The Economic Complexity Observatory: An Analytical Tool for Understanding the Dynamics of Economic Development

Alexander J. G. Simoes and **César A. Hidalgo**
MIT Media Lab, Macro Connections
20 Ames Street, Cambridge, MA 02139 USA
{simoes, hidalgo}@mit.edu

Abstract

We introduce The Economic Complexity Observatory, a tool for helping users understand the evolution of countries' productive structures and trade partners. Here we bridge the gap of harnessing the raw computational power of cycling through thousands of entries of data with the analytical, decision making qualities of the human mind through the use of information visualization "apps". These allow for a richer understanding of large volumes of data and improve the quality of information readily available to decision makers. We present the data, process and different visualizations then, discuss the interface of The Observatory as a website built on standards with accessibility in mind.

Introduction

For most practical matters, information is most valuable when it is inside people's brains. Brains, however, are not solid state drives or tape recorders. Brains are not good at storing large amounts of raw data, like the ones that society is currently collecting, but are excellent at making sense of complex visual patterns. Ultimately, understanding comes from the development of theories that can help account for these patterns and make predictions regarding phenomena not yet observed.

Honing our understanding of the world requires developing analytical tools that can help compare our theories with observations about the world. This process is full of simplifications, since the information that exists in the world is too vast to be inputted into a brain. Hence, the process by which understanding is developed is constrained by the availability of representations that can help bridge the world of theories and the world of data. Ultimately, comparisons between theory and observations can only be made after we find these "bridge representations", in which both theory and data can be depicted such that the proper comparisons can be made.

The data deluge brought by the information age has increased the amount of data available, and the accessibility of this data, thus making the bridge representations that we use to develop understanding increasingly more narrow. A chief example of this is the field of economic growth, which has been limited by the use of simplified representations that abstract all the complexity of the world based on a few aggregate factors (Solow 1956), (Solow 1957), (Romer 1986), (Romer 1990). These representations have mistakenly cast the problem of economic development as a problem of accumulation, rather than a problem of increased diversity (Hidalgo and Hausmann 2009).

In this paper we introduce the first version of the Economic Complexity Observatory, a new tool that can be used to represent high volumes of highly disaggregated trade data. The goal of The Observatory is to help create a new bridge where economic growth development theory and data can meet. Ultimately, the creation of simple, yet accurate depictions of large volumes of data will provide new anchors for more disaggregate forms of development theory (Hausmann and Hidalgo 2010) and will help decision making in an industrial policy setting.

Data and Method

Trade Data

Every year billions of dollars worth of goods flow across the globe exchanging hands from one firm to another and often times crossing borders in between. Customs authorities make note of such records, since they are required for tariffs and national accounting purposes. Trade date is reported yearly to the United Nations and stored in the UN Comtrade database.

Here we use data in the Standard International Trade Classification (SITC) to the 4 digit resolution. For the 1962-2000 period we use the data compiled by (Feenstra et al. 2005), and for the 2001-2009 period we use data downloaded directly from the UN COMTRADE. Both datasets include bilateral trade data with information of the following:

- Country (reporting data)
- Country (trade partner)
- Product (SITC4 code)
- Year
- Direction of trade flow (import or export)
- Amount (Value in $US)

For the creation of The Observatory we organized this data in a relational database with the following 3 dimensions of variance:

1. Country
2. Product
3. Year

Additionally, we calculated the Revealed Comparative Advantage (RCA) that each country has on each product using the (Balassa 1965) definition of RCA and the proximity between pairs of products following (Hidalgo et al. 2007). Therefore along with the three variables listed above, other attributes such as export value, import value, RCA and more are also stored. As such a sample data point may be in the following format:

Country	**Product**	**Year**	Exp.	Imp.	RCA
Sudan	Cotton	1997	104284	308	158.4
Sudan	Cotton	1998	234284	291	161.2

The Economic Complexity Observatory

The Observatory consists of a variety of online tools, or apps, that allow users to create representations of the set of goods exported and imported by countries for all years for which we have data available. These are relevant representations because recent work shows that the level of income of countries tends to follow their productive structure (Hidalgo and Hausmann 2009). Moreover, countries are more likely to start exporting products that are adjacent in the Product Space to the ones that they are already exporting (Hidalgo et al. 2007), making the productive structure of a country a truly fundamental aspect of its economy.

Here we present three different apps that have been developed in depth.

1. TreeMaps

Based off of the original TreeMap algorithm by Ben Shneiderman (Shneiderman 1992) and improved upon by Martin Wattenberg's "Map of the Market" (Wattenberg 1999) we use this tiling algorithm to generate a rectangle that when taken as a whole represents 100% of either a particular country's trade in a given year or of a particular product's traders in a given year.

Figure 1 shows a treemap representing Sudan's exports in the year 1997. Here, each rectangle represents a product that Sudan exports with Revealed Comparative Advantage (RCA). For an expanded discussion on why we are only viewing exports at this particular cut off see (Hausmann, Hwang, and Rodrik 2007) and (Hidalgo and Hausmann 2009). As with all TreeMaps the area of each rectangle maps to a value in 2-dimensional space, in this example the value is the export share as a percentage of Sudan's total in 1997. Moreover, each rectangle's color represents its Leamer category, (Leamer 1984) an industry categorization based on imputed factor endowments. Finally, the alpha channel is used to represent the RCA value. Therefore, this visualization summarizes quickly Sudan's export structure and shows that a large fraction of Sudan's exports for that year come from a single product, Cotton which accounts for roughly 30% of Sudan's exports.

As a point of comparison in figure 2 we show the productive structure of Spain in 1997. Spain being a developed

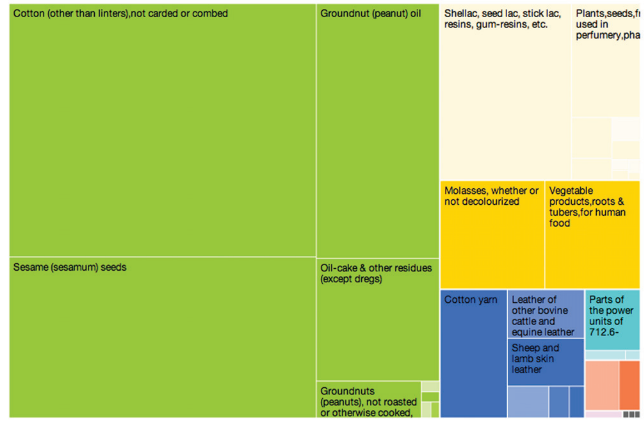

Figure 1: A TreeMap showing Sudan's Exports in the year 1997.

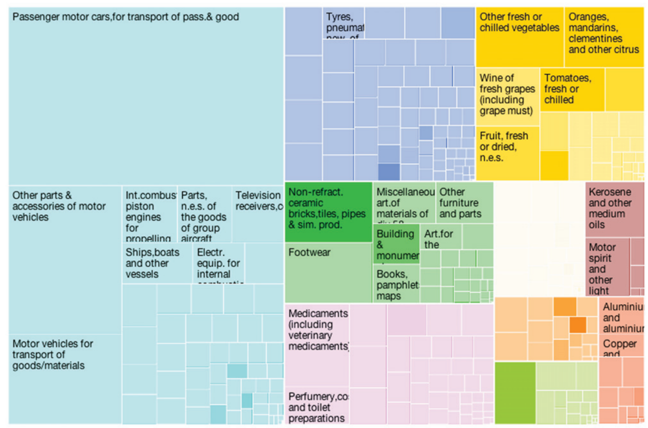

Figure 2: A TreeMap showing Spain's Exports in the year 1997.

country with a similar population to that of Sudan, has a productive structure that is not at all similar. This paints the picture that there are many things that go unseen using the more traditional forms of economic growth that rely on aggregate measures such as GDP.

2. Product Space

The Product Space maps are based on the research of (Hidalgo et al. 2007). Each node in the product space represents a product and links connect products that tend to be exported by the same countries. The products being exported by a country are denoted by the opacity of the nodes (full opacity indicates that this is a product exported with an RCA > 1).

The location of a country in the Product Space is predictive of the products that a country will make in the future. The product space is therefore a predictive tool that can help guide industrial policy. As an example of this, in figure 3 we show the Product Space for China in the years 1985, 1995, 2000 and 2009. Based on this visualization we are able to see how China's development changes from ex-

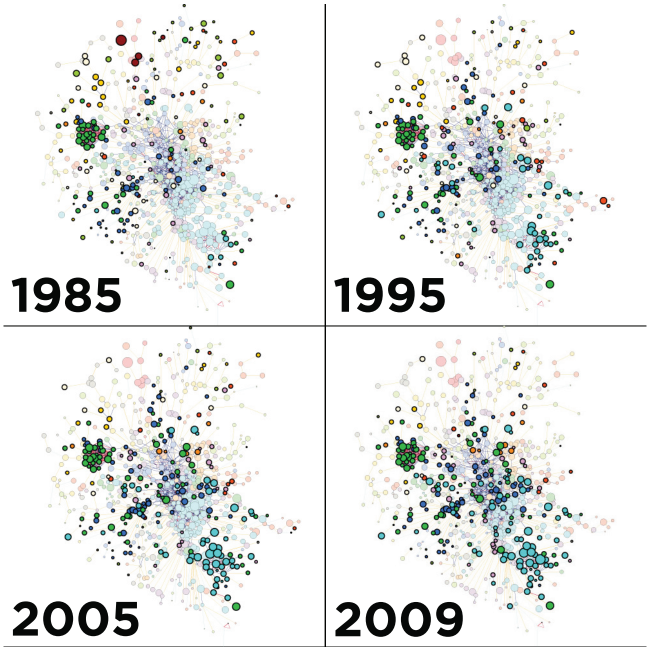

Figure 3: The progression of change in the Product Space of China over a 14 year period.

porting products mainly concentrated in the periphery of the Product Space in 1985 to moving towards exporting products in the more densely connected center.

3. Stacked Area Charts

The stacked area charts used are unique in their ability to show the same dataset we have been working with but now with values changing across time. One may view data as either shares of a total (as shown in figure 4) or in nominal values of $US (as shown in figure 5). We are able to plot here how a country's export diversification changes over time (what products they have an RCA > 1) or how a product's exporters change over time (which countries export a product with an RCA > 1).

Let us take again the exports of Sudan over a 35 year period (from 1965 – 2000) as an example. In figure 4 we are clearly able to see the diversification of exports change quite rapidly in the five year period from 1995 to 2000 and if we were to look into current events of that time concerning Sudan we would learn that they had just begun exporting oil in the late 1990s. Much of the knowledge that can be learned from simply gleaning at these visualizations would not be possible without spending hours poring over spreadsheets of data.

Process and Technologies Used

Since The Observatory is intended for a wide audience, including users in developing countries, it is important to use technologies that can facilitate this goal. As such we have fully embraced web standards, using semantic HTML5 for markup and JavaScript on all front-end development. Since

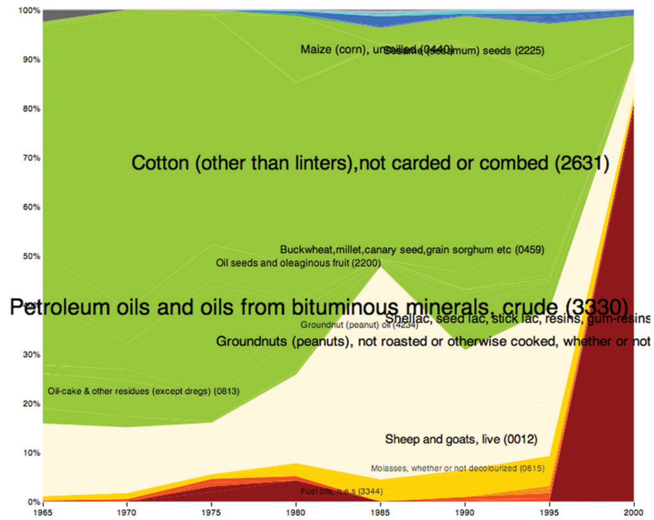

Figure 4: Stacked area chart of Sudan's shares of exports from 1965 to 2000

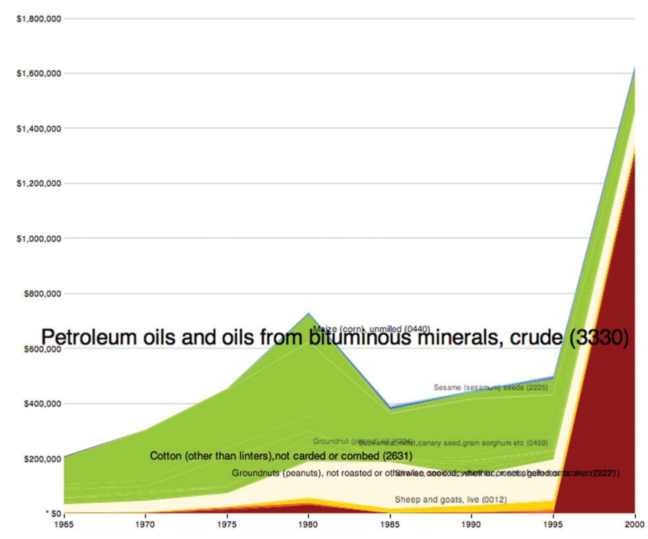

Figure 5: Stacked area chart of Sudan's exports in $US from 1965 to 2000

there is no use of Flash throughout the entire site it is also accessible from a number of mobile platforms including the iPhone and iPad. Two particular open source drawing libraries were utilized and have been customized for this project; Raphaël JavaScript Library and JavaScript InfoVis Toolkit. As for server-side technologies, MySQL is used as the database and Python using Django, a web framework is used to access the data.

Conclusion

The Economic Complexity Observatory is a tool that makes rich data visualizations accessible to a wide audience of economic development experts and practitioners. In particular, this first version of The Observatory provides some simple analytical tools that can be used to understand the evolution of countries' productive structures.

Ultimately, these tools provide representations of our world that are simple enough to be grasped by a wide audience, but complex enough to demand new theories that can be used to understand the mechanics of economic development.

References

Balassa, B. 1965. Trade liberalisation and "revealed" comparative advantage1. *The Manchester School* 33(2):99–123.

Feenstra, R. C.; Lipsey, R. E.; Deng, H.; Ma, A. C.; and Mo, H. 2005. World trade flows: 1962-200. *National Bureau of Economic Research Working Paper Series* (11041).

Hausmann, R., and Hidalgo, C. A. 2010. Country diversification, product ubiquity, and economic divergence. *CID Working Paper No. 201* (201).

Hausmann, R.; Hwang, J.; and Rodrik, D. 2007. What you export matters. *Journal of Economic Growth* 12:1–25. 10.1007/s10887-006-9009-4.

Hidalgo, C. A., and Hausmann, R. 2009. The building blocks of economic complexity. *Proceedings of the National Academy of Sciences of the United States of America* 106(26).

Hidalgo, C. A.; Klinger, B.; Barabasi, A. L.; and Hausmann, R. 2007. The product space conditions the development of nations. *Science* 317(5837):482.

Leamer, E. E. 1984. *Sources of International Comparative Advantage*. The MIT Press.

Romer, P. M. 1986. Increasing returns and long-run growth. *The Journal of Political Economy* 94(5):pp. 1002–1037.

Romer, P. M. 1990. Endogenous technological change. *The Journal of Political Economy* 98(5):pp. S71–S102.

Shneiderman, B. 1992. Tree visualization with tree-maps: 2-d space-filling approach. *ACM Trans. Graph.* 11:92–99.

Solow, R. M. 1956. A contribution to the theory of economic growth. *The Quarterly Journal of Economics* 70(1):65–94.

Solow, R. M. 1957. Technical change and the aggregate production function. *The Review of Economics and Statistics* 39(3):pp. 312–320.

Wattenberg, M. 1999. Visualizing the stock market. In *CHI '99 extended abstracts on Human factors in computing systems*, CHI EA '99, 188–189. New York, NY, USA: ACM.