Science AAAS

The Product Space Conditions the Development of Nations

C. A. Hidalgo, *et al.*

Science **317**, 482 (2007);
DOI: 10.1126/science.1144581

ing epibiotic and pelagic communities. These icebergs can be compared to estuaries that supply surrounding coastal regions with nutrients. In that respect, icebergs may be thought of as "Lagrangian estuaries," drifting through the Southern Ocean while enriching the surrounding pelagic zone. Our preliminary studies suggest that free-drifting icebergs and their associated communities could serve as areas of increased production and sequestration of organic carbon to the deep sea, a process unaccounted for in current global carbon budgets (*33*).

References and Notes

1. I. Velicogna, J. Wahr, *Science* **311**, 1754 (2006).
2. R. Thomas *et al.*, *Science* **306**, 255 (2004).
3. D. W. J. Thompson, S. Solomon, *Science* **296**, 895 (2002).
4. E. Rignot, S. S. Jacobs, *Science* **296**, 2020 (2002).
5. D. G. Vaughan, G. J. Marshall, W. M. Connolley, J. C. King, R. Mulvaney, *Science* **293**, 1777 (2001).
6. A. J. Cook, A. J. Fox, D. G. Vaughan, J. G. Ferrigno, *Science* **308**, 541 (2005).
7. T. A. Scambos, C. Hulbe, M. Fahnestock, J. Bohlander, *J. Glaciol.* **46**, 516 (2000).
8. R. N. Williams, W. G. Rees, N. W. Young, *Int. J. Remote Sens.* **20**, 3183 (1999).
9. O. Orheim, *Ann. Glaciol.* **11**, 205 (1988).
10. M. Kristensen, *Prog. Phys. Geogr.* **7**, 313 (1983).
11. G. Stone, *Nat. Geog. Mag.,* December 2001, pp. 36–52.
12. K. R. Arrigo, G. L. van Dijken, D. G. Ainley, M. A. Fahnestock, T. Markus, *Geophys. Res. Lett.* **29**, 10.1029/2001GL014160 (2002).
13. H. J. W. de Baar *et al.*, *Nature* **373**, 412 (1995).
14. R. S. Kaufmann *et al.*, *Mar. Biol.* **124**, 387 (1995).
15. D. G. Ainley, E. F. O'Connor, R. J. Boekelheide, *Ornithol. Monogr*. **32**, 1 (1984).
16. C. R. Joiris, *Polar Biol.* **11**, 415 (1991).
17. C. A. Ribic, D. G. Ainley, W. R. Fraser, *Antarct. Sci.* **3**, 181 (1991).
18. Materials and methods are available as supporting material on *Science* Online.
19. Shipboard acoustic doppler current profiler data taken during the cruise were analyzed by T. Chereskin (University of California, San Diego; Scripps Institution of Oceanography) to provide a description of surface currents surrounding icebergs A-52 and W-86.
20. B. M. Loscher, H. J. W. de Baar, J. T. M. de Jong, C. Veth, F. Dehairs, *Deep-Sea Res. II* **44**, 143 (1997).
21. H. J. W. de Baar, J. T. M. de Jong, in *The Biogeochemistry of Iron in Seawater*, D. R. Turner, K. A. Hunter, Eds. (Wiley, New York, 2001), pp. 123–253.
22. W. O. Smith, D. M. Nelson, *Science* **227**, 163 (1985).
23. K. H. Coale *et al.*, *Science* **304**, 408 (2004).
24. P. W. Boyd *et al.*, *Science* **315**, 612 (2007).
25. E. A. Pakhomov, P. W. Froneman, R. Perissinotto, *Deep-Sea Res. II* **49**, 1881 (2002).
26. G. Sugihara, L.-F. Bersier, T. R. Southwood, S. L. Pimm, R. M. May, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5246 (2003).
27. W. G. Sunda, S. A. Huntsman, *Mar. Chem.* **50**, 189 (1995).
28. M. L. Wells, N. G. Zorkin, A. G. Lewis, *J. Mar. Res.* **41**, 731 (1983).
29. H. W. Rich, F. M. M. Morel, *Limnol. Oceanogr.* **35**, 652 (1990).
30. T. D. Waite, F. M. M. Morel, *J. Colloid Interface Sci.* **102**, 121 (1984).
31. D. A. S. Finden, E. Tipping, G. H. M. Jaworski, C. S. Reynolds, *Nature* **309**, 783 (1984).
32. S. Kraemer, A. Butler, P. Borer, J. Cervini-Silva, *Rev. Mineral. Geochem.* **59**, 53 (2005).
33. I. Marinov, A. Gnanadesiker, J. R. Toggweiler, J. L. Sarmiento, *Nature* **441**, 964 (2006).

34. We thank all the shipboard scientific personnel on the research vessel *Laurence M. Gould* cruise LMG05-14A for excellent support, including R. Wilson, K. Reisenbichler, R. Sherlock, J. Ellena, M. Vardaro, K. Osborn, D. Chakos, J. Derry, L. Ekern, J. Kinsey, C. Koehler, and K. Noble. Captain R. Verret and his crew made sampling around icebergs a reality even under the most difficult conditions. The Raytheon Polar Services support group of S. Suhr-Sliester, J. Spillane, P. Fitzgibbons, K. Pedigo, J. Dolan, E. Roggenstein, and D. Elsberg provided excellent deck and laboratory support. D. Long (Brigham Young University) provided timely QuikSCAT images of the location of iceberg A-52 during our cruise. RADARSAT images of our study area were forwarded to the ship through Palmer Station. This research was supported by NSF grants ANT-0529815, ANT-0650034, and OCE-0327294, and by the David and Lucile Packard Foundation. We thank P. Penhale (NSF, Polar Programs) for having the foresight and courage to fund this speculative project. W. Moore and C. Hexel contributed to the ^{224}Ra analysis and data synthesis. C. Stoker of NASA/Ames Research Center loaned us the ROV, and H. Thomas at MBARI trained us in its operation.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1142834/DC1
Materials and Methods
Figs. S1 to S3
Table S1
References

21 March 2007; accepted 4 June 2007
Published online 21 June 2007;
10.1126/science.1142834
Include this information when citing this paper.

The Product Space Conditions the Development of Nations

C. A. Hidalgo,[1]*† B. Klinger,[2]* A.-L. Barabási,[1] R. Hausmann[2]

Economies grow by upgrading the products they produce and export. The technology, capital, institutions, and skills needed to make newer products are more easily adapted from some products than from others. Here, we study this network of relatedness between products, or "product space," finding that more-sophisticated products are located in a densely connected core whereas less-sophisticated products occupy a less-connected periphery. Empirically, countries move through the product space by developing goods close to those they currently produce. Most countries can reach the core only by traversing empirically infrequent distances, which may help explain why poor countries have trouble developing more competitive exports and fail to converge to the income levels of rich countries.

Does the type of product that a country exports matter for subsequent economic performance? The fathers of development economics held that it does, suggesting that industrialization creates spillover benefits that fuel subsequent growth (*1*–*3*). Yet, lacking formal models, mainstream economic theory has been unable to incorporate these ideas. Instead, two approaches have been used to explain a country's pattern of specialization. The first focuses on the relative proportion between productive factors (i.e., physical capital, labor, land, skills or human capital, infrastructure, and institutions) (*4*). Hence, poor countries specialize in goods intensive in unskilled labor and land, whereas richer countries specialize in goods requiring infrastructure, institutions, and human and physical capital. The second approach emphasizes technological differences (*5*) and has to be complemented with a theory of what underlies them. The varieties and quality ladders models (*6*, *7*) assume that there is always a slightly more advanced product, or just a different one, that countries can move to, disregarding product similarities when thinking about structural transformation and growth.

Think of a product as a tree and the set of all products as a forest. A country is composed of a collection of firms, i.e., of monkeys that live on different trees and exploit those products. The process of growth implies moving from a poorer part of the forest, where trees have little fruit, to better parts of the forest. This implies that monkeys would have to jump distances, that is, redeploy (human, physical, and institutional) capital toward goods that are different from those currently under production. Traditional growth theory assumes there is always a tree within reach; hence, the structure of this forest is unimportant. However, if this forest is heterogeneous, with some dense areas and other more-deserted ones, and if monkeys can jump only limited distances, then monkeys may be unable to move through the forest. If this is the case, the structure of this space and a country's orientation within it become of great importance to the development of countries.

In theory, many possible factors may cause relatedness between products, that is, closeness between trees; such as the intensity of labor, land, and capital (*8*), the level of technological sophistication (*9*, *10*), the inputs or outputs involved in a product's value chain (e.g., cotton, yarn, cloth, and garments) (*11*), or requisite insti-

[1]Center for Complex Network Research and Department of Physics, University of Notre Dame, Notre Dame, IN 46556, USA. [2]Center for International Development, Kennedy School of Government, Harvard University, Cambridge, MA 02139, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: chidalgo@nd.edu

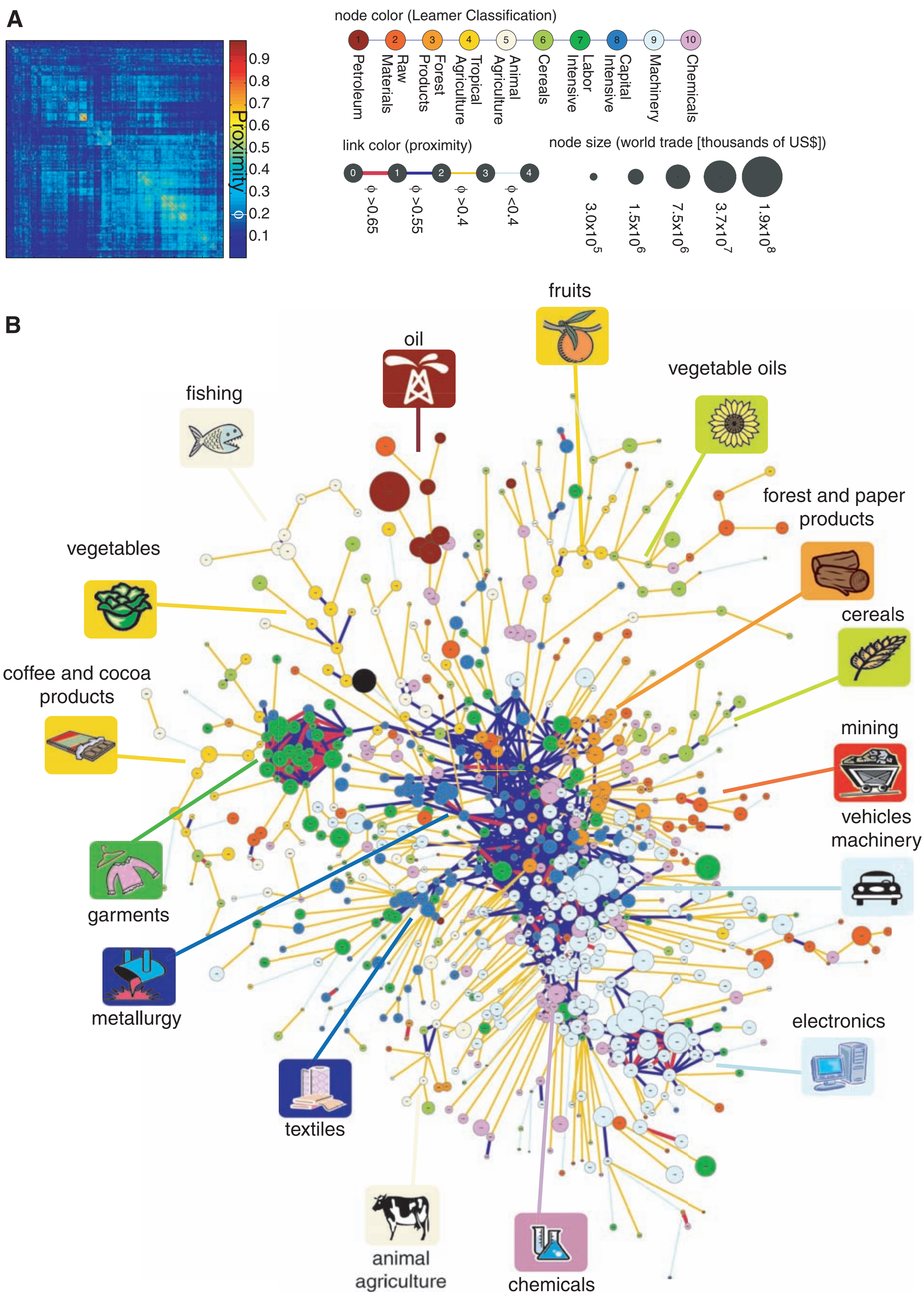

Fig. 1. The product space. (**A**) Hierarchically clustered proximity (φ) matrix representing the 775 SITC-4 product classes exported in the 1998–2000 period. (**B**) Network representation of the product space. Links are color coded with their proximity value. The sizes of the nodes are proportional to world trade, and their colors are chosen according to the classification introduced by Leamer.

tutions (*12*, *13*). All of these are a priori notions of what dimension of similarity are most important and assume that factors of production, technological sophistication, or institutional quality exhibit little specificity. Instead, we take an agnostic approach and use an outcomes-based measure, based on the idea that, if two goods are related because they require similar institutions, infrastructure, physical factors, technology, or some combination thereof, they will tend to be produced in tandem, whereas dissimilar goods are less likely to be produced together. We call this measure "proximity," which formalizes the intuitive idea that the ability of a country to produce a product depends on its ability to produce other products. For example, a country with the ability to export apples will probably have most of the conditions suitable to export pears. They would certainly have the soil, climate, packing technologies, and frigorific trucks. In addition, they would have skilled agronomists, phytosanitary laws, and trade agreements that could be easily redeployed to the pear business. If instead we consider a different product such as copper wires or home appliance manufacture, all or most of the capabilities developed for the apple business render useless. We introduce proximity as the concept that captures this intuitive notion.

The concept of proximity. Formally, the proximity ϕ between products i and j is the minimum of the pairwise conditional probabilities of a country exporting a good given that it exports another.

$$\phi_{i,j} = \min\{P(RCAx_i|RCAx_j), P(RCAx_j|RCAx_i)\}$$

Where RCA stands for revealed comparative advantage (*14*)

$$RCA_{c,i} = \frac{\dfrac{x(c,i)}{\sum_i x(c,i)}}{\dfrac{\sum_c x(c,i)}{\sum_{c,i} x(c,i)}}$$

which measures whether a country c exports more of good i, as a share of its total exports, than the "average" country ($RCA > 1$ not $RCA < 1$).

We used international trade data, cleaned and made compatible (*15*) through a National Bureau of Economic Research (NBER) project lead by R. Feenstra (*16*), disaggregated according to the Standardized International Trade Code at the four-digit level (SITC-4), providing for each country the value exported to all other countries for 775 product classes. With these data, we calculated the 775-by-775 matrix of revealed proximities between every pair of products by using the equation above.

A hierarchically clustered version of the matrix is shown (Fig. 1A). A smooth and homogeneous product space would imply uniform values (homogenous coloring), whereas a product-ladder model (*7*) would suggest a matrix with high values (or bright coloring) only along the diagonal. Instead the product space of Fig. 1A appears to be modular (*17*, *18*), with some goods highly connected and others disconnected. Furthermore, as a whole the product space is sparse, with ϕ_{ij} distributed according to a broad distribution (fig. S2) with 5% of its elements equal to zero, 32% of them smaller than 0.1, and 65% of the entries taking values below 0.2. These substantial number of negligible connections call for a network representation (*19*), allowing us to explore the structure of the product space together with the proximity between products of given classifications and participation in world trade. To offer a visualization in which all 775 products are included, we reached all nodes by calculating the maximum spanning tree, which includes the 774 links maximizing the tree's added proximity (fig. S4) and superposed on it all links with a proximity larger than 0.55 (figs. S5 and S6). This set of 1525 links is used to visualize the structure of the full proximity matrix, which is far from homogenous and appears to have a core-periphery structure (Fig. 1B). The core is formed by metal products, machinery, and chemicals, whereas the periphery is formed by the rest of the product classes. The products in the top of the periphery belong to fishing, tropical, and cereal agriculture. To the left there is a strong peripheral cluster formed by garments and another one belonging to textiles, followed by animal agriculture. The bottom of the network shows a large electronics cluster, followed to the right by mining, forest, and paper products.

The network shows clusters of products somewhat related to the classification introduced by

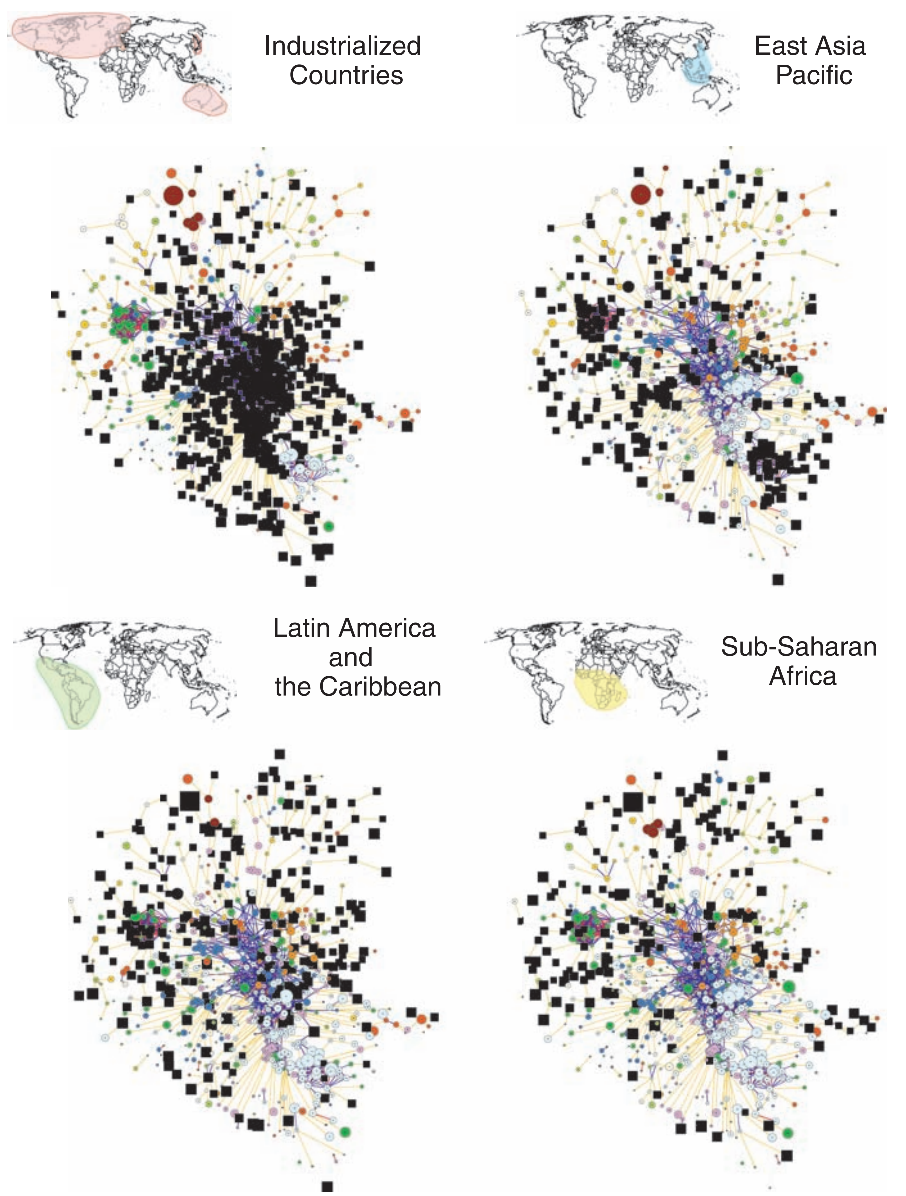

Fig. 2. Localization of the productive structure for different regions of the world. The products for which the region has an *RCA* > 1 are denoted by black squares.

Leamer (*8*), which is based on relative factor intensities (table S1 and fig. S8), that is, the relative amount of capital, labor, land, or skills required to produce each product. Although the classification performed by Leamer was done with a different methodology, the agreement between it and the structure of the product space is striking. Yet it also introduces a more detailed split of some product classes. For example, machinery is naturally split into two clusters, one consisting of vehicles and heavy machinery and another one belonging to electronics. The machinery cluster is interwoven with some capital-intensive metal products but is not tightly connected to similarly classified products such as textiles.

The map obtained can be used to analyze the evolution of a country's productive structure. For this purpose we held the product space fixed and studied the dynamics of production within it, although changes in the product space represent an interesting avenue for future research (*20*).

The pattern of specialization for four regions in the product space is shown in Fig. 2 (*21*). Products exported by a region with $RCA > 1$ are shown with black squares. Industrialized countries occupy the core, composed of machinery, metal products, and chemicals. They also participate in more peripheral products such as textiles, forest products, and animal agriculture. East Asian countries have developed RCA in the garments, electronics, and textile clusters, whereas Latin America and the Caribbean are further out in the periphery in mining, agriculture, and the garments sector. Lastly, sub-Saharan Africa exports few product types, all of which are in the far periphery of the product space. These results indicate that each region has a distinguishable pattern of specialization clearly visible in the product space. Links to the maps for the 132 countries included in the study can be found in the Supporting Online Material (SOM) text.

Next, we show how the structure of the product space affects a country's pattern of specialization. Figure 3A shows how comparative advantages evolved in Malaysia and Colombia between 1980 and 2000 in the electronics and the garments sectors, respectively. Both countries follow a diffusion process in which comparative advantage move preferentially toward products close to existing goods: garments in Colombia and electronics in Malaysia.

Testing diffusion. Beyond this graphical illustration, is it true that countries develop comparative advantage preferentially in nearby goods? We used two different approaches to this question. First, we measured the average proximity of a new potential product j to a country's current productive structure, which we call "density" and define as

$$\omega_j^k = \frac{\sum_i x_i \phi_{ij}}{\sum_i \phi_{ij}}$$

where ω^k_j is the density around good j given the export basket of the kth country and $x_i = 1$ if $RCA_{ki} > 1$ and 0 otherwise. A high density value means that the kth country has many developed products surrounding the jth product. To study the evolution of comparative advantage, we considered "transition products" as those with an $RCA_{c,i} < 0.5$ in 1990 and an $RCA_{c,i} > 1$ in 1995. As a control, we considered "undeveloped products" those that in 1990 and 1995 had an $RCA_{c,i} < 0.5$ and disregarded those cases not fitting any of these two criteria. Figure 3B shows how density is distributed around transition products (yellow) and compares it to densities around undeveloped products (red). Clearly, these distributions are very distinct, with a higher density around transition products than among undeveloped ones [analysis of variance (ANOVA) $P < 10^{-30}$].

At the single product level, we considered the ratio between the average density of all countries in which the jth product was a transition product and the average density of all countries in which the jth product was not developed. Formally, we define the "discovery factor" H_j as

$$H_j = \frac{\sum_{k=1}^{T} \omega^k_j / T}{\sum_{k=T+1}^{N} \omega^k_j / (N-T)}$$

where T is the number of countries in which the jth good was a transition product and N is the total number of countries. Figure 3C shows the frequency distribution of this ratio. For 79% of products, this ratio is greater than 1, indicating that ω_j^k is greater in countries that transitioned into the j^{th} good than in those that did not, often substantially.

An alternative way of illustrating that countries develop RCA in goods close to those they

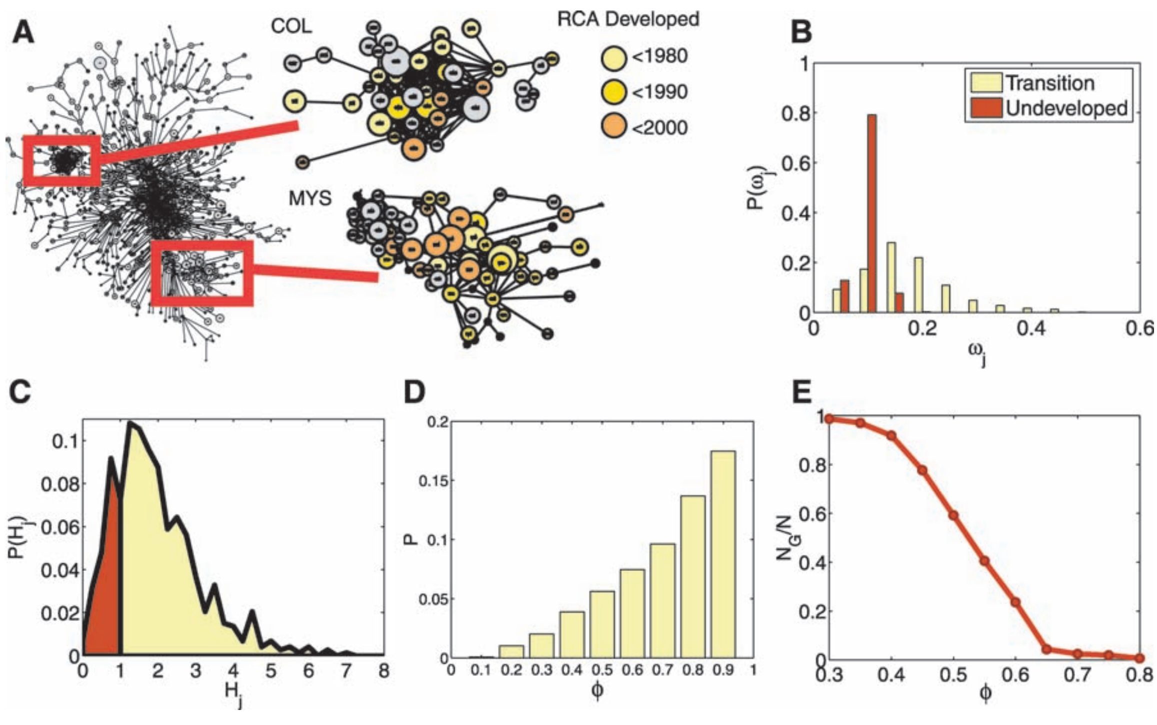

Fig. 3. Empirical evolution of countries. (**A**) Examples of RCA spreading for Colombia (COL) and Malaysia (MYS). The color code shows when this countries first developed $RCA > 1$ for products in the garments sector in Colombia and in the electronics cluster for Malaysia. (**B**) Distribution of density (ω) for transition products and undeveloped products (**C**) Distribution for the relative increase in density for products undergoing a transition with respect to the same products when they remain undeveloped. (**D**) Probability of developing RCA given that the closest connected product is at proximity ϕ. (**E**) Relative size of the largest connected component N_G with respect to the total number of products in the system N as a function of link ϕ.

already had is to calculate the conditional probability of transitioning into a product given that the nearest product with $RCA > 1$ is at a given ϕ. There is a monotonic relationship (Fig. 3D) between the proximity of the nearest developed good and the probability of transitioning into it. Although the probability of moving into a good at $\phi = 0.1$ in the course of 5 years is almost nil, the probability is about 15% if the closest good is at $\phi = 0.8$ (*22*).

Because production shifts to nearby products, we asked whether the product space is sufficiently connected that given enough time, all countries can reach most of it, particularly the richest parts. Lack of connectedness may explain the difficulties faced by countries trying to converge to the income levels of rich countries: they may not be able to undergo structural transformation because proximities are just too low. A simple approach is to calculate the relative size of the largest connected component as a function of ϕ. At $\phi \geq$

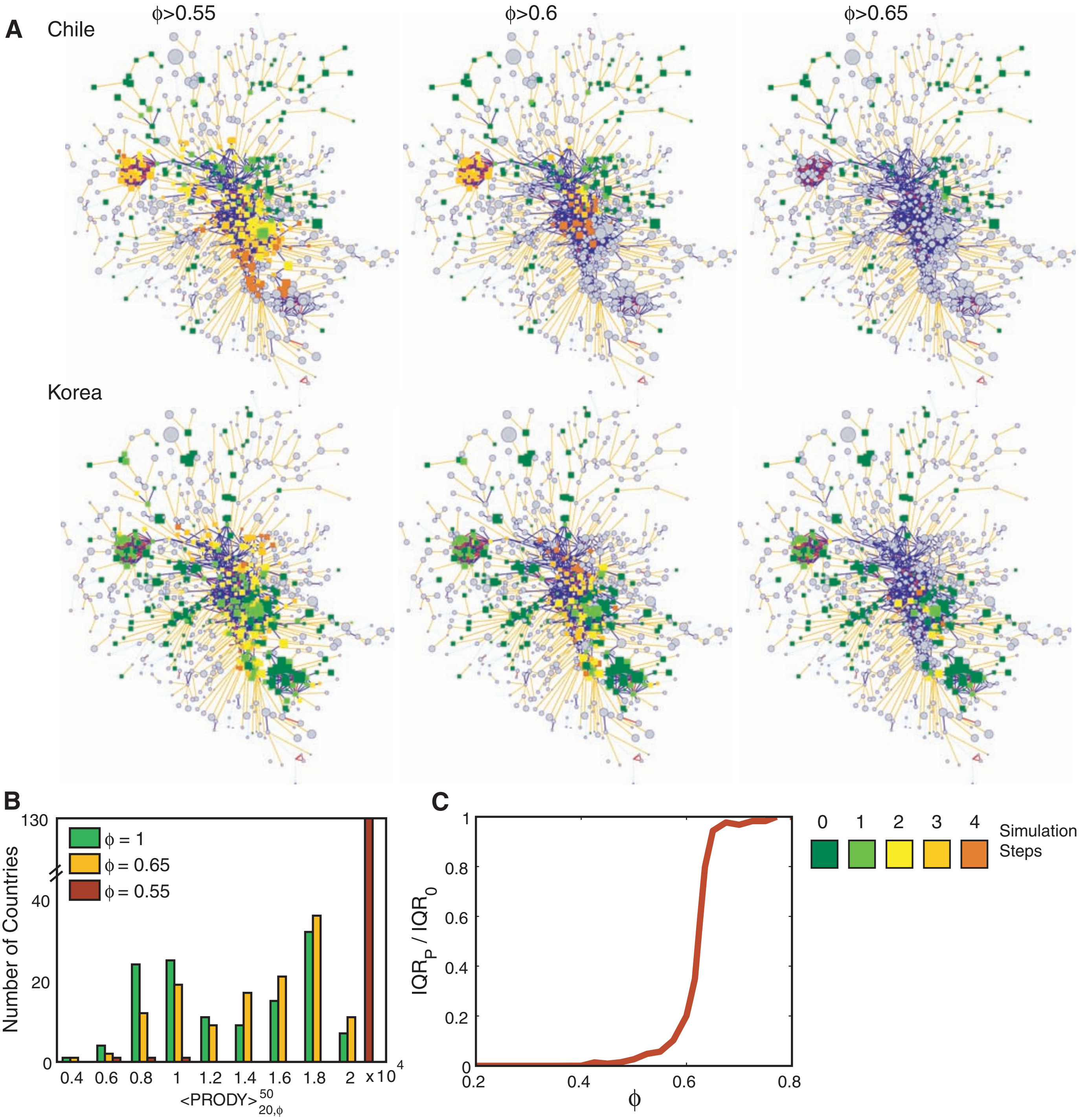

Fig. 4. Simulated diffusion process and inequality. (**A**) Simulated diffusion process for Chile and Korea in which we allowed countries to develop RCA in all products closer than ϕ values of 0.55, 0.6, and 0.65. The number of steps required to develop RCA can be read from the color code on the bottom right corner. (**B**) Distribution for the average PRODY of the best 50 products in a countries basket before and after 20 rounds of diffusion. The original distribution is shown in green, whereas the one associated with the distribution after 20 diffusion rounds with $\phi = 0.65$ is presented in yellow and $\phi = 0.55$ in red. (**C**) IQR of the distribution of the best 50 products after diffusing with a given ϕ normalized by the IQR of the best 50 products in absence of diffusion.

0.6, the largest connected component has a negligible size compared with the total number of products (Fig. 3E), whereas for $\phi \leq 0.3$ the product space is almost fully connected, meaning that there is always a path between two different products.

We studied the impact of the product space structure by simulating how the position of countries evolve when allowed to repeatedly move to products with proximities greater than a certain ϕ_o. If countries diffuse to nearby products and these are sufficiently connected to others, then after several iterations, 20 in our exercise, countries would be able to reach richer parts of the product space. On the other hand, if the product space is disconnected, countries will not be able to find paths to the richer part of the product space, independently of how many steps they are allowed to make.

The results of our simulation for Chile and Korea are presented in Fig. 4A. At a relatively low proximity (ϕ_o = 0.55), both countries are able to diffuse through to the core of the product space; however, Korea is able to do so much faster, thanks to its positioning in core products. For higher proximities, the question becomes whether a country can spread at all. At $\phi_o = 0.6$, Chile is able to spread slowly throughout the space, whereas Korea is still able to populate the core after four rounds. At $\phi_o = 0.65$, Chile is not able to diffuse, lacking any close-enough products, whereas Korea develops RCA slowly to a few products close to the machinery and electronics cluster.

To generalize this analysis for the whole world, we needed a measure to summarize the position of a country in the product space. We adopted a measure based on Hausmann, Hwang, and Rodrik (*23*), which involves a two-stage process. First, for every product we assigned a value, which is the weighted gross domestic product (GDP) per capita of countries with comparative advantage in that good, called PRODY (*23*). We then averaged the PRODYs of the top N products that a country has access to after M iterations at ϕ_o and denoted it by $< PRODY >^N_{M\phi_o}$. Figure 4B shows the distribution of $< PRODY >^N_{M\phi_o}$ for $N = 50$, $M = 20$, and $\phi_o = 1$ (green), $\phi_o = 0.65$ (yellow), and $\phi = 0.55$ (red). The distribution for $\phi_o = 1$ allows us to characterize the current distribution of countries in the product space, which shows a bimodal distribution, a signature of a world divided into rich and poor countries with few countries occupying the center of the distribution. When we allow countries to diffuse up to $\phi_o = 0.65$, this distribution does not change significantly: it shifts slightly to the right because of the acquisition of a limited number of sophisticated products by some countries. This diffusion process, however, stops after a few rounds, and the world maintains a degree of inequality similar to its current state. Contrarily, when we consider $\phi_o = 0.55$, most countries are able to diffuse and reach the most sophisticated basket in the long run. Only a few countries are left behind, which unsurprisingly make up the poorest end of the income distribution.

To quantify the level of convergence we calculated the interquartile range (IQR) for the $< PRODY >^{50}_{20\phi}$ distribution and normalized this quantity by dividing it with the IQR for the original distribution. Figure 4C shows that the convergence of the system goes through an abrupt transition and that convergence is possible if countries are able to diffuse to products located at a proximity $\phi > 0.65$.

The bimodal distribution of international income levels and a lack of convergence of the poor toward the rich has been explained by using geographic (*24*) and institutional (*12*, *13*) arguments. Here, we introduced another factor to this discussion: the difficulties involved in moving through the product space. The detailed structure of the product space is shown here and, together with the location of the countries and the characteristics of the diffusion process undergone by them, strongly suggests that not all countries face the same opportunities when it comes to development. Poorer countries tend to be located in the periphery, where moving toward new products is harder to achieve. More interestingly, among countries with a similar level of development and seemingly similar levels of production and export sophistication, there is significant variation in the option set implied by their current productive structure, with some on a path to continued structural transformation and growth and others stuck in a dead end.

These findings have important consequences for economic policy, because the incentives to promote structural transformation in the presence of proximate opportunities are quite different from those required when a country hits a dead end. It is quite difficult for production to shift to products far away in the space, and therefore policies to promote large jumps are more challenging. Yet it is precisely these long jumps that generate subsequent structural transformation, convergence, and growth.

References and Notes

1. A. Hirschman, *The Strategy of Economic Development* (Yale Univ. Press, New Haven, CT, 1958).
2. P. Rosenstein-Rodan, *Econ. J.* **53**, 202 (1943).
3. K. Matsuyama, *J. Econ. Theory* **58**, 317 (1992).
4. E. Heckscher, B. Ohlin, *Heckscher-Ohlin Trade Theory*, H. Flam, M. Flanders, Eds. (MIT Press, Cambridge, MA, 1991).
5. P. Romer, *J. Polit. Econ.* **94**, 5 (1986).
6. P. Aghion, P. Howitt, *Econometrica* **60**, 2 (1992).
7. G. Grossman, E. Helpman, *Rev. Econ. Stud.* **58**, 1 (1991).
8. E. Leamer, *Sources of Comparative Advantage*: *Theory and Evidence* (MIT Press, Cambridge, MA, 1984).
9. S. Lall, *Oxf. Dev. Stud.* **28**, 337 (2000).
10. R. Caballero, A. Jaffe, *NBER Macroeconom. Ann.* **8**, 15 (1993).
11. E. Dietzenbacher, M. Lahr, *Input-Output Analysis: Frontiers and Extensions* (Palgrave, New York, 2001).
12. D. Rodrik, A. Subramanian, F. Trebbi, *NBER Work. Pap. 9305* (2002).
13. D. Acemoglu, S. Johnson, J. A. Robinson, *Am. Econ. Rev.* **91**, 1369 (2001).
14. We use the Balassa definition (*25*) of revealed comparative advantage (Materials and Methods).
15. Because one country's exports are another country's imports, national statistics can be reconciled this way, and missing data from nonreporting countries can be completed.
16. R. R. Feenstra, H. D. Lipsey, A. Ma, H. Mo, *NBER Work. Pap. 11040* (2005).
17. E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, A.-L. Barabási, *Science* **297**, 1551 (2002).
18. G. Palla, I. Derenyi, I. Frakas, T. Vicsek, *Nature* **435**, 814 (2005).
19. Good introductions to networks are (*26*, *27*).
20. The network shown here represents the structure of the product space as determined from the 1998–2000 periods. Holding the product space as fixed is a good first approximation, because the dynamics of the network is much slower than the one of countries. The Pearson correlation coefficient (PCC) between the proximity of all links present in this network and the ones obtained from the same network in 1990 and 1985 are 0.69 and 0.66, respectively (SOM text). This indicates that, although the network changes over time, after 15 years the strength of past links still predicts the strength of the current links to a considerable extent.
21. An alternative approach in which the network of trade relationships was studied was undertaken by (*28*–*30*).
22. We repeated the same exercise with the rank of proximity instead of proximity itself in order to assess whether what matters is absolute or relative proximity. We found that absolute distance appears to be what matters most. Although transition probability increases linearly with proximity, they decay with rank as a power law. Moreover, the rank effect is stronger for products in sparser parts of the product space, where transitions are also less frequent. Thus, densely connected products can develop RCA through more paths than sparsely connected ones, indicating the importance of absolute proximity.
23. We follow the methodology developed in Hausmann, Hwang, and Rodrik (*31*), which weighs the GDP per capita of each country exporting that product by the RCA that the country has in that good.
24. J. Gallup, J. Sachs, A. Mellinger, *Int. Reg. Sci. Rev.* **22**, 179 (1999).
25. B. Balassa, *Rev. Econ. Stat.* **68**, 315 (1986).
26. R. A. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
27. G. Caldarelli, *Scale-Free Networks: Complex Webs in Nature and Technology* (Oxford Univ. Press, Oxford, 2007).
28. M. A. Serrano, M. Boguña, *Phys. Rev. E* **68**, 015101 (2003).
29. D. Garlaschelli, M. I. Loffredo, *Phys. Rev. Lett.* **93**, 188701 (2004).
30. D. Garlaschelli, T. Di Matteo, T. Aste, G. Caldarelli, M. I. Loffredo, http://arxiv.org/abs/physics/0701030.
31. R. Hausmann, J. Hwang, D. Rodrik, *NBER Work. Pap. 11905* (2006).

32. We would like to thank the following for valuable comments: P. Aghion, L. Alfaro, O. Blanchard, R. Caballero, O. Galor, E. Helpman, A. Khwaja, J. Lahey, R. Lawrence, D. Lederman, L. Pritchett, R. Rigobon, D. Rodrik, A. Rodriguez-Clare, C. Sabel, E. Stein, F. Sturzenegger, and D. Weil. C.A.H. acknowledges support from the Kellogg Institute at Notre Dame. C.A.H. and A.-L.B. acknowledge support from NSF grants ITR DMR-0426737 and IIS-0513650 and from the James McDonald Foundation 220020084.

Supporting Online Material

www.sciencemag.org/cgi/content/full/317/5837/482/DC1
Materials and Methods
SOM Text
Figs. S1 to S18
Table S1
References

2 May 2007; accepted 5 July 2007
10.1126/science.1144581

Supplementary Material

The Concept of Proximity

The Intuitive Definition

The concept of proximity formalizes the intuitive idea that the ability of a country to produce a product depends on its ability to produce other ones. For example, a country with the ability to export apples will probably have most of the conditions suitable to export pears. They would certainly have the soil and the climate, together with the appropriate packing technologies, frigorific trucks and containers. They would also have the human capital, particularly the agronomists that could easily learn the pear business. However, when we consider a different business such as mining, textiles or appliance manufacture, all or most of the capabilities developed for the apple business render useless. Unfortunately this intuitive definition of proximity is, very cumbersome to measure. It requires quantifying the overlap between the set of markets related to each product. Thus, we measure proximity by using an outcome based method founded on the assumption that similar products are more likely to be exported in tandem.

Details of the measurement

First, a stringent measure of exports is needed. We do not want to consider marginal exports, and thus we say that a country exports a product whenever they have Revealed Comparative Advantage (RCA) in it. We use the Balassa[1] definition of RCA which is given by

$$RCA(c,i) = \frac{\dfrac{x(c,i)}{\sum_i x(c,i)}}{\dfrac{\sum_c x(c,i)}{\sum_{i,c} x(c,i)}}$$

where ***x(c,i)*** is the value of the exports of country ***c*** in the ***i***'th good. Basically RCA is larger than one when the share of exports of country on a given product is larger than the share of that product on the global trade. This definition of RCA allows us to set a hard threshold for a countries export. When RCA(c,i) is greater or equal to 1 we say that country ***c*** exports product ***i***, and when RCA(c,i)<1 that country is not an effective exporter of that product.

Using RCA as an indication of a country effectively exporting a good, we define the proximity between goods ***i*** and ***j*** as:

$$\phi_{ij} = min\left\{ P(RCA_i|RCA_j), P(RCA_j|RCA_i) \right\}$$

where ***P(RCAi|RCAj)*** is the conditional probability of exporting good ***i*** given that you export good ***j***. In this definition we consider the minimum between both conditional probabilities because in the case that a country is the sole exporter of a particular good we would have that the conditional probability of exporting any other good given that one would be equal to one for all other goods exported by that country. The converse is not true and by taking the minimum we get rid of this problem and at the same time symmetrize the proximity matrix.

More details about the motivation of proximity and the option value associated with it were covered in the work of Hausmann and Klinger [2].

References

[1] B. Balassa, *The Review of Economics and Statistics*, **68**, 315 (1986).
[2] Ricardo Hausmann and Bailey Klinger, *Structural Transformation and Patterns of Comparative Advantage in the Product Space*, CID Working Paper No. 128, August 2006, -abstract-

Source Data

International trade data is taken from Feenstra, Lipsey, Deng, Ma, & 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 build 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

Basic Statistics

We define the product space as the set of all proximity measures. We now concentrate on the product space built using trade flow data from 1998-2000, which consists of a 1006x1006 matrix whose entries are the proximity between products. Each row and column of this matrix represents a particular product and each off-diagonal element represents the proximity between a pair of products. Figure 1sA shows the proximity matrix where columns are sorted using its sitc4 code name. Figure 1sB shows the same matrix sorted using an average linkage clustering algorithm, revealing its modular structure and the many empty rows and columns which belong to untraded products. In fact only 775 products conforms the actual product space.

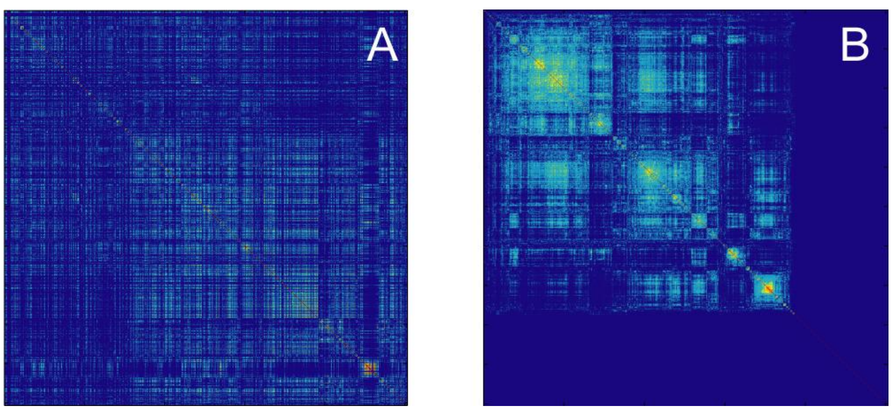

Figure 1s. The product space matrix representation. A. The product space matrix sorted in incresing order of the Sitc4 numerical code. B. The product space hierarchically clustured shows a modular structure and also revelas that only 775 products actually belong to the space.

Proximity values follow a broad, approximately log-normal distribution. Figure 2s A shows the number of links below a certain threshold. Figure 2s B. shows the frequency distribution of proximities. This broad, heterogeneous distribution evidences that the product space has a few strong links and many marginal links, which are not significant and represent the background of the proximity measure.

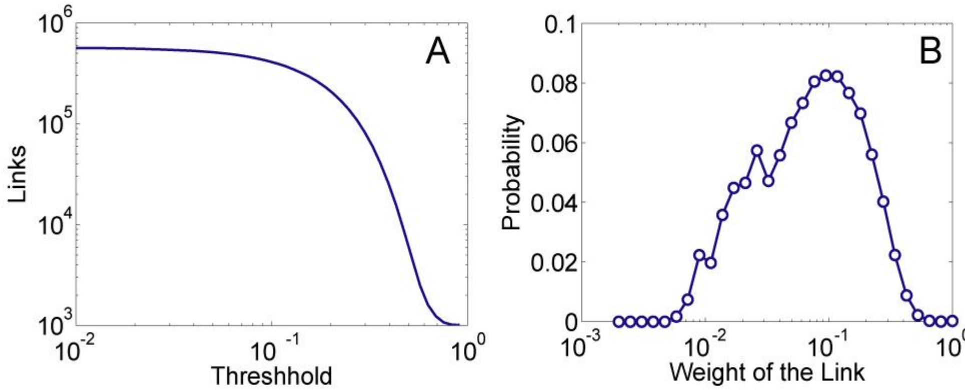

*Figure 2s. Distribution of Proximity Values. A**.** Cumulative Distribution of Proximity Values. B. Density Distribution of Proximity Values.*

At a value of 0.5 the proximity matrix has a giant connected component. Figure 3s shows a rough network representation of the proximity matrix in which only links above 0.5 have been considered. This visualization is not the most suitable and shows that a simple threshold criterion does not reveal much of the structure. We invite you to look at our more sophisticated visualization technique in the next section.

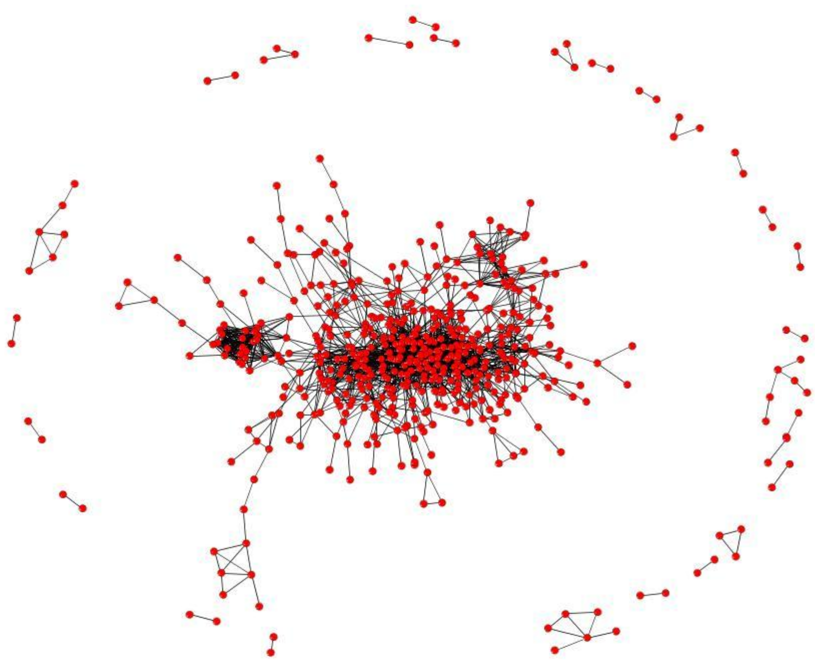

*Figure 3s**.** Network representation of the proximity matrix that considers only the proximity values above 0.5.*

Network Representation of the Product Space

We generated a network representation of the proximity matrix to help us develop intuition about its structure as well as to visualize and study the dynamics of countries on it. The matrix representing the product space has many small values which represent weak connections between products. That is why a network representation becomes an adequate way to layout the products, giving us a quick visual way to show the relevant links and to determine were countries are located and where they could be headed.

Maximum Spanning Tree (MST)

To include all products in our network we generated a "skeleton" for it: the Maximum Spanning Tree (MST). This is nothing more but the tree containing a sum of weights which is maximal. In other words, it is the set of N-1 links (N being the number of nodes) that connect all nodes in the network and maximizes the sum of the proximities in it.

We generated the MST by considering the strongest non-diagonal value of the proximity matrix and then considered the strongest link connected to that dyad. We then picked up the strongest link connecting a new node to our triad and continued adding links until all the nodes on the network were considered (Figure 4s).

Figure 4s. Earliest version of the MST representing the "skeleton" of the product space.

We also wanted to consider the strongest links which are not necessarily in the MST. We did this by considering the MST plus all the links above a certain threshold. A suitable visualization was obtained by keeping all links with a proximity value of 0.55 or larger (Fig. 5s). This resulted in a network with 775 nodes and 1525 links. Lower proximity values gave rise to crowded network representations while higher values resulted in sparse networks. As a rule of thumb, a good network visualization can be achieved with an average degree equals to 4. This is when the number of links is twice the one of nodes, which is the case for the 0.55 threshold.

Figure 5s. Representation of the product space based on the MST plus all links with a proximity above 0.55.

Network Layout

Good network visualization requires an appropriate layout. This is why we lay out the network using a force spring algorithm. Here nodes are represented as equally charged particles and links are assumed to be springs. The layout is determined by the relaxed positions.

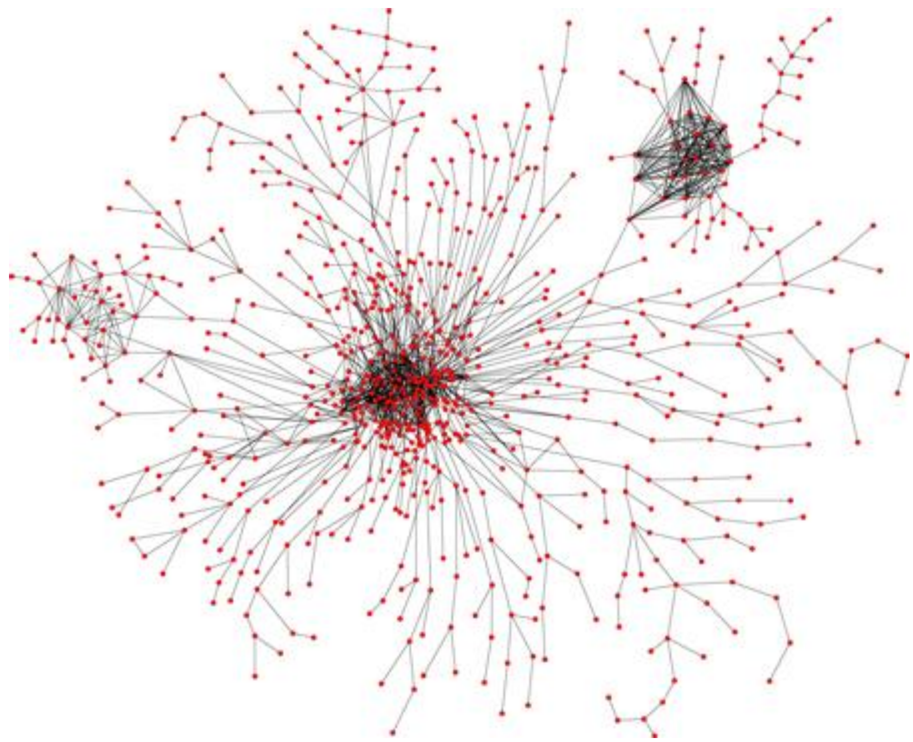

Figure 6s .Network representation of the product space. Layout uses a force spring algorithm.

The force spring layout is not the ultimate solution, but it brings us close to a good one. That is why we retouched the layout manually to avoid overlapping links and untangle dense clusters.

Node Sizes and Colors

An advantage of using a network representation is that we can simultaneously look at the structure of the space and other covariates. In our case we painted the network using the product classifications performed by Leamer[1], and made the size of the nodes proportional to the money moved by that particular industry or World Trade. To give a sense of the proximity of the links involved in our network representation we color coded them by using dark red and blue for strong links; and yellow and light blue for weaker ones.

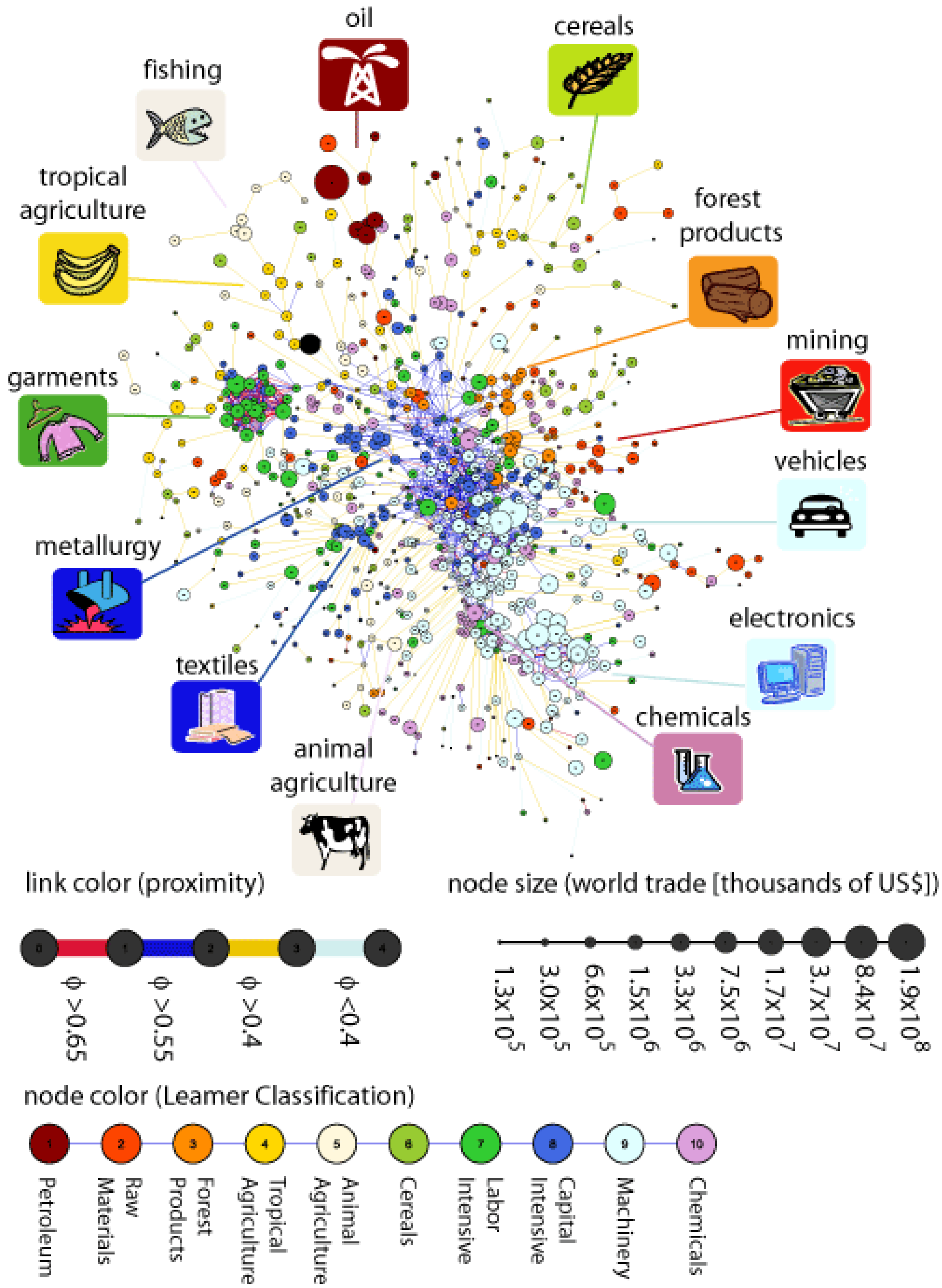

Figure 7s. Final version of the product space in which node size represents its world trade, node color shows its classification as proposed by Leamer and link color indicates a range in the proximity values.

(1) E. Leamer , *Sources of Comparative Advantage: Theory and Evidence* (MIT Press, Cambridge MA, 1984).

Product Space Properties

Using a network representation for the products space we can not only see which products are close to each other and the groups they form, but also their classifications and values. However, the network representation is nothing more than a powerful visualization technique and we still need to study the space properties using the entire proximity matrix complemented.

The Product Space Can Classify Products

The first property we study is the ability of the product space to classify goods into different classes. We compare our network representation with the clusters introduced by Leamer, as it is shown in figure 1, by using a different color for each product class. We see that the product space is not colored at random. Products in the same classes lie close to each other and tend to form clusters.

Although the classification performed by Leamer was done used a different methodology, the agreement between it and the structure of the product space is striking. Beyond the intuitive proof of Figure 7s we can tests the strength of these correlations by taking the average proximity between and within the products belonging to one of the clusters defined by Leamer (table 1s).

1998	Petroleum	Raw Materials	Forest Products	Tropical Agriculture	Animal Products	Cereals	Labor Intensive	Capital Intensive	Machinery	Chemical	Within / Between
Petroleum	**0.28**	0.15	0.16	0.16	0.15	0.13	0.14	0.16	0.11	0.15	0.89
Raw Materials		**0.17**	0.15	0.13	0.14	0.13	0.13	0.15	0.12	0.14	0.24
Forest Products			**0.26**	0.16	0.17	0.13	0.18	0.21	0.17	0.17	0.57
Agriculture				**0.21**	0.18	0.15	0.17	0.17	0.11	0.14	0.57
Animal Products					**0.20**	0.15	0.16	0.17	0.12	0.16	0.29
Cereals						**0.14**	0.13	0.15	0.11	0.14	0.04
Labor Intensive							**0.22**	0.22	0.17	0.16	0.38
Capital Intensive								**0.26**	0.19	0.20	0.43
Machinery									**0.24**	0.21	0.62
Chemical										**0.24**	0.46

Table 1s. Average strength of the links between and within products as classified by Leamer.

Table 1s shows that the average proximity of products belonging to the same cluster is always higher than the proximity for products belonging to different clusters. But not all clusters have the same size, thus we look at the distribution of proximities for all links connecting products with the same or different Leamer classifications. Figure 8s shows the distribution of proximity for links connecting nodes with the same Leamer classification (blue) and for links connecting nodes annotated differently. It is clear from the figure that nodes with the same classification are connected by links with higher proximity values, and because of the large number of links present in the system (L>200'000), the difference between these two distributions is highly significant (log(P-value)<-300 ANOVA)

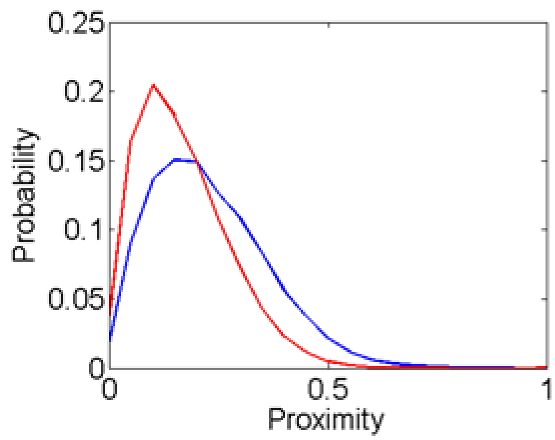

Figure 8s. Distribution of proximity for links connecting products with the same Leamer classification (blue) and with a different one (red).

Correlations Between the Position and Value of the Goods.

All products have a value, which in this work we consider as the average income per-capita associated with that good or PRODY. It follows to ask: Are rich goods located in particular parts of the product space? By looking at its network representation and setting the size of the nodes proportional to the PRODY of a product (figure 9s), we see that the largest nodes are located either in the center or the down most portion of the network. At a first glance, we can say that there is a rich region of the product space, composed by machinery, electronics and chemicals, and a poor, peripheral region, made of some agricultural and labor intensive goods.

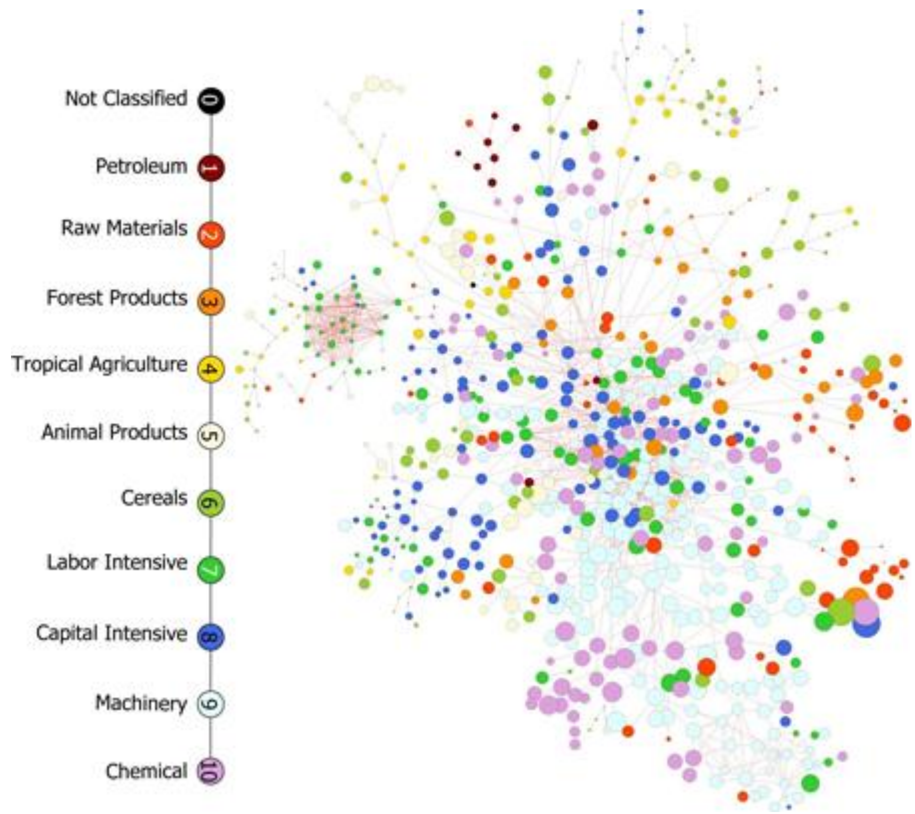

Figure 9s. Network representation of the product space in which node sizes are proportional to PRODY.

We con look beyond the actual value of products and study the value of goods as a function of their distance between them. Basically we ask: Is this particular product at the top or at the bottom of the PRODY sophistication scale? To answer this we study the average PRODY of products at a given distance of a particular node. We define distance as *-log(Proximity).* Figure 10s shows six examples of products, three of them at the bottom of the sophistication scale (Footwear, Cotton Undergarments and Coats and Jackets) which belong to the labor intensive cluster and thus products far from them are richer or more attractive. On the other hand, chemicals such as organo sulphur compounds, phenols and cyclic alcohols appear at the top of the sophistication scale and see all other products as less sophisticated.

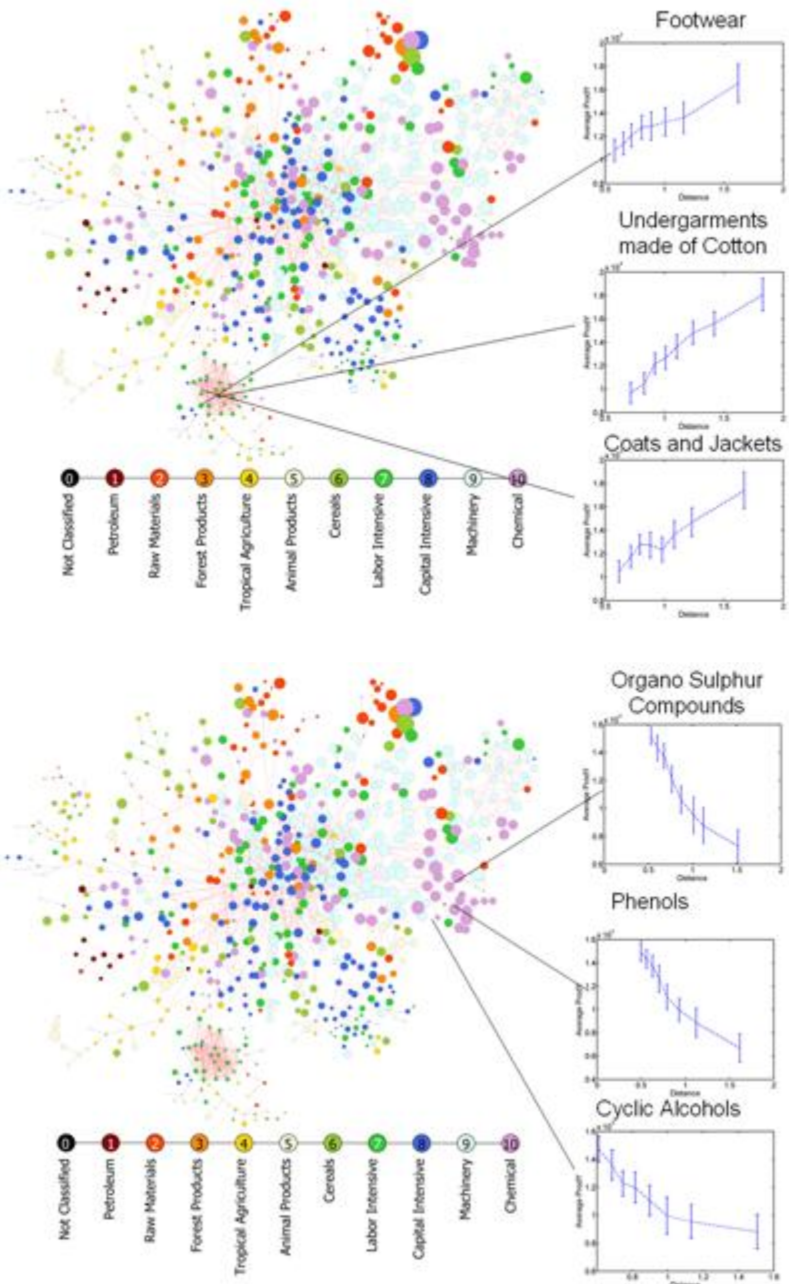

Figure 10s. Prody as a function of distance for six different products in the space. Plots were calculated using the full proximity matrix.

We performed the same analysis for each product class and found that there are products at the top of the scale, at the bottom and in local maxima (Figure 11s). If the structural transformation only moves countries to more sophisticated goods, a local maximum would trap countries. Examples of these are cereals and animal agriculture products which are goods located in the periphery of the product space but have a relatively large PRODY compared to their neighbors.

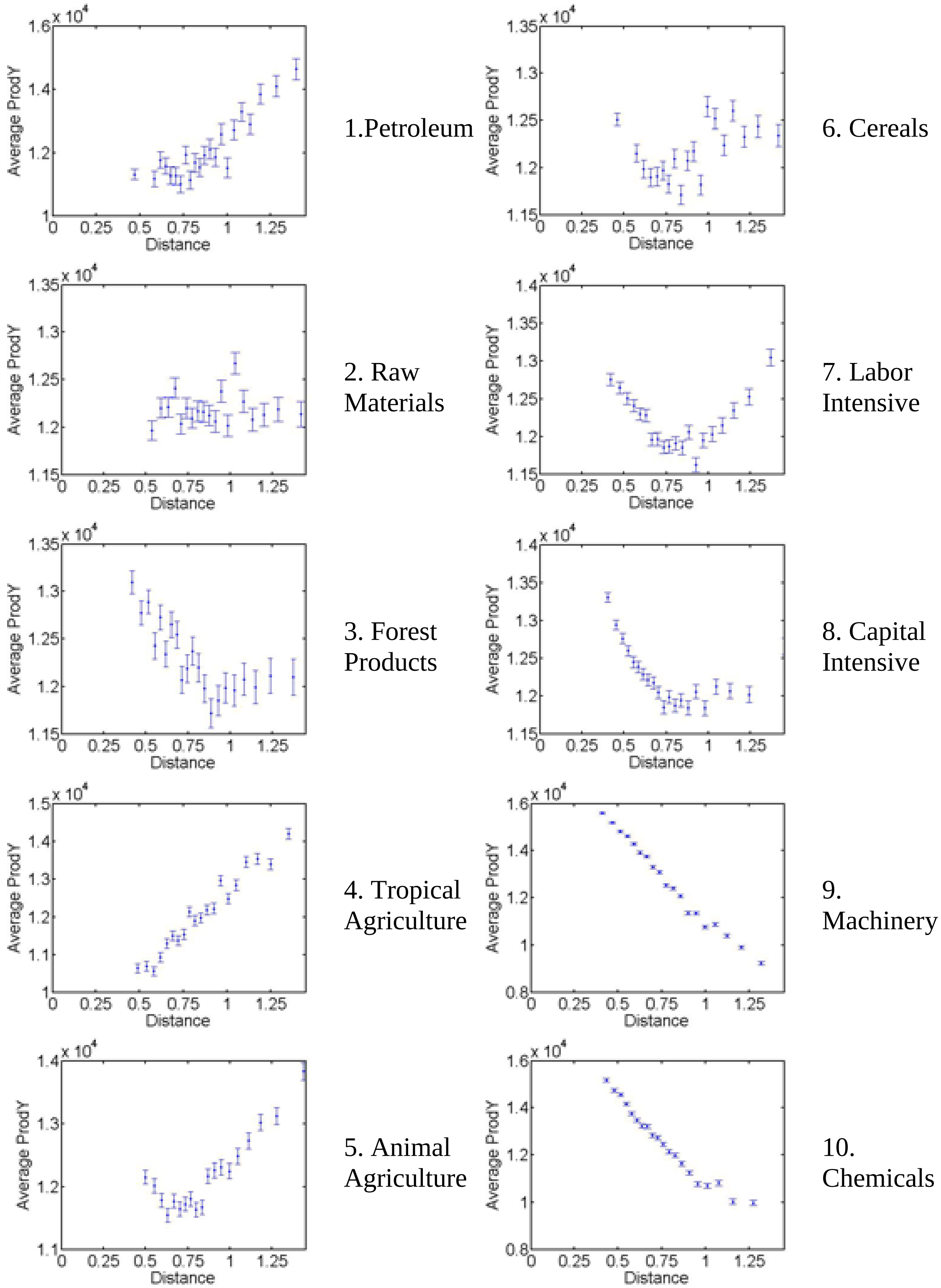

Figure 11s. Average PRODY as a function of the distance for products with a given Leamer Annotation.

Changes in Time

How fast does the product space changes in time? We can take a simple look at these by calculating the Pearson's Correlation Coefficient (PCC) between the matrices representing the product space in 1985, 1990 and 1998. Table 2s shows that the structure of the product space appears to be stable and that although links do change in time, after 10 or 13 years strong links remain strong and weak links remain weak. Thus products that are close tend to remain close and the ones that are far tend to stay far. The correlation was calculated over each pair of corresponding proximities between different time periods. Proximity values equal to zero were excluded from the calculation.

PCC	**1985**	**1990**	**1998**
1985	**1**	**0.702**	**0.696**
1990		**1**	**0.616**
1998			**1**

Table 2s. Pearson's Correlation Coefficient between the product spaces generated with data from 1985, 1990 and 1998.

Empirical Diffusion

Looking at pictures

Once the product space has been created and visualized, it becomes relatively easy to visualize the structural transformations of countries and how they are conditioned by the space itself. This study begins by visualizing where countries are located at different times, an amazing visual experience able to summarize the productive structure of a nation that preserves a significant level of detail. Figure 12s shows the products for which Malaysia has developed RCA with black squares. Figure 13s shows the same for Colombia. The versions presented here come from the beginning of our research, in which the layout was a slightly different, but still conserves the same color code and overall position. Node sizes are still proportional to world trade. We can appreciate that Malaysia had an impressive spread over the electronics and forest products cluster while during these same time period Colombia was able to spread through the garments sector.

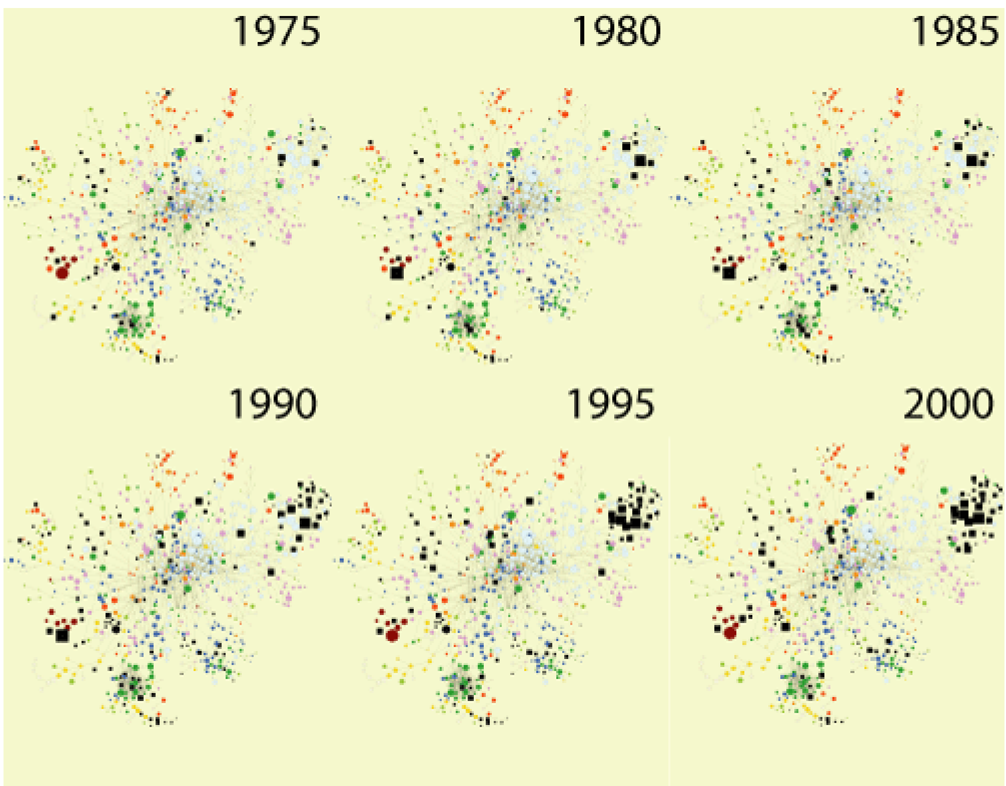

Figure 12s. Evolution of Malaysia (High Resolution Image) The products for which Malaysia has developed RCA are shown with black squares. Vector Image .ai.

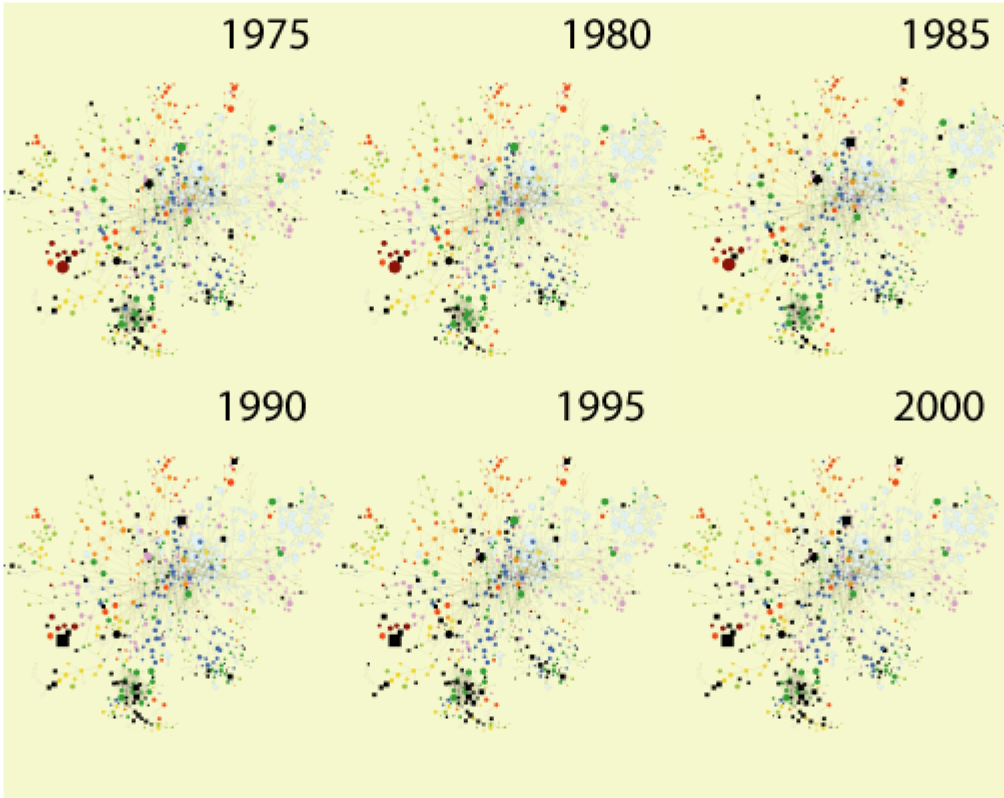

Figure 13s. Evolution of Colombia (High Resolution Image) The products for which Colombia has developed RCA are shown with black squares. Vector Image. ai.

Possible transitions between products

The visual examples presented above help develop our intuition and explain in a simple way how countries undergo structural transformations. They show that there is a tendency for countries to develop RCA close to products for which RCA was already developed, but are not a proof of this. For simplicity, we call a product for which a country has developed RCA, an *occupied product* (O), and one for which it has not an *unoccupied product* (U). When we compare two time points there are 4 possible transitions (U->U,U->O,O->U,O->O), and in our case, we are concerned with the second one which takes unoccupied products to occupied ones. Additionally, we call a product undergoing this particular transition: *transition products*. We now ask: are transition products closer to occupied products than to unoccupied ones? If this is significantly the case, it would be evidence supporting that countries perform structural transformations by *jumping* from occupied products to nearby ones.

To proof this we need to define some quantities. First we define density as the weighted fraction of the space which appears to be occupied from the point of view of a product in a particular country. Mathematically density ω can be written as:

$$\omega_j = \sum_i x_i \phi_{ij} / \sum_i \phi_{ij}$$,

where ϕ_{ij} is the proximity between the *i'th* and *j'th* product and x_i is 1 when the *i'th* product is occupied and zero otherwise. To measure this quantity empirically we consider as undeveloped products all those that had an RCA<0.5 on 1990. Starting from this definition, transition products are the ones that had an RCA >1 on 1995 and undeveloped products are the ones that remain with an RCA<0.5 on 1995. The ones that had an RCA between 0.5 and 1 in 1995 were regarded as inconclusive.

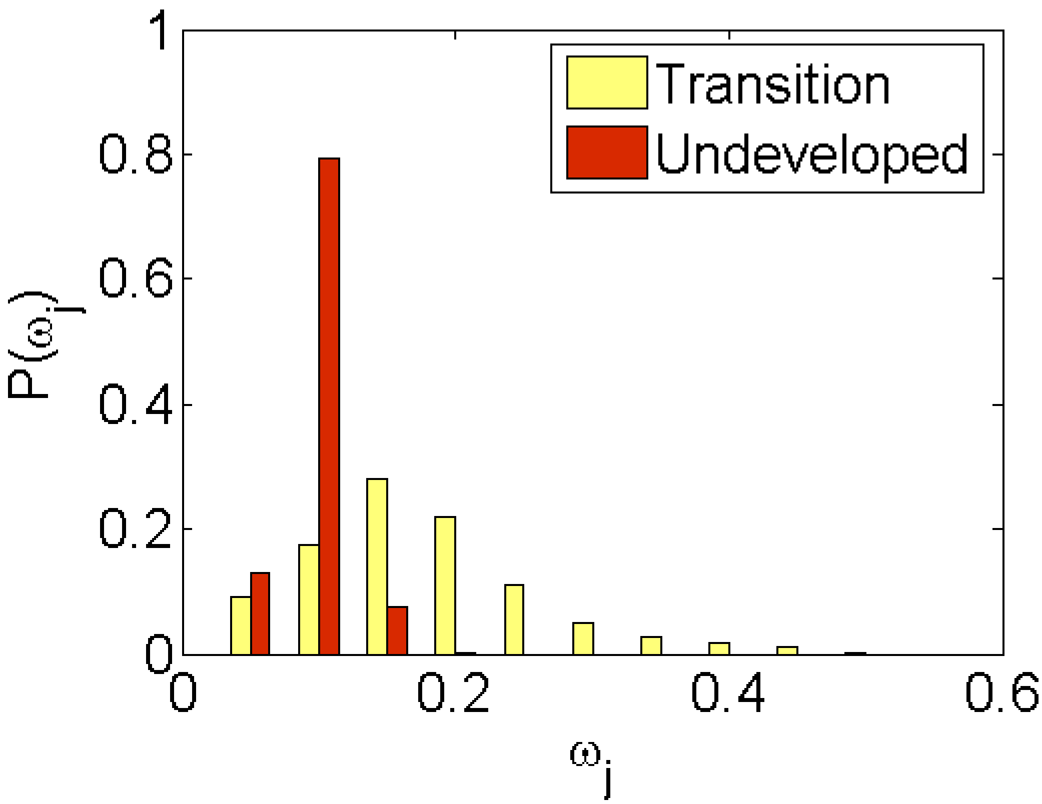

Figure 14s. Density distribution for products that underwent a transition and those remaining undeveloped.

Figure 14s shows that the density distribution for transition products takes significantly larger values than for products that remained undeveloped, suggesting that density predicts a transition. To further characterize this we can take the ratio between the average density of products on countries were they underwent a transition and compare it to the average density for countries were they remain undeveloped. We call this the discovery factor (*H*) which can be written as

$$H_j = \frac{\sum_{k=1}^{T} \omega_j^k \Big/ T}{\sum_{k=T+1}^{N} \omega_j^k \Big/ (N-T)}$$

where the top summation goes over the T countries where the *j'th* product underwent a transition and the bottom one over the N-T countries were the product remain undeveloped. Figure 15s shows that in fact for more ~80% of the goods this ratio is larger than one, illustrating again that density tends to be higher for transition products.

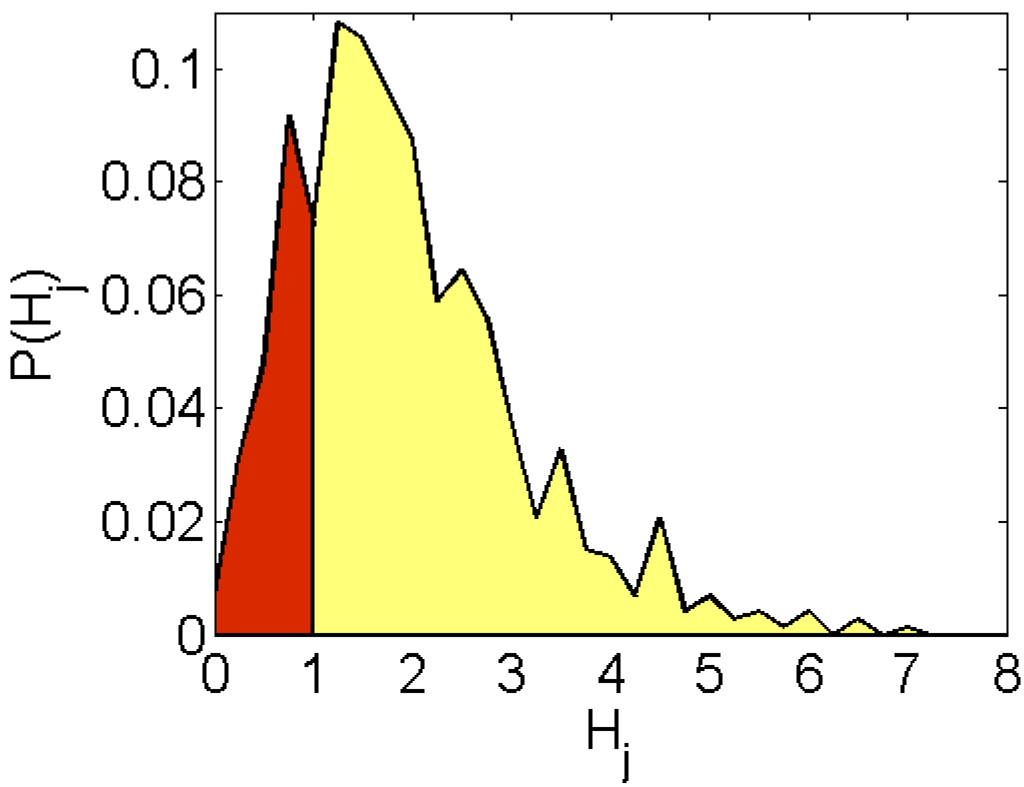

Figure 15s. Distribution of the discovery factor H.

Yet another way to show this is to consider the probability for a product to develop given that the closest developed product is at proximity ϕ. Figure 16s shows that this is a monotonically increasing function of ϕ. In fact further inspection shows that P has a quadratic dependence on f. Thus the chances for a country to develop a product increase enormously when that product is close to an already developed one.

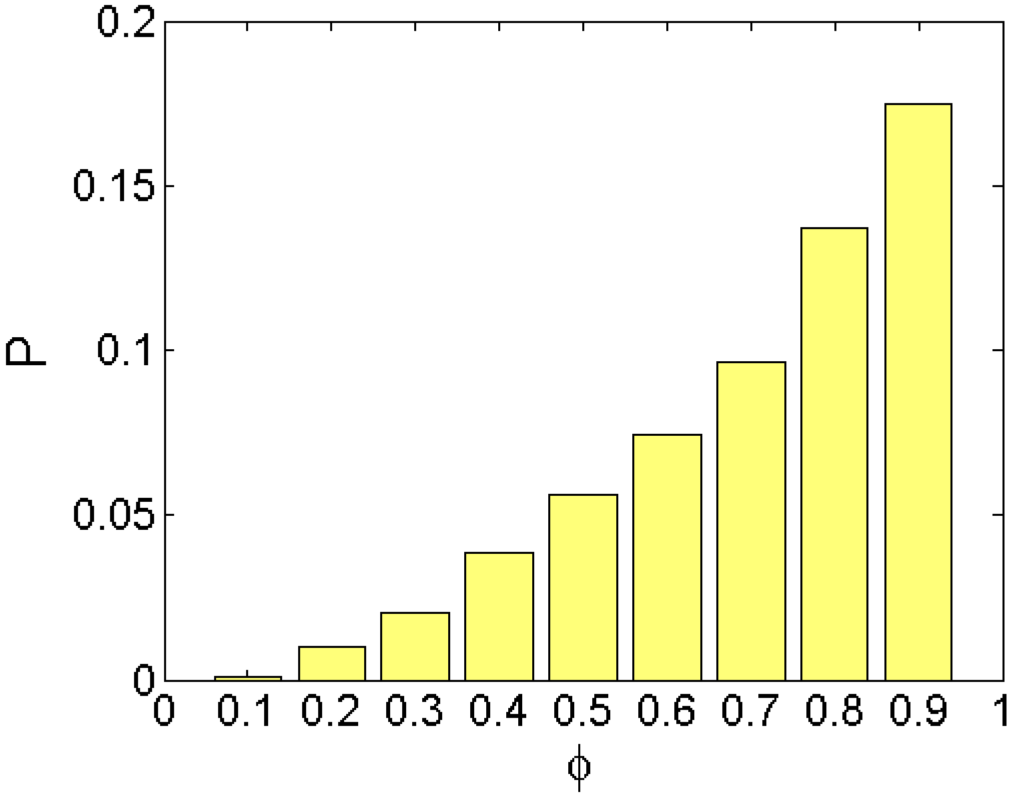

Figure 16s. Probability of undergoing a transition given that the closest occupied product is at proximity ϕ.

Simulated Diffusion

One diffusion step

Empirically, we showed using examples and statistics that products in which countries develop RCA tend to lie close to other products for which these countries have already developed RCA.

Using these we try to anticipate how a country will diffuse across the product space. As an example, we show figure 17s, in which we highlighted with black squares all products at a given proximity of the ones already developed by Chile and Korea. We refer to this example as one diffusion step.

In this case we tuned the proximity of the jump and show that for high proximities the set of options available is small while for low proximities is large, however different.

The available options are strongly conditioned by current exports. Korea is a country that has developed RCA in several branches of machinery and therefore can diffuse from the center of the space. At proximity of 0.5 its options include the entire core of the network plus the entire electronics and garments clusters, among other things. Chile diffuses from the periphery and to achieve a similar set of options needs to diffuse as far as proximities of 0.3.

In summary we find that the set of options available for a country are strongly conditioned by its position in the product space and its ability to diffuse into products up to given proximities.

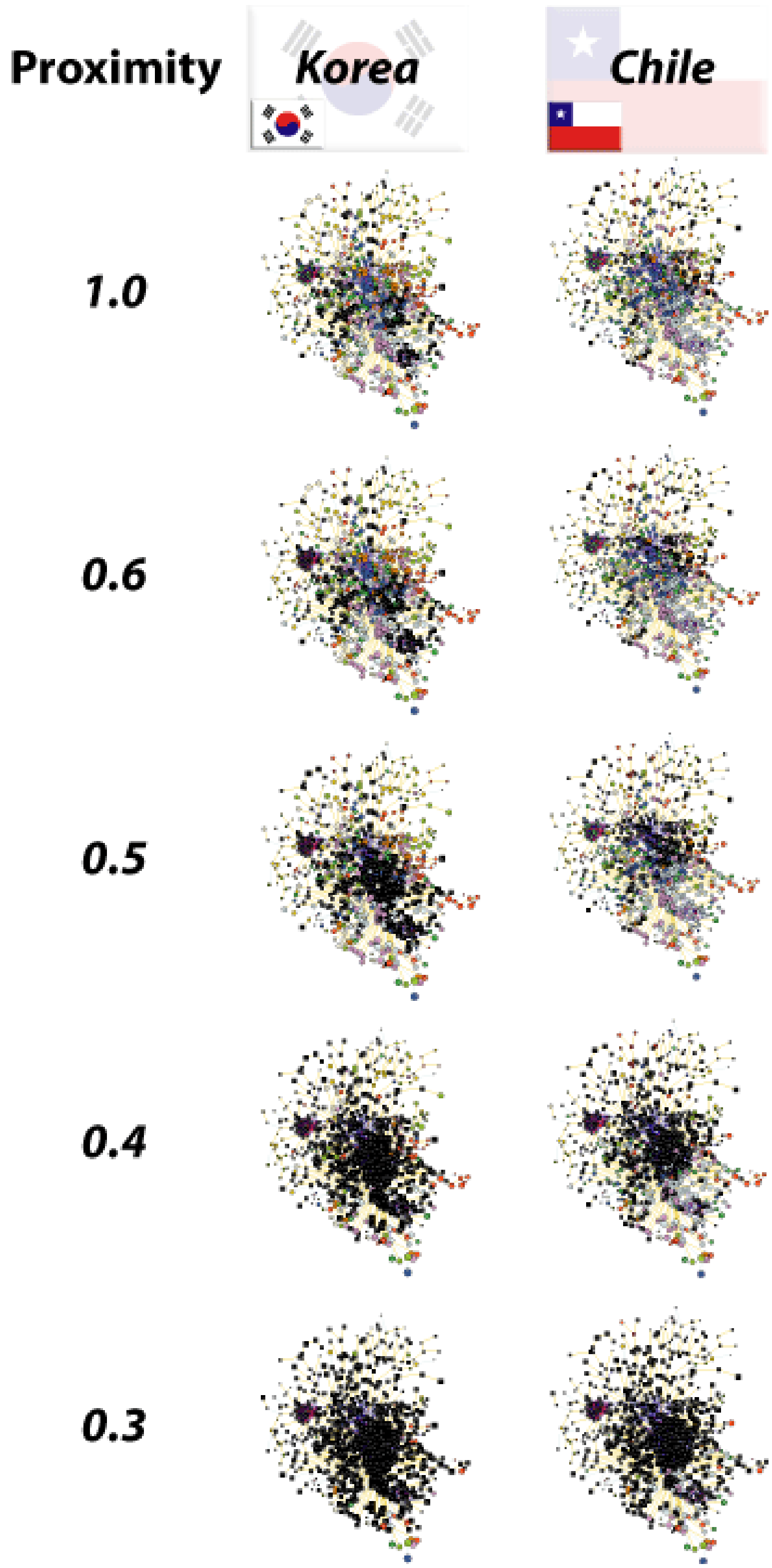

Figure 17s. One step diffusion process for Korea and Chile. The black squares denote all products closer than a given proximity considering their exports baskets in the year 2000.

Iterated diffusion

We can refine the diffusion process presented above by choosing a particular proximity and iterate the one step diffusion process. This represents a set of products potentially available to countries after diffusing to close products iteratively. At this point we ask ourselves: Is there a critical value of proximity at which countries will be able to diffuse across the product space? To explore this question we simulate a diffusion process in which a country "jumps" to all goods reachable from its current export basket, such that the proximity to them is larger or equal than a given value. Figure 18s illustrates through a color code the products available to Chile and Korea after diffusing iteratively at different proximities for 4 time steps. We observe that at relatively low proximities (ϕ = 0.55) both countries are able to diffuse, however Chile does so much slower and reaches the core in the second and third rounds, compared to Korea which does so on the first and second. At larger proximities the diffusion process halts. At ϕ = 0.65 Chile is unable to diffuse at all, while Korea slowly does so close to the core of the product space.

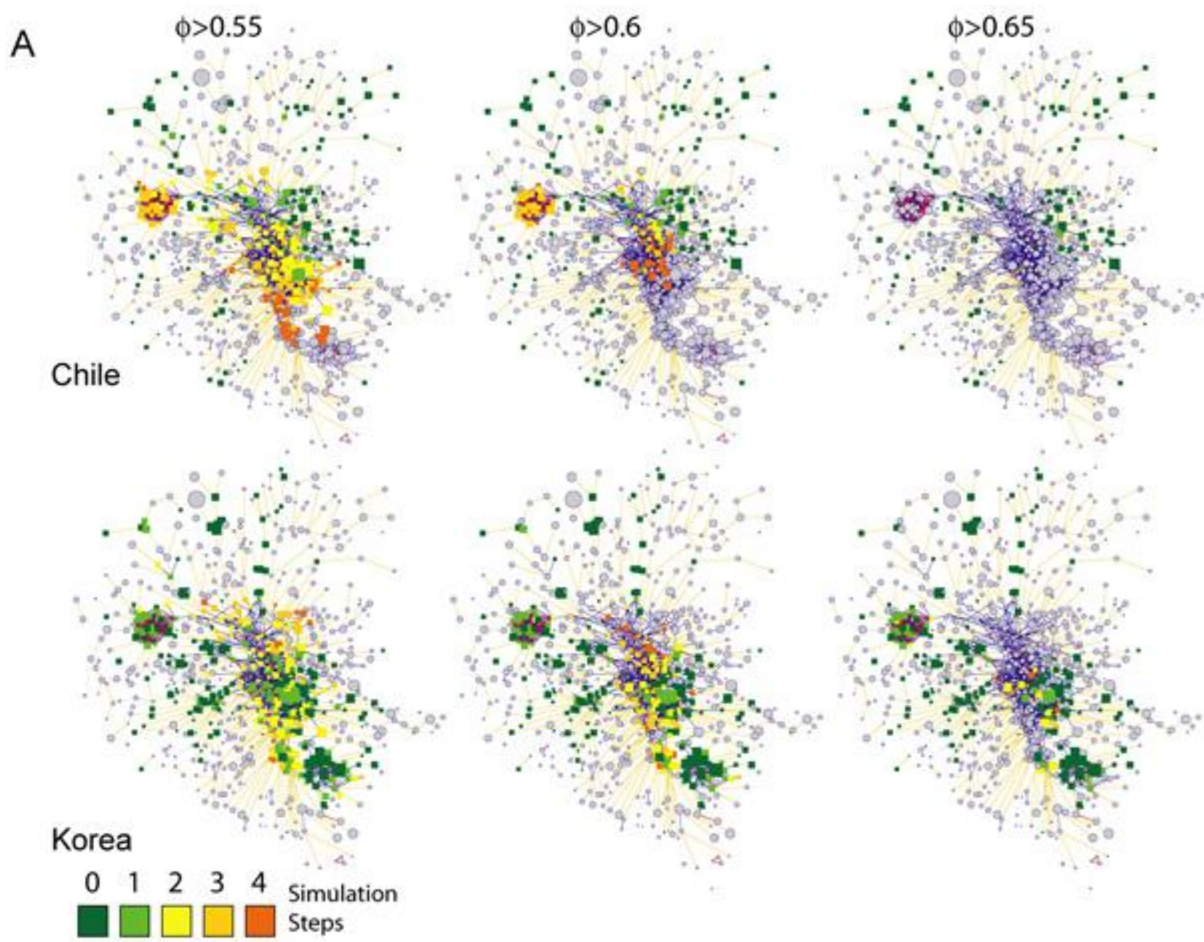

Figure 18s. Iterated diffusion process for Chile and Korea.

Economic Convergence

We characterize the value of a certain configuration by considering the value of its top products. We can assign value to a good by following the work of Hausmann, Hwang and Rodrick in which the value or sophistication of a good is equal to the average GDP per capita associated with that good. This quantity is called PRODY and in our particular example we consider the average PRODY of the top N products of a countries export basket after M diffusion steps with proximity ϕ. We denote this quantity by $< PRODY >_{M\phi}^{N}$.

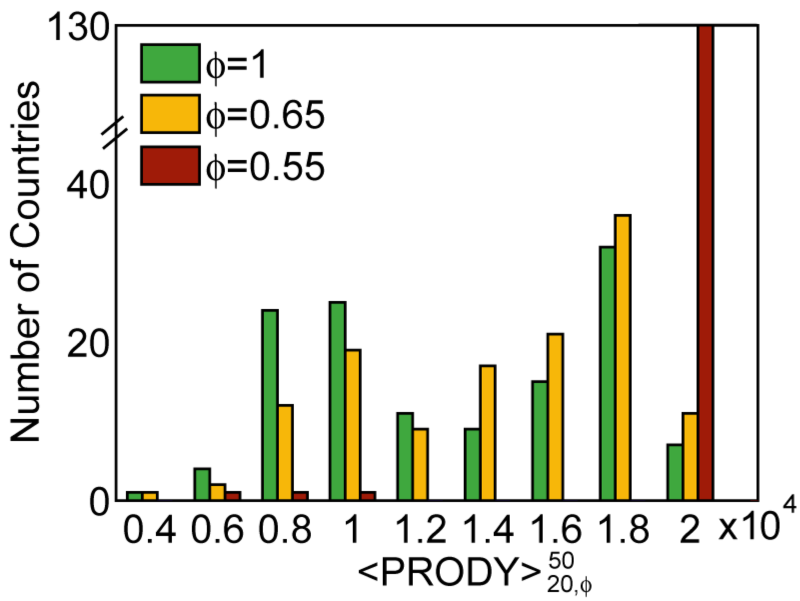

Figure 19s. Distribution for the average PRODY of the top 50 products reached after 20 diffusion steps at three different proximities.

Figure 19s shows that the original distribution of $< PRODY >_{M\phi}^{N}$ is bimodal. Indicating a world in which countries are divided into those producing sophisticated goods and unsophisticated ones. If we allow countries to diffuse in this space to acquire only goods that are really close by (ϕ=0.65). This distribution remains practically unchanged evidencing the structural constrains imposed by the product space. Whereas, if we allow countries to diffuse into products at relatively large proximities (ϕ=0.55) we find that after a large number of rounds most countries are able to reach the most attractive parts of the space, except for a few of them that remain stuck in the lowest bracket of this distribution.