ASI/UNOOSA Open Universe initiative:
Expert Meeting
Rome, 11 April 2017

Andrea Zaccaria
Institute for Complex Systems, CNR, Rome, Italy

The Economic Complexity Approach to Big Data: an Application to the Technological Impact of Astrophysics

In collaboration with
G. Cimini, M. Cristelli, G.L. Chiarotti, A. Gabrielli, E. Pugliese,
A. Tacchella, and L. Pietronero

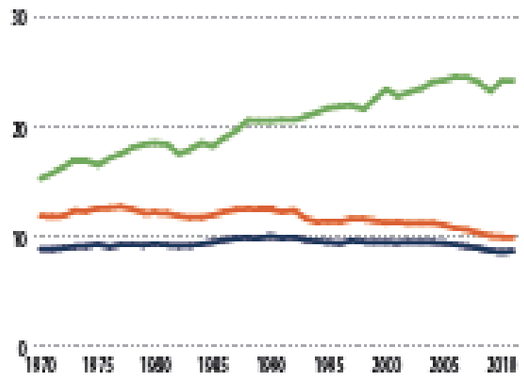
Figure 1.2 How Sub-Saharan Africa fares in relation to eight earlier transformers

The figures here show how Sub-Saharan Africa is performing in relation to eight earlier transformers on various indicators of depth.

ACET 15
Sub-Saharan Africa
Earlier transformers

a Diversity: production

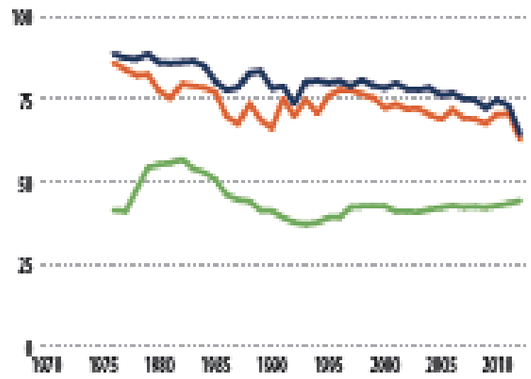
% of manufacturing in GDP



Source: World Development Indicators (database).

b Diversity: exports

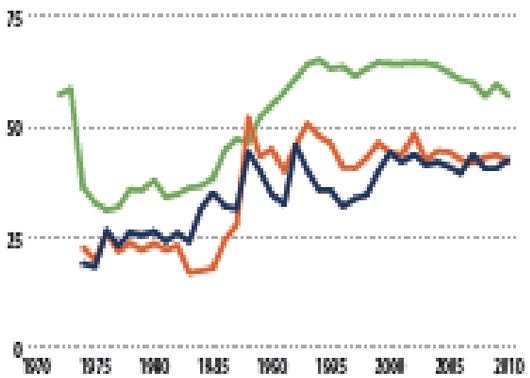
% of top five exports in total exports



Source: UN Comtrade, Revision 2, Digit 3.

c Diversity: exports of manufactures and services

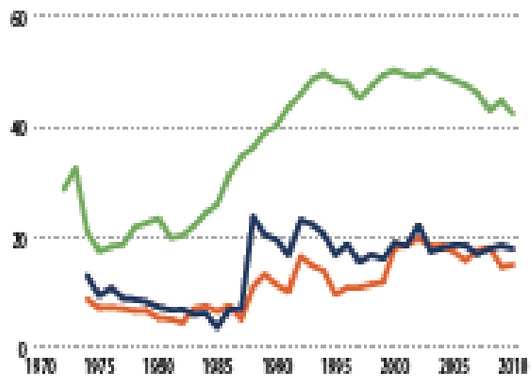
% of total goods and services exports



Source: World Bank staff estimates; World Trade Organization; IMF.

d Diversity: exports of manufactures

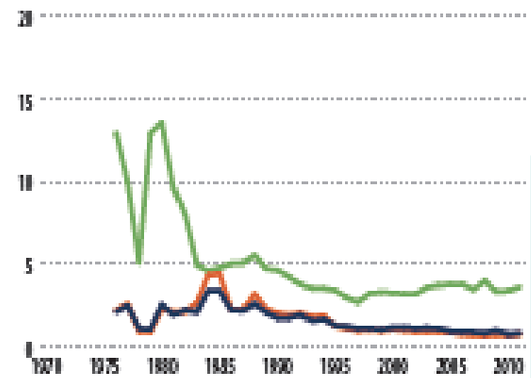
% of total goods and services exports



Source: World Bank staff estimates; World Trade Organization; IMF.

e Export competiveness: export market share without extractives

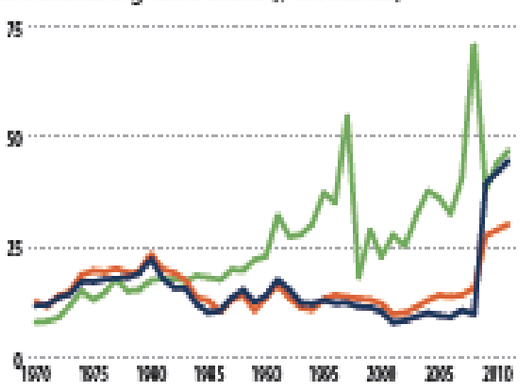
% of exports in GDP relative to world average (without extractives)



Source: World Development Indicators (database); UN Comtrade, Revision 2, Digit 3.

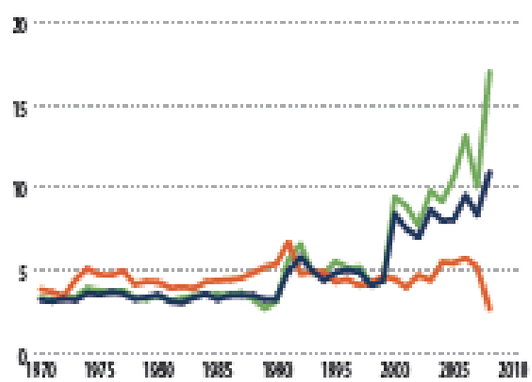
f Productivity: manufacturing value added per worker

Manufacturing value added ($ thousands)



Source: UNIDO, Revision 3, Digit 2.

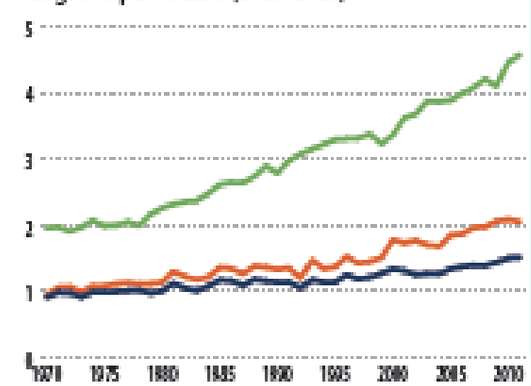
g Productivity: ratio of labor productivity to the average wage in manufacturing



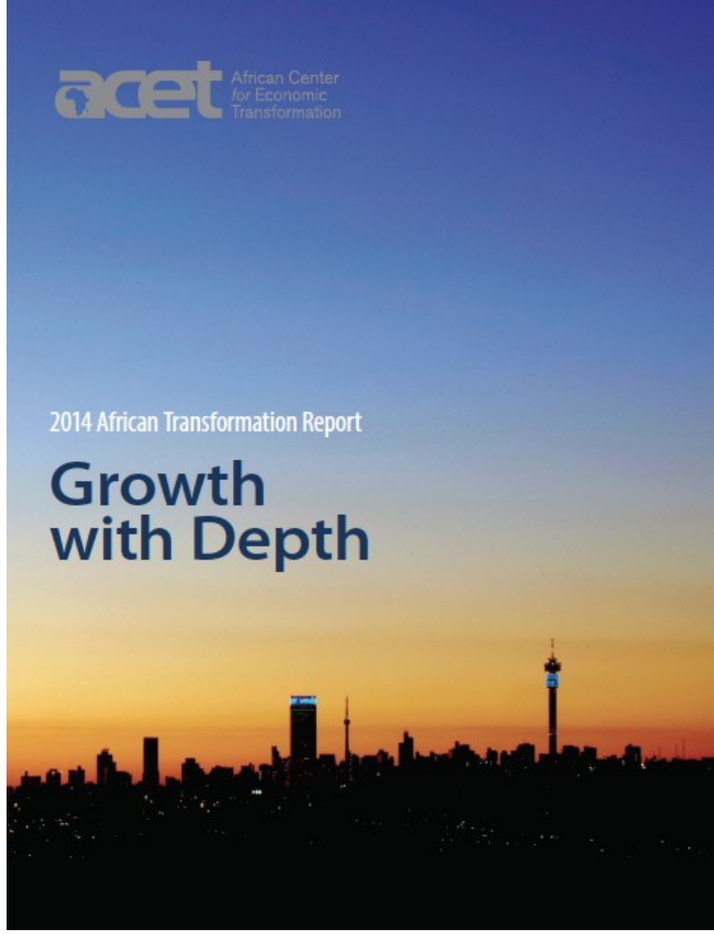
Source: UNIDO, Revision 3, Digit 2.

h Productivity: cereal yields

Kilograms per hectare (thousands)

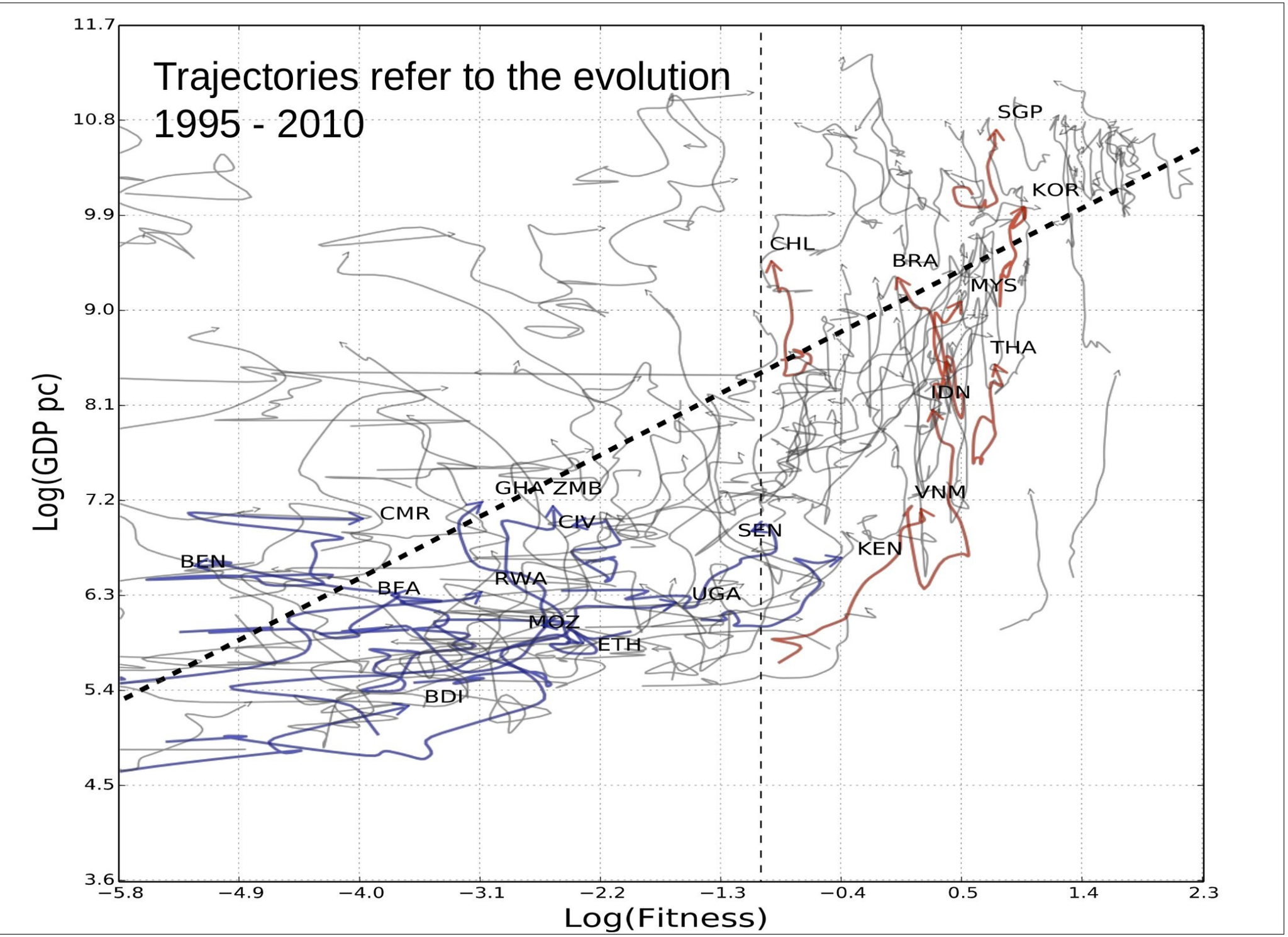


Source: World Development Indicators (database).



- ACET report on Industrialization
- 221 pages
- Comparison among dozens of indicators & countries
- Experts needed
- No clear conclusion!

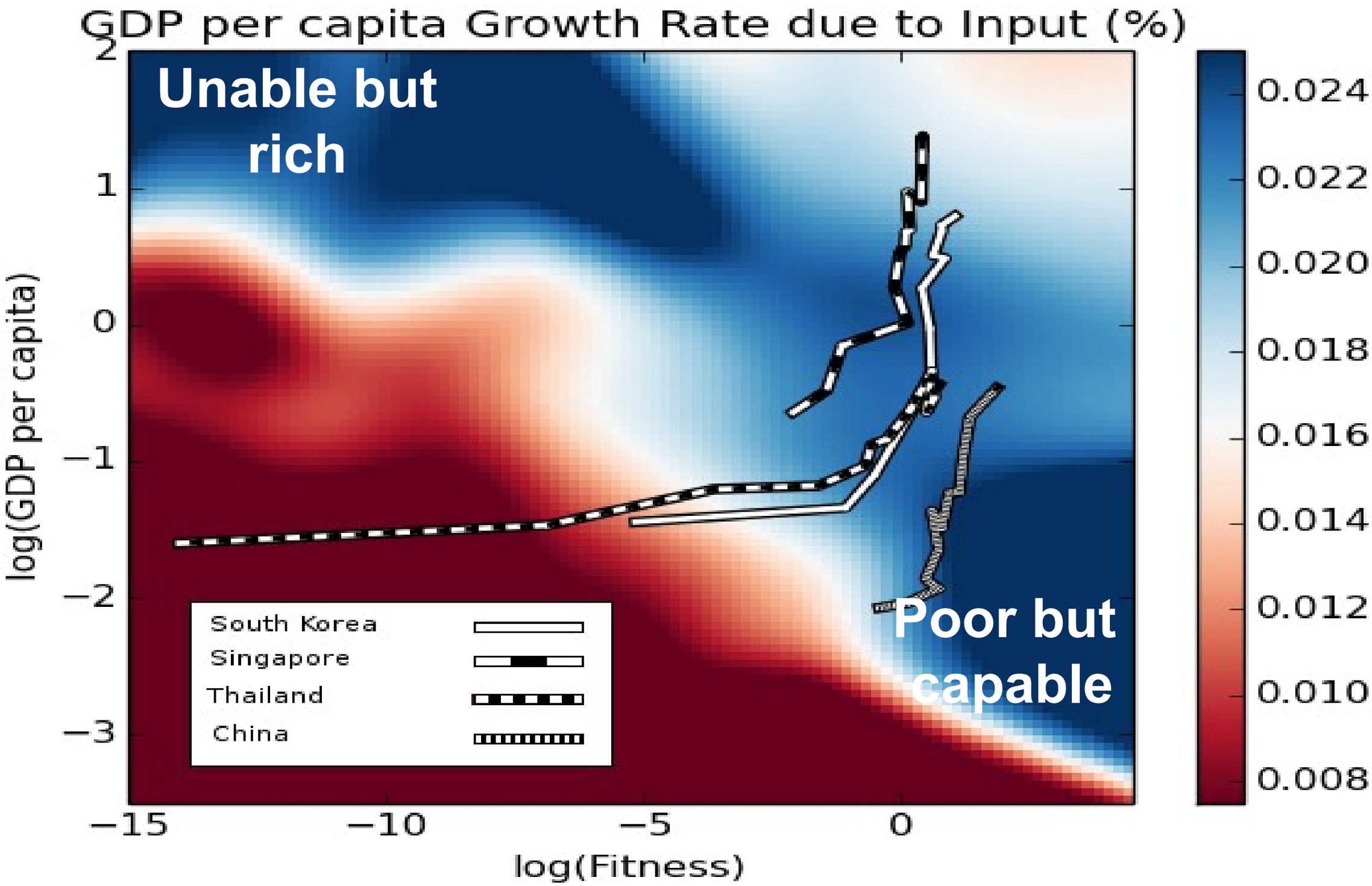
Economic Complexity Approach



- **New synthetic concepts**
- **Individual countries trajectories in a new space**
- **Clear interpretation**

A Tacchella, M Cristelli, G Caldarelli, A Gabrielli, and L. Pietronero
A New Metrics for Countries' Fitness and Products' Complexity.
Scientific Reports 2, 273 (2012)

Poverty trap: a lateral exit

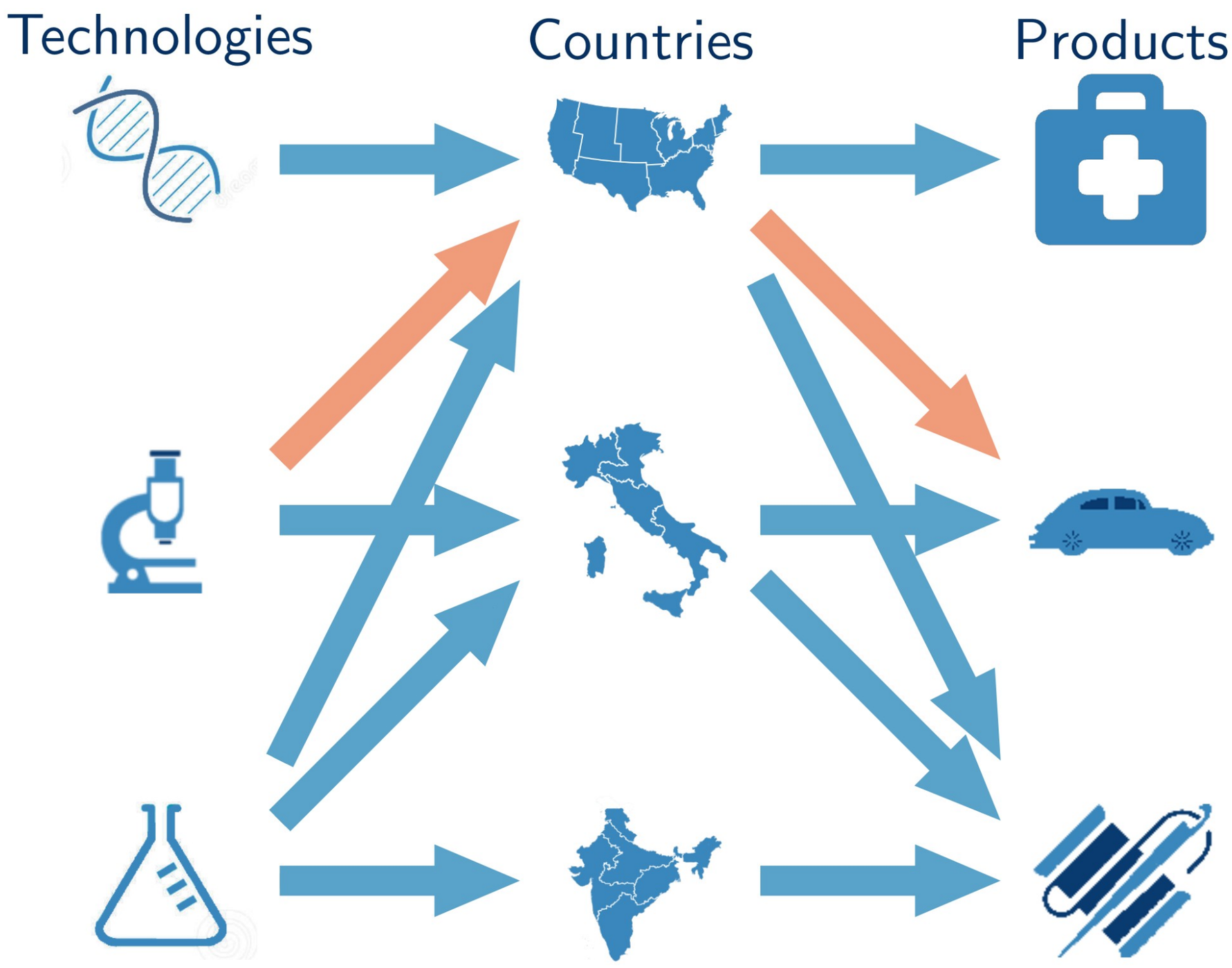


E Pugliese, G L Chiarotti, A Zaccaria, and L Pietronero. Complex economies have a lateral escape from the poverty trap. PloS ONE, 12(1):e0168540, 2017.

Database

- Export of products: UN-COMTRADE data of trade flows between countries, 1963-2015. About 4000 (6 digits) products
- Patenting activity: PATSTAT~ 100M patents from mid 19th century, about 70K technological codes with country of origin of the applicant
- Scientific activity: SciVal platform, an API collecting aggregating data from Scopus: production of articles in journals, proceedings, books, including citations. 254 scientific fields, about 20 years

The multilayer space of development



From data to tools

$N_{\mathcal{C}} \times N_{\mathcal{S}}$ "country-science" matrix

$$M^{\mathcal{S}}_{cs}(y) = \begin{cases} 1 & \text{if } \dfrac{w_{cs}(y)}{\sum_{s'} w_{cs'}(y)} \Big/ \dfrac{\sum_{c'} w_{c's}(y)}{\sum_{c's'} w_{c's'}(y)} \geq 1 \\ 0 & \text{otherwise.} \end{cases}$$

$N_{\mathcal{C}} \times N_{\mathcal{P}}$ "country-product" matrix $M^{\mathcal{P}}_{cp}(y)$

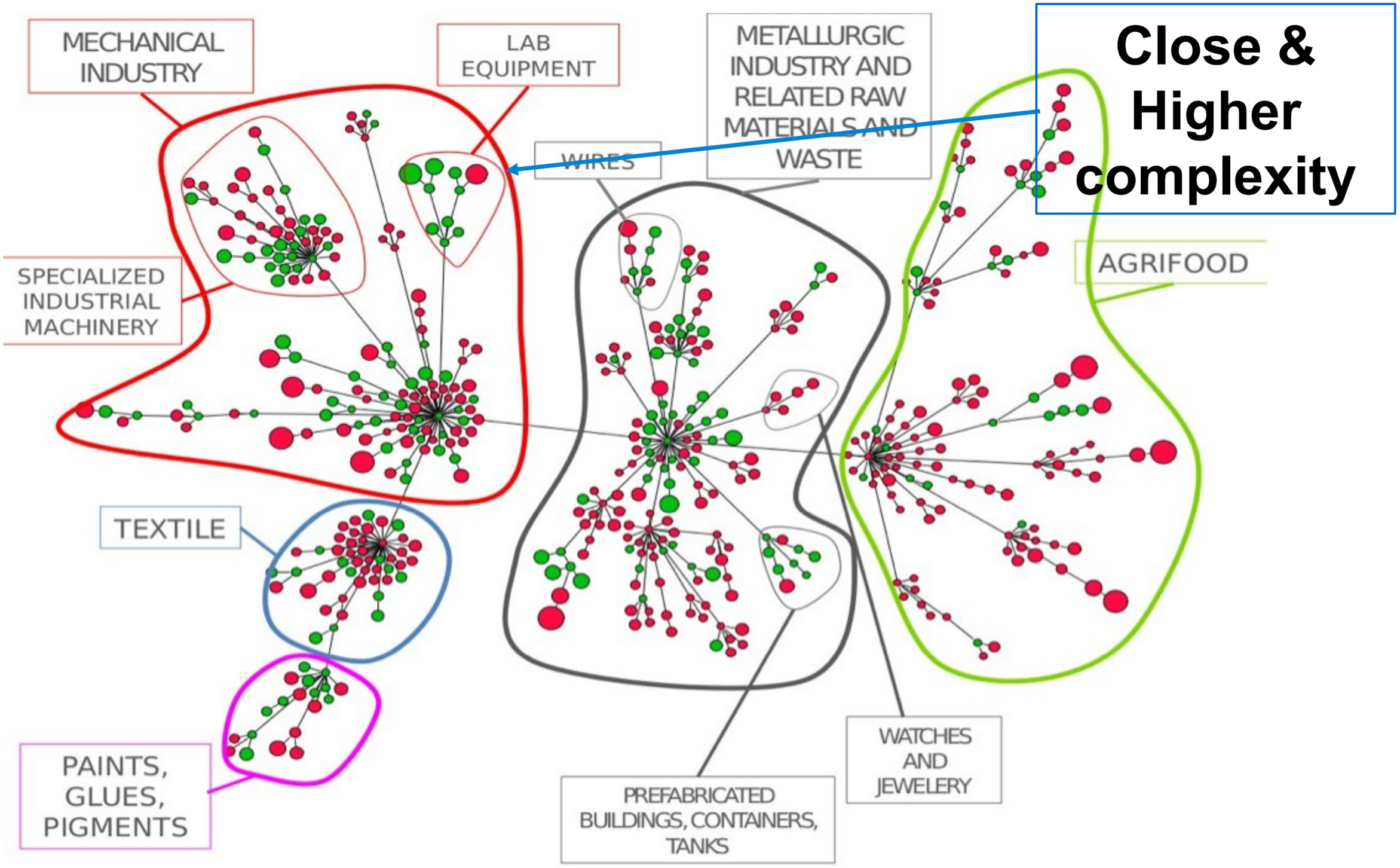
$N_{\mathcal{C}} \times N_{\mathcal{T}}$ "country-technology" matrix $M^{\mathcal{T}}_{ct}(y)$

Balassa's RCA on number of citations, export volumes in US $, number of patent families applied

"Mass" or "resource" diffusion: equal distributions from and to the C,S,P,T layers

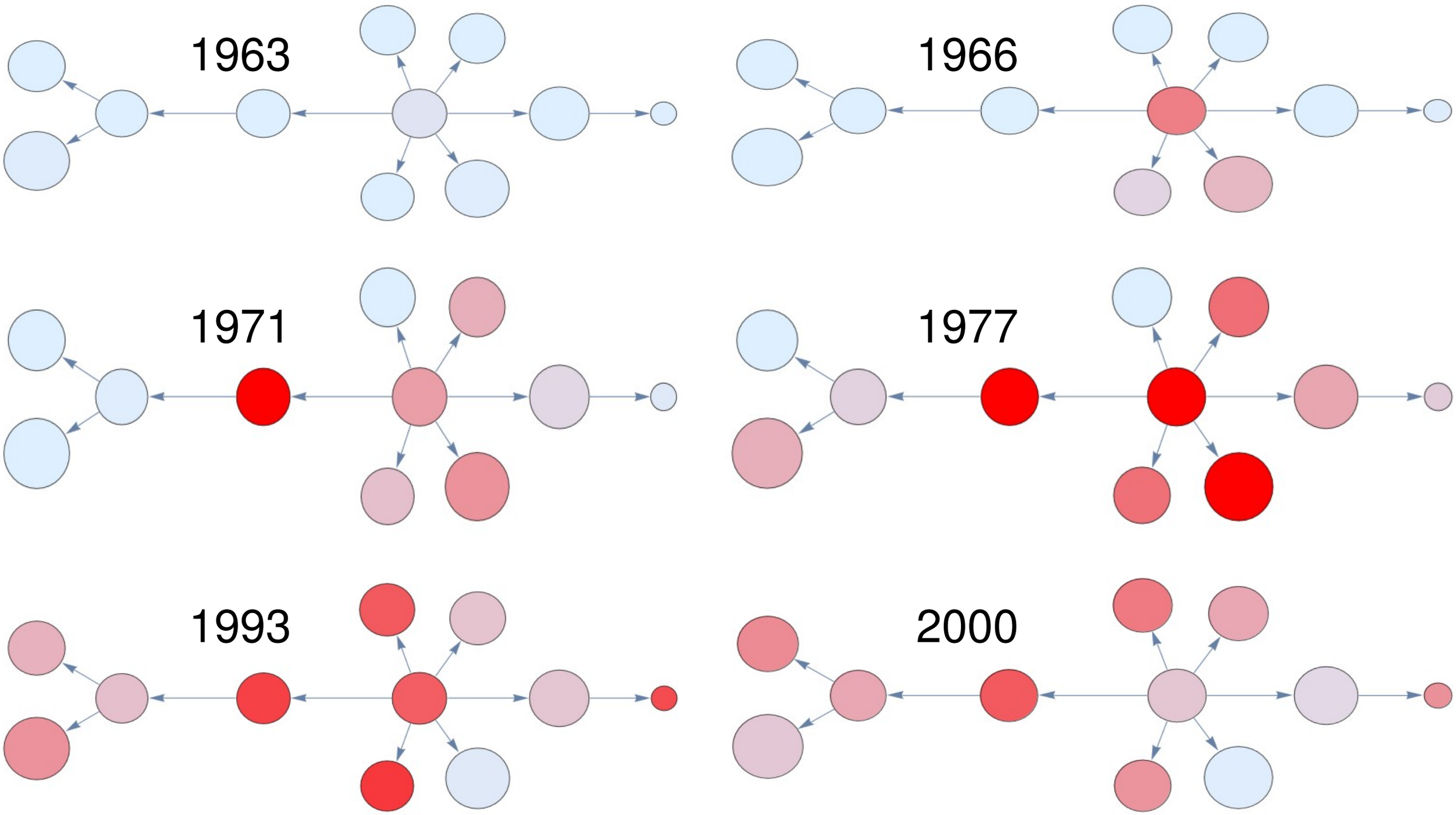
$$B^{\mathcal{L}\to\mathcal{K}}_{l\to k}(y_1, y_2) = \sum_{c\in\mathcal{C}} \mu^{\mathcal{L}\to\mathcal{C}}_{l\to c}(y_1)\mu^{\mathcal{C}\to\mathcal{K}}_{c\to k}(y_2) = \frac{1}{u^{\mathcal{L}}_l(y_1)} \sum_{c\in\mathcal{C}} \frac{M^{\mathcal{L}}_{cl}(y_1)M^{\mathcal{K}}_{ck}(y_2)}{d^{\mathcal{K}}_c(y_2)}$$

Product Progression Network

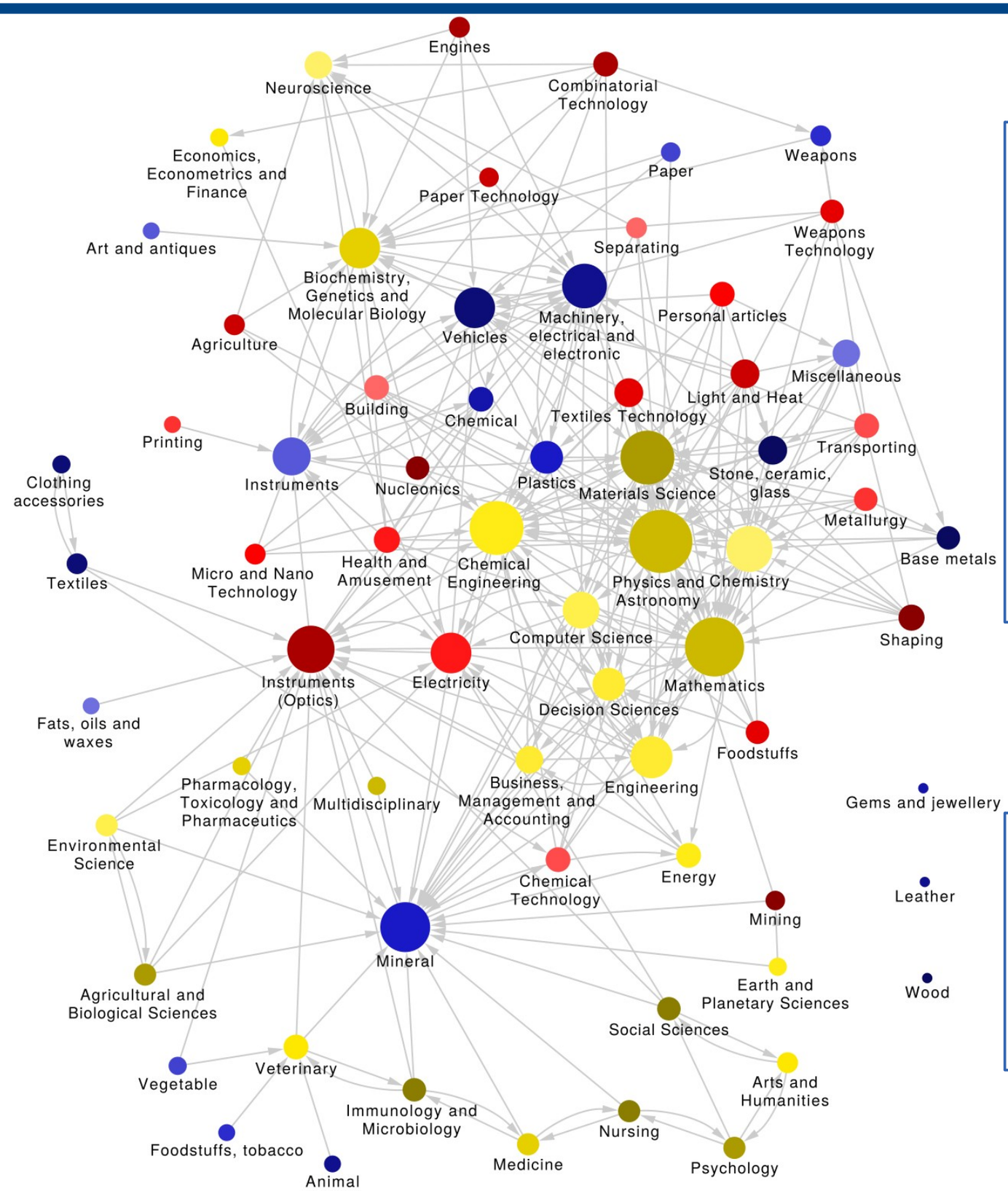


A Zaccaria, M Cristelli, A Tacchella, and L. Pietronero
How the Taxonomy of Products Drives the Economic Development of Countries.
PLoS ONE 9(12): e113770(2014)

Results: South Korea progression



Multilayer Network

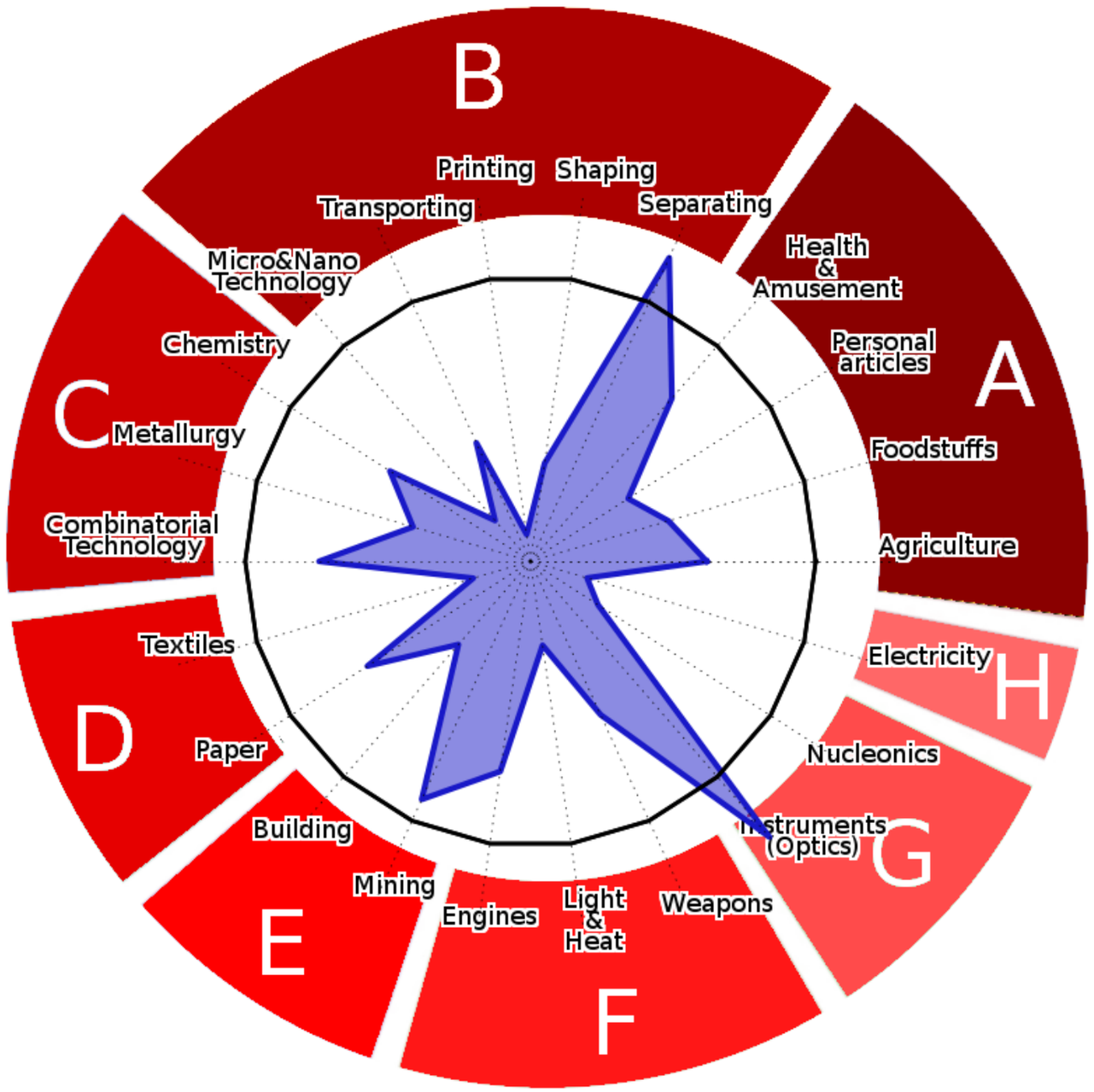


23 Scientific categories (yellow)
25 technological codes (red)
40 industrial sectors (blue)

Links obtained at 99.999% significance level wrt randomization

Astrophysics → Technologies

Starting nodes: Scopus codes 3103 (Astronomy and Astrophysics) and 1912 (Space and Planetary Science)



Recipient nodes:
E21B (Earth drilling)
B01J (Chemical and physical processes, e.g. catalysis)
G06Q (Data processing systems)
F04D (Non positive displacement pumps)
+ ...

Final comments

- **Economic Complexity: a new approach to "Big Data"**
- **Application 1: trade data, network of products**
- **Application 2: multilayer with science and technologies**

Open Universe initiative:

- **Will hopefully produce an increase of scientific production in astrophysics, triggering a cascade of positive feedback on technologies**
- **Hopefully this initiative will be extended also to economic databases**
- **Our experience with economic data: homogeneity is crucial. Same things should be called with the very same name**

Thank you for your attention!

A Zaccaria, M Cristelli, A Tacchella, and L. Pietronero
How the Taxonomy of Products Drives the Economic Development of Countries.
PLoS ONE 9(12): e113770 (2014)

E. Pugliese, GL Chiarotti, A Zaccaria, and L Pietronero *Complex economies have a lateral escape from the poverty trap,*PloS one, 12(1), e0168540 (2017)

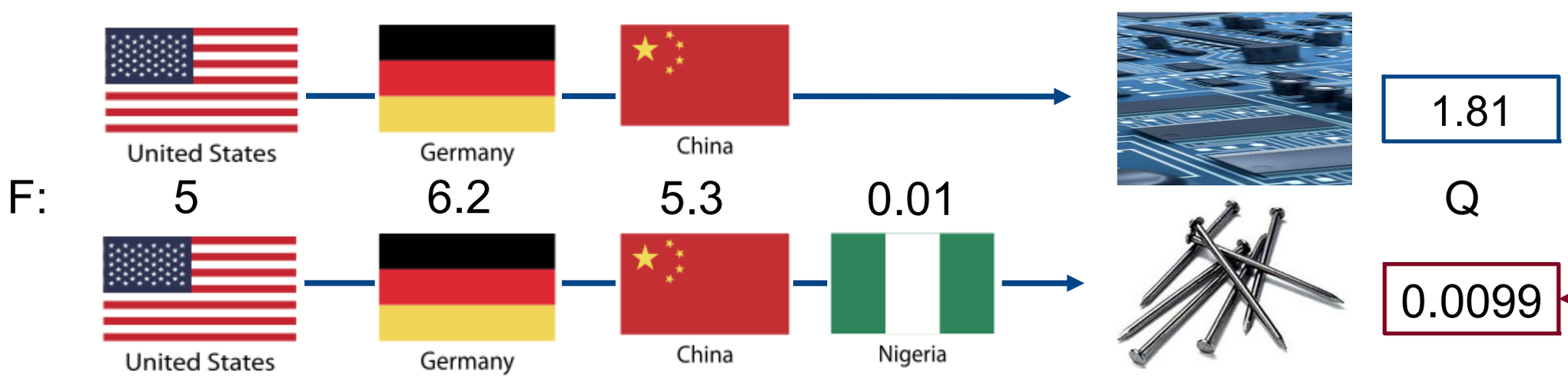
Fitness & Complexity Algorithm

$$\tilde{F}_c^{(n)} = \sum_p M_{cp} Q_p^{(n-1)} \qquad \tilde{Q}_p^{(n)} = \frac{1}{\sum_c M_{cp} \frac{1}{F_c^{(n-1)}}}$$

$$F_c^{(n)} = \frac{\tilde{F}_c^{(n)}}{\langle \tilde{F}_c^{(n)} \rangle_c} \qquad Q_p^{(n)} = \frac{\tilde{Q}_p^{(n)}}{\langle \tilde{Q}_p^{(n)} \rangle_p}$$

F: diversification weighted by complexity

Q: Extremal non-linear complexity of products. A single low fitness producer implies low complexity

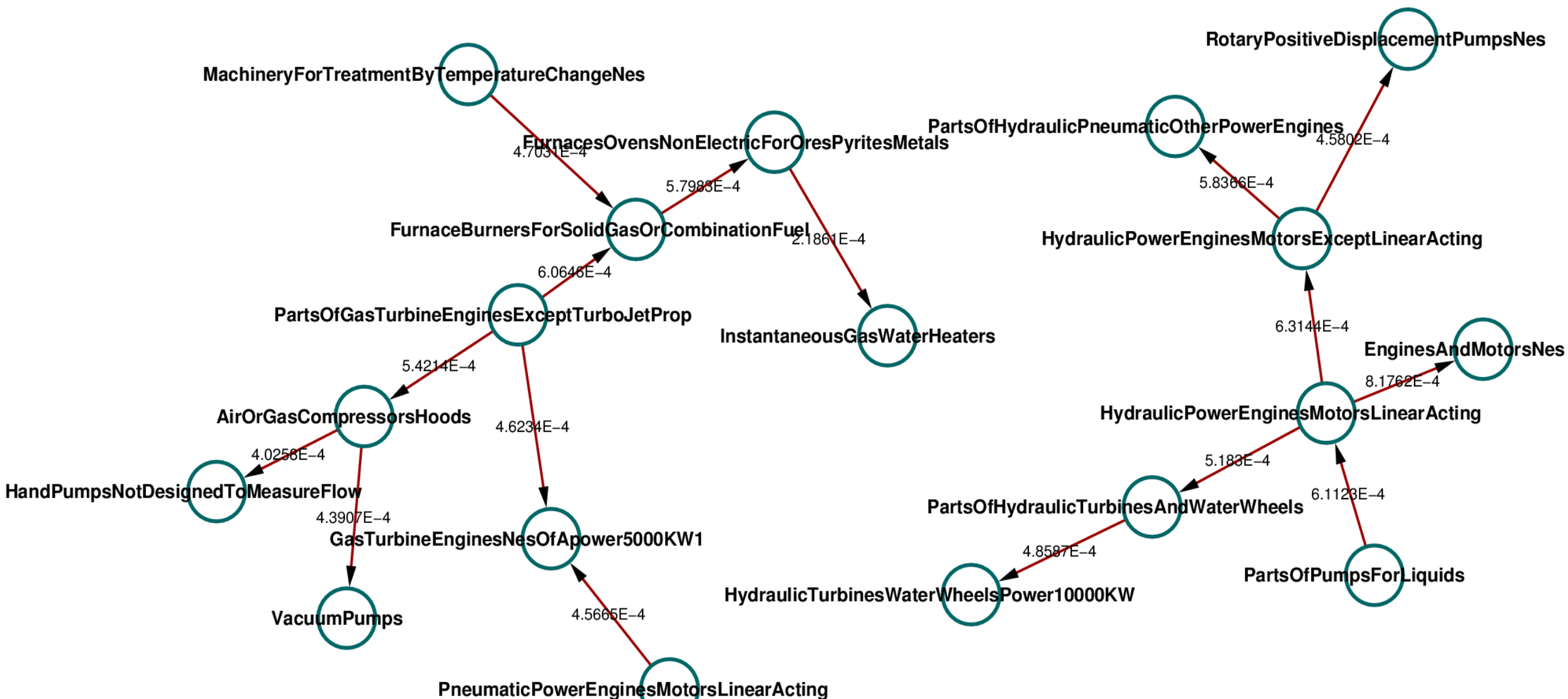


A Tacchella, M Cristelli, G Caldarelli, A Gabrielli, and L. Pietronero
A New Metrics for Countries' Fitness and Products' Complexity.
Scientific Reports 2, 273 (2012)

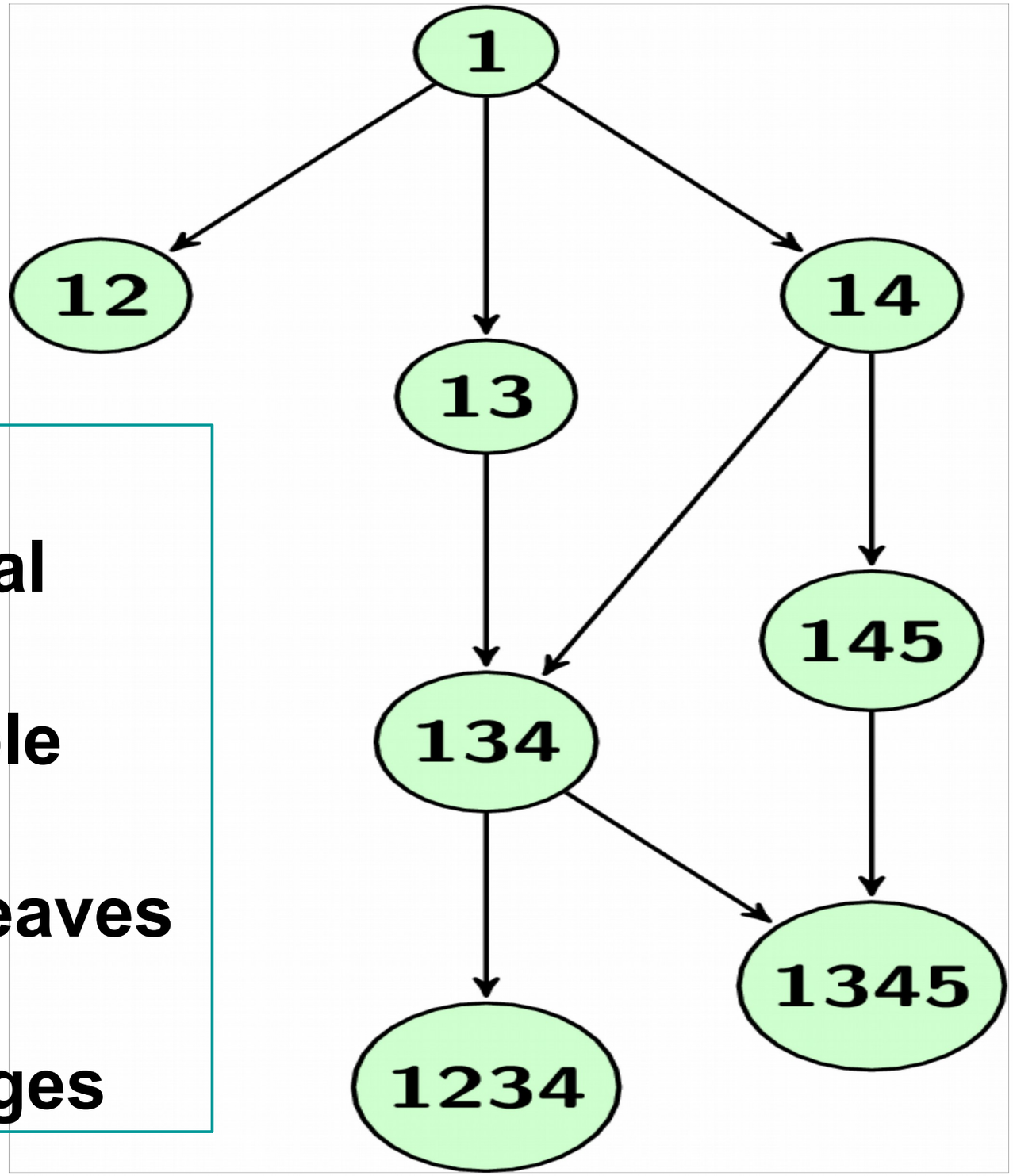
Product Progression Network

Example: components regarding hydraulic turbines, water wheels etc.

Validated & used for consulting by the World Bank: it reflects real processes and supply chains



Philosophy: the product progression



- Directed
- Hierarchical structure
- From simple roots to complex leaves
- Loops
- Double edges

MORE COMPLEX

Step 1/2: projection

Bipartite

Diversification

$$d_c = \sum_p M_{cp}$$

Ubiquity

$$u_p = \sum_c M_{cp}$$

$$B_{pp'} = \frac{1}{\max(u_p, u_{p'})} \sum_c \frac{M_{cp} M_{cp'}}{d_c}$$

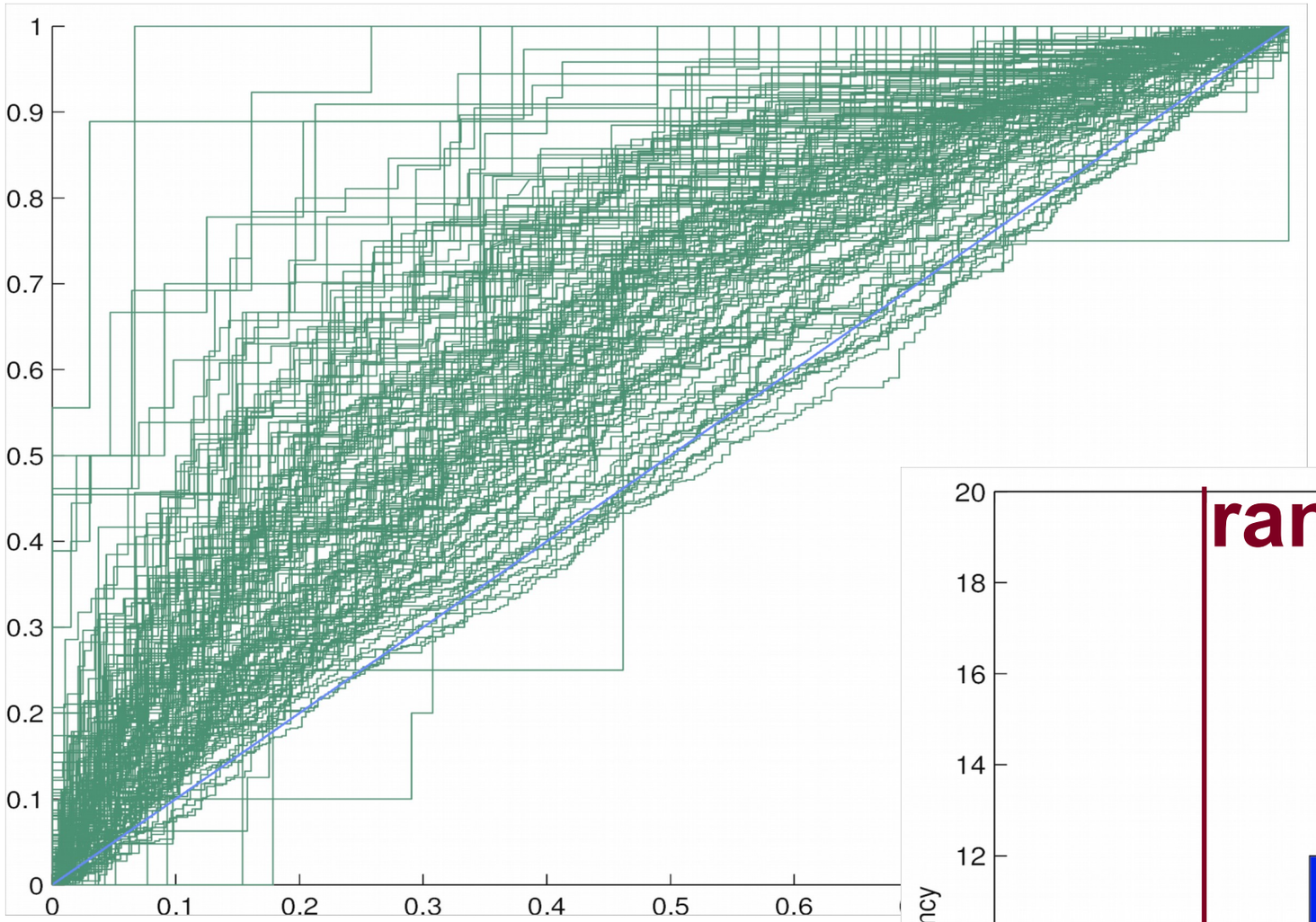
- **Conditional probability**
- **Nestedness**

Monopartite product-product undirected almost fully connected network

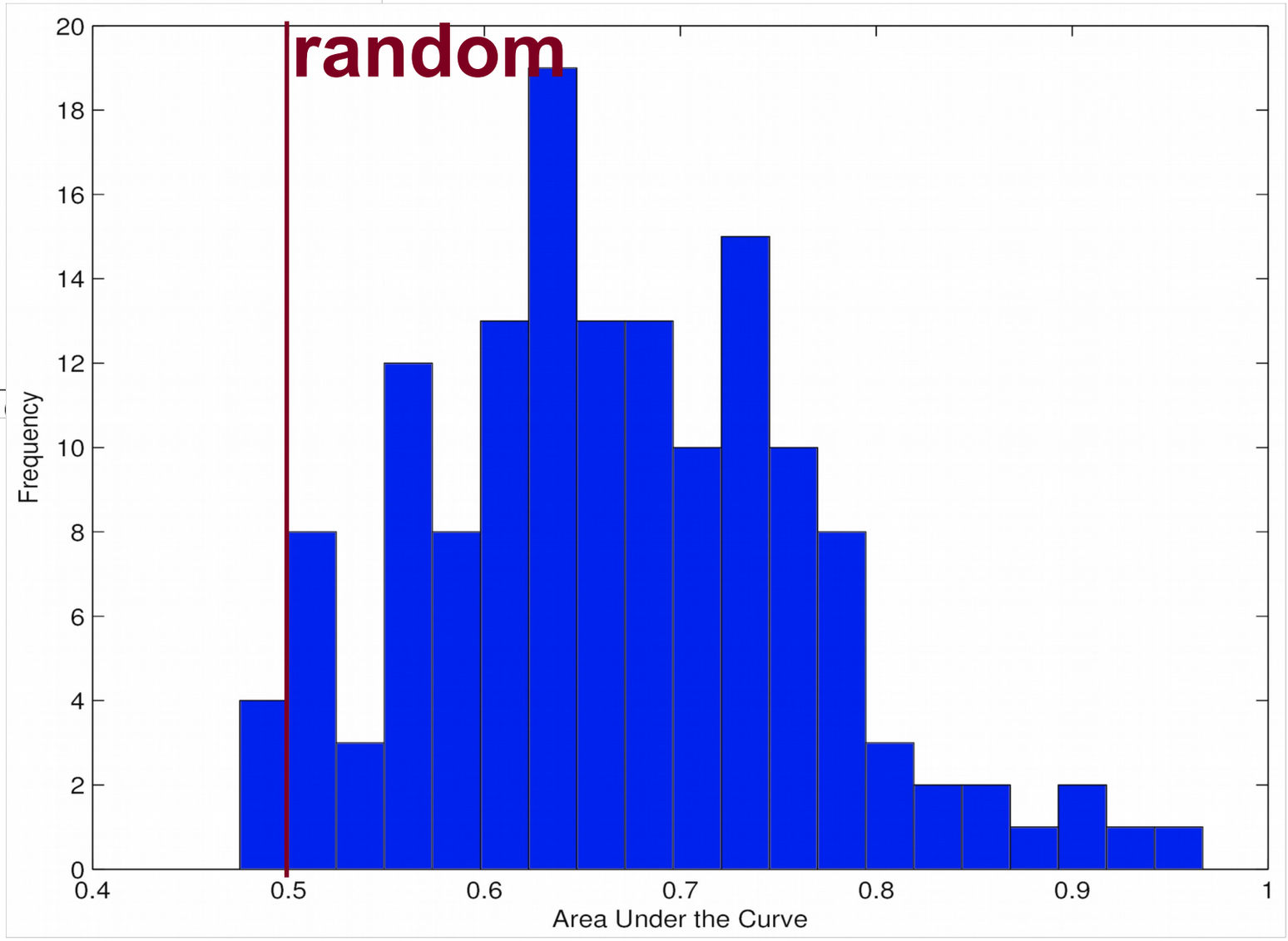
Step 2/2: directed filter

- Select i-th node
- Select maximum (no diagonal), and add (undirected) link (if more than one Max is present: keep the larger wrt to its column, i.e. the smaller column sum (with diagonal)
- Direction: from the node whose row sum is larger to the one whose row sum is lower (the source is small wrt its row)
- Repeat for each i

Predictions



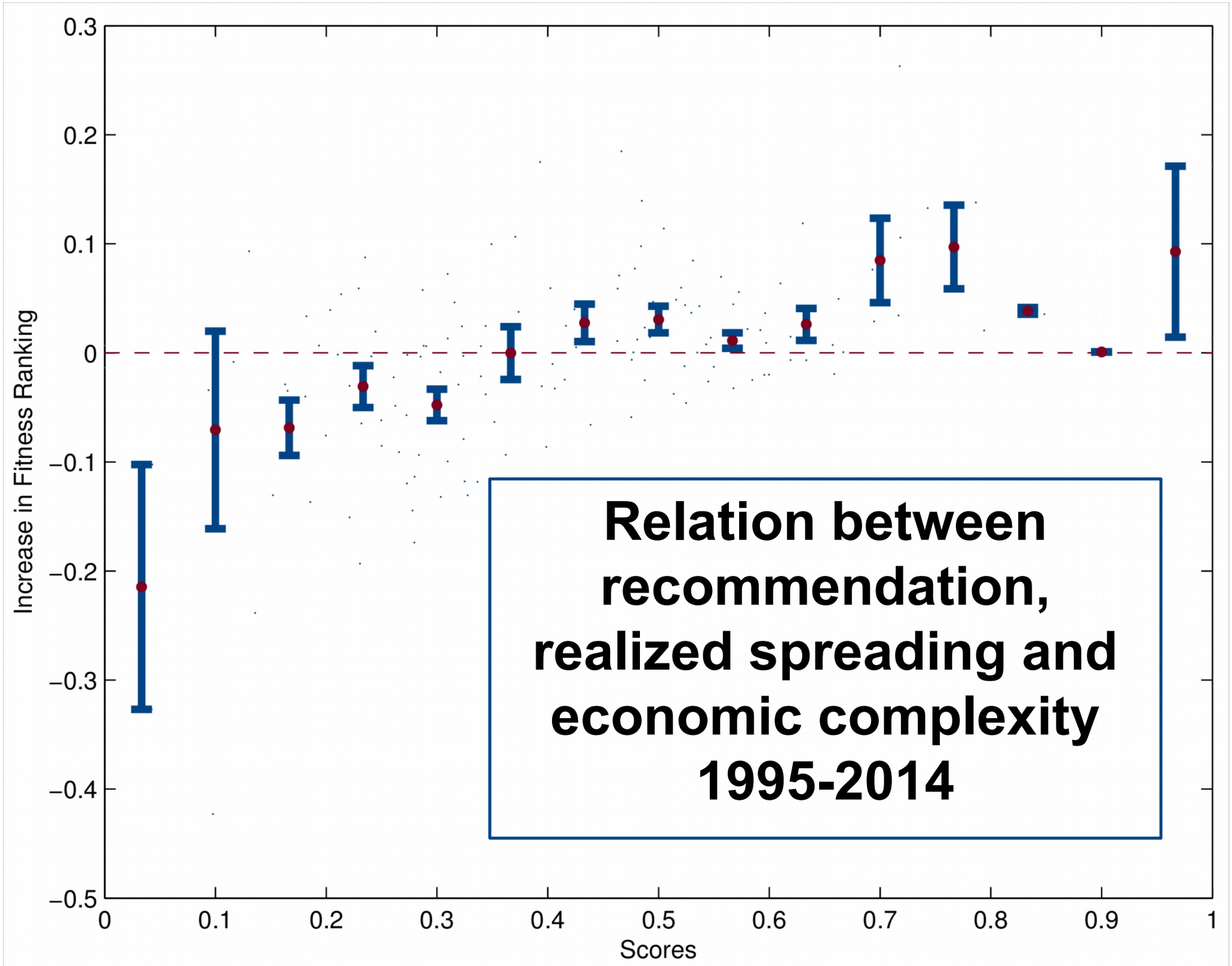
Given a country, sum the weights of the links from exported products



ROC curves systematically above the random threshold

Product Progression	Product Space
• Directed network: gives information about causality and time evolution. • Projection formula for nested bipartite networks • Dedicated filter, tested on simple matrices and capabilities-based toy model • Recommendations to lower the barrier to escape from the poverty trap and enter into the laminar regime	• Proximity-based, only similarity between products is taken into account: undirected • General formula based on conditional probabilities • The filtered network is used only for visualization. Why minimal spanning tree? • Proximity of products can be used for predictions C Hidalgo, B Klinger, AL Barabasi, and R Hausmann *The product space conditions the development of nations* Science 317(5837), 2007

Following the advice



Relation between recommendation, realized spreading and economic complexity 1995-2014

Conclusions

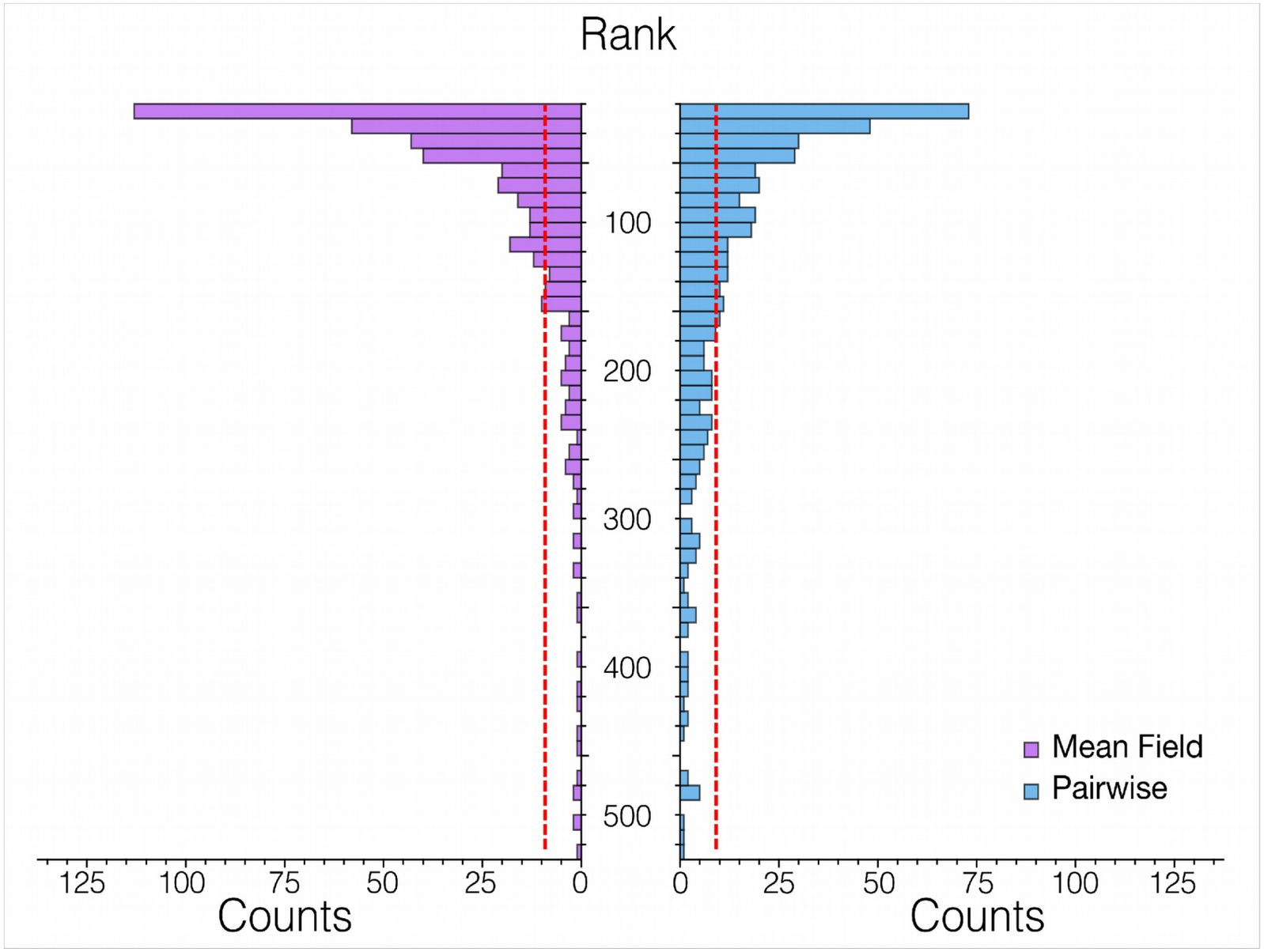
- Projection formula for **nested** bipartite networks
- **Directed filter** to extract information about **progression** from highly connected monopartite networks
- Application: trade data, network of **products**
- **Prediction** of new and already exported products
- **Recommendations**: a tool to lower the barrier to escape from the poverty trap

Thank you for your attention!

A Zaccaria, M Cristelli, A Tacchella, and L. Pietronero
How the Taxonomy of Products Drives the Economic Development of Countries.
PLoS ONE 9(12): e113770(2014)

E. Pugliese, GL Chiarotti, A Zaccaria, and L Pietronero *Complex economies have a lateral escape from the poverty trap,* arXiv:1511.08622v1 (2015)

Taxonomy → Largest assist



Our choice of links correspond to the largest elements of the specific row of the Assist matrix

Important test to verify link between static network properties and development of countries

The transpose of the adjacency matrix performs much worse: importance of the directed links wrt the symmetric network