takes into account both the number of countries that are able to export such product and how evenly its offer on the global market is distributed. To do so, we define an *export* share $\zeta_{cp}^0 = X_{cp}/\sum_{c'} X_{c'p}$ which is normalized with respect to the overall export of that same product at a global level. The bare entropic indicator for the ubiquity of products will therefore be:

$$H_p^0 = -\sum_{c=1}^{N_c} \zeta_{cp}^0 \log\left(\zeta_{cp}^0\right) \tag{3}$$

These quantifiers will be at the basis of our measure constructions.

Iterative scheme

The entropic bare measures of Eqs. (2) and (3), although already rather sound, are not making full use of the information at our disposal. For instance, the formulation of H_c^0 is indifferent to interchanges among products, meaning that swapping the exports of two products would leave the entropy unaltered. Analogously, swapping the exports of two countries for the same product p would preserve the associated ubiquity H_p^0. Nevertheless, each product (and country) should enter differently in the evaluation of the measures. This is also required by the basic assumptions of the economic complexity approach, which through a comparative analysis aims at distinguishing among products and among producing countries. In general, one would intuitively expect "important" products to weight more in a country's wealth indicator. Such importance should be determined by how many developed countries are exporting these products. We can introduce such dependencies by reweighing the shares entering Eqs. (2) and (3) with weights self-consistently related to diversities and ubiquities. The imposition of such self-consistency can be achieved by the construction of an iterative algorithm in which at each step we evaluate finer measures of diversity and ubiquity by using weights determined by the same quantities evaluated at the previous step. A general formulation of such algorithm reads:

$$\begin{cases} H_c^{(k+1)} = -\sum_p \xi_{cp}^{(k)} \log\left(\xi_{cp}^{(k)}\right) \\ H_p^{(k+1)} = -\sum_c \zeta_{cp}^{(k)} \log\left(\zeta_{cp}^{(k)}\right) \end{cases} \tag{4}$$

with shares at the k-th step defined as

$$\xi_{cp}^{(k)} = \frac{X_{cp} f\left(H_p^{(k)}\right)}{\sum_{p'} X_{cp'} f\left(H_{p'}^{(k)}\right)} \; ; \; \zeta_{cp}^{(k)} = \frac{X_{cp} g\left(H_c^{(k)}\right)}{\sum_{c'} X_{c'p} g\left(H_{c'}^{(k)}\right)} \tag{5}$$

Here f and g are two functions that, respectively, take in as argument H_p and H_c. Inspecting Eqs. (4) and (5) one can easily appreciate the role played by the weights introduced to modify the bare shares in the Shannon function. For the sake of argument and to simplify notation, let us look at the f's and the g's in Eq. (5) as to independent variables with labels referring to the corresponding product p ($f(H_p^{(k)}) \to f_p$) or country c ($g(H_c^{(k)}) \to g_c$). Then, let us consider for example a country with a very special basket having one single dominant product p ($X_{cp} \simeq 1$) and all the other products $p' \neq p$ with $X_{cp'} \simeq 0$. Due to the shape of the function $-x\log(x)$ in the interval $[0, 1]$, an increase of f_p would lead to a decrease of $H_c^{(k+1)}$ both because the weighted importance of product p increases (increase of $\xi_{cp}^{(k)}$ towards 1) and because the weighted importance of the other products p' decreases ($\xi_{cp'}^{(k)}$ become closer to 0). On the contrary, a decrease of f_p would lead to an increase of $H_c^{(k+1)}$. So, an unbalanced basket dominated by a single export becomes even less efficient if that export is further valued.
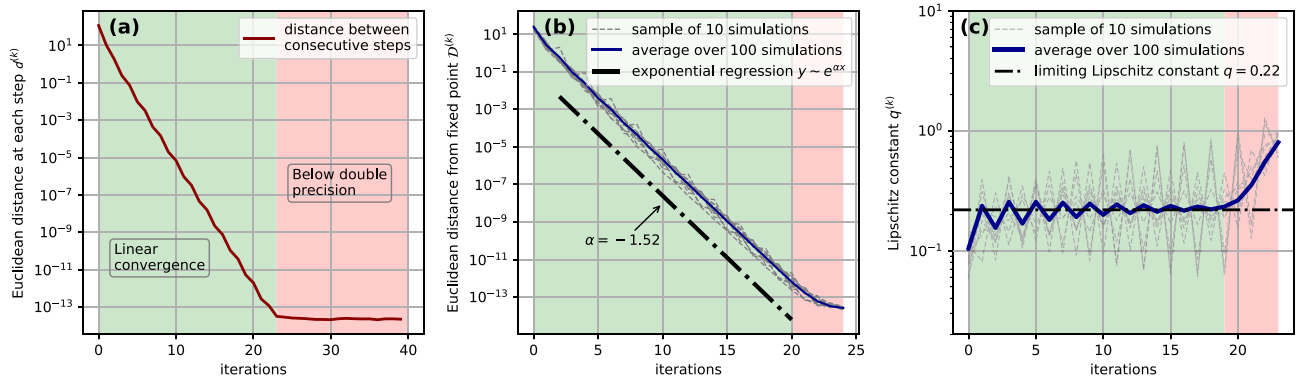
Figure 3. Convergence of the iterative scheme (**a**) Euclidean distance $d^{(k)}$ evaluated between consecutive steps of the iterative algorithm of Eq. (4). It clearly exhibits linearly fast convergence (green area of the plot), while after 23 iterations double precision limit is reached (red zone in the plot). (**b**) Euclidean distance $\mathscr{D}^{(k)}$ from the fixed point of 100 iterations initiated with random initial conditions. Linear convergence to the same fixed point is observed, proving its uniqueness and global convergence of the algorithm. (**c**) Evaluation of the Lipschitz constant q as the ratios of distances from the fixed point of consecutive iterations. A sample of the 100 iterations starting from random initial conditions are colored in grey, whereas in blue we see the average performed over the whole data sample. The dashed black line represents the estimation of the Lipschitz constant obtained by the regression illustrated in the previous panel.

Thus, the logic at the basis of using weighted shares is such to allow distinction among products and to favor an optimal balance among all productions at the same time.

The choice of the functions f and g in Eq. (4) is clearly not unique, and in Ref.[5] a particular form of self-consistency was preliminarily proposed. In any case the choice is dictated by the aims of the analysis. Thus, it needs to reflect the fact that the more a product is ubiquitous the less it should contribute to the diversity of a basket of a country. Also, the more a country is wealthy and developed, the less it should count in determining the ubiquity of a product. Therefore we need to define two functions f and g that invert the concept of diversity and ubiquity, respectively. Thanks to the boundedness of the Shannon entropy function, a simple and convenient way to achieve such inversion is through the following simple linear relations:

$$\begin{aligned} f(H_p) &= \log(N_c) - H_p \\ g(H_c) &= \log(N_p) - H_c \end{aligned} \tag{6}$$

These positive weights, with their continuity, grant existence of a fixed point[32] and in fact stability of the iterative algorithm.

Indeed, the scheme of Eq. (4) can be seen as a map φ of a closed, compact set in itself

$$\varphi : [0, \log N_p]^{N_c} \otimes [0, \log N_c]^{N_p} \to [0, \log N_p]^{N_c} \otimes [0, \log N_c]^{N_p} \tag{7}$$

Such a peculiar feature is guaranteed by the boundedness of the Shannon's entropy function, which (together with the continuity property of φ allows us to apply Brower's fixed point theorem[32] to ultimately prove the existence of a fixed point $\{H_c, H_p\}_{p=1...N_p}^{c=1...N_c}$ for the map φ. In Fig. 3a we show indeed how such fixed point is reached exponentially fast by iterating numerically the scheme of Eq. (4). Evaluating the Euclidean distance between two consecutive steps, defined for every $k \in \mathbb{N}$ as:

$$d^{(k)} = \left(\sum_c \left(H_c^{(k+1)} - H_c^{(k)} \right)^2 + \sum_p \left(H_p^{(k+1)} - H_p^{(k)} \right)^2 \right)^{1/2} \tag{8}$$

Numerically we also tested the uniqueness of such fixed point (to check if the algorithm is globally convergent) by iterating the scheme starting from different randomly chosen initial conditions. We evaluated at each step the distance with respect to the previously obtained fixed point, defined as

$$\mathscr{D}^{(k)} = \sum_c \left(\left(H_c^{(k)} - H_c \right)^2 + \sum_p \left(H_p^{(k)} - H_p \right)^2 \right)^{1/2} \tag{9}$$

and found for it an exponential decay with $k \to \infty$: $\mathscr{D}^{(k)} \propto e^{\alpha k}$ with $\alpha = -1.52$ (Fig. 3b). With such parameter we were therefore able to estimate the Lipschitz constant[33], defined as $q = \lim_{k \to \infty} \mathscr{D}^{(k+1)} / \mathscr{D}^{(k)} = 0.22$ (see Fig. 3c). These estimates allow to classify the algorithm as globally convergent with a linear rate of convergence ($q < 1$ means that the map associated with the algorithm is a contraction).

The scheme therefore always ultimately converges to the same fixed point for the entropies of countries and ubiquities of products, for which the following consistency relations hold
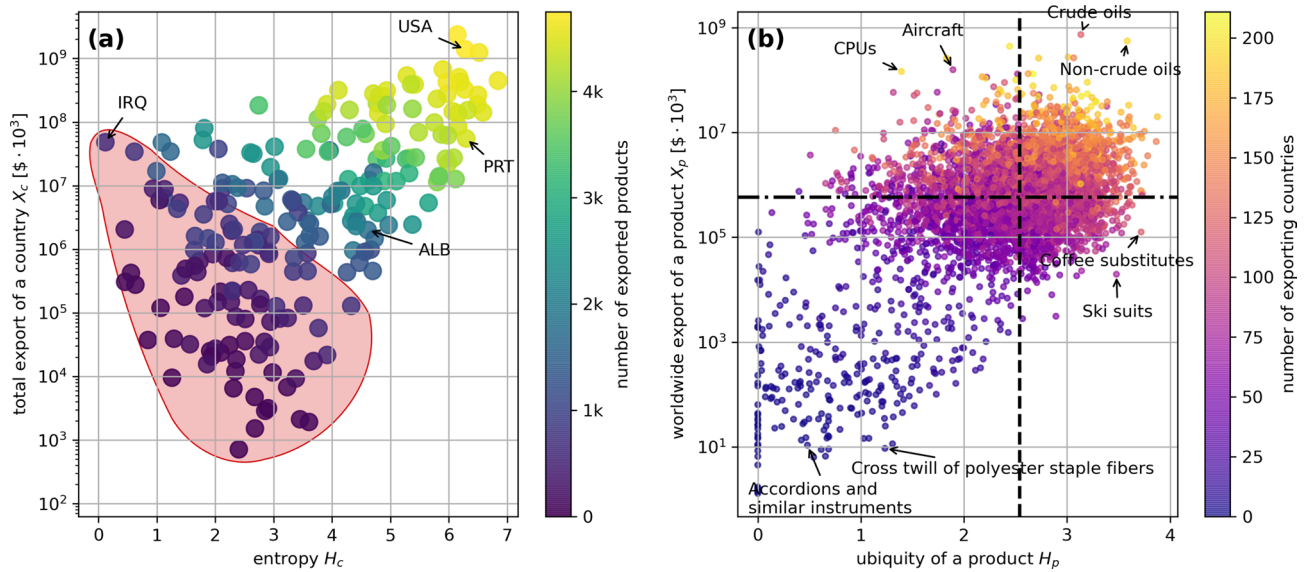
Figure 4. (**a**) Fixed point of the entropic measure H_c for the countries vs. the overall export X_c. The colors used are related to the number of exported products. We can see how the countries that belong to the "poverty trap" line ($N < 1000$) introduced in Fig. 2b are here scattered over a wide area (highlighted in red). (**b**) Ubiquity of the products H_p vs. World aggregated export of that same product. The color scale reflects the number of countries exporting such product. The black lines mark the medians of the ubiquity and of the amount of worldwide export. They allow us to classify products in 4 different main categories.

$$\begin{cases} H_c = -\sum_p \xi_{cp} \log\left(\xi_{cp}\right) \\ H_p = -\sum_c \zeta_{cp} \log\left(\zeta_{cp}\right) \end{cases} \tag{10}$$

where we introduced the weighted shares normalized with respect to a country's export

$$\xi_{cp} = \frac{X_{cp} f\left(H_p\right)}{\sum_{p'} X_{cp'} f\left(H_{p'}\right)} \; ; \; \zeta_{cp} = \frac{X_{cp} g(H_c)}{\sum_{c'} X_{c'p} g(H_{c'})} \tag{11}$$

The iteration procedure, through its fixed point, solves the mathematical problem of providing the solution of the self-consistency conditions in Eqs. (10) and (11).

From an economic point of view, the iteration scheme can be interpreted as an attempt to progressively establish the weights of the factors playing a role in the production network of a good. Trying to establish the relevance of all the intermediate steps that lead to the realization of a product reveals to be an extremely challenging task. Production chains are indeed the result of multiple complex interactions, often involving products realized across different countries at a global scale. Our approach naturally overcomes the limitations induced by analyzing a single national value chain: through exports, all the national productivity chains are effectively interconnected with one another, merging into a global value chain. One therefore needs to take into account such connections, keeping in mind that every country is nevertheless required to offer a highly diversified basket of products in order to retain an adequate self-sustainability as well as the ability to compete as a driving force in the global economy. The iterative scheme endogenously captures the intrinsic relevance of both countries and products, establishing weights for these structural ingredients of the global value chain.

Results

In this section, we present some of the main results emerging from the analysis performed with our algorithm.

In Fig. 4a, we report on the horizontal axis our fixed point entropic measures H_c of the 223 countries for the year 2015. On the vertical axis is the total export of each country in the same year, and the colors vary according to total number of exports. Countries with maximal entropic efficiency are in fact those with the highest total export value on the top right corner. Not surprisingly, also in view of what we anticipated in the second section (Fig. 2b), the countries in this corner are also those with the largest numbers of productions. Here we can identify in particular the four countries pointed out in Fig. 2a–d as representative of four different types of economy. While USA and PRT are in the top right corner, although with different levels of total export and number of products, IRQ is in the top left corner, of countries which are based almost exclusively on export of natural resources. ALB clearly differentiates from IRQ with its higher entropic measure, largely due to the relatively broader distribution of shares in the basket. Notably, this entropic measure successfully captures the distance from the "poverty trap" highlighted in Fig. 2b and here (Fig. 4a) represented by the region in which the countries with less then 1000 exported products lay. The countries in this trap are now spread on a rather large region colored in red. The entropic measure also provides a better estimator of country efficiency than the
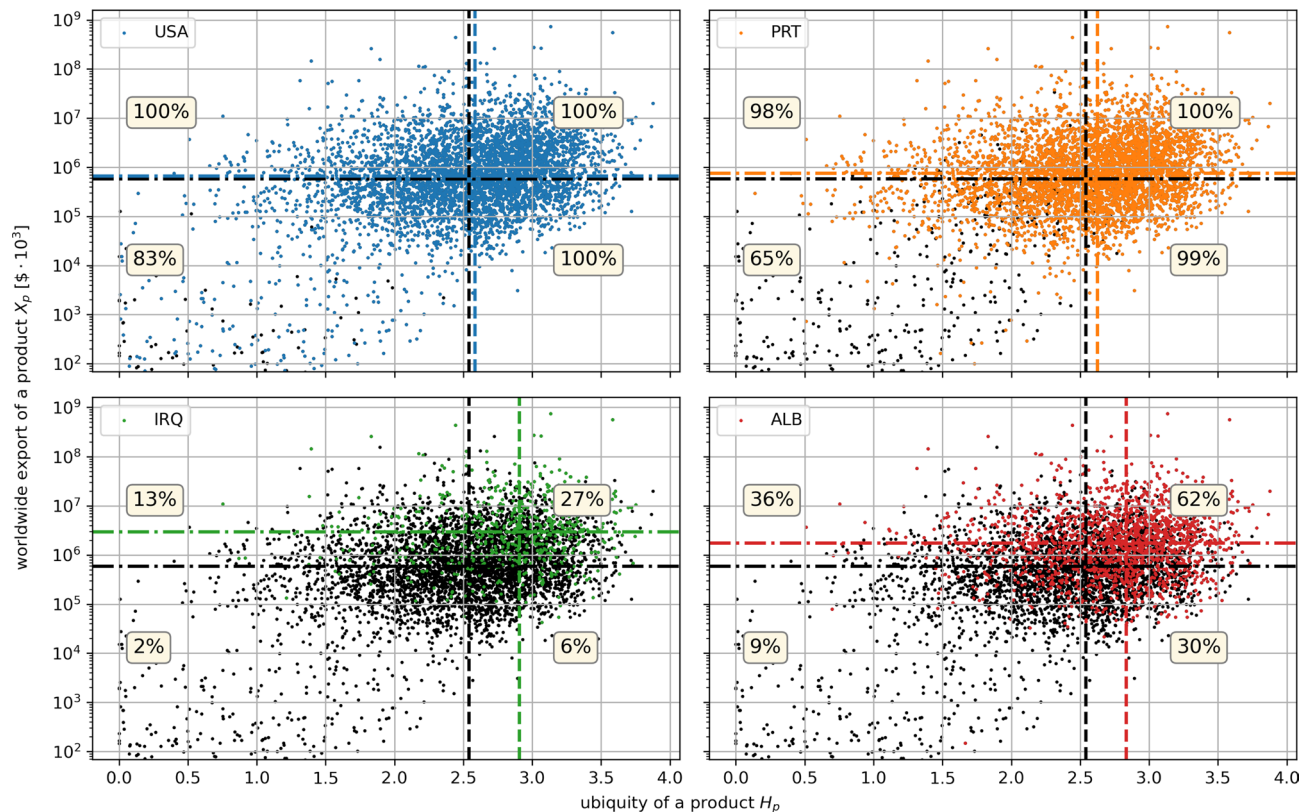
Figure 5. Prouducts exported by USA (top-left), Portugal (top-right), Iraq (bottom-left) and Albania (bottom-right) in the worldwide export vs. ubiquity plane of Fig. 4b. Black dots represent all the existing products, while the colored ones are only those actually exported by each country. It is clear how least developed countries tend to have medians (colored lines) leaning towards more ubiquitous and highly exported products (the black lines are the medians for all the existing products).

median of the bare basket shares distribution, according to which ALB, PRT and USA appear to be equidistant with respect to the equipartition line.

The entropic ubiquities of products are shown in Fig. 4b, where they are reported on the horizontal axis. The year of reference is again 2015. On the vertical axis is reported the worldwide amount of export of each product and the scale of colors is in accordance with the number of exporting countries. The horizontal and vertical lines simply represent median values of the corresponding quantities. In this way one partitions the products in four categories: in the top right corner fall the products with large ubiquity and with large volume of global export. These products are generally exported by a large number of countries, which is often also responsible for the large value of the global exports. Products like crude and non-crude oils are also falling in this category. The products in the top left corner are those with low ubiquity and include computer processor units (CPUs) and aircrafts. In this case, while the amount of global export is often very large, the number of exporting countries is small, as appropriate for very specialized productions. In the right bottom corner the are exports with high ubiquity exported in variable but moderate total amounts by a number of countries definitely lower than that of nations contributing to the upper right corner. Finally, in the lower left corner there are products which are very marginal in the global economy. These products have low ubiquity, small total amount of exports, and few exporting countries.

It is of particular interest to find out how specific countries are positioned with their exports compared to the global situation represented in Fig. 4b. In Fig. 5a–d we do this comparison for USA, PRT, IRQ and ALB. For instance, in the case of USA we reproduce Fig. 4b with all products in black, coloring in blue only those that are actually exported by the USA. Remarkably, we see that almost all points are blue, meaning that the USA are exporting almost all products present in the global basket. To quantitatively appreciate how the exports of the USA are divided in the four sectors identified in Fig. 4b, we report (in blue) the medians pertaining to only the blue dots. The percentages reported in each sector give a measure of the overlap between USA export basket and global basket in that sector. For the USA the only percentage lower than 100% is in the lower left corner.

Also for PRT all the sectors are rather well occupied, but one can already appreciate a slight shift to the right of the horizontal median, indicating an average higher ubiquity of exports with respect to the USA. In the case of IRQ and ALB we see instead a very sensible shift of the medians, and, especially in the case of IRQ considerably low percentages, indicating that these economies are definitely far from reproducing the global export basket. The shifts indicate that such countries are forced to export products that are not only on average more ubiquitous, but also more dominant in the world aggregated basket of exports. This is not surprising, since the

export aggregated at a global level exactly coincides with the worldwide import. This means that poorly developed countries, lacking the possibility to export elaborated products, concentrate their efforts in goods that are relatively easy to produce and highly demanded in the global panorama.

A peculiarity that emerges from this analysis is how developed countries seem to export all kinds of products, including those with extremely high ubiquity. Least developed countries, on the other hand, are lacking an enormous percentage of such products. This suggests that a country, in order to increase its development, cannot aim to just produce the most complex and/or heavily exported products on the market. There are indeed articulated constraints linked to the possibility of producing such goods. Such constraints in many cases remain even hidden (related with the so called "intangibles"). This analysis sheds light on how the lack of apparently irrelevant products in the basket might heavily impact the overall development itself of a country down the line. In other words, we understood how a developed country can hardly refrain from continuing to keep its exports as diverse as possible, including ubiquitous and poorly demanded products.

It is interesting to compare our rankings with those one obtains with the approach of Ref.[2]. Here the rankings of that approach were obtained by use of the full 6-digit code of the HS7 product classification For sure the fitness method, at this level of resolution can be anticipated to suffer of convergence problems[11], which in fact were already encountered at the four digit level[10]. Instability issues of the algorithm impose to cut somehow arbitrarily the number of iterations in this case. The problem arises from the implementation of non linearity in the algorithm[10], which leads some products' measures (therein called *complexity*) to progressively approach the limiting value 0. Such behavior denotes an instability of the scheme, extremely marked for the complexities but less visible for the fitness measure of countries. Indeed, evaluating the Spearman correlator[34] between the countries' fitness F_c resulting after hundred iterations and our diversity H_c of Eq. (10) yields the rather high value $\rho_S\{H_c, F_c\} = 0.94$ (6 digits code of the HS07 classification, year 2015). The Spearman correlator between products' complexity Q_p and our ubiquity H_p of Eq. (10) yields $\rho_S\{H_p, Q_p\} = -0.32$. The minus sign comes from the fact that ubiquity is conceptually an inverse of complexity. The remarkable distance of the obtained correlation from the value -1 indicates that the two approaches give rather different rankings for the products.

Coarse-grained analysis, and intra- vs inter-sectorial contributions

The entropy function satisfies a special summation rule when clustering states[15,17]. The HS07 classification of products[30] is intrinsically structured as a multilayered nomenclature, where the leftmost digits are shared by products belonging to a more general macro-category. So far our analysis was carried out at the finest possible level of details offered by the dataset, consisting in 6 numerical digits identifying a total of 5016 product categories. According to the convention of the HS07 classification, one can group together products sharing the same 4 leftmost digits resulting in a total of 1241 macro-categories P. Referring to the fixed point shares ξ_{cp} of Eq. (10) we introduce the coarse-grained shares $\xi_{cP} = \sum_{p \in P} \xi_{cp}$, for which one can evaluate a *coarse-grained* Shannon entropy:

$$H_c^{\text{CG}} = -\sum_P \xi_{cP} \log\left(\xi_{cP}\right) \tag{12}$$

One can easily see that the inequality $H_c > H_c^{\text{CG}}$ holds, with the equality holding only in the case in which for every coarse-grained category P there is one and only one fine-grained category p. The difference between the two entropies is indeed related to the average *intra-sectorial* entropy of a macro-category P defined as

$$H_{cP} = -\sum_{p \in P} \frac{\xi_{cp}}{\xi_{cP}} \log\left(\frac{\xi_{cp}}{\xi_{cP}}\right). \tag{13}$$

Straightforward calculations show that the difference between the two level entropies is the following weighted average:

$$\Delta H_c = H_c - H_c^{\text{CG}} = \sum_P \xi_{cP} H_{cP}. \tag{14}$$

This summation rule allows us to regard the coarse-grained entropy of Eq. (12) as *inter-sectorial* and ΔH_c in Eq. (14) as *intra-sectorial* contributions to the total entropic measure of country c. As we see in Fig. 6a, the intra-sectorial contribution is more substantial for countries with high entropic measure, indicating the importance of organization within sectors. In Fig. 6b, we report also the total export of each country as a function of ΔH_c. One sees that the countries with the largest export, both in terms of total amount, and in terms of number of products, are those with highest intra-sectorial contribution. Germany (DEU) reaches the top of the list in this case. At the same time, most of the countries belonging to the "poverty trap" show a low ΔH_c, which is another indication of the scarce inner organization of their economic structure.

One can of course perform intra- and inter-secorial decompositions also with reference to larger sectors of the economies. Such decompositions were already considered in the economic literature with reference to bare entropic indicators[16]. This opens novel perspectives of analysis within the complexity approach.

Conclusions

In this report we showed how the Shannon entropy function can be used to develop a consistent and rapidly-convergent method for the evaluation of economic complexity measures. In view of the universally accepted meaning of the Shannon function, our construction gives concrete support to the expectation that diversity can be assumed as a basic ingredient of such measures.
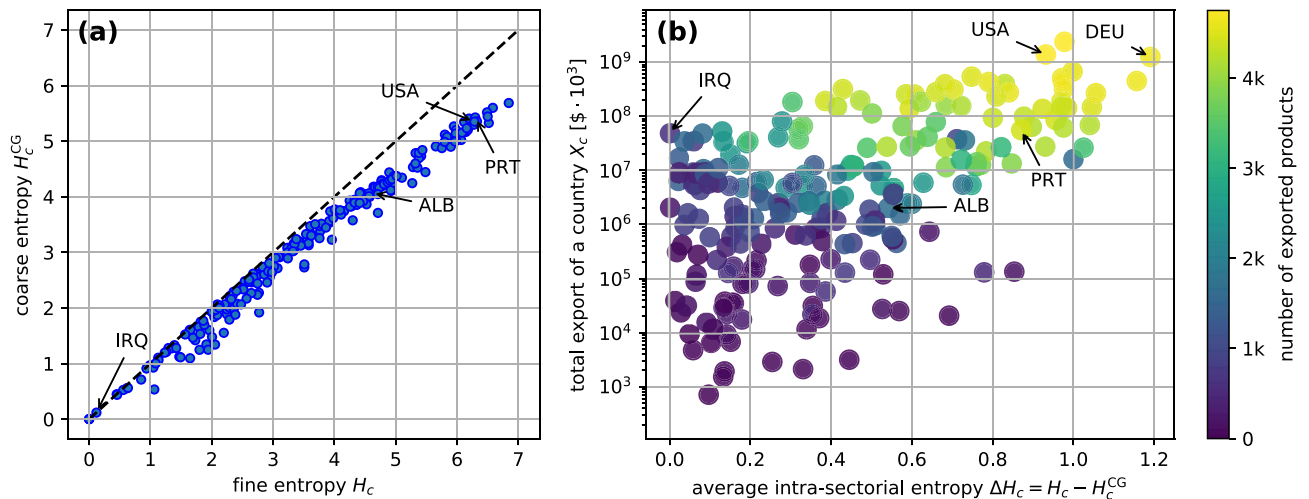
Figure 6. (**a**) Comparison between the fine-grained (6 digit HS07 classification) entropies H_c of the countries for Eq. (10) and the coarse-grained (4 digits HS07 classification) entropies H_c^{CG} of Eq. (12). The quadrant bisector marks the inequality $H_c > H_c^{CG}$ relating the two. (**b**) Average intra-sectorial entropy ΔH_c of Eq. (14) plotted against the overall export of a country X_c. Germany (DEU) emerges as the country with the highest ΔH_c, implying an extremely articulated structure of the export shares. The gradient of the color of every point is related to the overall number of exported products by each country in the 6 digit HS07 classification.

From a mathematical and numerical point of view, the success of our approach relies on the continuity and boundedness of this function, which guarantees both the existence of and the convergence to a fixed point in our iterative scheme. Here our method was proven to be effective on a weighted bipartite network between two layers composed of 5016 (products) and 223 (countries) nodes, and with the links' weight spanning values across more than 9 decades. The proven stability of our algorithm therefore acquires additional significance in the perspective of more general applications, giving confidence to obtain consistent and meaningful results also outside the particular context and database considered here.

Results show that, using our refined entropic measure, it is possible to establish a meaningful and unambiguous ranking of countries, with the most developed ones also scoring the highest entropic measures. On the other hand, countries that have a poorly developed economy or simply rely on the exploitation of their raw materials are characterized by low values of our entropic index. In a similar fashion, products can be ranked according to their ubiquity. The more ubiquitous is a product, the more it is exported (or imported) in the global market. Developed countries characterize themselves as exporter of products in the whole range of ubiquities.

Another key mathematical property is the possibility to decompose the function into parts which can be ascribed to intra- and inter-sectorial contributions[17,18,20]. This gives the possibility to quantitatively analyze the complexity measure of countries in terms of the interplay among different categories in which one may wish to partition the production, and in terms of the the importance of each individual category. It emerges that developed countries are the ones displaying the largest intra-sectorial entropies.

While the analysis presented here referred only to a particular year of the dataset, a comparison of the results for different years opens to an analysis of the evolution dynamics of the complexity measures. This would allow to argue further information both on the different growth potential of each country and on the evolution of the hierarchy of importance of various products in the global economy. A preliminar analysis along these lines[5,20] has already highlighted a strong bond connecting the entropic diversity of a country and the hierarchy of products to a dynamical model regulating the yearly evolution of single exports for every country. For the time evolution of the basket compositions, the present authors have recently developed a statistical mechanics model[4–6], calibrated on the same database used here, which clarifies the non-equilibrium character of its dynamics. The combined use of this model with the entropic complexity analysis offers in perspective possibilities of predictions going beyond standard regression analysis. As we showed in "Coarse-grained analysis, and intra- vs inter-sectorial contributions", the entropic measures allow to keep track of the whole information of the most detailed data, without dispersing it along a coarse-graining process. The control of the effects of coarse-graining on the entropy production is indeed one major issue in the general context of statistical mechanics out of equilibrium[20,35,36].

Data availability

The data used in this work are extracted from the BACI database[28], which is a refined version of the freely accessible COMDTRADE database[29] redacted by the United Nations. Reference tables of the Harmonized System 2007 nomenclature and ISO-3166 country codes can be found respectively at Refs.[30,31].

References

1. Hidalgo, C. & Hausmann, R. The building blocks of economic complexity. *Proc. Natl. Acad. Soc. USA* **106**, 10570. https://doi.org/10.1073/pnas.0900943106 (2009).
2. Tacchella, A., Cristelli, M., Caldarelli, G., Gabrielli, A. & Pietronero, L. A new metrics for countries' fitness and products' complexity. *Sci. Rep.* **2**, 723. https://doi.org/10.1038/srep00723 (2012).
3. Hausmann, R. *et al. The Atlas of Economic Complexity: Mapping Paths to Prosperity* (Mit Press, ***, 2014).
4. Caraglio, M., Baldovin, F. & Stella, A. L. Export dynamics as an optimal growth problem in the network of global economy. *Sci. Rep.* https://doi.org/10.1038/srep31461 *(2016)*.
5. Teza, G., Caraglio, M. & Stella, A. L. Growth dynamics and complexity of national economies in the global trade network. *Sci. Rep.* **8**, 1–8. https://doi.org/10.1038/s41598-018-33659-6 (2018).
6. Teza, G., Caraglio, M. & Stella, A. L. Data driven approach to the dynamics of import and export of g7 countries. *Entropy* **20**, 735. https://doi.org/10.3390/e20100735 (2018).
7. Grossman, G. & Helpman, E. Quality ladders in the theory of growth. *Rev. Econ. Stud.* **58**, 43. https://doi.org/10.2307/2298044 (1991).
8. Aghion, P. & Howitt, P. *Quality Ladders in the Theory of Growth* (MIT Press, ***, 1998).
9. Kemp-Benedict, E. An interpretation and critique of the method of reflections. *Munich Personal RePEc Archive.* https://mpra.ub.uni-muenchen.de/60705/ (2014).
10. Morrison, G. *et al.* On economic complexity and the fitness of nations. *Sci. Rep.* **7**, 1–11. https://doi.org/10.1038/s41598-017-14603-6 (2017).
11. Servedio, V., Buttà, P., Mazzilli, D., Tacchella, A. & Pietronero, L. A new and stable estimation method of country economic fitness and product complexity. *Entropy* **20**, 783. https://doi.org/10.3390/e20100783 (2018).
12. Balassa, B. Trade liberalisation and "revealed" comparative advantage 1. *Manchester Sch.* **33**, 99–123 (1965).
13. Mealy, P., Farmer, J. D. & Teytelboym, A. Interpreting economic complexity. *Sci. Adv.* https://doi.org/10.1126/sciadv.aau1705 *(2019)*.
14. Spellerberg, I. F. & Fedor, P. J. A tribute to claude shannon (1916–2001) and a plea for more rigorous use of species richness, species diversity and the 'shannon-wiener' index. *Glob. Ecol. Biogeogr.* **12**, 177–179. https://doi.org/10.1073/pnas.09009431060 (2003).
15. Jacquemin, A. P. & Berry, C. H. Entropy measure of diversification and corporate growth. *J. Ind. Econ.* **XXVII**, 359–369. https://doi.org/10.2307/2097958 (1979).
16. Saviotti, P. P. & Frenken, K. Export variety and the economic performance of countries. *J. Evol. Econ.* **18**, 201–218. https://doi.org/10.1073/pnas.09009431061 (2008).
17. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423. https://doi.org/10.1073/pnas.09009431062 (1948).
18. Weaver, W. The mathematics of communication. *Sci. Am.* **181**, 11–15. https://doi.org/10.1073/pnas.09009431063 (1949).
19. Jost, L. Entropy and diversity. *Oikos* **113**, 363–375. https://doi.org/10.1073/pnas.09009431064 (2006).
20. Teza, G. Out of equilibrium dynamics: From an entropy of the growth to the growth of entropy production. Ph.D. thesis, University of Padova (2020).
21. Guillaume, J.-L. & Latapy, M. Bipartite structure of all complex networks. *Inf. Process. Lett.* **90**, 215–221. https://doi.org/10.1073/pnas.09009431065 (2004).
22. Newman, M. *Networks: An Introduction* (Oxford University Press, ***, 2010).
23. Dormann, C. F., Fründ, J., Blüthgen, N. & Gruber, B. Indices, graphs and null models: Analyzing bipartite ecological networks. *Open Ecol. J.* https://doi.org/10.2174/1874213000902010007 *(2009)*.
24. Chacoff, N. P. *et al.* Evaluating sampling completeness in a desert plant-pollinator network. *J. Anim. Ecol.* **81**, 190–200. https://doi.org/10.1111/j.1365-2656.2011.01883.x (2012).
25. Allesina, S. & Pascual, M. Network structure, predator-prey modules, and stability in large food webs. *Theoret. Ecol.* **1**, 55–64. https://doi.org/10.1007/s12080-007-0007-8 (2008).
26. Ramasco, J. J., Dorogovtsev, S. N. & Pastor-Satorras, R. Self-organization of collaboration networks. *Phys. Rev. E* **70**, 036106. https://doi.org/10.1073/pnas.09009431068 (2004).
27. Corel, E. *et al.* Bipartite network analysis of gene sharings in the microbial world. *Mol. Biol. Evol.* **35**, 899–913. https://doi.org/10.1073/pnas.09009431069 (2018).
28. Gaulier, G. & Zignago, S. Baci: International trade database at the product-level. the 1994-2007 version. Working Papers 2010-23, CEPII (2010). https://doi.org/10.2139/ssrn.1994500.
29. UN. *Commodity Trade Statistics Database* (2017). https://comtrade.un.org. Accessed 24 June 2020.
30. WCO. *Harmonic System Nomenclature* (2007). http://www.wcoomd.org. Accessed 24 June 2020.
31. ISO. *ISO 3166 Country Codes* (2021). https://www.iso.org/iso-3166-country-codes.html. Accessed 24 June 2020.
32. Kellogg, R. B., Li, T.-Y. & Yorke, J. A constructive proof of the brouwer fixed-point theorem and computational results. *SIAM J. Numer. Anal.* **13**, 473–483. https://doi.org/10.1038/srep007233 (1976).
33. Evans, L. C. & Gariepy, R. F. *Measure Theory and Fine Properties of Functions* (CRC Press, ***, 2015).
34. Hogg, R. & Craig, A. *Introduction to Mathematical Statistics* (Macmillan, ***, 1995).
35. Seifert, U. From stochastic thermodynamics to thermodynamic inference. *Annu. Rev. Condensed Matter Phys.* **10**, 171–192. https://doi.org/10.1146/annurev-conmatphys-031218-013554 (2019).
36. Teza, G. & Stella, A. L. Exact coarse graining preserves entropy production out of equilibrium. *Phys. Rev. Lett.* **125**, 110601. https://doi.org/10.1038/srep007235 (2020).

Acknowledgements

Author contributions

G.T. collected and organized the data and carried out the numerical analyses. G.T., M.C. and A.L.S. devised the research, analyzed the data and wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.L.S.

Reprints and permissions information is available at www.nature.com/reprints.