The Political Economy of Economic Complexity: Theory, Data, Methods

Section 2

Introducing the Economic Complexity Index (ECI)

Claudius Gräbner [1,2,3]

[1] University of Duisburg-Essen
Institute for Socioeconomics

[2] Johannes Kepler University Linz
Institute for Comprehensive Analysis of the Economy

[3] ZOE. Institute for Future-Fit Economies

EAEPE Pre-Conference
September 11-12 2019

Outline

What is economic complexity?

- Distinction between *complexity economics* and *economic complexity* as used here
  - Complexity economics considers the economy as a complex system and uses corresponding methods from various sciences
    - More a *school of thought* or a *research program* a la Lakatos
  - Economic complexity as used here refers to the ability of regions perform certain activities
    - More a *property of entities*, although a general approach to development (CAD) is built upon this
  - While there are *important complementarities and relations* we deal with the second interpretation -- and its variants -- here
- The precise definition and interpretation of `complexity' differs among methodological approaches

Measures

- There are a number of measures that run under to label `economic complexity'
- Many of them are used to quantify the complexity of technologies
- We focus on the *Economic Complexity Index* -- a slightly different thing

Measures of complexity - examples

The invention complicatedness(Fleming and Sorenson, 2001)

- Comes from innovation research and quantifies difficulty of making technologies
- Technologies as combinations of ideas, the latter being proxied by patents
- The less frequent we observe a technology, the more difficult it is to make

Structural complexity (Broekel, 2017)

- Originally applied to *regions*, but invariant w.r.t. scale
- Quantifies the structure of knowledge as a network of ideas
- Relies on patent data

The Economic Complexity Index

- Developed in the context of *countries* and their *productive capabilities*
- Underlying data: exports of products (as obtained by e.g. UN COMTRADE)
- Main interest: relationship between complexity, development and inequality

Introductory remarks

- The *Economic Complexity Index* is the result of applying the *method of reflections* to international trade data
    - Resulting measure of complexity at the heart of a broader theory of development
- A whole research program with micro- and macroeconomic research is built around this concept
    - We will refer to this as the *complexity approach to economic development* (CAD)

The fundamental questions of the CAD is straightforward:

Why are some countries rich, and some countries poor?

Why do some poorer countries catch up, and other not?

The CAD is yet another – but promising – attempt to answer these questions.

What the CAD is not (mainly) about

- The CAD contends having identified a main source for the dynamics just sketched
- This source...
  - ... is not physical capital or land
  - ... is not human capital or education
  - ...is not power
  - ... is not (strictly) about export growth or diversification
- Rather, the CAD relates the path to prosperity to ***collective knowledge***
- Societies prosper if they manage to facilitate collective learning such that members do things that require many *person bytes*

The complexity approach - quick history I

- At Harvard, Ricardo Hausmann and Rodrik (2003) develop the idea of *development as a discovery process*
- Ricardo Hausmann and Klinger (2006) extend this to the idea of *the product space* as a formalization of *knowledge spillovers*
- At that time, Cesar Hidalgo was a physicist doing his PhD in network science with Albert-Làszlò Barabàsi in Notre Dame
- Originally he looks for ways to visualize trade data as a network
- Thereby he started to work with Ricardo Hausmann and developed the idea of the *product space* (Cesar A Hidalgo et al., 2007)
- This work gets extended into a general measure of economic complexity (C A Hidalgo and Hausmann, 2009)

Cesar Hidalgo (MIT)

The complexity approach - quick history II

- Hausmann and Hidalgo stopped collaborating a few years ago
- Hidalgo continues to work on complexity, but in a broader way and mainly with economic geographers
- There much empirical work done on the microfoundations and policy implications of complexity at Hausmann's CID
- The method as has been constructively criticized (and somehow re-interpreted) by Tacchella et al. (2012)
- Now very popular in interdisciplinary discourse and economic geography, less so in economics

Ricardo Hausmann (Harvard)

The fundamental building blocks

- We approach the ECI via its fundamental theoretical building blocks:

1. The idea that person bytes are a major determinant of development
2. The idea of how to measure person bytes
3. The idea of diversity and ubiquity of products and ideas
4. The idea of the product space
5. The definition of the ECI and the PCI

The fundamental building blocks

- We approach the ECI via its fundamental theoretical building blocks:

1. **The idea that person bytes are a major determinant of development**
2. The idea of how to measure person bytes
3. The idea of diversity and ubiquity of products and ideas
4. The idea of the product space
5. The definition of the ECI and the PCI

Fundamental idea I: Information, person bytes, and development

- Originally, Hidalgo was interested in how (physical) information in an economy and economic development relate (Cesar A Hidalgo, 2015)
- He argues that a good measure for the information accumulated by an economy are the products the economy can make
  - Products are `***crystallized information***' because they store the knowledge required for their creation
- But why bother about this information?

- As it is common, HH consiter technology to be essential for development, but for them technology consists of knowledge, recipes and tools
- So HH claim that the source for development lies in the *collective knowledge* accumulated in an economy
- It is *not* about individual knowledge:

- Individual knowledge of a person is referred to as a *person byte*
  - If this were decisive, better education would imply positive development
- But the key is that people learn how to create and use knowledge together,
  - This allows them to do things that require more than one person byte

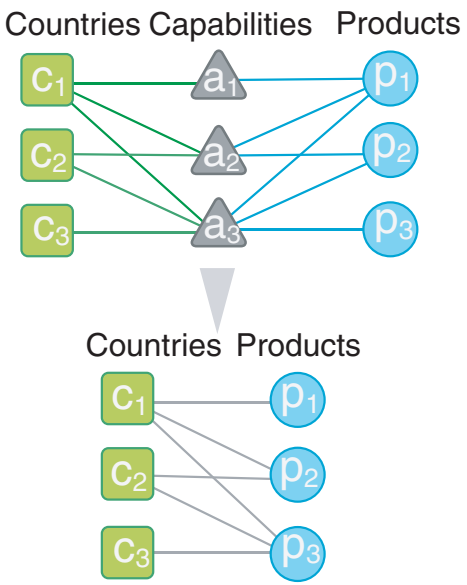
The fundamental building blocks

- We approach the ECI via its fundamental theoretical building blocks:

1. The idea that person bytes are a major determinant of development
2. **The idea of how to measure person bytes**
3. The idea of diversity and ubiquity of products and ideas
4. The idea of the product space
5. The definition of the ECI and the PCI

Fundamental idea II: Measuring person bytes

- How can the person bytes of an economy be measured?
  - Anecdotal evidence: median firm sizes
- HH argue that looking at the outcomes of exploiting collective knowledge is informative
  - Products are `***crystallized information***': store knowledge required for their creation
  - Products produced in a society provide information about collective knowledge



Source: C A Hidalgo and Hausmann (2009, p. 10571)

- Data on the production of products basically unavailable → export data
- For each product *p*: has country *c* have a *Revealed Comparative Advantage*

The Revealed Comparative Advantage

- Is share of p in export basket of c larger than in the total exports of the world market?
- Let P be the set of all products, and C the set of all countries

$$\frac{X_{cp}}{\sum_{p' \in P} X_{cp'}} \tag{1}$$

  - The share of product $p \in P$ in the export basket of country $c \in C$

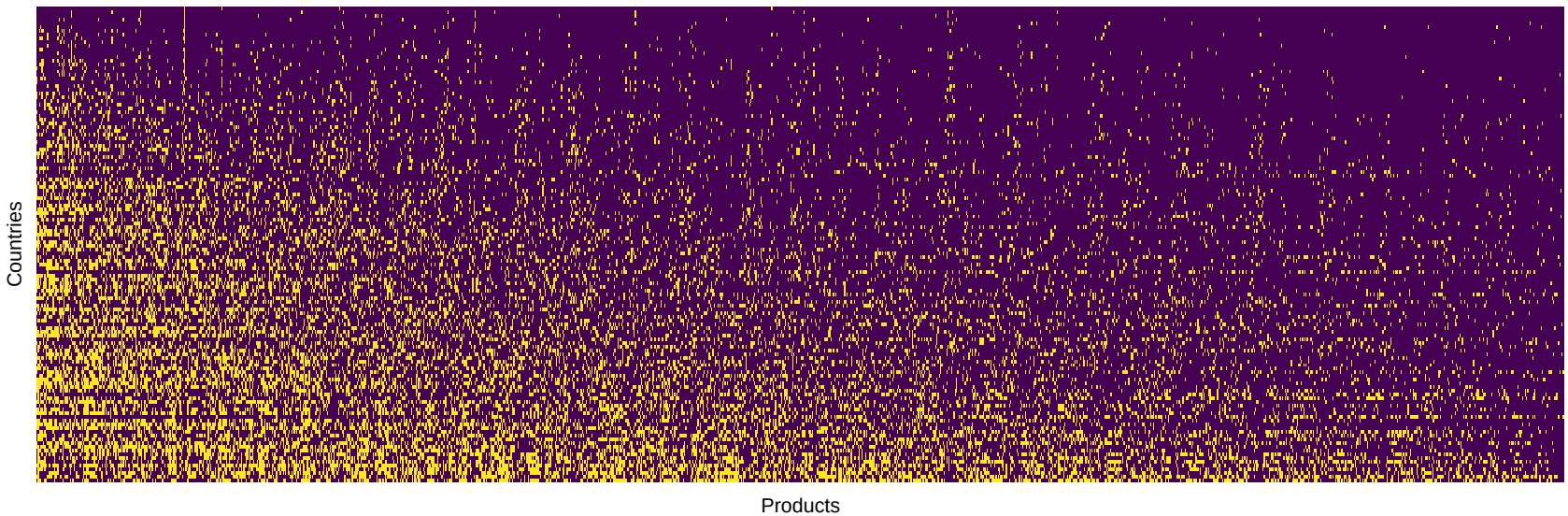
$$\frac{\sum_{c' \in C} X_{c'p}}{\sum_{c' \in C} \sum_{p' \in P} X_{c'p'}} \tag{2}$$

  - Share of the product in total exports in the world
- *RCA* of country c in product p is given by:

$$RCA_{cp} = \frac{X_{cp} / \sum_{p' \in P} X_{cp'}}{\sum_{c' \in C} X_{c'p} / \sum_{c' \in C} \sum_{p' \in P} X_{c'p'}} \tag{3}$$

- If $RCA_{cp} > 1$, c has a RCA in a product p and is a `notable' exporter of p

- How do you think the RCA is distributed across countries and products?
- Consider a matrix ***M*** with rows denoting countries, and columns products
  - $m_{cp} = 1 \leftrightarrow RCA_{cp} > 1$



- We call this a *lower-triangular* matrix
  - What would be the Ricardian prediction?
  - A *block-diagonal matrix*

The fundamental building blocks

- We approach the ECI via its fundamental theoretical building blocks:

1. The idea that person bytes are a major determinant of development
2. The idea of how to measure person bytes
3. **The idea of diversity and ubiquity of products and ideas**
4. The idea of the product space
5. The definition of the ECI and the PCI

Fundamental idea III: Diversity and ubiquity

- The MoR is build upon the concepts of *diversity* and *ubiquity*
- The ***diversity*** of a ***country*** refers to the number of products it exports with RCA
  - Row sum of matrix ***M***:

$$k_{c,0} = \sum_{p} M_{cp} \tag{4}$$

  - Diversity of a country's export basket
- The ***ubiquity*** of a ***product*** refers to the number of countries that export it with RCA
  - Column sums of matrix ***M***:

$$\kappa_{p,0} = \sum_{c} M_{cp} \tag{5}$$

  - Ubiquity of a product
- Its a good idea to compute these on your own (although packages for R and Python are available on the course homepage)

Fundamental idea III: Diversity and ubiquity

- The ***diversity*** of a ***country*** refers to the number of products it exports with RCA
- The ***ubiquity*** of a ***product*** refers to the number of countries that export it with RCA
- How do the two relate?

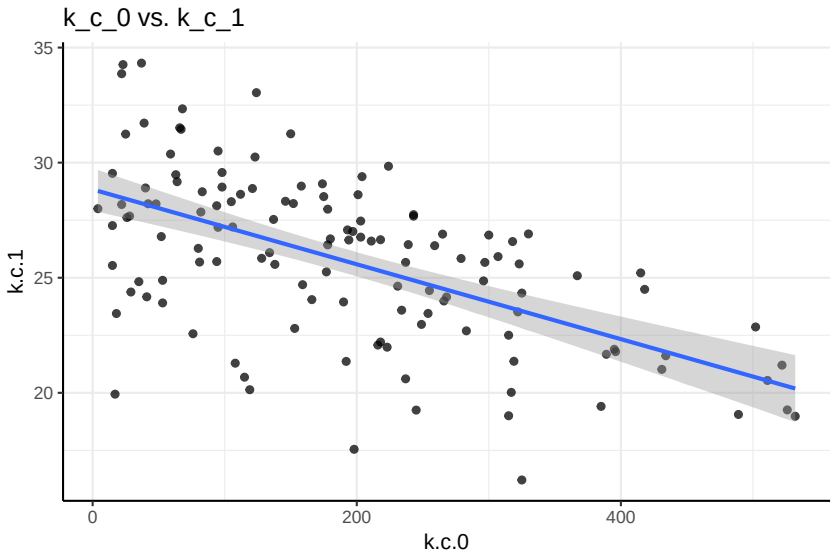
$$k_{c,1} = \frac{1}{k_{c,0}} \sum_p M_{cp} \kappa_{p,0} \tag{6}$$

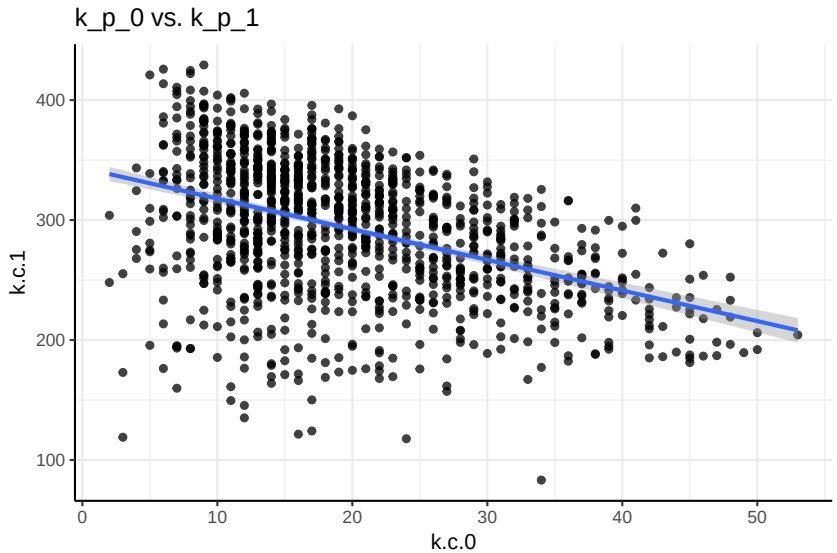
$$\kappa_{p,1} = \frac{1}{\kappa_{p,0}} \sum_c M_{cp} k_{c,0} \tag{7}$$

- For the first iteration remember the clear-cut interpretations:

	Explanation
$k_{c,0}$	Number of products exported by country c.
$k_{c,1}$	Average ubiquity of the products exported by country c.
$\kappa_{p,0}$	Number of countries exporting product p.
$\kappa_{p,1}$	Average diversification of the countries exporting product p.

- The ***diversity*** of a ***country*** refers to the number of products it exports with RCA
- The ***ubiquity*** of a ***product*** refers to the number of countries that export it with RCA
- How do the two relate?





Negative relationship between $k_{c,0}$ and $k_{c,1}$

▶ More diversified countries tend to export less ubiquitous products.

Negative relationship between $k_{p,0}$ and $k_{p,1}$

▶ More ubiquitous products tend to be exported by less diversified countries.

An important lesson

- This approach is data-driven and rough, so derivations from relationships occur frequently.
- However, such derivations are usually interesting and make us think!

Example

- Two countries that are similarly diversified are **Finland** ($k_{FIN,0} = 252$) and **Kenya** ($k_{KEN,0} = 250$)
  - Yet the products exported by Finland are exported on average by $k_{FIN,1} = 24$ countries
  - The products exported by Kenya are exported on average by $k_{KEN,1} = 36$ countries
  - Products exported by Finland are exported by fewer countries than those exported by Kenya
  - By considering $k_{FIN,3}$ we can then make the statement that products exported by Finland are exported by *more diversified* countries than those exported by Kenya - this is where the ECI measure is heading to
- So, diversification *in itself* does not seem to be very attractive
- Rather, it is important, *what kind of* products you produce
  - Ricardo Hausmann, Hwang, and Rodrik (2007): ``What you export matters``

The fundamental building blocks

- We approach the ECI via its fundamental theoretical building blocks:

1. The idea that person bytes are a major determinant of development
2. The idea of how to measure person bytes
3. The idea of diversity and ubiquity of products and ideas
4. **The idea of the product space**
5. The definition of the ECI and the PCI

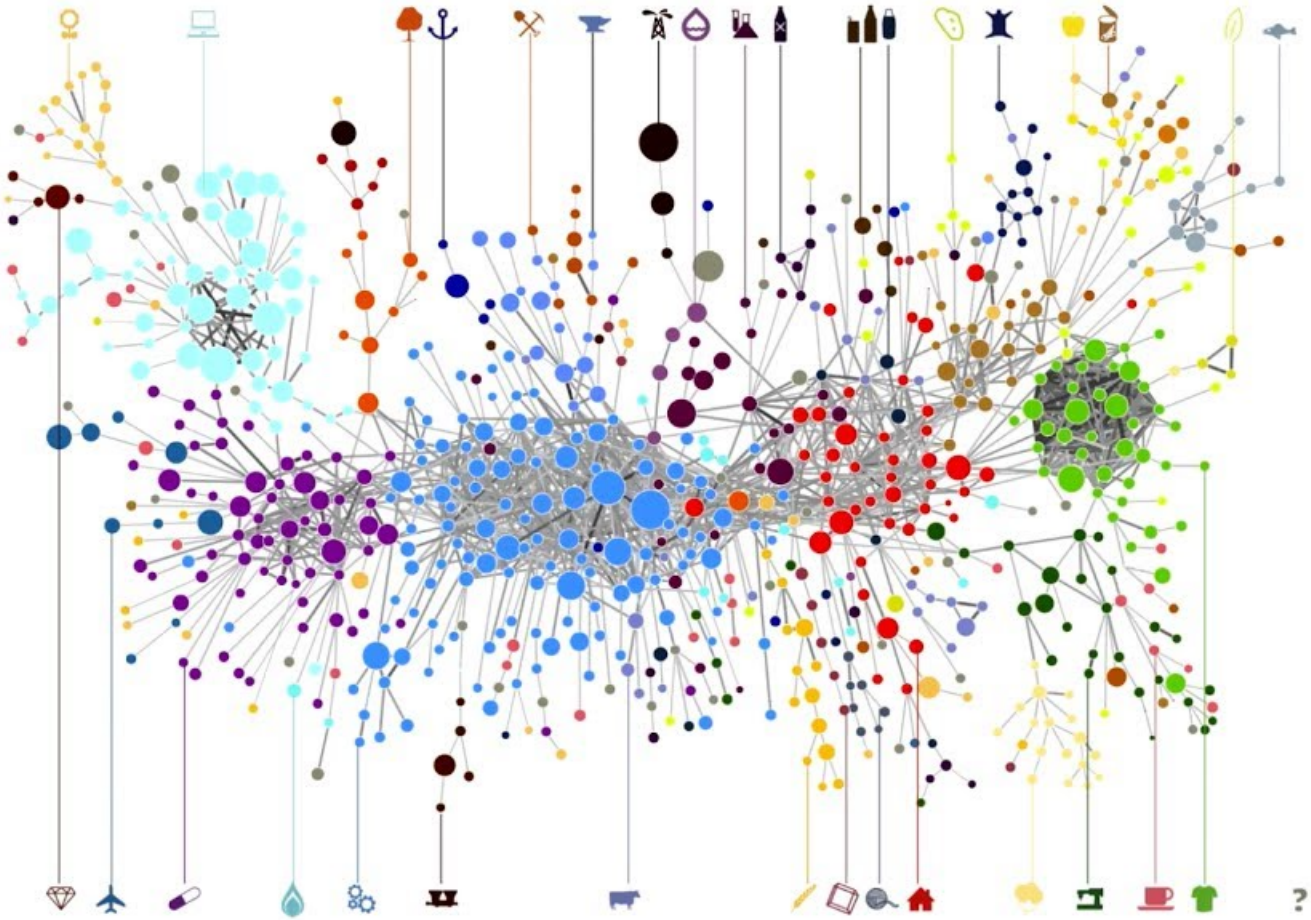
Fundamental idea IV: The product space

- The *product space* is a network where...
  - ... each node is a product
  - ... each link captures the proximity of two products
- What is meant by *proximity*? $\rightarrow$ Boschma (2005)
- For two products *i* and *j*, proximity is given by:

$$\phi_{i,j} = \min\left(\mathbb{P}(RCA_i|RCA_j), \mathbb{P}(RCA_j|RCA_i)\right) \tag{8}$$

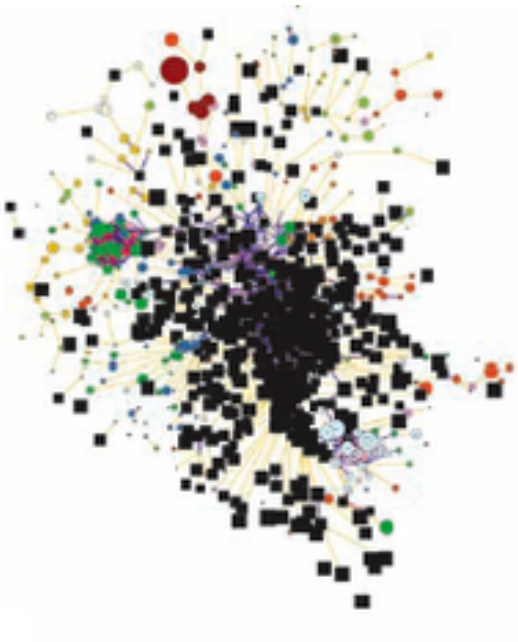
- $\phi_{i,j}$ is high if...
  - ...countries that have an RCA for product *i* also have an RCA for product *j*
  - Example: apples and oranges, or microprocessors and transistors
- $\phi_{i,j}$ is low if...
  - ...countries that have an RCA for product *i* are rather unlikely to also have an RCA for *j*
  - Example: apples and microprocessors, or oranges and transistors
- Cesar A Hidalgo et al. (2007) represent this as a network
- Interpretation: close products require similar knowledge

The product space

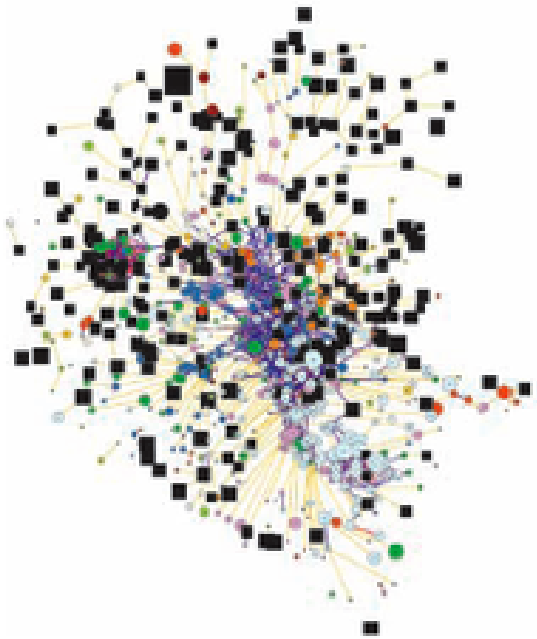


Source: Ricardo Hausmann, Cesar A Hidalgo, et al. (2014)

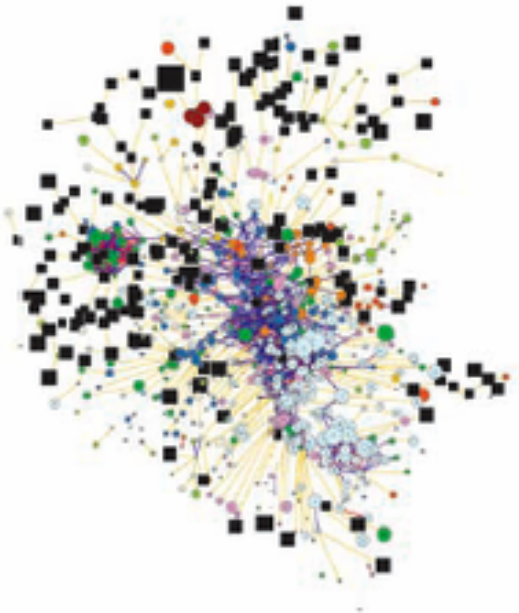
The product space



Industrialized countries



Latin American countries



African countries

Source: Ricardo Hausmann, Cesar A Hidalgo, et al. (2014)

Some remarks on the product space

- The product space is *outcome-based*
  - Two products are close if they require similar inputs
- It says little about what these inputs are and how they can be acquired
  - Infrastructure, institutions, human capital...
  - Capabilities of individual people...
  - ...and ability of organizations to bring them together
- How such capability accumulation takes place is one of the major research frontiers of the CAD
- Taking the product space seriously implies accepting a number of political economy issues:
  - Do countries in the core really have an incentive to let other countries enter the core?
  - Isn't international trade much more zero-sum than commonly believed?
  - Doesn't this imply that dependency theories, infant industry protection, and technology gap research has much more merit than commonly believed?

The fundamental building blocks

- We approach the ECI via its fundamental theoretical building blocks:

1. The idea that person bytes are a major determinant of development
2. The idea of how to measure person bytes
3. The idea of diversity and ubiquity of products and ideas
4. The idea of the product space
5. **The definition of the ECI and the PCI**

An ECI and an PCI

- When talking about the ECI we are actually talking about two distinct measures
- First, the **ECI** is a measure for countries
- Second, the **PCI** is a measure for products
- In theory, it takes a complex country to manufacture a complex product...
- ...and complex products are those that can only be manufactured in complex countries
- This mutual dependency entails the key to compute these measures
- But in practice they can be used for quite distinct purposes

Fundamental idea V: The ECI and PCI

- The *economic complexity index* (ECI) and the *product complexity indec* (PCI) wrap up all the above-set into single numbers
- We begin with our measures for *ubiquity* and *diversity*

$$k_{c,N} = \frac{1}{k_{c,0}} \sum_p M_{cp} \kappa_{p,N-1} \tag{9}$$

$$\kappa_{p,N} = \frac{1}{\kappa_{p,0}} \sum_c M_{cp} k_{c,N-1} \tag{10}$$

- For the first iterations remember the clear-cut interpretations:

	Explanation
$k_{c,0}$	Number of products exported by country c.
$k_{c,1}$	Average ubiquity of the products exported by country c.
$k_{c,2}$	Average diversification of countries with an export basket similar to country c.
$\kappa_{p,0}$	Number of countries exporting product p.
$\kappa_{p,1}$	Average diversification of the countries exporting product p.
$\kappa_{p,2}$	Average ubiquity of the products exported by countries that export product p.

Deriving the ECI

- We can drive the recursions of $k_{c,N}$ and $\kappa_{p,N}$ to the limit:
- Inserting equation (10) into (9):

$$k_{c,N} = \frac{1}{k_{c,0}} \sum_{p} M_{cp} \frac{1}{\kappa_{p,0}} \sum_{c'} M_{c'p} k_{c',N-2} \tag{11}$$

- which corresponds to:

$$k_{c,N} = \sum_{c'} k_{c',N-2} \sum \frac{M_{cp} M_{c'p}}{k_{c,0} \kappa_{p,0}}. \tag{12}$$

Deriving the ECI

- We can drive the recursions of $k_{c,N}$ and $\kappa_{p,N}$ to the limit:
- Inserting equation (10) into (9):

$$k_{c,N} = \frac{1}{k_{c,0}} \sum_p M_{cp} \frac{1}{\kappa_{p,0}} \sum_{c'} M_{c'p} k_{c',N-2} \tag{13}$$

- which corresponds to:

$$k_{c,N} = \sum_{c'} k_{c',N-2} \sum_p \frac{M_{cp} M_{c'p}}{k_{c,0} \kappa_{p,0}}. \tag{14}$$

- Setting $\tilde{M}_{cc'} = \sum_p \frac{M_{cp}M_{c'p}}{k_{c,0}\kappa_{p,0}}$:

$$k_{c,N} = \sum_{c'} k_{c',N-2}\tilde{M}_{cc'}. \tag{15}$$

- Recursion (15) reaches an equilibrium whenever $k_{c,N} = k_{c,N-2} = 1$.
- Taking the eigenvector $\vec{K}$ that corresponds to the second-largest eigenvalue of $\tilde{M}_{cc'}$ in equilibrium yields *after normalization*:

$$ECI = \frac{\vec{K} - mean(\vec{K})}{sd(\vec{K})} \tag{16}$$

- Equivalent reasoning and assuming that $\vec{Q}$ is the product equivalent to $\vec{K}$:

$$PCI = \frac{\vec{Q} - mean(\vec{Q})}{sd(\vec{Q})} \tag{17}$$

- If not directly digestible, I suggest you calculate one of the examples outlined in C A Hidalgo and Hausmann (2009)
- Have a look at the R and Python code on the course homepage

Overview

- The ECI and PCI are outcome-based measures
- They did not emerge from a clearly demarcated theoretical framework
- Given their focus on knowledge and discovery they are inspired by evolutionary theories
- At the same time, the lack of concrete theoretical underpinning is one of the greatest weaknesses of the approach
- Here I want to highlight three interesting complementary stream of literature:
    1. Hidalgo's work on the role of information in economics
    2. The theoretical affinity to Latin American structuralism
    3. The affinity to the evolutionary theory of technology gaps

An underlying theory?

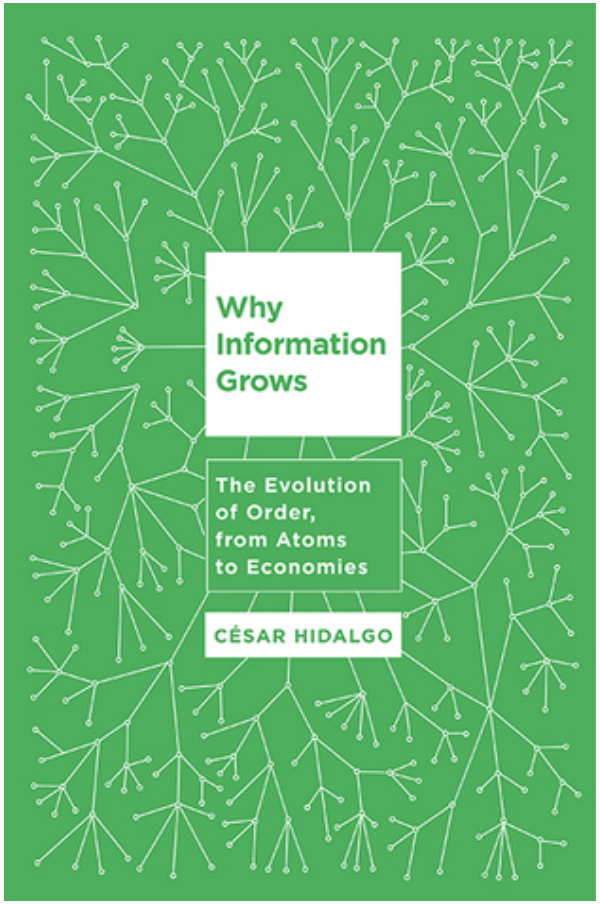
1. **Hidalgo's work on the role of information in economics**

2. The theoretical affinity to Latin American structuralism

3. The affinity to the evolutionary theory of technology gaps

Hidalgo on economic growth as growth of information I

- Cesar A Hidalgo (2015) presents an sketch of a theory underlying complexity
- Inspired by Hidalgo's education as a physicist it rethinks economic growth as growth of information
- For him, the fundamental question of economics (and social sciences in general) is: where does the order in our society come from?
- According to the second law of thermodynamics, we should observe increasing disorder
- The counter-forces are information and computation
- Information is the order embodied in codified sequences such as DNA
- To grow, information needs energy to emerge, matter to be stored, and computational abilities of matter to adapt and evolve



Hidalgo on economic growth as growth of information II

- Economies grow and evolve through the embodiment of increasingly large amounts of information into increasingly complex physical objects
- But a single human has limited computational power - that of a *person byte*
- But as social animals humans can distribute their computational power in networks
- They can also `cristalyze' their computations and imaginations in products -- which are considered ``crystalized imagination''
- This implies that the stronger the ties between humans in a society, the more they can counter-act the natural tendency to disorder
- Complex activities tend to be concentrated to regions with a diverse set of knowledge
- Key conclusion: ``over the long run a region's level of income will approach the complexity of its economy'' (p. 180).
- Unfortunately, what remains open is the answer to the most important questions:
- Why are knowledge and knowhow distributed in such a specific way?

An underlying theory?

1. Hidalgo's work on the role of information in economics
2. **The theoretical affinity to Latin American structuralism**
3. The affinity to the evolutionary theory of technology gaps

The affinity between the CAD & structualism I

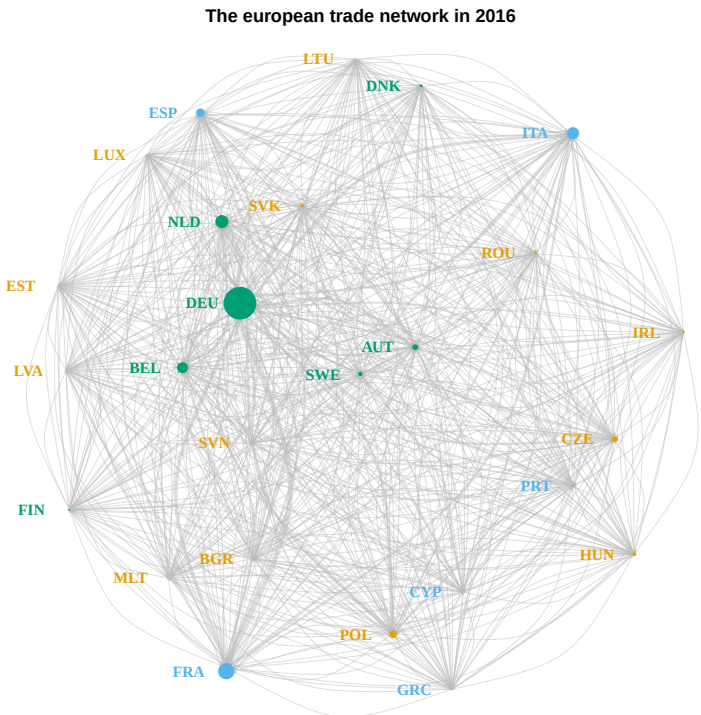
- Structualism as a theoretical framework emerged in Latin America in the 1950s
- Continues to thrive today, particularly within the *Economic Commission for Latin America and the Caribbean* (ECLAC)
- Posits that the industrial structure determines the direction of economics development
- Development process reallocation from low productivity to high productivity sectors where increasing returns to scale
- Globally, there are dominant *cores* and dependent *peripheries*, related to each other via global trade and payment flows
- The only way for the peripheries to escape their situation is technological upgrading
- The theoretical claims have been very difficult to test

The affinity between the CAD & structualism II

- Just as Hidalgo, the structualists emphasize the role of production structures and the unequal distribution of technological capabilities
- The product space seems to be *the* vehicel to test stuctualist conjectures
  - The product space is a core-periphery network
  - The richer countries are located in the core, the poorer countries in the periphery
- The struggle to reach the centre is one of technological upgrading
- But it is also a struggle of interests -- with asymmetric power relations playing an essential role
- Moreover, structualism is a very *systemic* theory, Hidalgo's theory is not

The affinity between the CAD & structualism III

- The combination of complexity and structualism seems particularly promising for the European Union
- Core-periphery thinking has some tradition here
- Today the core-periphery thinking enjoys a revival (Simonazzi, Ginzburg, and Nocella, 2013; Gräbner et al., 2017; Celi et al., 2018; Gräbner et al., 2019, e.g.)
- Also indicates that the dichotomy between core and periphery becomes difficult to sustain (Gräbner et al., 2019, e.g.)
- The ability to produce complex products seems to be *one* important explanation for the polarization between rich and poor
- Discuss in more detail during applications section



An underlying theory?

1. Hidalgo's work on the role of information in economics

2. The theoretical affinity to Latin American structuralism

3. **The affinity to the evolutionary theory of technology gaps**

- Technological change and innovation has been a central topic in evolutionary economics
- The idea of knowledge as a central diver of economic development led to a critical view on the idea of comparative advantage
- It has frequently been argued for the relevance of *absolute* advantages and the relevance of *technology gaps* (Dosi, Grazzi, and Moschella, 2015)
- The resulting theories of economic development were mostly applicable on the regional level
- Today strong overlap with economic geography
- Current focus is on empirics of knowledge flows and spillovers
- The relation to the theory of economic complexity are natural and bear the potential for a clearer theoretical foundation
- Less attention is put on the macroeconomic level and the political economy dimension, though

Summary

- There are two related measures associated with the CAD:
  - The Economic Complexity Index (ECI)
  - The Economic Complexity Index (PCI)
- The ECI measures the amount of technological capabilities accumulated in a country (or region...)
- The PCI measure the amount of tech capabilities required to manufacture a product
- One starts with an export matrix, with the rows being countries and the columns products
- The nb of countries exporting a product denote its *ubiquity*
- The nb of products a country exports denotes its *diversity*
- By correcting ubiquity by the diversity of the exporting countries and vice versa we eventually come the ECI and PCI
- The structure of the matrix can also be formalized as a network - the product space
- Products in the center are most attractive -- countries exporting these countries tend to be rich
- The theoretical underpinning of the approach is poor, but there are affinities to structualist and evolutionary theory

References I

Ron Boschma. “Proximity and Innovation: A Critical Assessment”. In: *Regional Studies* 39.1 (Feb. 2005), pp. 61–74.

Tom Broekel. “Measuring technological complexity - Current approaches and a new measure of structural complexity”. In: *arxiv* (2017).

Giuseppe Celi et al. *Crisis in the European Monetary Union*. A Core-Periphery Perspective. London: Routledge, 2018.

Giovanni Dosi, Marco Grazzi, and Daniele Moschella. “Technology and costs in international competitiveness: From countries and sectors to firms”. In: *Research Policy* 44.10 (Dec. 2015), pp. 1795–1814.

Lee Fleming and Olav Sorenson. “Technology as a complex adaptive system: evidence from patent data”. In: *Research Policy* 30.7 (Aug. 2001), pp. 1019–1039.

Claudius Gräbner et al. “Is Europe disintegrating? Macroeconomic divergence, structural polarization, trade and fragility”. In: *ICAE Working Paper* 64 (July 2017).

Claudius Gräbner et al. “Structural change in times of increasing openness: assessing path dependency in European economic integration”. In: *Journal of Evolutionary Economics* forthcoming (2019).

Ricardo Hausmann, Cesar A Hidalgo, et al. *The Atlas of Economic Complexity*. Cambridge, MA: Puritan Press, 2014.

C A Hidalgo and R Hausmann. “The building blocks of economic complexity”. In: *Proceedings of the National Academy of Sciences* 106.26 (2009), pp. 10570–10575.

Ricardo Hausmann, Jason Hwang, and Dani Rodrik. “What you export matters”. In: *Journal of Economic Growth* 12.1 (2007), pp. 1–25.

Cesar A Hidalgo et al. “The Product Space Conditions the Development of Nations”. In: *Science* 317.7 (2007), pp. 482–487.

Cesar A Hidalgo. *Why Information Grows*. The Evolution of Order, from Atoms to Economies. New York, NY: Basic Books, 2015.

References II

Ricardo Hausmann and Bailey Klinger. "Structural Transformation and Patterns of Comparative Advantage in the Product Space". In: *CID Working Papers* (Aug. 2006).

Ricardo Hausmann and Dani Rodrik. "Economic development as self-discovery". In: *Journal of Development Economics* 72.2 (Dec. 2003), pp. 603–633.

Annamaria Simonazzi, Andrea Ginzburg, and Gianluigi Nocella. "Economic Relations Between Germany and Southern Europe". In: *Cambridge Journal of Economics* 37.3 (2013), pp. 653–675.

Andrea Tacchella et al. "A New Metrics for Countries' Fitness and Products' Complexity". In: *Scientific Reports* 2.723 (Oct. 2012), pp. 1–7.