SUPPLEMENTARY MATERIAL FOR:

THE BUILDING BLOCKS OF ECONOMIC COMPLEXITY

Cesar A. Hidalgo, Ricardo Hausmann
Center for International Development and Harvard Kennedy School, Harvard University

TABLE OF CONTENTS

SECTION 1: SOURCE DATA

All of the figures presented in the main text of this paper were constructed using International trade data taken from Feenstra, Lipsey, Deng, Ma and Mo's "World Trade Flows: 1962-2000" dataset. This dataset consists of imports and exports both by country of origin and by destination, with products disaggregated to the SITC revision 4, four-digit level. The authors built this dataset using the United Nations COMTRADE database. The authors cleaned that dataset by calculating exports using the records of the importing country, when available, assuming that data on imports is more accurate than data from exporters. This is likely, as imports are more tightly controlled in order to enforce safety standards and collect customs fees. In addition, the authors correct the UN data for flows to and from the United States, Hong Kong, and China. We focus only on export data and do not disaggregate by country of destination. More information on this dataset can be found in NBER Working Paper #11040, and the dataset itself is available at www.nber.org/data. and http://cid.econ.ucdavis.edu/data/undata/undata.html

We checked the validity of our results by using two additional datasets: COMTRADE classified according to the Harmonized System at the 4-digit level (1241 products, 103 countries) and the North American Industry Classification System (NAICS) (318 products, 150 countries). We found that our results are not affected by the use of data at these different levels of aggregation. We chose to work with the Feenstra dataset because, of the three datasets available, it is the one only one that has been cleaned and checked thoroughly as part of a dedicated research project.

The labor data used to construct figure 2d was downloaded from the US Bureau of Labor and Statistics at http://www.bls.gov/data/

SECTION 2: REVEALED COMPARATIVE ADVANTAGE (RCA)

One way to empirically estimate whether a country is a significant exporter of a product is to calculate the Revealed Comparative Advantage (*RCA*) that that country has in a particular product. *RCA* is a measure constructed to inform whether a country's share of a product's world market, is larger or smaller than the product's share of the entire world market. Mathematically, we can rewrite the above sentence by introducing S_{cp}, as the share that country c has of the world market for product p, and T_p as the total share of product p of the world market. Using this notation, RCA can be written as

$$RCA_{cp} = S_{cp} / T_p \qquad (1)$$

where

$$T_p = \sum_c S_{cp} \qquad (2)$$

RCA CUTOFFS, EXPORTS AND COUNTRIES' LEVEL OF DIVERSIFICATION

The natural cutoff used to determine whether a country has revealed comparative advantage in a product is $RCA \geq 1$. At this point the country's share of that product's market is equal or larger than the product's share of the world market. The benchmark here is a world in which countries export an amount of each product equal to the share of that product in the world market times the size of its economy.

From an empirical perspective, we can study the number of products ($k_{c,0}$) for which a country has *RCA* as a function of the *RCA* cutoff. By performing this exercise we find that the RCA_{cp}=1 cutoff lies on the phase transition of a softened step function (Figure S1).

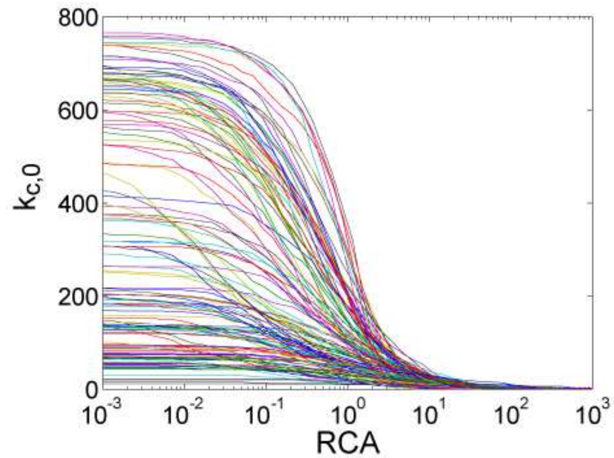

Fig S 1 Diversification ($k_{c,0}$) as a function of the RCA cutoff for all countries in the study

What is interesting about looking at $k_{c,0}(RCA)$ from this empirical perspective is that we can see that there are a few countries that had exports in almost all of the 772 products exported in the year 2000. For example, Germany exported 758 products with an $RCA \geq 0.01$, and 707 products with $RCA \geq 0.1$, a profile similar to that of other industrialized countries like the U.K., U.S.A and Italy. Hence lowering the RCA threshold shows that industrialized countries manufacture and export products in almost all of the SITC-4 categories, and that specialization patterns are empirically driven by the lack of diversification of less developed countries, rather than by the absence of more productive economies in comparatively less sophisticated sectors.

SECTION 3: THE COUNTRY-PRODUCT NETWORK

Fig S 2 shows a simple visualization of the country product network for the year 2000 in which countries are located at the center of the figure and products are grouped into root SITC-4 categories along the edges of the image. This network consists of 129 countries, 772 products and 13,470 links connecting countries and products when $RCA_{cp} \geq 1$. The large number of links in the network limits our ability to create a useful visualization of the entire set of connections.

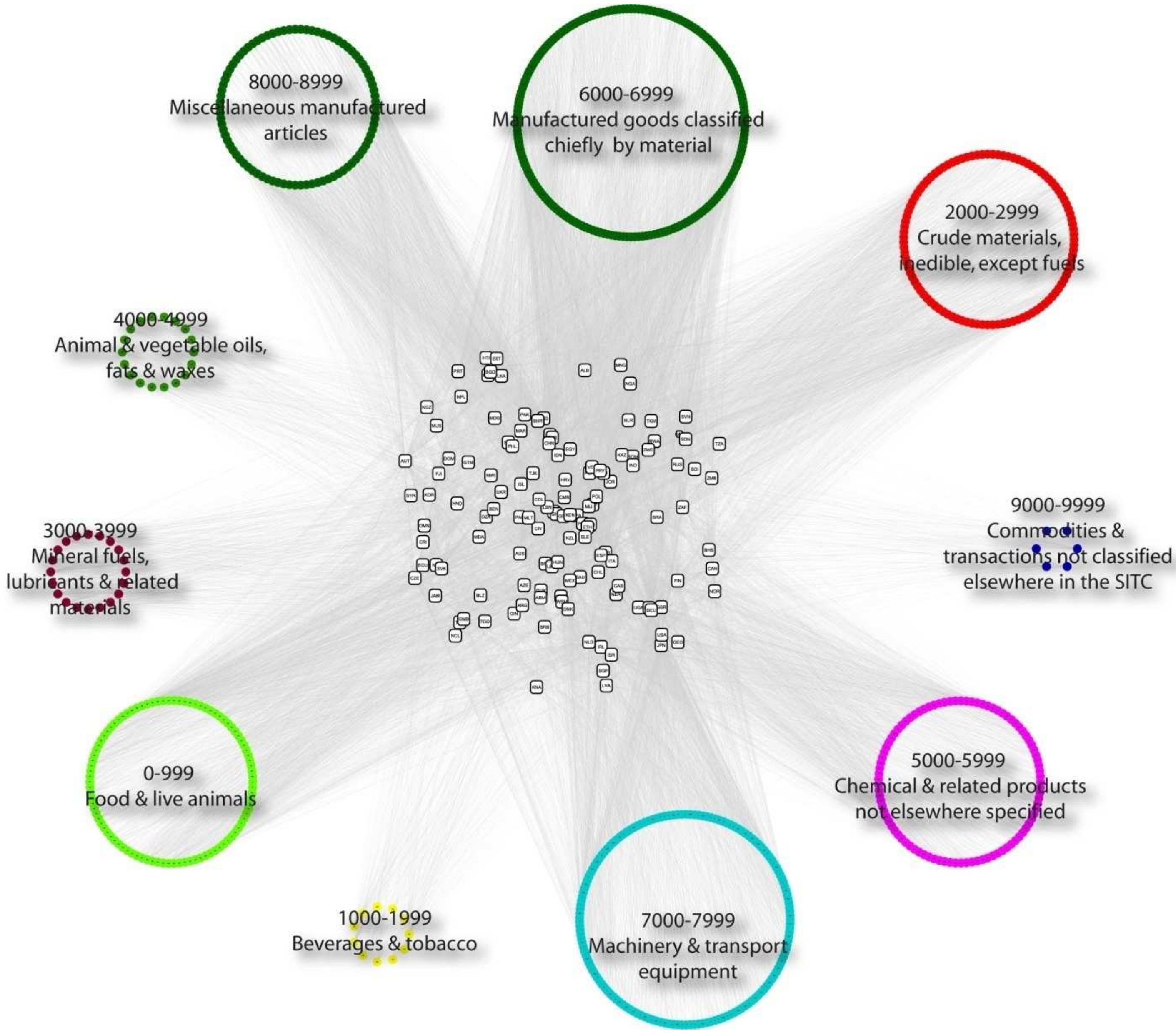

Fig S 2 Visualization of the country product network in which all exports with an RCA>1 are shown.

SECTION 4: BIPARTITE NETWORK ANALYSIS

A bipartite graph or network is a set of nodes and links in which nodes can be separated into two groups, or partitions, such that links only connect nodes in different partitions. While in principle many networks can be separated into different partitions (for example every tree is a bipartite graph), here we concentrate on examples that are bipartite, by definition, rather than as a property. One example of naturally occurring bipartite networks are publication networks, where nodes are researchers and papers, and links connect researchers to the papers they have authored. Another example is the movie-actor network in which nodes are actors and movies, and links connect actors to the movies in which they have starred..

With the exception of a few studies [[1,2,3,4]], bipartite networks have mostly been investigated by projecting the network into one of its partitions [[5,6,7,8,9,10,11,12,13,14]], typically by considering nodes to be connected if they share a neighbor in the opposite partition [[5,6,7,8,9,10,11,12,13,14]]. For example, co-authorship networks link scientists that have co-authored one or more papers [[8,9,10,11]], whereas movie-actor networks connect actors that have appeared together in one or more movies.

While valuable information can be obtained from these projections, there is important information that is left out by reducing the bipartite network into either one of its partitions, regardless of the sophistication of the projection method. Here we present a method to characterize the structure of a bipartite network by iteratively considering the properties of neighboring nodes.

THE METHOD OF REFLECTIONS

In this section we explain in detail the method of reflections as a general technique to study the structure of bipartite networks. To shorten the math we adopt a different notation than the one used for the particular example of countries and products. Going forward, we indicate all variables that are related to nodes in each partition by either Latin or Greek characters.

Consider a bipartite network *M* described by the adjacency matrix $M_{a\alpha}$, where $M_{a\alpha}$ =1 if node *a* is connected to node α and zero otherwise.

We define the method of reflections as the recursive set of observables

$$k_{a,N} = \frac{1}{k_{a,0}} \sum_{\alpha} M_{a\alpha} \kappa_{\alpha,N-1} \tag{3}$$

$$\kappa_{\alpha,N} = \frac{1}{\kappa_{\alpha,0}} \sum_{a} M_{a\alpha} k_{a,N-1} \tag{4}$$

for *n*>0,with

$$k_{a,0} = \sum_{\alpha} M_{a\alpha} \tag{5}$$

$$\kappa_{\alpha,0} = \sum_{a} M_{a\alpha} \tag{6}$$

Following these definitions, the degree of nodes in the bipartite network is given by k_0 and κ_0 (in this notation we can drop the *a* and α indices when referring to the general concept described by the variable as the alphabet already indicates if the variables refers to one partition or the other –countries or products-). In the example of the main text these variables are the diversification ($k_{a,0}$) of countries and the ubiquity ($k_{p,0}$) of products. Following from (3) and (4), the average ubiquity of a country's exports is given by k_1 whereas the average diversification of a product's exporters is given by κ_1. The recursive nature of the method of reflections allows us to characterize the structure of the bipartite network by defining *N* variables for each one of its partitions. For example, continuing the characterization of the country-product network into a third layer of analysis in which k_2, the average κ_1 of a country's exports, and κ_2 ,the average k_1 of a product's exporter, is considered, allows us to characterize countries and products through a three dimensional phase space spanned by k_0 , k_1 , k_2 and κ_0 , κ_1 , κ_2 .

In principle we can use the method of reflections to characterize countries and products by *N* variables. The method of reflections can be generalized by choosing different values for k_0 and κ_0 and iterating over them using (3) and (4). In fact, the measure of product sophistications PRODY [15] can be seen as a special case of the method of reflections in which $k_{a,0}$ is the GDP(PPP) of a country and $M_{a\alpha}$ is a matrix of RCAs. In such a case then PRODY=$k_{\alpha,1}$. When these variables were constructed, however, the authors were not aware that their methods were combining income information with the structure of a bipartite network.

THE VARIABLES FOR THE FIRST THREE LEVELS

Table S 1 shows how we interpret the first three pairs of variables describing the country-product network through the method of reflections:

Definition	Working Name	**Description:** Short summary Question Form
$k_{a,0}$	Diversification	Number of products exported by country a. How many products are exported by country *a*?
$\kappa_{\alpha,0}$	Ubiquity	Number of countries exporting product α. How many countries export product α?
$k_{a,1}$	$k_{c,1}$	Average ubiquity of the products exported by country *a*. How common are the products exported by country *a*?
$\kappa_{\alpha,1}$	$k_{p,1}$	Average diversification of the countries exporting product α. How diversified are the countries that export product α?
$k_{a,2}$	$k_{c,2}$	Average diversification of countries with an export basket similar to country *a* How diversified are countries exporting goods similar to those of country *a*?
$\kappa_{\alpha,2}$	$k_{p,2}$	Average ubiquity of the products exported by countries that export product α. How ubiquitous are the products exported by product's α exporters?

Table S 1 Interpretation of the bipartite network description obtained from the method of reflections.

INTERPRETING HIGHER REFLECTIONS

As we iterate the method of reflections, it becomes increasingly harder to interpret the variables generated by it. We can gain insight into what higher reflection variables stand for by analytically solving the recursion formulas presented in (3)-(6). Analytically solving the recursion requires us to be able to express $\vec{k}_N$ and $\vec{\kappa}_N$ as a function of the initial conditions, $\vec{k}_0$ and $\vec{\kappa}_0$. Mathematically (3)-(4) we search for solutions of the form:

$$k_{a,N} = \sum_b C_{ab,N}(\vec{k}_0, \vec{\kappa}_0) k_{b,0} \quad , \quad \kappa_{\alpha,N} = \sum_\beta C_{\alpha\beta,N}(\vec{k}_0, \vec{\kappa}_0) \kappa_{\beta,0} \tag{7}$$

To illustrate this we calculate the elements $\vec{k}_2$ as an example. According to the definitions of the method shown in (3)-(6) the elements of $\vec{k}_2$ can be expressed as:

$$k_{a,2} = \frac{1}{k_{a,0}} \sum_\alpha M_{a\alpha} \kappa_{\alpha,1} = \frac{1}{k_{a,0}} \sum_{\{a\}_\alpha} \kappa_{\alpha,1} \tag{8}$$

Where $\{a\}_\alpha$ is the set of the α neighbors of *a*. We can use (4) to rewrite (8) as

$$k_{a,2} = \frac{1}{k_{a,0}} \sum_{\{a\}_\alpha} \frac{1}{\kappa_{\alpha,0}} \sum_{\{\alpha\}_b} k_{b,0} \tag{9}$$

Which can be taken into the form (7) by permuting the sums and changing the index of the first summation to a sum over the second neighbors of *a,* and the index of the second summation to a sum over the neighbors of *a* and *b*.

$$k_{a,2} = \frac{1}{k_{a,0}} \sum_{\{\{a\}\}_b} \sum_{\{a \cap b\}_\alpha} \frac{1}{\kappa_{\alpha,0}} k_{b,0} \tag{10}$$

Which satisfies the form presented in (7) with

$$C_{ab,2}(\vec{k}_0, \vec{\kappa}_0) = \frac{1}{k_{a,0}} \sum_{\{a \cap b\}_\alpha} \frac{1}{\kappa_{\alpha,0}} \tag{11}$$

We can interpret $k_{a,2}$ from the form presented in (10) by noticing that $k_{a,2}$ is a linear combination of the elements of $\vec{k}_0$ with coefficients given by product of the degrees of all nodes lying in the path connecting nodes *a* and *b,* including node a but not node b. Hence the coefficients $C_{ab,2}(\vec{k}_0, \vec{\kappa}_0)$ can be interpreted as the probability that a random walker that started at *a* ends up at *b* after two steps.

The random walker interpretation of the method of reflections is true not only for $\vec{k}_2$ but for any *N*. Fig S 3 shows an example of a three node network in which some of the coefficients associated with *N=4* are presented explicitly. Hence the method of reflections is a way to express the properties of a node in a network as a combination of the properties of all its neighbors, the coefficients of the linear combination being the probability that two nodes are connected by a random walker after *N* steps.

The coefficients of the expansion can be interpreted as a measure of similarity between the nodes in the network, which is context dependent, as what matters in the expansion is the relative weight of these coefficients when compared to each other.

$$k_{a,N} = \sum_b C_{ab,N}(\vec{k}_0, \vec{\kappa}_0) k_{b,0}$$

$$= k_a \left(\frac{1}{k_b}\frac{1}{\kappa_\beta}\frac{1}{k_a}\frac{1}{\kappa_\alpha} + \ldots + \frac{1}{k_b}\frac{1}{\kappa_\gamma}\frac{1}{k_b}\frac{1}{\kappa_\gamma} + \ldots\right) + k_b(\ldots.) + k_c(\ldots)$$

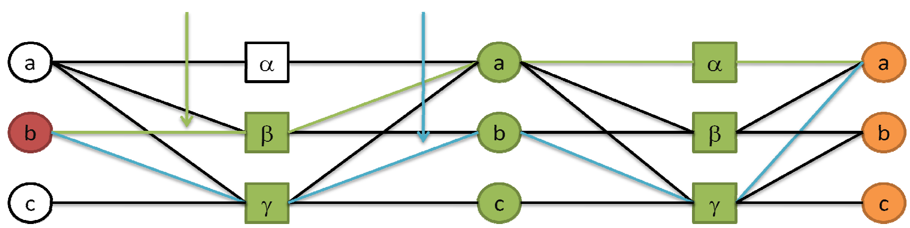

Fig S 3 Example showing how the method of reflections can be seen as an expansion of the properties of a node as a function of the properties of other nodes in the network with weights given by the product of the inverse of the degrees of each node traversed in the path connecting them.

Finally, we would like to mention that while higher order reflections do extract increasingly more relevant information about the productive structure of a country, as measured by how they are related to income and growth, it is important to mention that as N-> ∞ all variables will progressively converge to the a similar value. Surprisingly, we find the tiny deviations of these values to be extremely informative.

A SIMPLE EXAMPLE

In this section we explain the method of reflections using a simple example in which a network composed of four countries and four products is considered (Fig S 4).

Fig S 4 A simple network used to exemplify the method of reflections.

In this example, the diversification of countries and the ubiquity of products is given by:

$k_{c1,0}$=4
$k_{c2,0}$=1
$k_{c3,0}$=2
$k_{c4,0}$=1

$k_{p1,0}$=1
$k_{p2,0}$=2
$k_{p3,0}$=2
$k_{p4,0}$=3

Next, we calculate higher reflections of the method (or iterations). The first reflection consists of the average ubiquity of country's products and of the average diversification of a product's exporters and is given by:

$k_{c1,1}$=(1/4)(1+2+2+3)=2
$k_{c2,1}$=(1/1)(2)=2
$k_{c3,1}$=(1/2)(2+3)=2.5
$k_{c4,1}$=(1/1)(3)=3

$k_{p1,1}$=(1/1)(4)=4
$k_{p2,1}$=(1/2)(4+1)=2.5
$k_{p3,1}$=(1/2)(4+2)=3
$k_{p4,1}$=(1/3)(4+2+1)=2.33

The second reflection is given by the average first reflection values of a node's neighbors.

$k_{c1,2}$=(1/4)(4+2.5+2.25+2.5)=2.9583
$k_{c2,2}$=(1/1)(2.5)=2.5
$k_{c3,2}$=(1/2)(3+2.333)=2.66
$k_{c4,2}$=(1/1)(2.333)=2.33

$k_{p1,2}$=(1/1)(2)=2
$k_{p2,2}$=(1/2)(2+2)=2
$k_{p3,2}$=(1/2)(2+2.5)=2.25
$k_{p4,2}$=(1/3)(2+2.5+3)=2.5

We can use this example to illustrate how the method of reflections is able to differentiate between different countries based only on information regarding which country exports which product. In this example, the most diversified country is c1, which exports all four products while there are two countries, c2 and c4, that only export a single product. The sole export of c2 however, is a relatively non ubiquitous product that is exported only by c1, the most diversified country, while the sole export of c4 is a product that is exported by all countries except c2.

As we iterate the method we find that there is important information encoded in the relative position of countries and products relative to one another. For example, when we look at the values characterizing countries after the second reflection ($k_{c,2}$) we can see that country c1 comes up ahead, followed by country c3, c2 and c4. The method places country c2 ahead of c4 because by the second reflection it is already considering that country c2 produces a non ubiquitous product that is found only in diversified countries, probably signaling that country c2 has a relatively good endowment of capabilities and produces a small number of products because of other reason, such as being of relatively small size. On the contrary, c4 produces a product that is ubiquitous and it is found in diversified and non diversified countries, probably indicating that is a simple product which is accessible to countries with relatively simple productive structures. Hence while both, c2 and c4 produce the same number of products, the method can differentiate between them and considers c2 to have a more complex productive structure than c4.

While small in size this example illustrates how the method of reflections can be used to characterize the structure of a bipartite network and how this can be applied to help the understanding of the productive structure of countries and the sophistication of products.

SECTION 5: BIPARTITE NETWORK STRUCTURE MEASURED IN OTHER DATASETS

In this section we present two additional $k_{c,0}$-$k_{c,1}$ diagrams constructed using data aggregated according to the Harmonized system and according to the North American Industry Classification System (NAICS).

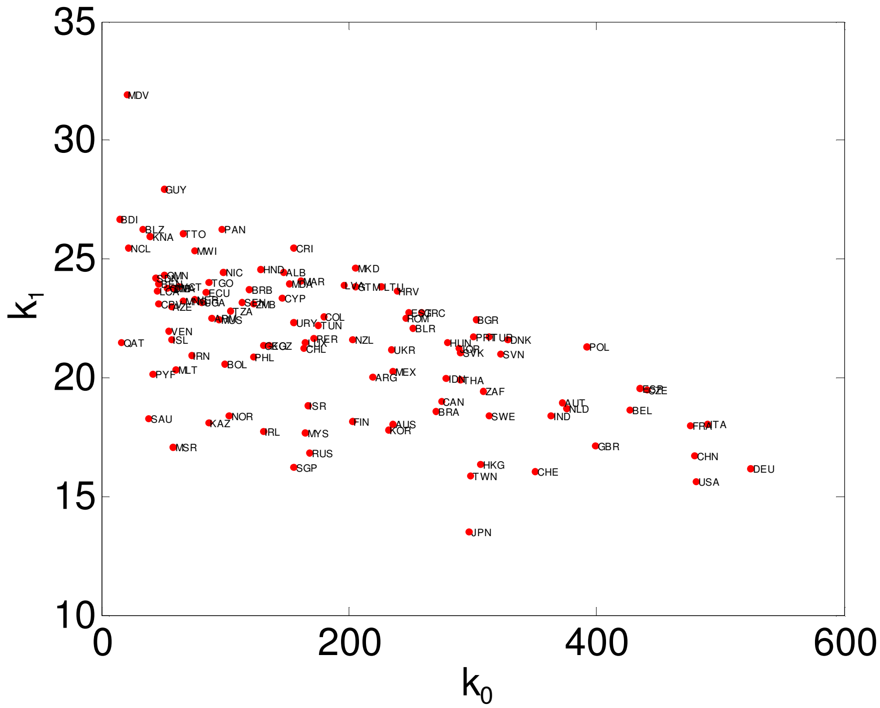

Fig S 5 ***$k_{c,0}$-$k_{c,1}$*** **diagram constructed using data containing 103 countries and 1241 products aggregated according to the Harmonized System.**

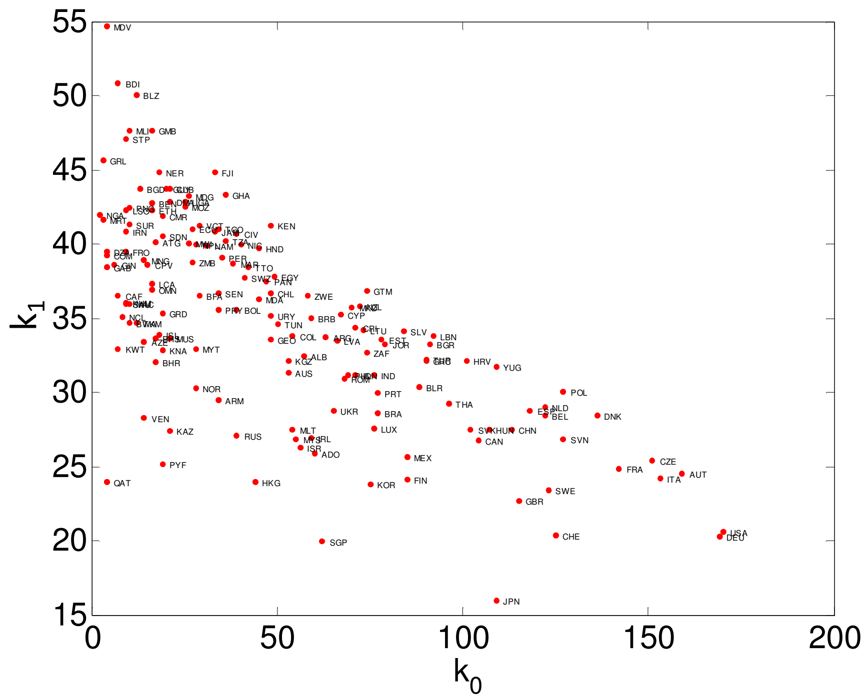

Fig S 6 $k_{c,0}$-$k_{c,1}$ diagram constructed using data containing 150 countries and 318 products aggregated according to the NAICS.

SECTION 6: RANDOMIZING A BIPARTITE NETWORK

To decide whether the structure of a network is trivial,* we need to compare it to an appropriate null model. The four null models we introduce in this section are an extension of the randomization algorithms introduced by Maslov and Sneppen [16] to analyze degree correlations in protein interaction networks. Our case differs from theirs in that we are dealing with a bipartite network rather than with a simple graph.

The idea behind the randomization procedure is that we can create a null model starting from the data we want to analyze by shuffling the links of the network while conserving some of its statistical properties. The most popular version of this randomization procedure, which was designed for simple graphs†, consists of randomizing the links in the network by permuting the nodes at the end of a pair of links. For example, if we consider a simple graph containing the links {*a,b*} and {*c,d*}, then an allowed randomization step would consist of replacing these two links by the pairs {*a,d*} and {*b,c*}, given that the {*a,d*} and {*b,c*} links were not already part of

---

*Expected from chance

† Simple Graph is a network in which there is only one type of nodes, and connections are strictly binary (0 or 1).

the network. The randomization procedure described above conserves the number of links in the network as well as its degree[‡] sequence and degree distribution. This is because the randomization procedure conserves the exact number of connections of each node, making it a good null model to compare properties of a network while controlling for the degree of nodes, which is the most fundamental property of a network.

In the case of a bipartite network, we have two separate degree sequences, one for each of its partitions. Here we introduce four null models to control for all possible combinations of degree sequences. Null Model 1 is a network with the same number of nodes and links as the original network, yet in Null Model 1 connections have been randomly assigned. Null Model 1 is the less stringent of our Null Models and represents a network with the same number of links as the original network, but with a random degree sequence for both partitions. Null Model 2 controls for the degree sequence of one partition of the network, while randomizing the target of those links in the other partition. Null Model 2 represents a network with a diversification sequence matching the one in the observed data, yet in Null Model 2 the products exported by a country have been randomly assigned. Null Model 2 also conserves the total number of links in the network. Null Model 3 is symmetric to Null Model 2 in the sense that it represents a network with the same ubiquity distribution as the one observed in the data, but where the exporters of each product have been randomly assigned. Finally, Null Model 4 is a model obtained by permuting links in the network such that the diversification of countries and the ubiquity of products are exactly the same as those observed in the empirical data.

It is important to notice that as Null Models become more stringent, the number of possible permutations that can be performed in the randomization procedure drops substantially. The possible number of permutations that can be performed in a randomization procedure does not only depend on the stringency of the null model, but also on the structure of the original network. For example, if we consider a bipartite network that can be represented by a triangular adjacency matrix (for simplicity assume that the number of

[‡] Degree: The number of links a node has. Degree Sequence: List containing the degrees of all nodes in the network.

products is equal to the number of countries and that M_{cp}= 1 $c<p$; M_{cp}=0 otherwise), then there is not a single possible permutation that could be performed using the fourth null model. For such a case, Null Model 4 is equivalent to the original network.

NULL MODEL SUMMARY

Null Model	Number of links	$k_{c,0}$ sequence	$k_{p,0}$ sequence	$<k_{c,0}>$	$<k_{c,1}>$	$< k_{p,0}>$	$< k_{p,1}>$
Null Model 1	$= M_{cp}$	$\neq M_{cp}$	$\neq M_{cp}$	$= M_{cp}$	$\neq M_{cp}$	$= M_{cp}$	$\neq M_{cp}$
Null Model 2	$= M_{cp}$	$= M_{cp}$	$\neq M_{cp}$	$= M_{cp}$	$\neq M_{cp}$	$= M_{cp}$	$\neq M_{cp}$
Null Model 3	$= M_{cp}$	$\neq M_{cp}$	$= M_{cp}$	$= M_{cp}$	$\neq M_{cp}$	$=M_{cp}$	$\neq M_{cp}$
Null Model 4	$= M_{cp}$	$= M_{cp}$	$= M_{cp}$	$= M_{cp}$	$\neq M_{cp}$	$=M_{cp}$	$\neq M_{cp}$

Table S 2 Summary null model behavior. <> stands for the average of a quantity.

SECTION 7: THE $K_{P,0}$-$K_{P,1}$ DIAGRAM

We compare the $k_{p,0}$-$k_{p,1}$ diagram obtained from our data with the one from our four null models (Fig S 7), finding that the structure of the country-product network is characterized by a strong negative correlation between $k_{p,0}$-$k_{p,1}$ and a wide range of $k_{p,1}$ values that cannot be explained by any of the four null models. This result becomes even more evident when we study higher order reflections of the method (see SM section 7). Products from different sectors are colored according to the ten root categories in the SITC-4 classification, showing that while there is a correspondence between the $k_{p,0}$-$k_{p,1}$ diagram and the SITC-4 classification, there are important variations among similarly classified products. For example, this graph shows that natural resource-based products, such as minerals and fuels, exhibit a wide range of ubiquities ($k_{p,0}$) at approximately constant diversification of its exporters ($k_{p,1}$), meaning that raw materials are on average exported by poorly diversified countries regardless of being relatively ubiquitous like coniferous wood ($k_{p,0}$=43, $k_{p,1}$=115), or rare as tin ore ($k_{p,0}$=8, $k_{p,1}$ =109). On the other hand, products classified as machinery show variation in the level of diversification of their exporters ($k_{p,1}$) at relatively low ubiquities ($k_{p,0}$). Hence the $k_{p,0}$-$k_{p,1}$ diagram can separate simple machines produced in less-diversified countries, such as handheld calculators, ($k_{p,0}$ =7, $k_{p,0}$ =144) from more complex machines produced in diversified countries such as motorcycles ($k_{p,0}$ =5, $k_{p,1}$ =270).

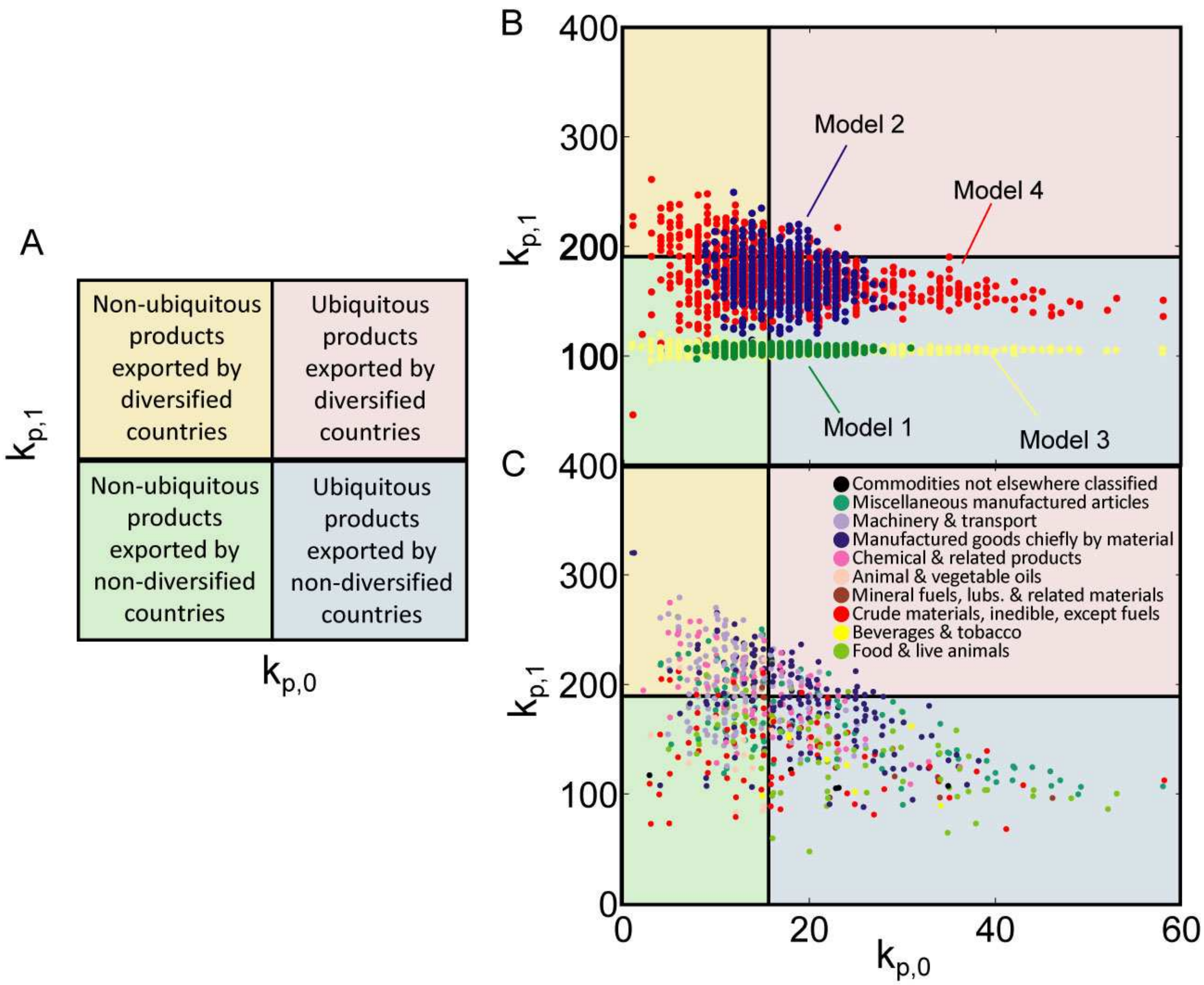

Fig S 7 Method of reflections and products characteristics. A, Schematic explanation of the $k_{p,0}$– $k_{p,1}$ space to characterize products. B, $k_{p,0}$– $k_{p,1}$ diagram for null models. C, $k_{p,0}$– $k_{p,1}$ diagram for the empirically observed exports data.

SECTION 8: A THIRD REFLECTION VIEW OF THE STRUCTURE OF THE COUNTRY-PRODUCT NETWORK

Here we continue the analysis presented in the manuscript to a third layer of analysis in which we show figures characterizing countries by $k_{c,0}$,$k_{c,1}$,$k_{c,2}$ and products by $k_{p,0}$,$k_{p,1}$,$k_{p,2}$ (Fig S 8-Fig S 11).

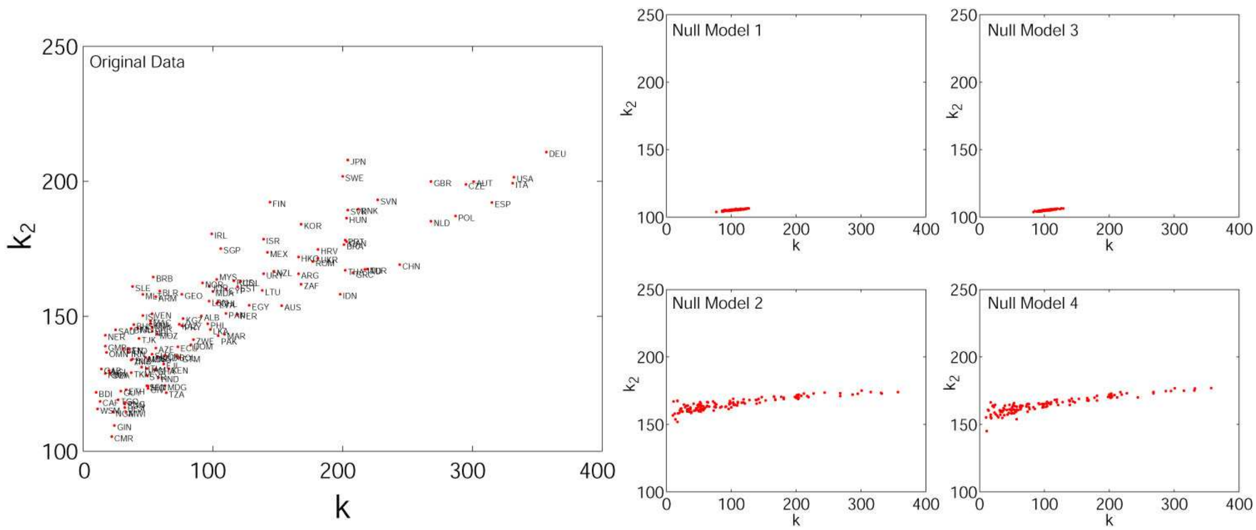

Fig S 8 Scatter plot for $k_{c,0}$ and $k_{c,2}$ for the original data in the year 2000 and the four null models.

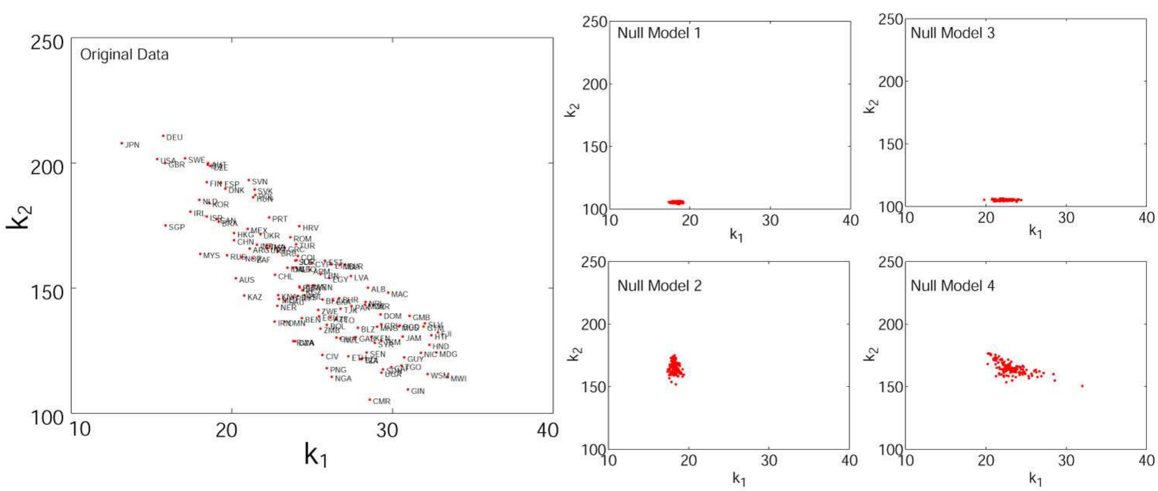

Fig S 9 Scatter plot for $k_{c,1}$ and $k_{c,2}$ for the original data in the year 2000 and the four null models.

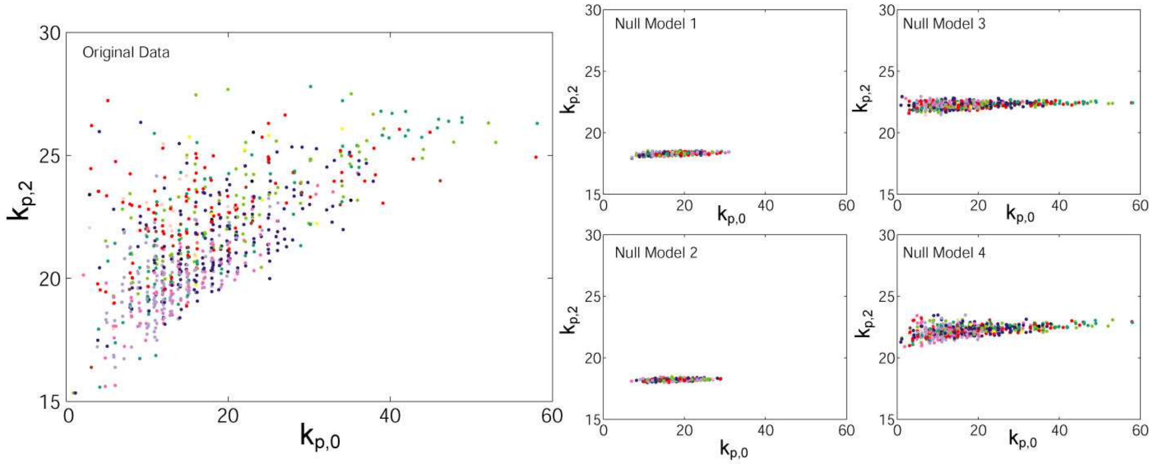

Fig S 10 Scatter plot for κ and κ_2 for the original data in the year 2000 and the four null models.

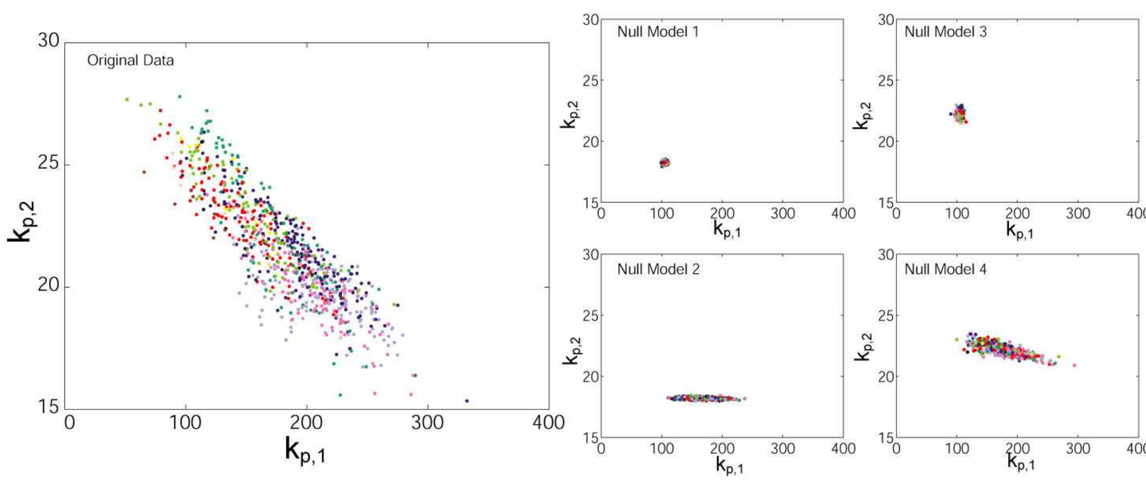

Fig S 11 Scatter plot for κ_1 and κ_2 for the original data in the year 2000 and the four null models.

SECTION 9: NULL MODELS AND GDP

In this section we present scatter plots between GDP per capita and the first two variables of the method of reflections characterizing the structure of bipartite networks created from our four null models (Fig S 12, Fig S 13).

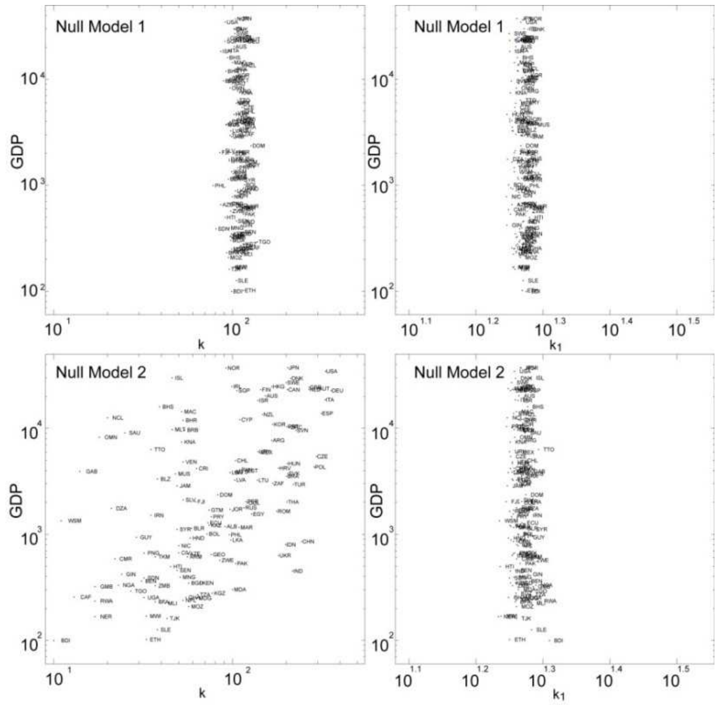

Fig S 12 Scatter plot between GDP and bipartite network properties for countries ($k=k_{c,0}$, $k_1=k_{c,1}$) and Null Models 1 and 2

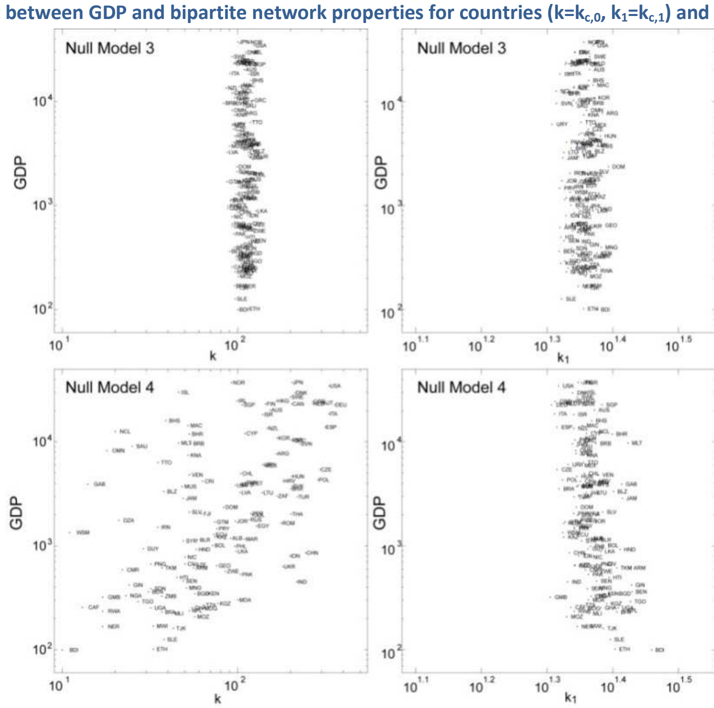

Fig S 13 Scatter plot between GDP and bipartite network properties for countries ($k=k_{c,0}$, $k_1=k_{c,1}$) and Null Models 3 and 4

SECTION 10: THE METHOD OF REFLECTIONS AND COUNTRY RANKINGS (YEAR 2000)

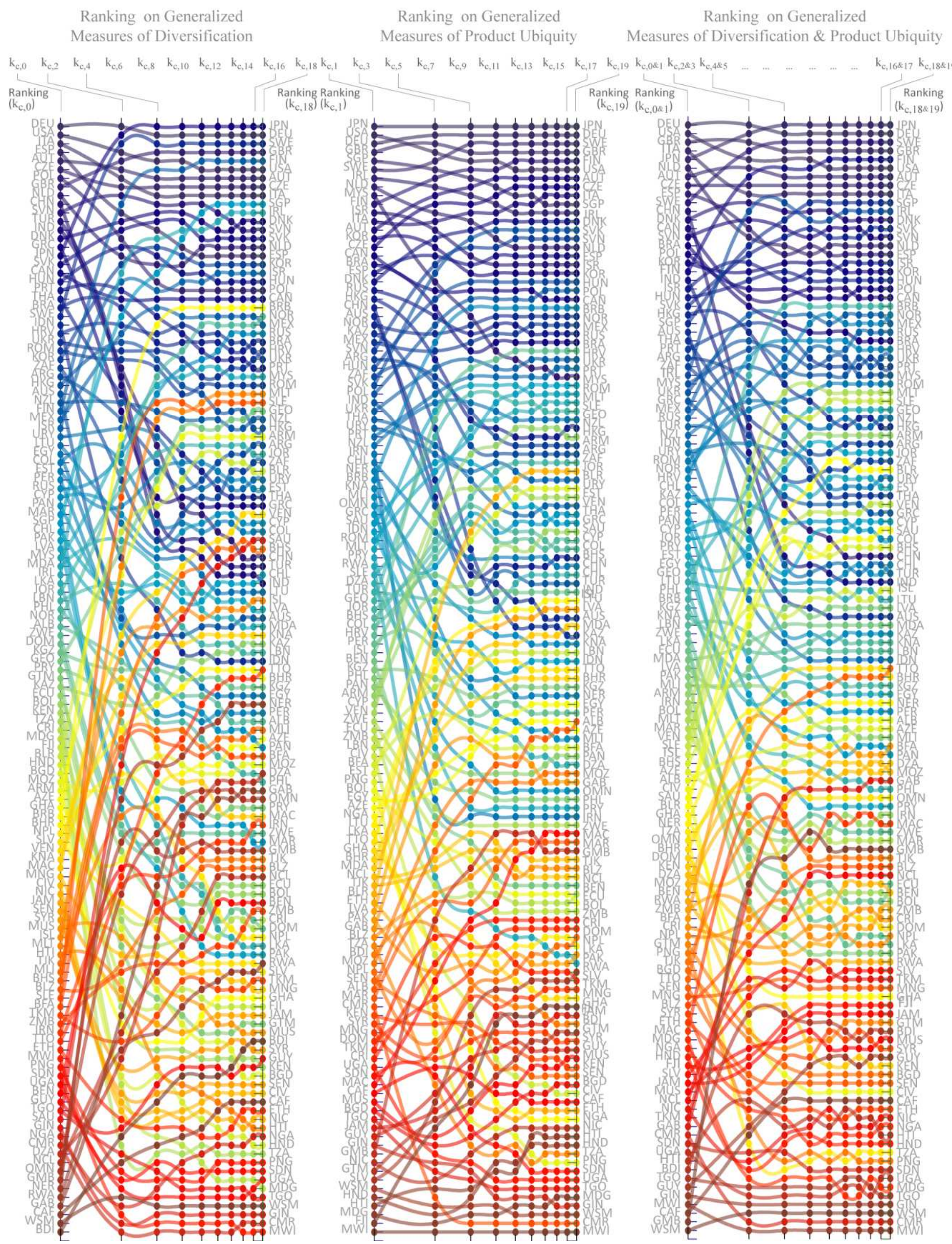

Fig S 14 Relative ranking of countries based on the Method of Reflections for the year 2000

SECTION 11: THE METHOD OF REFLECTIONS AND POPULATION

Economic output is usually measured in per capita terms, as the goal of development is to generate and distribute wealth in the most democratic way possible. Yet there are some other variables in which the per capita idea does not apply as directly as it does for income. One example is diversification, which in our formalism is represented by $k_{c,0}$. While in principle we might be tempted to consider the per capita level of diversification, as a good indicator of the diversification that can be attributed to each individual in a population, it is important to consider that such normalization assumes that the level of diversification grows linearly with the number of people. This, however, would not be a careful way of measuring the amount of diversification that should be attributed to each individual in a population, as the number of different products a group of people can make might well depend on the possible number of interactions, and hence go as the square of the population, or could depend on a more complex function that is hitherto unknown. Normalizing diversification by the number of individuals in a population can therefore be considered naïve, as it assumes a linear functional form as the correct normalization for a variable that does not necessarily depends linearly in the population.

The diversification of a country $k_{c,0}$, however, does depend on a country's population (Table S 3 column 1). Hence, we still need a variable that would give us a measure of the level diversification of a country that is independent of its number of inhabitants. In Table S 3 we present the dependence of our first four measures of diversification ($k_{c,0}$,$k_{c,2}$,$k_{c,4}$,$k_{c,8}$) on population, showing that higher order reflections of the method generate measures of diversification that are independent of a country's population, and are therefore good indicators of the level of diversification of a country that is due to the complexity of its economy rather than to its population.

VARIABLES	*Log* $k_{c,0}$	*Log* $k_{c,0}$	*Log* $k_{c,4}$	*Log* $k_{c,8}$
Log Population	0.190***	0.0168**	0.00343	0.000267
t-test	(4.812)	(2.168)	(1.488)	(1.198)
Constant	1.272**	4.708***	5.004***	5.081***
t-test	(2.005)	(37.63)	(134.7)	(1415)
Observations	127	127	127	127
Adjusted R^2	0.150	0.029	0.010	0.003

Table S 3 Correlation between population and successive generations of measures of diversification constructed from the method of reflections (statistically significant at the 5% level, *** statistically significant at the 1% level).**

SECTION 12: SHARES OF PRODUCTS IN THE WORLD

One critique of our methods that can be raised is that the SITC-4 classification is more disaggregated for goods produced by richer countries, as rich countries are the ones that created the classification system. A classification bias in that direction would overstate the level of diversification of rich countries and understate that of poor countries.

We have shown that our results do not depend on the level of aggregation by considering two additional datasets aggregated according to different classification systems, which summarize all tradable goods using a different number of product classifications. Here we complement this test of the validity of our methods by looking at the share in world trade associated with each product in the SITC-4 classification (Fig S 15), finding that, contrary to the critique presented above, industrialized country products have large shares in total trade, indicating that they are not more narrowly classified than agricultural products and raw materials (except oil) when benchmarked by their share in world trade. In simpler terms, if we were to further disaggregate products into categories to achieve more homogenous shares in world trade, we would have to disaggregate cars into classes, like SUVs, sedans and compacts rather than melons into different types, indicating that the data behaves in the opposite way than what the critique suggests.

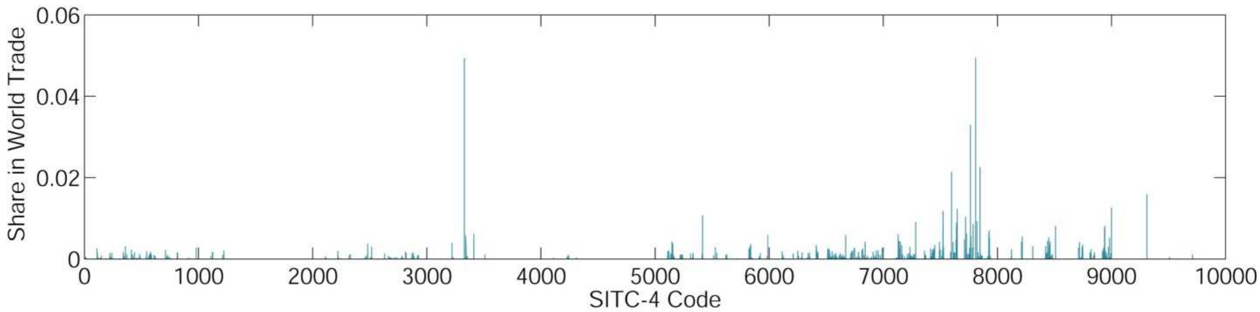

Fig S 15 Share in world trade for products sorted by SITC-4 code.

Table S 4 and Table S 5 respectively show the five products with smallest, largest share in world trade.

SITC-4 Code	Product Names	World Market Share in the year 2000 (Total World Trade = 1)
6553	Knitted/crocheted fabrics elastic or ruberized	3.2×10^{-8}
19	Live animals of a kind mainly used for human food	5.3×10^{-8}
6344	Wood-based panels N.E.S.	1.7×10^{-7}
3415	Coal gas, water gas, producer gas & similar gases	5.5×10^{-7}
2652	True hemp, raw or processed, not spun; tow and waste	8.0×10^{-7}

Table S 4 The five products with the smallest world share in the year 2000.

SITC-4 Code	Product Names	World Market Share in the year 2000 (Total World Trade = 1)
7810	Passenger motor cars, for transport of pass. & good	0.0494
3330	Petroleum oils & crude oils obt. from bitumen minerals	0.0493
7764	Electronic microcircuits	0.0329
7849	Other parts and accessories of motor vehicles	0.0225
7599	Parts and accessories suitable for calculating and data processing machines	0.0214

Table S 5 The five products with the largest world share in the year 2000

SECTION 13: NETWORK STRUCTURE, INCOME AND GROWTH

In this section we present regressions showing how the structure of the bipartite network is connected to income and economic growth. We also compare the performance of our structural measures to two other measures of diversity: the Hirschaman-Herfindahl (H-H) index and Entropy.

The HH index is a measure of market concentration commonly used for antitrust purposes, yet it has also been used as a measure of diversification. The H-H index (*H*) is defined as:

$$H_c = \sum_p \left(S_{cp}\right)^2 \qquad (12)$$

where S_{cp} is the share of product *p* in the export basket of country *c*. An alternative method to measure the diversification of a country's export basket is to consider its entropy, which is defined as:

$$E_c = -\sum_p S_{cp} \log\left(S_{cp}\right) \qquad (13)$$

High entropy values are characteristic of diversified export baskets, whereas low entropy values are associated with export baskets that are concentrated in a small number of products.

We present the results of our regressions as tables (Table S 6-Table S 9). To help the reader understand the information contained in these tables, we have created a figure explaining how to read these regression tables (Fig S 16):

Understanding a regression table *(Read figure by following the numbers)*

(1) Each column represent a different regression

(2) The regression is a fit between the row variables and the column variable.

(3) The variables involved in the regression can be read from those whose coefficient appear in the table. For example, column *(1)* regresses growth against GDP per capita ppp and entropy. All other variables are not used in column's *(1)* regression

(4) The time frames involved in the analysis are presented in bold font underneath the name of the variables. For example, in this exercise data from the years 1985,1990, 1995 and 2000 were used to predict growth in the 85-90,90-95,95-00 and 00-05 periods respectively.

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*	*(11)*
Predicted Variable	Growth **(85-90-95-00-05)**	Growth **(85-90-95-00-05)**	Growth **(85-90-95-00-05)**	Growth **(85-90-95-00-05)**	Growth **(85-90-95-00-05)**	Growth **(85-90-95-00-05)**	Growth **(85-90-95-00-05)**	Growth **(85-90-95-00-05)**	Growth **(85-90-95-00-05)**	Growth **(85-90-95-00-05)**	Growth **(85-90-95-00-05)**
Predictors											
GDP per capita ppp **(85,90,95,00)**	-0.00269* (-1.732)	0.000292 (0.211)	-0.00286* (-1.785)	0.000361 (0.220)	-0.00393** (-2.431)	0.00224* (1.767)	0.00326** (2.553)	0.00310** (2.479)	-0.00257* (-1.686)	0.000281 (0.207)	-0.00275* (-1.749)
Entropy **(85,90,95,00)**	0.00798*** (6.280)		0.00885*** (3.760)						0.00759*** (6.060)		0.00851*** (3.680)
Herfindahl **(85,90,95,00)**		-0.0373*** (-4.970)	0.00602 (0.440)							-0.0351*** (-4.765)	0.00636 (0.474)
k **(85,90,95,00)**				0.000121*** (5.351)							
k_1 **(85,90,95,00)**				0.000953*** (2.853)							
k_4 **(85,90,95,00)**					0.00113*** (7.074)						
k_5 **(85,90,95,00)**					0.00260*** (5.694)						
k_8 **(85,90,95,00)**						0.00102*** (3.474)					
k_9 **(85,90,95,00)**						0.00312*** (5.504)					
k_{18} **(85,90,95,00)**							-0.000265 (-1.147)	0.000632** (2.131)	0.000673** (2.359)	0.000647** (2.234)	0.000676** (2.365)
k_{19} **(85,90,95,00)**								0.00280*** (4.671)	0.00259*** (4.494)	0.00265*** (4.523)	0.00260*** (4.493)
Constant	0.0166 (1.497)	0.0236* (1.910)	0.0142 (1.144)	-0.0160 (-0.933)	-0.173*** (-7.646)	-0.224*** (-4.198)	0.0349 (0.889)	-0.163*** (-2.846)	-0.142** (-2.576)	-0.132** (-2.355)	-0.145*** (-2.611)
Observations	451	451	451	451	451	451	451	451	451	451	451
Adjusted R^2	0.090	0.062	0.089	0.071	0.136	0.071	0.013	0.057	0.127	0.101	0.125

(4) Each cell of the regression table contains two values the regression coefficient and the t-statistic associated with it.

(5) An indicator of the significance associated with the regression coefficient is given by the *'s.
* p-value<0.1
** p-value<0.05
*** p-value<0.001

(6) The total variance in the column variable explained by each column value is reported in the R^2.

(7) We arranged the tables to show regressions between (i) the predicted variable, entropy and Herfindahl index in the first three columns, (ii) the predicted value and variables from the method of reflections in the 4th to 8th columns, (iii) the predicted value, entropy, Herfindahl index and the variables from the method of reflection in the 9th to 11th columns.

Fig S 16 How to read regression tables

In this section we present regression tables between E, H, $k_{c,0}$, $k_{c,1}$, $k_{c,4}$, $k_{c,8}$, $k_{c,12}$, $k_{c,18}$ and income per capita adjusted by power-purchasing parity (Table S 6) and E, H, $k_{c,0}$, $k_{c,1}$, $k_{c,4}$, $k_{c,5}$, $k_{c,8}$, $k_{c,9}$, $k_{c,18}$, $k_{c,19}$ and economic growth for a 20 year period (Table S 7), two ten year periods (Table S 8) and four five year periods (Table S 9). Additionally, we present regression results for four five year periods with fixed country effects (Table S 10). A fixed country effect regression means that dummy variables were introduced to capture all the variation between countries, hence the quantity we look for here is the within R^2, which is the variation in growth explained by the productive structure after controlling for all between-country variations. Technically dummy variables are defined as 0 for all countries except one. In fixed effect regressions we introduce one of these variables per country considered.

Table S 5 studies the relationship between the level of income in 2000, as measured by the log of GDP per capita at purchasing power parity, and different measures of productive structure. Columns 1 and 2 use pre-existing measures of diversification, in particular the entropy and the H-H index. The first can explain 37.7 percent of the variance in income per capita, while the second can only account for 17.6 percent, as shown by the R^2 of the regression. Columns 3 to 8 use successive iterations of our method. Diversification $k_{c,0}$ explains 34.5 percent of the variance; $k_{c,1}$ explains 37.8 percent, and subsequent variables converge to 53 percent by the 8th reflection, with higher order variables adding little additional power. Columns 9 to 11 show a "horse race" between $k_{c,18}$ and the pre-existing measures taken one at the time or simultaneously. It shows that $k_{c,18}$ contains much more information than the others do, as reflected in the fact that adding them increases the R^2 very little vis a vis column 8 but much more vis a vis columns 1 and 2. Table S 6 does a cross-country regression of growth between 1985 and 2005 and initial values of productive structure indicators. Columns 1–3 use the entropy indicator, the H-H index and the two combined. Columns 4–8 use successive pairs of k variables. Columns 9-11 present a horse race between the $k_{c,18}$-$k_{c,19}$ pair and the traditional measures of productive structure, both separately and taken together. All regressions also control for the initial level of GDP per capita. The results are similar to those of the previous table. The variables we introduce do a better job at predicting the pattern of future growth and higher reflections of the method have the largest predictive power. Interestingly, there is complementary information in successive measures of our variables so that both appear significant in the regression. $k_{c,18}$-$k_{c,19}$ contain more information than the traditional measures and beat them in a horse race (equations 9-11).

Table S 7 repeats these regressions, splitting the sample into two periods of 10 years, 1985-95 and 1995-05, and finds similar results: pairs of k variables do a better job of explaining growth than do the traditional variables, and the quality of the fit increases with each iteration. A horse race between traditional and k variables shows that the bulk of the explanatory power comes from the *k* variables, although the traditional variables have some residual information that is statistically significant, although small. Table S 8 repeats the analysis using four 5-year periods between 1985 and 2005 and finds similar results.

Table S 9 presents an equivalent set of regressions but controls for average fixed country characteristics by including a dummy variable per country. This regression bases its identification only in the within-country variation in growth and finds similar but even stronger results. Our preferred specification – column 8 – is able to explain 33.72 percent of the within-country variance, while adding the traditional variables only increases the explanatory power to 35 percent. The two traditional variables on their own (column 3) explain only 21.72 percent of the within-country variance, indicating that the fit increases much more when adding the *k* variables to the traditional variables (contrast of columns 3 and 11) than when adding the traditional variables to the *k* variables (contrast column 8 and 11).

INCOME (YEAR 2000)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	**(11)**
Predicted Variable	Log GDP per capita ppp (2000)	Log GDP per capita ppp (2000)	Log GDP per capita ppp (2000)	Log GDP per capita ppp (2000)	Log GDP per capita ppp (2000)	Log GDP per capita ppp (2000)	Log GDP per capita ppp (2000)	Log GDP per capita ppp (2000)	Log GDP per capita ppp (2000)	Log GDP per capita ppp (2000)	Log GDP per capita ppp (2000)
Predictors											
Entropy (2000)	0.552***								0.157**		0.202
	(8.712)								(1.991)		(1.275)
Herfindahl (2000)		-2.554***								-0.639	0.270
		(-5.250)								(-1.552)	(0.329)
$k_{c,0}$ (2000)			0.00859***								
			(8.147)								
$k_{c,1}$ (2000)				-0.159***							
				(-8.740)							
$k_{c,4}$ (2000)					0.116***						
					(11.48)						
$k_{c,8}$ (2000)						1.201***					
						(11.93)					
$k_{c,12}$ (2000)							12.21***				
							(11.99)				
$k_{c,18}$ (2000)								392.6***	324.0***	365.8***	315.6***
								(11.99)	(6.854)	(9.923)	(5.859)
Constant	6.696***	8.914***	7.603***	12.34***	-9.796***	-185.6***	-1968***	-63581***	-52466***	-59234***	-51109***
	(30.38)	(71.88)	(55.88)	(27.49)	(-6.147)	(-11.41)	(-11.94)	(-11.99)	(-6.853)	(-9.921)	(-5.858)
Observations	125	125	125	125	125	125	125	125	125	125	125
Adjusted R^2	**0.377**	**0.176**	**0.345**	**0.378**	**0.513**	**0.533**	**0.535**	**0.535**	**0.546**	**0.541**	**0.543**

Table S 6 Regression coefficients for income per capita

20 YEAR GROWTH

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	**(11)**
Predicted Variable	Growth (85,05)	Growth (85, 05)	Growth (85, 05)	Growth (85, 05)	Growth (85, 05)	Growth (85, 05)	Growth (85, 05)	Growth (85, 05)	Growth (85, 05)	Growth (85, 05)	Growth (85, 05)
Predictors											
GDP per capita ppp (1985)	-0.00176	0.000993	-0.00206	-0.00154	-0.00249	-0.00223	-0.00470	-0.00233	-0.00244	-0.00238	-0.00242
	(-0.794)	(0.533)	(-0.882)	(-0.497)	(-0.735)	(-0.688)	(-1.478)	(-0.758)	(-0.849)	(-0.804)	(-0.831)
Entropy (1985)	0.00660***		0.00828**						0.00200		0.00322
	(3.650)		(2.600)						(0.931)		(0.896)
Herfindahl (1985)		-0.0273***	0.0116							-0.00414	0.00723
		(-2.765)	(0.760)							(-0.406)	(0.454)
$k_{c,0}$ (1985)				6.62e-05**							
				(2.080)							
$k_{c,1}$ (1985)				-0.000612							
				(-0.749)							
$k_{c,4}$ (1985)					0.00169***						
					(2.866)						
$k_{c,5}$ (1985)					0.0321*						
					(1.737)						
$k_{c,8}$ (1985)						0.0338***					
						(3.075)					
$k_{c,9}$ (1985)						0.890***					
						(2.713)					
$k_{c,18}$ (1985)							0.401***	38.88***	35.05**	37.26***	35.57***
							(3.453)	(2.952)	(2.618)	(2.849)	(2.643)
$k_{c,19}$ (1985)								1127***	1017**	1080***	1033***
								(2.928)	(2.603)	(2.829)	(2.632)
Constant	0.0114	0.0137	0.00650	0.0338	-0.735*	-19.29***	-69.21***	-23801***	-21475**	-22808***	-21801***
	(0.751)	(0.776)	(0.437)	(0.922)	(-1.883)	(-2.807)	(-3.454)	(-2.940)	(-2.610)	(-2.834)	(-2.633)
Observations	97	97	97	97	97	97	97	97	97	97	97
Adjusted R^2	**0.195**	**0.115**	**0.192**	**0.118**	**0.206**	**0.247**	**0.202**	**0.274**	**0.274**	**0.268**	**0.268**

Table S 7 Regression coefficients for a twenty year period of growth

10 YEAR GROWTH

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	**(11)**
Predicted Variable	Growth (85-95-05)	Growth (85-95-05)	Growth (85-95-05)	Growth (85-95-05)	Growth (85-95-05)	Growth (85-95-05)	Growth (85-95-05)	Growth (85-95-05)	Growth (85-95-05)	Growth (85-95-05)	Growth (85-95-05)
Predictors											
GDP per capita ppp (85,95)	-0.00235	0.000334	-0.00246	0.00112	-0.00395**	-0.00311*	0.00310**	-0.00119	-0.00346*	-0.00229	-0.00343*
	(-1.322)	(0.209)	(-1.349)	(0.595)	(-2.062)	(-1.695)	(2.188)	(-0.707)	(-1.899)	(-1.327)	(-1.850)
Entropy (85,95)	0.00699***		0.00759***						0.00458***		0.00434
	(4.962)		(2.985)						(3.002)		(1.648)
Herfindahl (85,95)		-0.0325***	0.00422							-0.0210**	-0.00162
		(-3.890)	(0.285)							(-2.494)	(-0.112)
$k_{c,0}$ (85,95)				9.75e-05***							
				(3.967)							
$k_{c,1}$ (85,95)				0.000916**							
				(2.543)							
$k_{c,4}$ (85,95)					0.00102***						
					(5.577)						
$k_{c,5}$ (85,95)					0.00329***						
					(5.971)						
$k_{c,8}$ (85,95)						0.00577***					
						(5.056)					
$k_{c,9}$ (85,95)						0.0152***					
						(5.594)					
$k_{c,18}$ (85,95)							-0.000660***	0.455***	0.310***	0.375***	0.311***
							(-3.577)	(4.306)	(2.705)	(3.428)	(2.695)
$k_{c,19}$ (85,95)								1.158***	0.789***	0.954***	0.792***
								(4.312)	(2.709)	(3.433)	(2.699)
Constant	0.0171	0.0226	0.0153	-0.0186	-0.168***	-1.178***	0.102***	-96.21***	-65.48***	-79.20***	-65.78***
	(1.356)	(1.602)	(1.087)	(-0.971)	(-6.137)	(-5.215)	(3.107)	(-4.308)	(-2.705)	(-3.428)	(-2.695)
Observations	221	221	221	221	221	221	221	221	221	221	221
Adjusted R^2	0.113	0.077	0.109	0.085	0.170	0.160	0.068	0.137	0.168	0.158	0.164

Table S 8 Regression coefficients for two ten year periods of growth

5 YEAR GROWTH

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	**(11)**
Predicted Variable	Growth (85-90-95-00-05)	Growth (85-90-95-00-05)	Growth (85-90-95-00-05)	Growth (85-90-95-00-05)	Growth (85-90-95-00-05)	Growth (85-90-95-00-05)	Growth (85-90-95-00-05)	Growth (85-90-95-00-05)	Growth (85-90-95-00-05)	Growth (85-90-95-00-05)	Growth (85-90-95-00-05)
Predictors											
GDP per capita ppp (85,90,95,00)	-0.00269*	0.000292	-0.00286*	0.000361	-0.00393**	0.00224*	0.00326**	0.00310**	-0.00257*	0.000281	-0.00275*
	(-1.732)	(0.211)	(-1.785)	(0.220)	(-2.431)	(1.767)	(2.553)	(2.479)	(-1.686)	(0.207)	(-1.749)
Entropy (85,90,95,00)	0.00798***		0.00885***						0.00759***		0.00851***
	(6.280)		(3.760)						(6.060)		(3.680)
Herfindahl (85,90,95,00)		-0.0373***	0.00602							-0.0351***	0.00636
		(-4.970)	(0.440)							(-4.765)	(0.474)
$k_{c,0}$ (85,90,95,00)				0.000121***							
				(5.351)							
$k_{c,1}$ (85,90,95,00)				0.000953***							
				(2.853)							
$k_{c,4}$ (85,90,95,00)					0.00113***						
					(7.074)						
$k_{c,5}$ (85,90,95,00)					0.00260***						
					(5.694)						
$k_{c,8}$ (85,90,95,00)						0.00102***					
						(3.474)					
$k_{c,9}$ (85,90,95,00)						0.00312***					
						(5.504)					
$k_{c,18}$ (85,90,95,00)							-0.000265	0.000632**	0.000673**	0.000647**	0.000676**
							(-1.147)	(2.131)	(2.359)	(2.234)	(2.365)
$k_{c,19}$ (85,90,95,00)								0.00280***	0.00259***	0.00265***	0.00260***
								(4.671)	(4.494)	(4.523)	(4.493)
Constant	0.0166	0.0236*	0.0142	-0.0160	-0.173***	-0.224***	0.0349	-0.163***	-0.142**	-0.132**	-0.145***
	(1.497)	(1.910)	(1.144)	(-0.933)	(-7.646)	(-4.198)	(0.889)	(-2.846)	(-2.576)	(-2.355)	(-2.611)
Observations	451	451	451	451	451	451	451	451	451	451	451
Adjusted R^2	0.090	0.062	0.089	0.071	0.136	0.071	0.013	0.057	0.127	0.101	0.125

Table S 9 Regression coefficients for four five year periods of growth.

5 YEAR GROWTH FIXED EFFECTS

	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*	**(11)**
Predicted Variables	Growth **(85-90-95-00-05)**	Growth **(85-90-95-00-05)**	Growth **(85-90-95-00-05)**	Growth **(85-90-95-00-05)**	Growth **(85-90-95-00-05)**	Growth **(85-90-95-00-05)**	Growth **(85-90-95-00-05)**	Growth **(85-90-95-00-05)**	Growth **(85-90-95-00-05)**	Growth **(85-90-95-00-05)**	Growth **(85-90-95-00-05)**
Predictors											
GDP per capita ppp **(85,90,95,00)**	-0.0585***	-0.0581***	-0.0595***	-0.0773***	-0.0863***	-0.0891***	-0.0651***	-0.0899***	-0.0868***	-0.0884***	-0.0867***
	(-7.911)	(-7.721)	(-8.072)	(-10.11)	(-11.28)	(-11.78)	(-8.337)	(-11.89)	(-11.39)	(-11.58)	(-11.40)
Entropy **(85,90,95,00)**	0.0134***		0.0247***						0.00706**		0.0142**
	(4.478)		(4.037)						(2.453)		(2.435)
Herfindahl **(85,90,95,00)**		-0.0390***	0.0585**							-0.0181	0.0360
		(-2.842)	(2.117)							(-1.435)	(1.410)
$k_{c,0}$ **(85,90,95,00)**				0.000223***							
				(3.710)							
$k_{c,1}$ **(85,90,95,00)**				0.00238***							
				(6.549)							
$k_{c,4}$ **(85,90,95,00)**					0.000537***						
					(2.922)						
$k_{c,5}$ **(85,90,95,00)**					0.00366***						
					(9.287)						
$k_{c,8}$ **(85,90,95,00)**						0.000611***					
						(2.801)					
$k_{c,9}$ **(85,90,95,00)**						0.00410***					
						(8.998)					
$k_{c,18}$ **(85,90,95,00)**							-0.000535***	0.000601***	0.000653***	0.000618***	0.000673***
							(-2.808)	(2.801)	(3.051)	(2.879)	(3.141)
$k_{c,19}$ **(85,90,95,00)**								0.00419***	0.00395***	0.00410***	0.00389***
								(8.799)	(8.164)	(8.521)	(8.031)
Constant	0.467***	0.514***	0.429***	0.594***	0.588***	0.589***	0.651***	0.596***	0.543***	0.585***	0.511***
	(7.427)	(8.147)	(6.592)	(9.546)	(8.165)	(8.070)	(8.203)	(8.310)	(7.315)	(8.133)	(6.583)
Observations	451	451	451	451	451	451	451	451	451	451	451
Within R^2	**0.2071**	**0.1784**	**0.2179**	**0.2991**	**0.3379**	**0.3373**	**0.1779**	**0.3372**	**0.3494**	**0.3415**	**0.3535**

Table S 10 Regression table for four five year periods of growth considering fixed country effects.

SECTION 14: ADDITIONAL RESULTS

PRODY AND EXPY

The variables *PRODY* and *EXPY* were introduced originally by Hausmann, Hwang and Rodrik [15] to characterize the sophistication of products and of countries' exports starting from trade and income data. *PRODY* and *EXPY* allow us to study the income of countries from a product-specific perspective.

DEFINITIONS

PRODY

The *PRODY* of a product is the average income per-capita associated with that product. We can calculate *PRODY* using trade data as

$$PRODY_p = \sum_c \frac{S_{cp}}{\sigma_p} G_c \tag{14}$$

Where S_{cp} is the share of product p in the export basket of country c, G_c is the income of country c measured as *GDP* per capita adjusted for power purchasing parity and $\sigma_p = \sum_c s_{cp}$.

EXPY

The *EXPY* of a country is the average *PRODY* of its exports.

$$EXPY_c = \sum_p S_{cp} \, PRODY_p \tag{15}$$

We notice that PRODY and EXPY mix income and network information as these variables have a similar definition than the first two reflections of the method with k_0=GDP per capita and M_{cp} related to the shares of products in the export baskets of countries.

EXPY, $K_{C,0}$, $K_{C,1}$

Here we complement our results on income by showing that k and k_1 correlate with a countries' EXPY (Fig S 17).

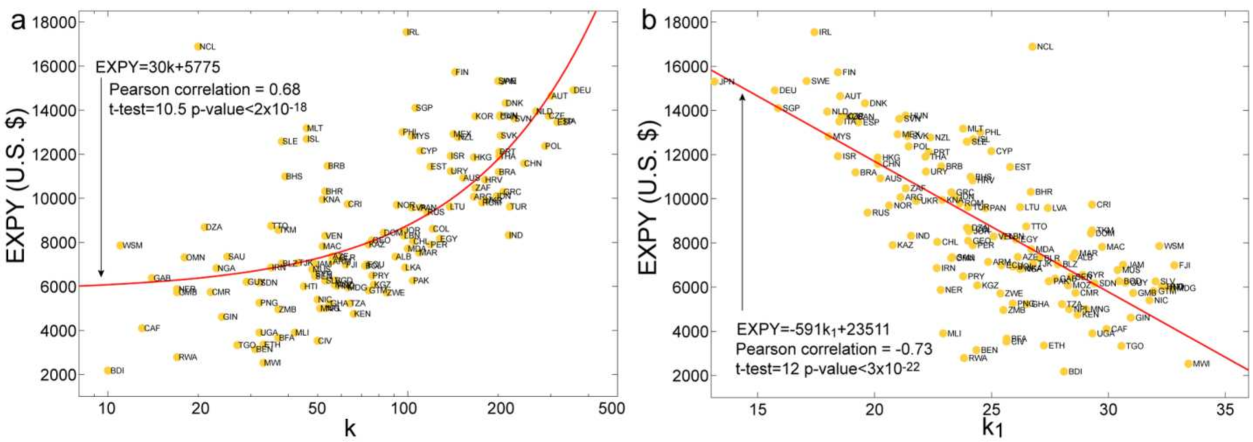

Fig S 17 *EXPY* and bipartite network structure. a, Diversification ($k_{c,0}$=k) versus *EXPY*. b, Average ubiquity of a country's products ($k_{c,1}$=k_1) versus *EXPY*.

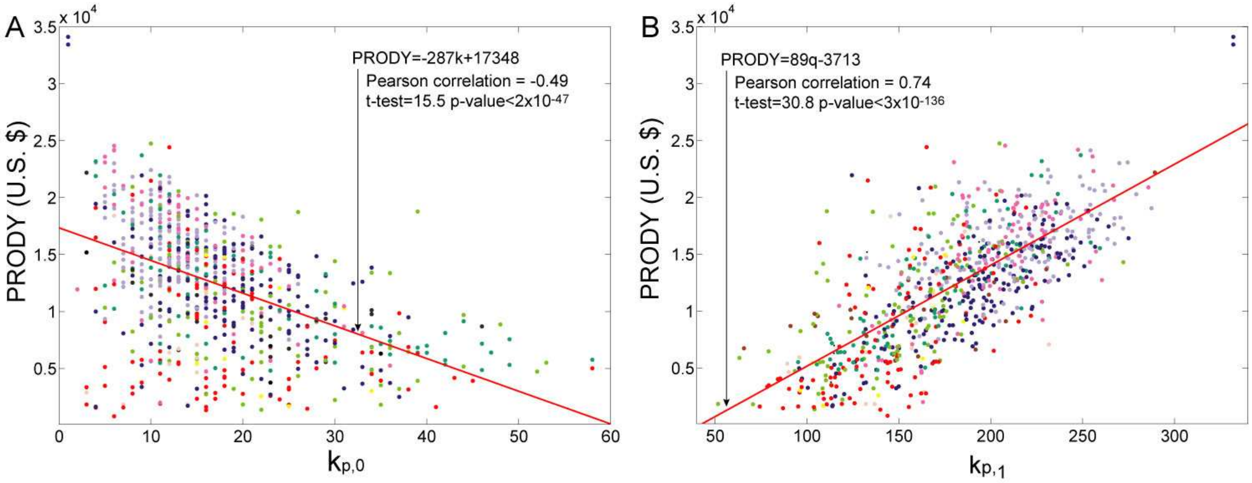

Fig S 18 *PRODY* and bipartite network structure. A, Ubiquity ($k_{p,0}$) versus *PRODY*. b, Average ubiquity of a country's products ($k_{p,1}$) versus *PRODY*.

NULL MODEL BEHAVIOR FOR PRODY AND EXPY, $K_{C,0}$, $K_{C,1}$

Here we present the null model behavior for the relationships found between *PRODY*, *EXPY* and the network structure (Fig S 19 - Fig S 22).

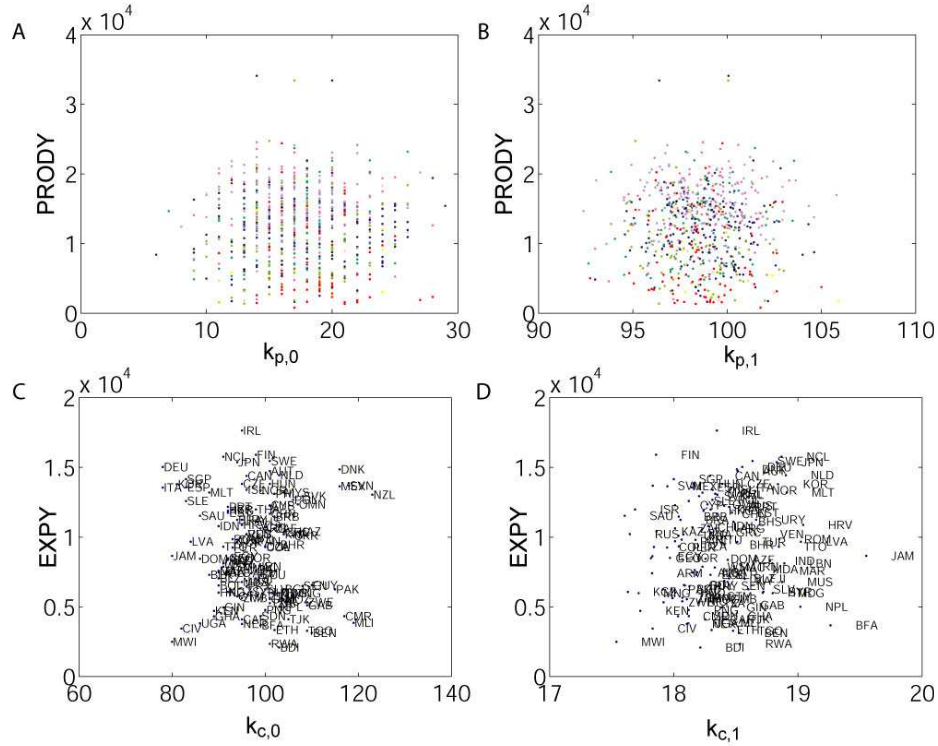

Fig S 19 Comparison between *PRODY* and *EXPY* with $k_{c,0}$, $k_{c,1}$, $k_{p,0}$ *and* $k_{p,1}$ for null model 1. A *PRODY* v/s $k_{p,0}$ B *PRODY* v/s $k_{p,1}$ C *EXPY* v/s $k_{c,0}$ D *EXPY* v/s , $k_{c,1}$

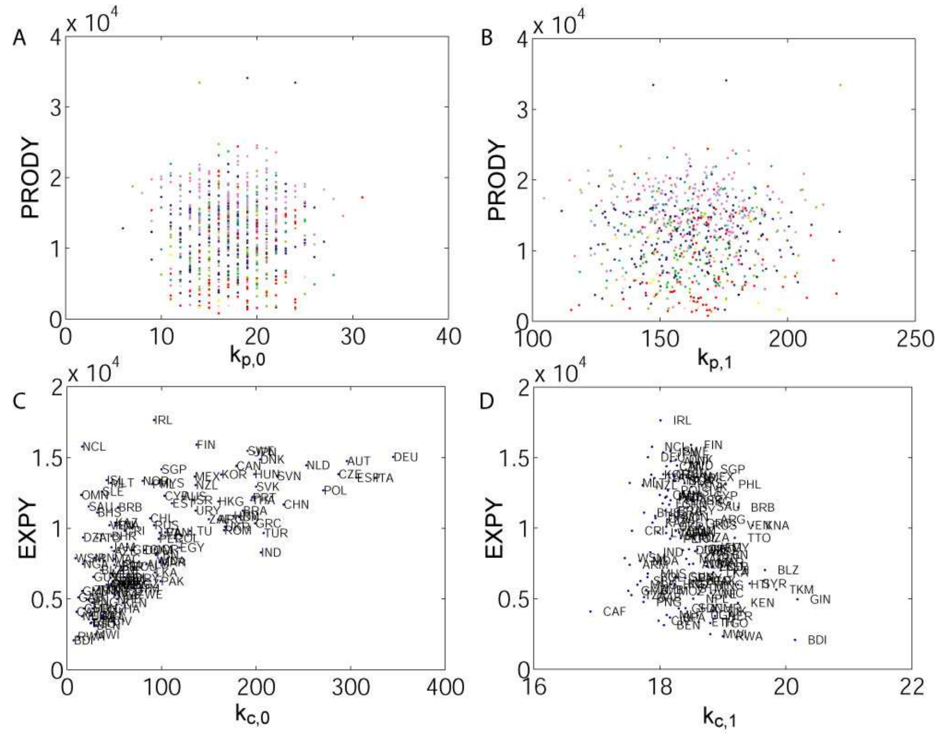

Fig S 20 Comparison between *PRODY* and *EXPY* with $k_{c,0}$, $k_{c,1}$, $k_{p,0}$ *and* $k_{p,1}$ for null model 2. A *PRODY* v/s $k_{p,0}$ B *PRODY* v/s $k_{p,1}$ C *EXPY* v/s $k_{c,0}$ D *EXPY* v/s , $k_{c,1}$

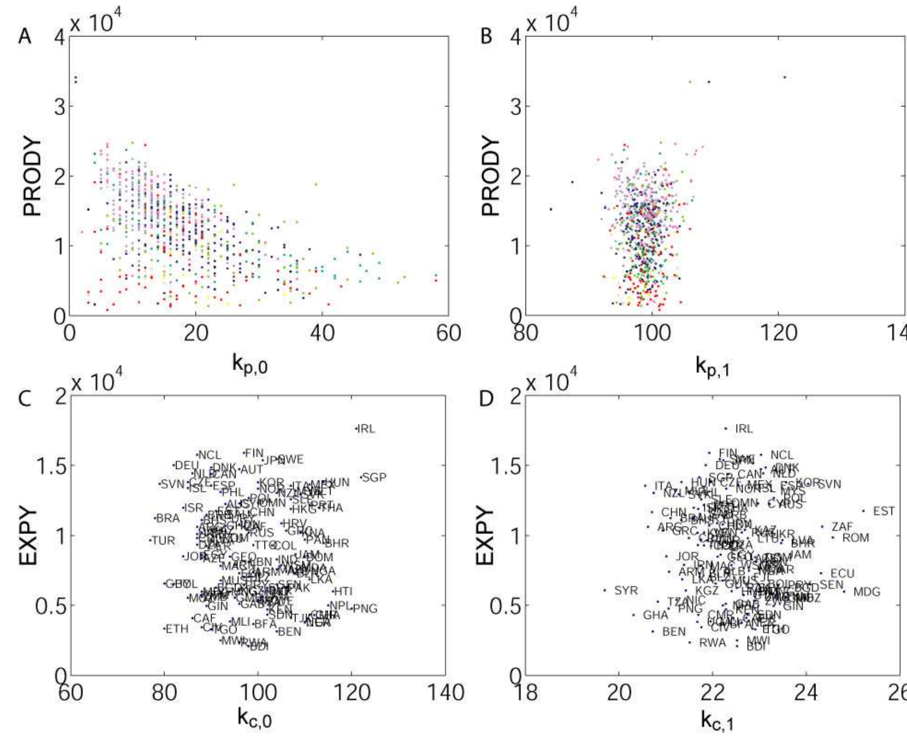

Fig S 21 Comparison between *PRODY* and *EXPY* with $k_{c,0}$, $k_{c,1}$, $k_{p,0}$ *and* $k_{p,1}$ for null model 3. A *PRODY* v/s $k_{p,0}$ B *PRODY* v/s $k_{p,1}$ C *EXPY* v/s $k_{c,0}$ D *EXPY* v/s , $k_{c,1}$

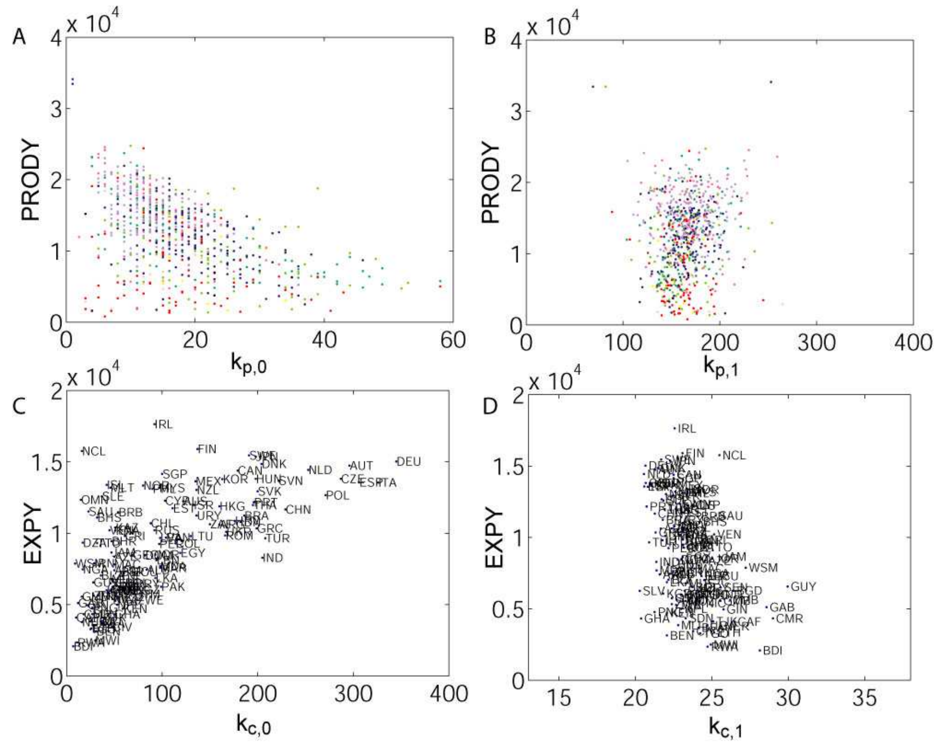

Fig S 22 Comparison between *PRODY* and *EXPY* with $k_{c,0}$, $k_{c,1}$, $k_{p,0}$ *and* $k_{p,1}$ for null model 4. A *PRODY* v/s $k_{p,0}$ B *PRODY* v/s $k_{p,1}$ C *EXPY* v/s $k_{c,0}$ D *EXPY* v/s , $k_{c,1}$

BIPARTITE NETWORK ANALYSIS AND PROXIMITY IN THE PRODUCT SPACE

We study the relationship between the analysis presented here and the proximity between products in the product space by asking if products that are close in the κ–θ diagram are proximate in The Product Space.

Proximity in the product space is defined as the minimum pair-wise conditional probability of co-exporting products p_1 and p_2. We can express this as a function of M as:

$$\phi_{p_1p_2} = \min\left\{\frac{\sum_c M_{cp_1}M_{cp_2}}{\sum_c M_{cp_1}}\middle|\frac{\sum_c M_{cp_1}M_{cp_2}}{\sum_c M_{cp_2}}\right\}. \qquad (16)$$

We expect pairs of products co-exported by a large fraction of countries (i.e. pairs of products having a large ϕ) to have a similar $k_{p,0}$ and $k_{p,1}$. We control for randomness by using our four null models, as these can be used to compare the relationship between $k_{p,0}$ and $k_{p,1}$ and ϕ for networks that are similar to M_{cp}. The four null-models allow us to study variations in the relationships between $k_{p,0}$, $k_{p,1}$ and ϕ that come from the network structure, rather than from their definition.

Proximity (ϕ) is a quantity associated with a pair of products. We compare ϕ to $k_{p,0}$ and $k_{p,1}$ by measuring the Euclidean distance in the $k_{p,0}$ and $k_{p,1}$ space:

$$\Delta_{p_1p_2} = \sqrt{(k_{p_1,0} - k_{p_2,0})^2 + (k_{p_1,1} - k_{p_1,1})^2}$$

$$\Delta_{p_1p_2} = \sqrt{\Delta k_{p,0}{}^2 + \Delta k_{p,1}{}^2}. \qquad (17)$$

We study the relationship between the distance in the $k_{p,0}$-$k_{p,1}$ space and ϕ (Fig S 23) and find that high proximity values are likely only among products close by in the $k_{p,0}$-$k_{p,1}$ diagram. We notice that the null models do not give rise to proximities as high as the ones observed in the original data, suggesting that the high observed co-production of some pairs of products cannot be expected from chance, and hence, high proximity values indicate similarities between the productive structures required to produce such pairs of products.

These results also show that a good ϕ threshold is to consider ϕ>0.5, as ϕ values above that threshold are extremely rare in any of the four null models.

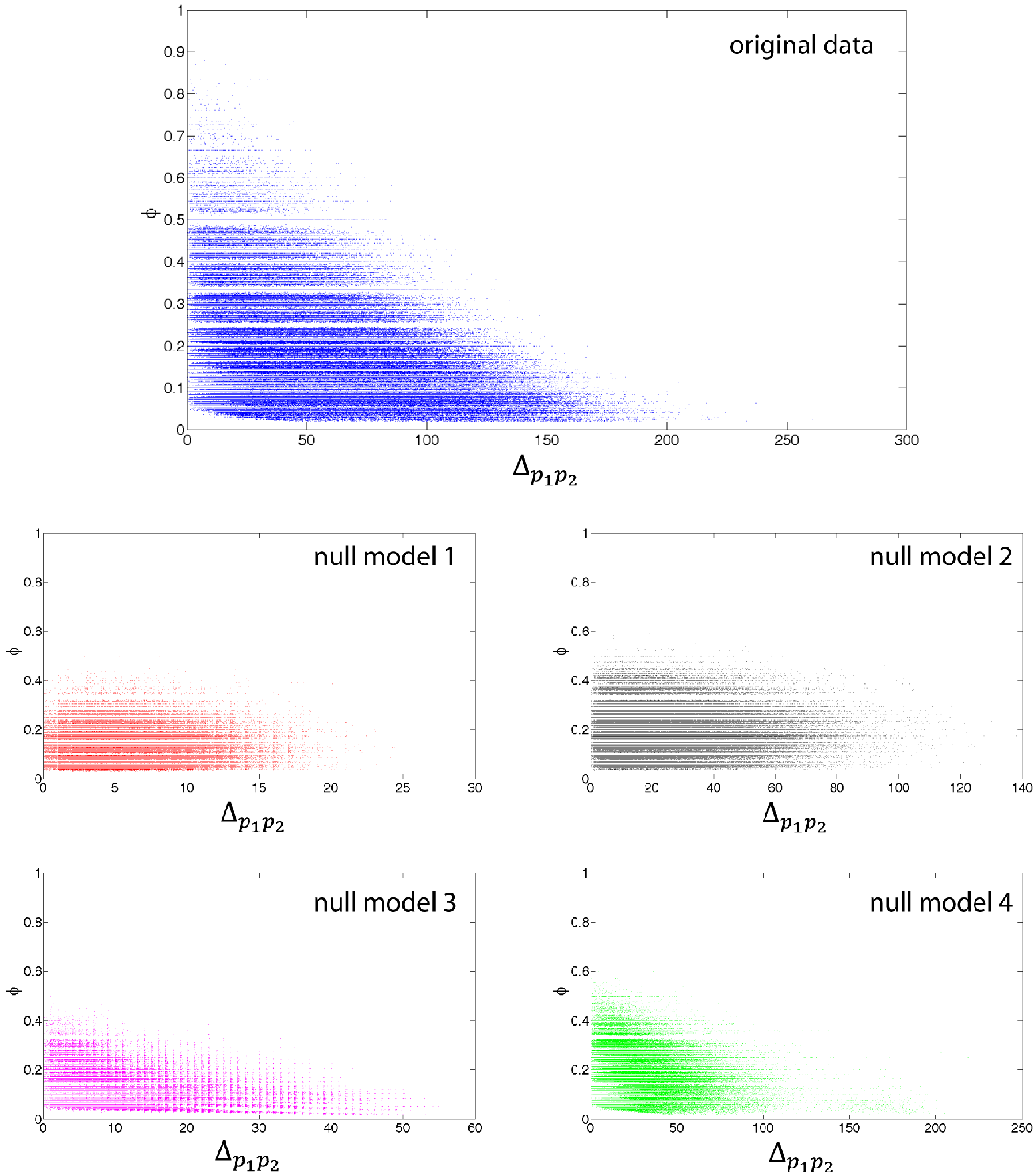

Fig S 23 Bipartite network structure and product proximity. The five plots show proximity as a function of the Euclidean distance between products in the $k_{p,0}$-$k_{p,1}$ diagram.

REFERENCES

[1] PG Lind, MC González, HJ Herrmann. Cycles and clustering in bipartite networks *Phys. Rev. E,* 72:056127 (2005)
[2] R Guimerà, M Sales-Pardo, LAN Amaral. Module identification in bipartite and directed networks *Phys. Rev. E*, 76:036102 (2007)
[3] S Lehmann, M Schwartz, LK Hansen. Bi-clique Communities *Phys. Rev. E,* 78:016108 (2008)
[4] K-I Goh et al. The Human Disease Network, *PNAS*, 104:8685-8690 (2007)
[5] W Souma, Y Fujiwara, H Aoyama Complex Networks and Economics *Physica A* 324:396-401 (2003)
[6] A.-L. Barabási, R. Albert. Emergence of scaling in random networks *Science* 286:509–512 (1999)
[7] Watts, D.J.; Strogatz, S.H..Collective dynamics of 'small-world' networks. *Nature* 393(6684): 409–10 (1998)
[8] MEJ Newman. The structure of scientific collaboration networks *PNAS,* 98-404-409 (2001)
[9] MEJ Newman. Scientific collaboration networks I Network Construction and Fundamental Results. *Phys. Rev. E*, 64:016131 (2001)
[10] MEJ Newman. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality *Phys. Rev. E*, 64:016132 (2001)
[11] AL Barabási et al. Evolution of the social network of scientific collaborations *Physica A* 311:590-614(2002)
[12] LAN Amaral et al. Classes of small-world networks *PNAS* 97:11149-11152 (2000)
[13] H Jeong, Z Neda, A-L Barabasi Measuring preferential attachment in evolving networks *Europhysics Letters* 61: 567-572 (2003) (2003)
[14] P Gleiser, L Danon Community Structure in Jazz arxiv/cond-mat/0307434 (2003)
[15] R. Hausmann, Hwang, D. Rodrik (2007) *Journal of Economic Growth*, 12(1):1-25 (2007)
[16] S Maslov, K Sneppen Specificity and stability in topology of protein networks, *Science*, 296**:**910-913 (2002)