

The building blocks of economic complexity

César A. Hidalgo¹ and Ricardo Hausmann

Center for International Development and Harvard Kennedy School, Harvard University, Cambridge, MA 02138

Edited by Partha Sarathi Dasgupta, University of Cambridge, Cambridge, United Kingdom, and approved May 1, 2009 (received for review January 28, 2009)

For Adam Smith, wealth was related to the division of labor. As people and firms specialize in different activities, economic efficiency increases, suggesting that development is associated with an increase in the number of individual activities and with the complexity that emerges from the interactions between them. Here we develop a view of economic growth and development that gives a central role to the complexity of a country's economy by interpreting trade data as a bipartite network in which countries are connected to the products they export, and show that it is possible to quantify the complexity of a country's economy by characterizing the structure of this network. Furthermore, we show that the measures of complexity we derive are correlated with a country's level of income, and that deviations from this relationship are predictive of future growth. This suggests that countries tend to converge to the level of income dictated by the complexity of their productive structures, indicating that development efforts should focus on generating the conditions that would allow complexity to emerge to generate sustained growth and prosperity.

economic development | networks

For Adam Smith, the secret to the wealth of nations was related to the division of labor. As people and firms specialize in different activities, economic efficiency increases. This division of labor, however, is limited by the extent of the market: The bigger the market, the more its participants can specialize and the deeper the division of labor that can be achieved. This suggests that wealth and development are related to the complexity that emerges from the interactions between the increasing number of individual activities that conform an economy (1–3).

Now, if all countries are connected to each other through a global market for inputs and outputs so that they can exploit a division of labor at the global scale, why have differences in Gross Domestic Product (GDP) per capita exploded over the past 2 centuries? (4, 5, *) One possible answer is that some of the individual activities that arise from the division of labor described above cannot be imported, such as property rights, regulation, infrastructure, specific labor skills, etc., and so countries need to have them locally available to produce. Hence, the productivity of a country resides in the diversity of its available nontradable “capabilities,” and therefore, cross-country differences in income can be explained by differences in economic complexity, as measured by the diversity of capabilities present in a country and their interactions.

During the last 20 years, models of economic growth have often included the assumption that the variety of inputs that go into the production of the goods produced by a country affects that country's overall productivity (3, 6). There have been very few attempts, however, to bring this intuition to the data. In fact, the most frequently cited surveys of the empirical literature do not incorporate a single reference to any measure of diversity of inputs or complexity (7).

We can create indirect measures of the capabilities available in a country by thinking of each capability as a building block or Lego piece. In this analogy, a product is equivalent to a Lego model, and a country is equivalent to a bucket of Legos. Countries will be able to make products for which they have all of the necessary capabilities, just like a child is able to produce a Lego model if the child's bucket contains all of the necessary Lego pieces. Using this analogy,

the question of economic complexity is equivalent to asking whether we can infer properties such as the diversity and exclusivity of the Lego pieces inside a child's bucket by looking only at the models that a group of children, each with a different bucket of Legos, can make. Here we show that this is possible if we interpret data connecting countries to the products they export as a bipartite network and assume that this network is the result of a larger, tripartite network, connecting countries to the capabilities they have and products to the capabilities they require (Fig. 1A). Hence, connections between countries and products signal the availability of capabilities in a country just like the creation of a model by a child signals the availability of a specific set of Lego pieces.

Note that this interpretation says nothing of the processes whereby countries accumulate capabilities and the characteristics of an economy that might affect them. It just attempts to develop measures of the complexity of a country's economy at a point in time. However, the approach presented here can be seen as a building block of a theory that accounts for the process by which countries accumulate capabilities. A detailed analysis of capability accumulation is beyond the scope of this article but the implications of our approach will be discussed briefly in *Discussion*.

In this article we develop a method to characterize the structure of bipartite networks, which we call the Method of Reflections, and apply it to trade data to illustrate how it can be used to extract relevant information about the availability of capabilities in a country. We interpret the variables produced by the Method of Reflections as indicators of economic complexity and show that the complexity of a country's economy is correlated with income and that deviations from this relationship are predictive of future growth, suggesting that countries tend to approach the level of income associated with the capability set available in them. We validate our measures of the capabilities available in a country by introducing a model and by showing empirically that our metrics are strongly correlated with the diversity of the labor inputs used in the production of a country's goods, approximated by using data on the use of labor inputs in the United States. Finally, we show that the level of complexity of a country's economy predicts the types of products that countries will be able to develop in the future, suggesting that the new products that a country develops depend substantially on the capabilities already available in that country.

Methods

We look at country product associations by using international trade data with products disaggregated according to 3 alternative data sources and classifications: First, the Standard International Trade Classification (SITC) revision 4 at the 4-digit level (see ref. 8; the data are available at www.nber.org/data, <http://cid.econ.udavis.edu/data/undata/undata.html>, and www.chidalgo.com/

Author contributions: C.A.H. and R.H. designed research, performed research, contributed new reagents/analytic tools, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

¹To whom correspondence should be addressed. E-mail: cesar.hidalgo@ksg.harvard.edu.

*In ref. 4, Maddison presents GDP per capita measures for 60 countries since 1820. In that year, the ratio of the 95th to the 5th percentile was 3.18 but it increased to 17.82 by the year 2000. Today, the U.S. GDP per capita is >60 times higher than Malawi's.

This article contains supporting information online at www.pnas.org/cgi/content/full/0900943106/DCSupplemental.

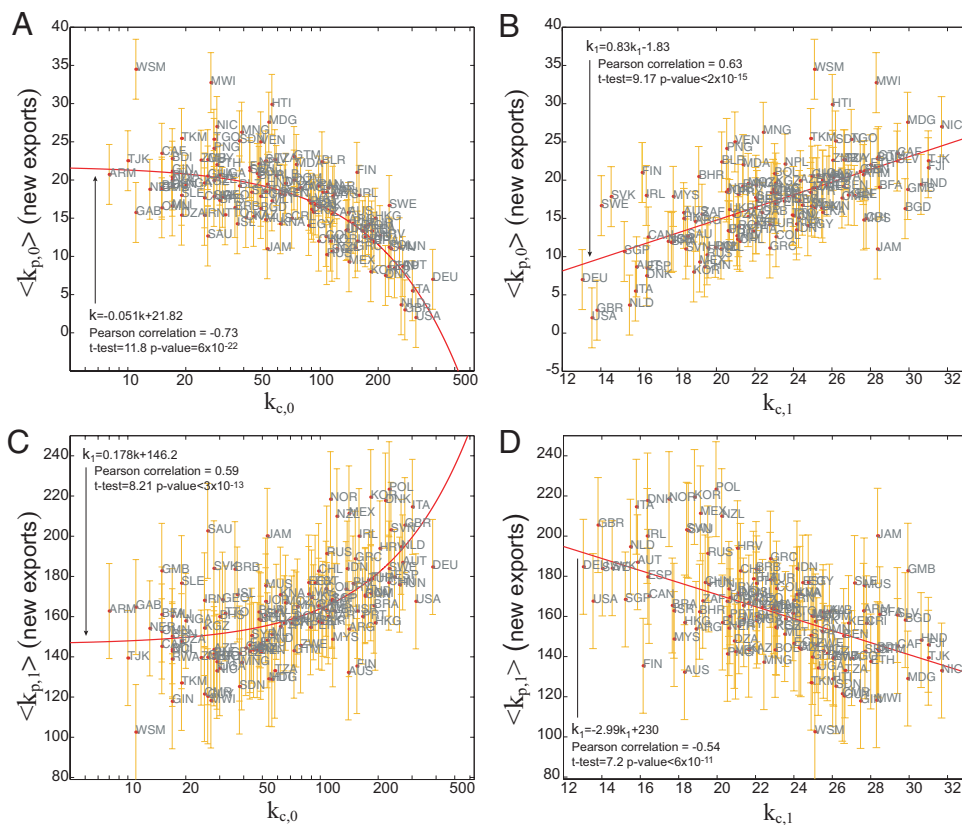


Fig. 4. Path dependent development. Average network properties ($\langle k_{p,0} \rangle$, $\langle k_{p,1} \rangle$; measured in 1992) of the new exports developed by a country between 1992 and 2000 as a function of the diversification of a country $k_{c,0}$ and the average ubiquity of its products $k_{c,1}$ measured in 1992. (A) $k_{c,0}$ vs. $\langle k_{p,0} \rangle$. (B) $k_{c,1}$ vs. $\langle k_{p,0} \rangle$. (C) $k_{c,0}$ vs. $\langle k_{p,1} \rangle$. (D) $k_{c,1}$ vs. $\langle k_{p,1} \rangle$.

number of employment categories going into the export basket of countries and our family of measures of diversification ($k_{c,0}, k_{c,2}, k_{c,4}, \dots, k_{c,2N}$). We also find a negative correlation between the average number of employment categories and measures of the ubiquity of products made by a country ($k_{c,1}, k_{c,3}, k_{c,5}, \dots, k_{c,2N+1}$) (Fig. 2D). This shows that more diversified countries indeed produce more complex products, in the sense that they require a wider combination of human capabilities, and that \bar{k}_c is able to capture this information.

Complexity of the Productive Structure, Income and Growth. We show that the information extracted by the method of reflections is connected to income by looking at the first 3 measures of diversification of a country ($k_{c,0}, k_{c,2}, k_{c,4}$) versus GDP per-capita adjusted for Purchasing Power Parity (PPP) (Fig. 3A–C). To make these 3 different measures comparable we have normalized them by subtracting their respective means ($\langle k_N \rangle$) and dividing them by their respective standard deviations ($\text{stdev}(k_N)$). As we iterate the method the relative ranking of countries defined by these variables shifts (Fig. 3D and SI Appendix, Fig. S14), making our measures of diversification and ubiquity increasingly more correlated with income (Fig. 3E and SI Appendix, Section 11). This can be illustrated by looking at the position, in the $k_{c,N}$ –GDP diagrams, of 3 countries that exported a similar number of products in the year 2000, albeit having large differences in income (Pakistan (PAK), Chile (CHL) and Singapore (SGP) Fig. 3A–C). Higher reflections of our method are able to correctly differentiate the income level of these countries because they incorporate information about the ubiquity of the products they export and about the diversification of other countries connected indirectly to them in M_{cp} , altering their relative rankings (Fig. 3D and SI Appendix, Fig. S14). For example, $k_{c,2}$ is

able to correctly separate Singapore, Chile and Pakistan, because it considers that in the bipartite network Singapore is connected to diversified countries mainly through nonubiquitous products, signaling the availability in Singapore of capabilities that are required to produce goods in diversified countries. In contrast, Pakistan is connected mostly to poorly diversified countries, and most of its connections are through ubiquitous products, indicating that Pakistan has capabilities that are available in most countries and that its relatively high level of diversification is probably due to its relatively large population, rather than to the complexity of its productive structure. Indeed, we find the method of reflections to be an accurate way to control for a country’s population, as correlations between \bar{k}_c and population decrease rapidly as we iterate the method (see SI Appendix, Section 11), whereas correlations between \bar{k}_c and GDP increase as we iterate the method. This is another piece of evidence suggesting that the information captured by our method is related to factors that affect the ability to generate per capita income.

Deviations from the correlation between \bar{k}_c and income are good predictors of future growth, indicating that countries tend to approach the levels of income that correspond to their measured complexity. We show this by regressing the rate of growth of income per capita on successive generations of our measures of economic complexity (i.e., $k_{c,0}, k_{c,1}$ or $k_{c,10}, k_{c,11}$) and on a country’s initial level of income

$$\log\left(\frac{\text{GDP}(t + \Delta t)}{\text{GDP}(t)}\right) = a + b_1 \text{GDP}(t) + b_2 k_{c,N}(t) + b_3 k_{c,N+1}(t),$$

finding that successive generations of the variables constructed in the previous section are increasingly good predictors of growth. In SI Appendix, Section 13, we present regression tables showing that these results are valid for a 20-year period (1985–2005), two 10-year

to use less labor inputs in the production of products than what would be reported from U.S. labor data, accentuating the effect presented in Fig. 2D.

