istituto dei sistemi complessi

Consiglio Nazionale delle Ricerche

The Scientific and Technological Competitiveness of Nations: a network analysis

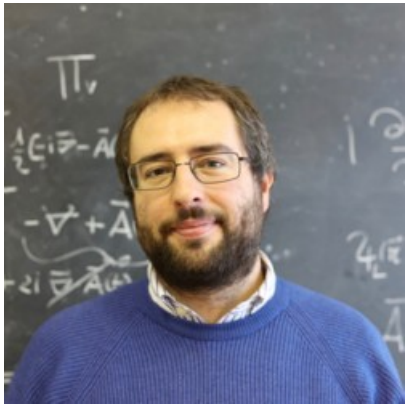
Andrea Gabrielli
Istituto dei Sistemi Complessi (ISC) – CNR, Rome, Italy
andrea.gabrielli@roma1.infn.it

CERN Knowledge Transfer

Collaborators

Giulio Cimini
(IMT – Lucca, Italy)

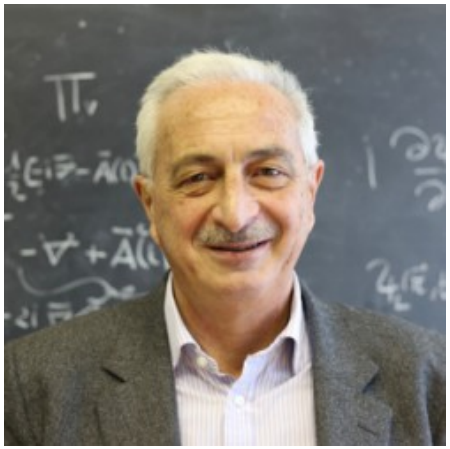
Francesco Sylos Labini
(Enrico Fermi Center, Rome, Italy)



Andrea Zaccaria
(ISC – CNR, Rome, Italy)

Aurelio Patelli
(CEA – Saclay, France)

Emanuele Pugliese
(JRC-IPTS, EC, Seville, Spain)



Luciano Pietronero
(Univ. Sapienza, Italy)

Summary

- In the international trade of products the nestedness of the export matrix is the main property
- Diversification of advanced countries and ubiquities of products are central features
- Their structure calls for a non-linear algorithm of classification of industrial competitiveness of countries and complexity of products
- Do science and technology behave similarly? Yes!
- What are the important parameters to predict development
- How scientific, technological and industrial development talk to each other?
- Multilayer space of innovation

Paul Romer (World Bank, Chief Economist)

Present standard Economy: “Church of Scientology” about approach and prediction

Equilibrium approach: Physics of XIX century unable to capture complexity

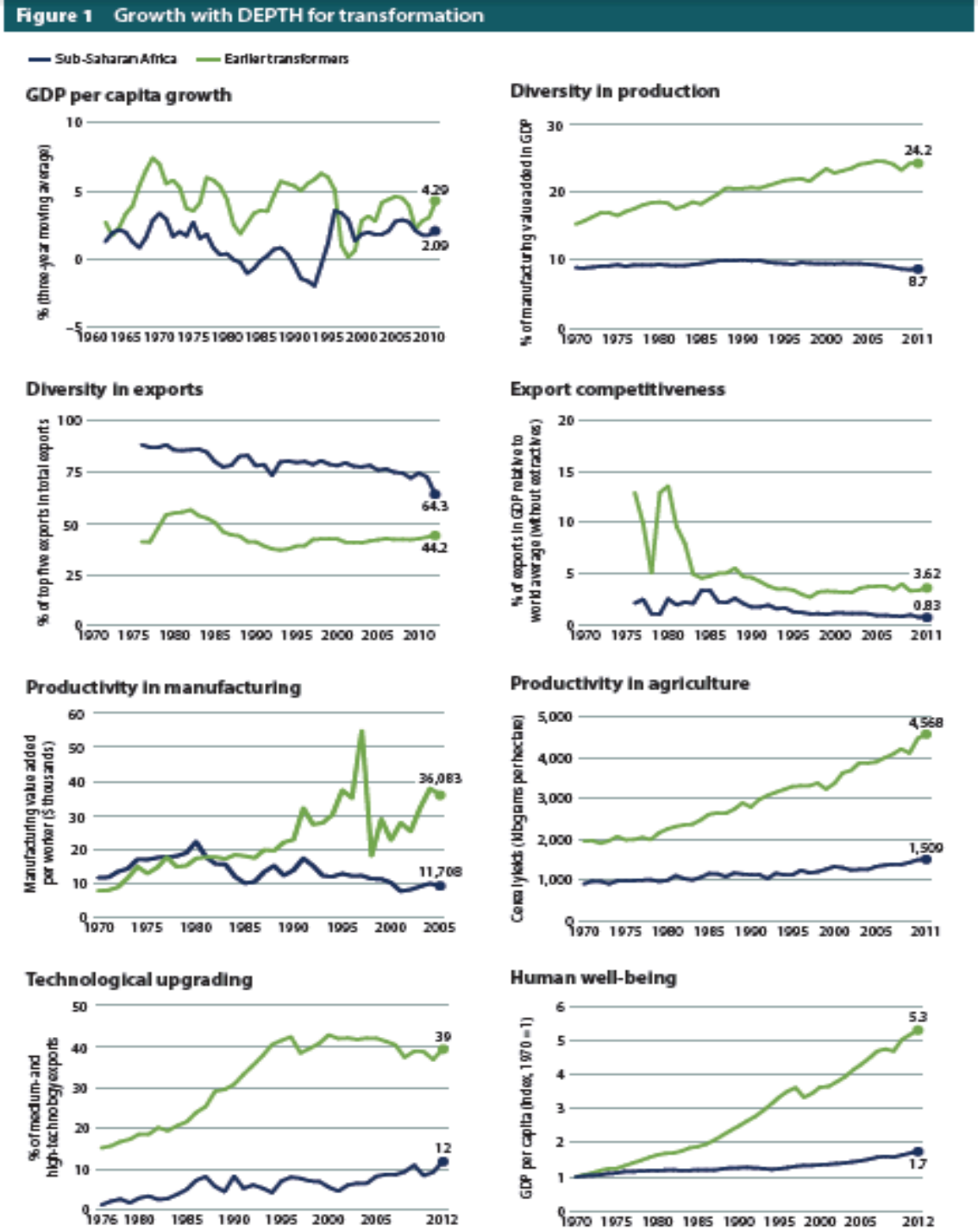
Fundamental concepts for a new economic theory: interactions, heterogeneity, scaling, self-organization, emerging properties → Complexity

USUAL APPROACH TO DEVELOPMENT AND GROWTH

- Aggregated data for the two groups of countries
- Interesting information but sometimes conflicting
- Difficult to get a unified comprehensive picture

More and more data but difficult to draw a clear conclusion: isgnal to noise ratio decreases

How to sum GDP with Technology?



DATABASES

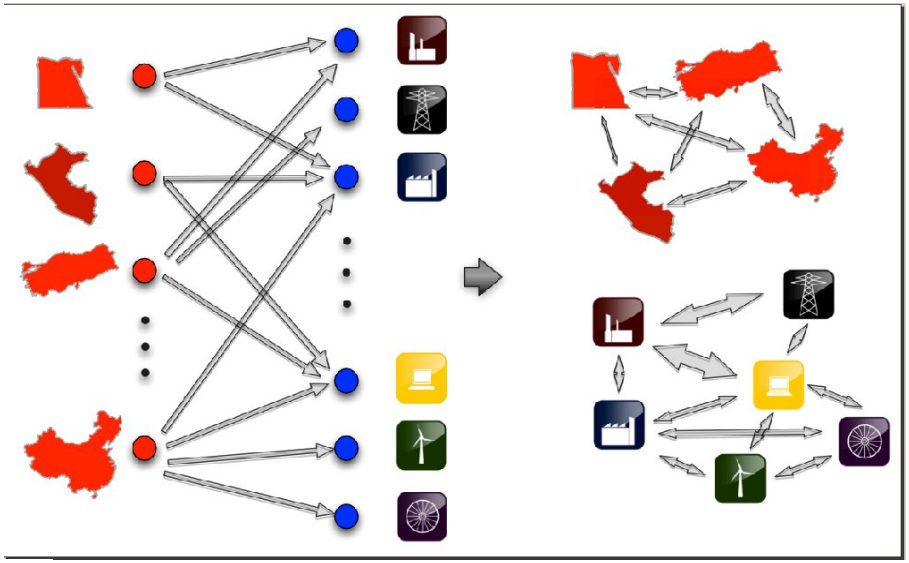
- Industrial Competitiveness: **UN COMTRADE** A database hosted by the United Nation with detailed data on countries and products export at different levels of aggregation, the most detailed being at 6 digits (6000 products). The complete data set is released to the public by UN once per year, with one year delay.
- Technologies: **EPO PATSTAT** A database organized by the European Patent Office, aggregating data from all the patent offices in the World (over 100 different offices). Patents are classified with respect to their technological field at various aggregation levels, the most detailed classification being sub-groups (70000 codes).
- Science: **SciVal** The SciVal platform aggregates data about scientific publications from Elsevier Scopus, which covers journals, trade publications, book series, conference proceedings, and books. Data cover years from 1996 to 2015, and each scientific publication is assigned to a category at two levels of aggregations: scientific sectors (28 codes) and subsectors (about 300). These data also provide a direct linkage between scientific publications and patents according to direct citations.

Bipartite country-product network

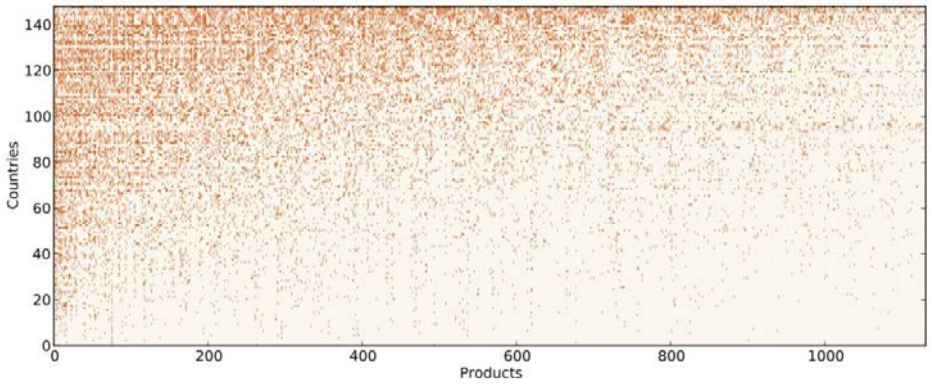
COMTRADE data – World Trade Web

$$RCA_{cp} = \frac{X_{cp}}{\sum_{p'} X_{cp'}} \Big/ \frac{\sum_{c'} X_{c'p}}{\sum_{c',p'} X_{c'p'}}$$

$$M_{cp} = \begin{cases} 1 & if \quad RCA_{cp} > R^* \\ 0 & if \quad RCA_{cp} < R^* \end{cases}$$



2010



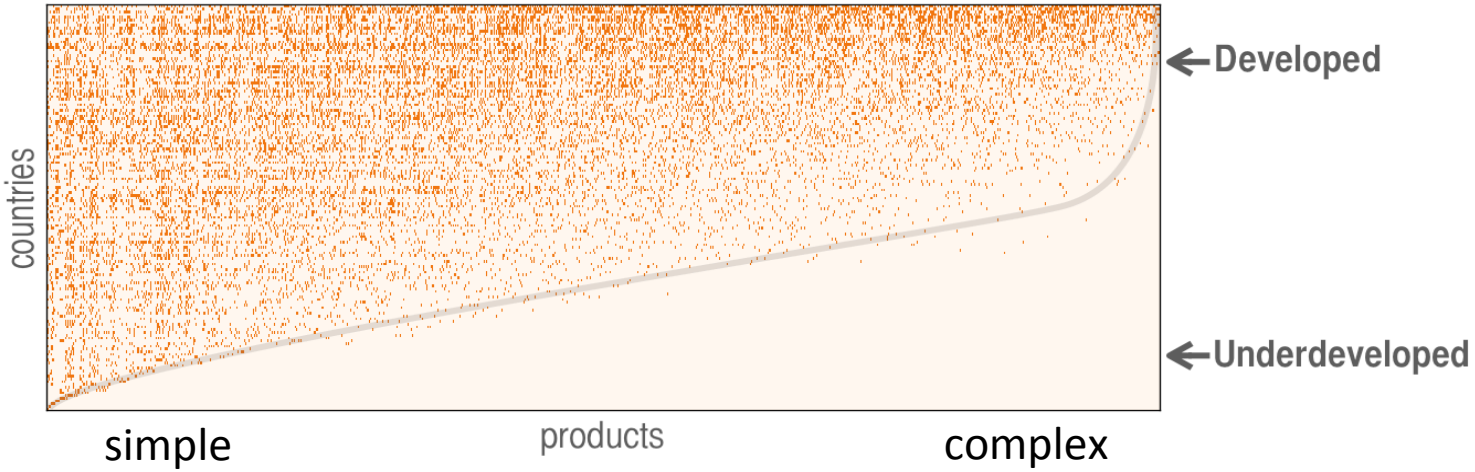
$$k_c = \sum_{p=1}^{N_p} M_{cp} = \text{diversification of country } c$$

$$k_p = \sum_{c=1}^{N_c} M_{cp} = \text{ubiquity of product } p$$

Triangularity of the export *M* matrix

PNAS **106**, 10570 (2009)

Extremely nested country – product matrix: ecosystem-like organization



Warning:
Specialization
↓
Block diagonal

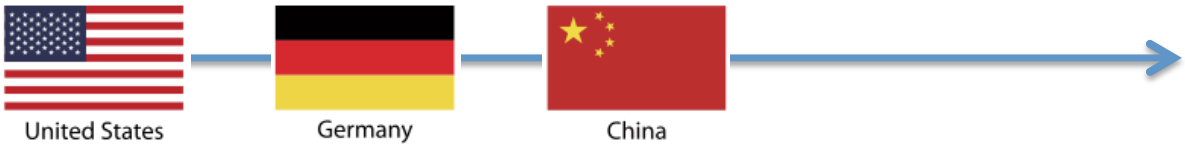
New paradigm:
diversification

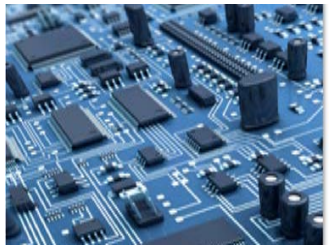
Scientific Reports **2**, 723 (2012) | PLoS ONE **8**, e7072 (2013) | Scientific Reports **6**, 30286 (2016).

PLoS ONE **10**(12), e0144564 (2015) | Scientific Reports **5**, 10595 (2015) | J. of Econ. Dyn. And Contr. **37**, 1683 (2013)

Measuring Intangible Properties

New metrics for Fitness of countries and Complexity of products

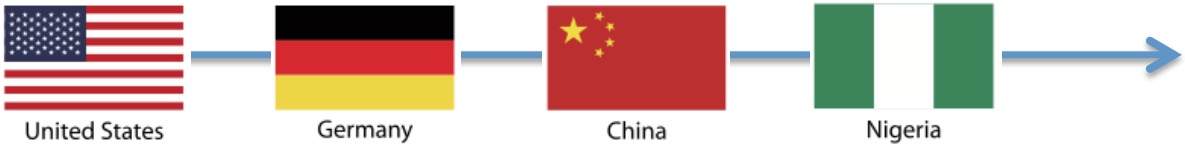




1.81

F: 5.4 6.2 5.3 0.01

Q



0.0099

Fitness:

$$\tilde{F}_c^{(n)} = \sum_p M_{cp} Q_p^{(n-1)}$$

$$F_c^{(n)} = \frac{\tilde{F}_c^{(n)}}{\langle \tilde{F}_c^{(n)} \rangle_c}$$

Complexity:

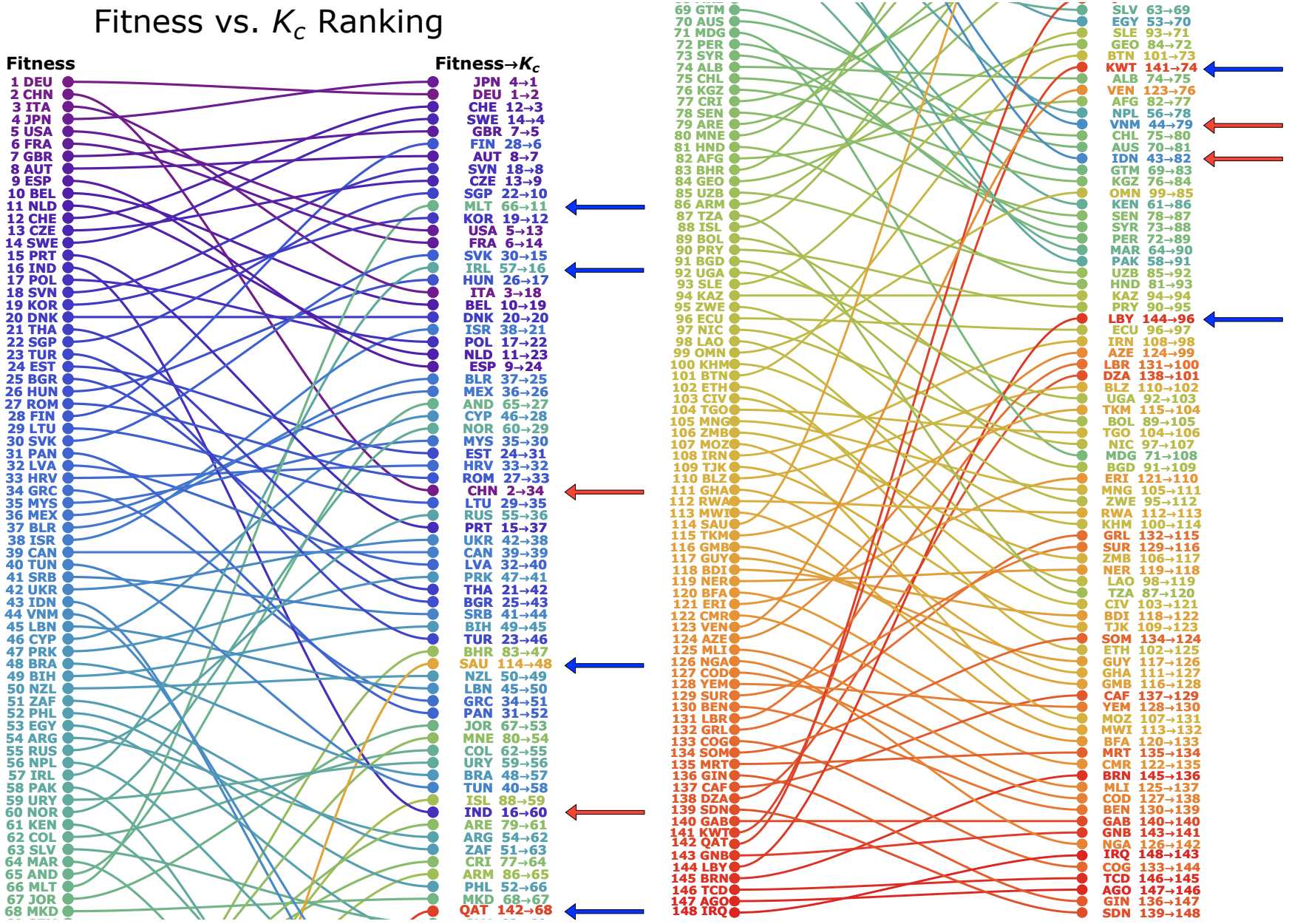
$$\tilde{Q}_p^{(n)} = \frac{1}{\sum_c M_{cp} \frac{1}{F_c^{(n-1)}}}$$

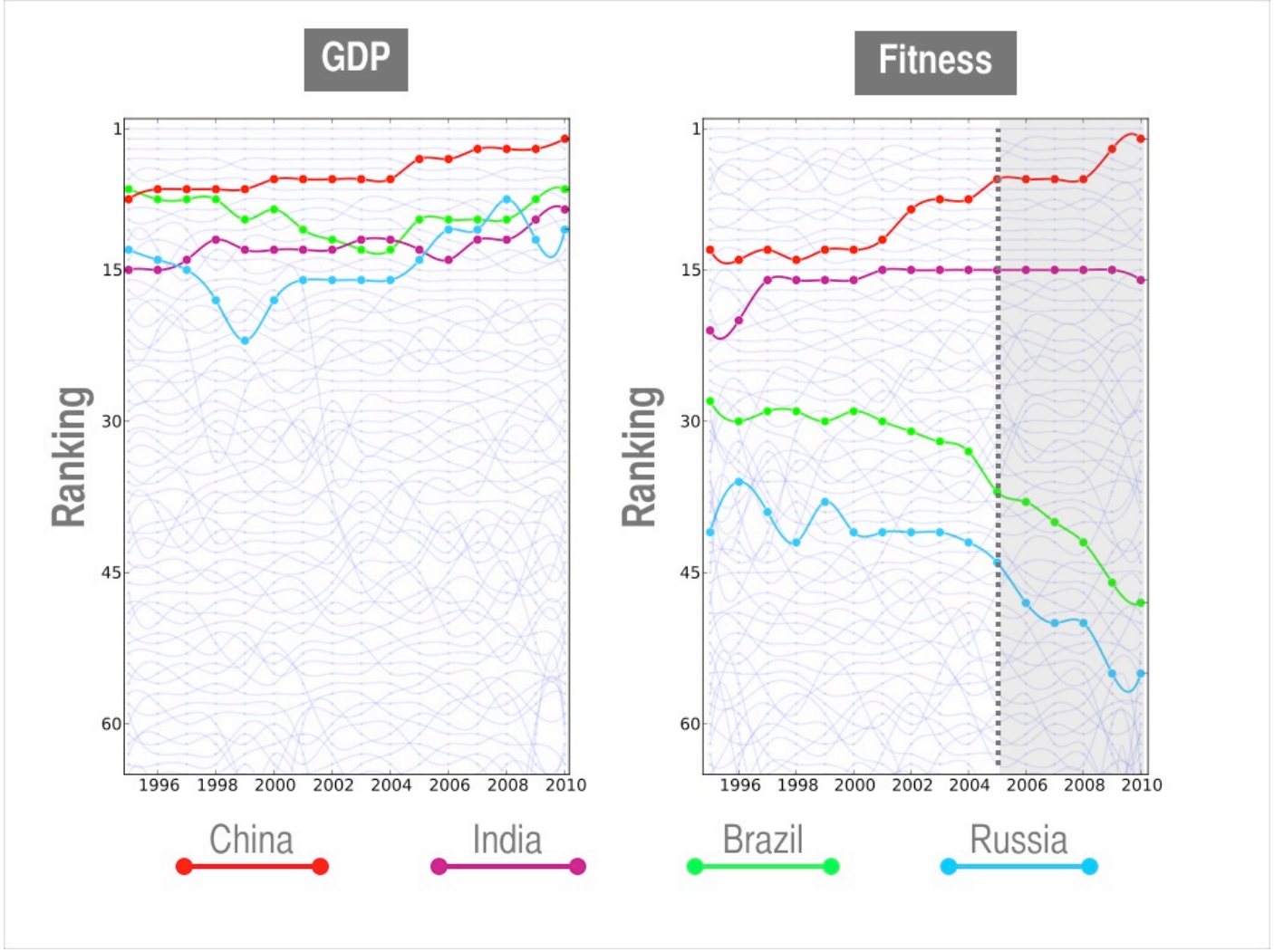
$$Q_p^{(n)} = \frac{\tilde{Q}_p^{(n)}}{\langle \tilde{Q}_p^{(n)} \rangle_p}$$

F_c: diversification weighted by complexity

Q_p: Almost extremal dependence on F_c's.
A **single low fitness producer** implies low complexity

COMPARISON OF THE RANKINGS

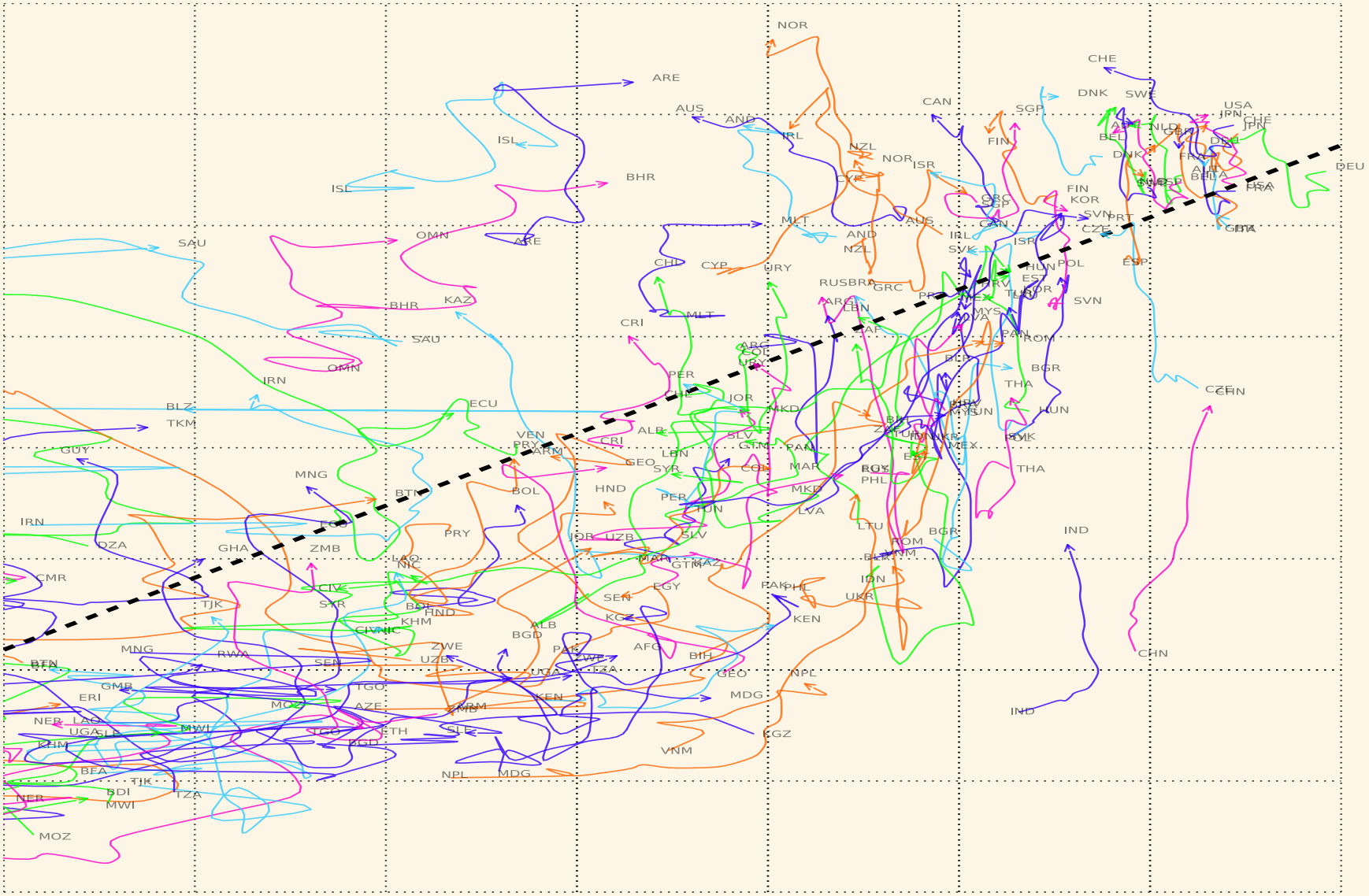




Goldman Sachs 2001:
BRIC countries will dominate
World economy for the next 50 years

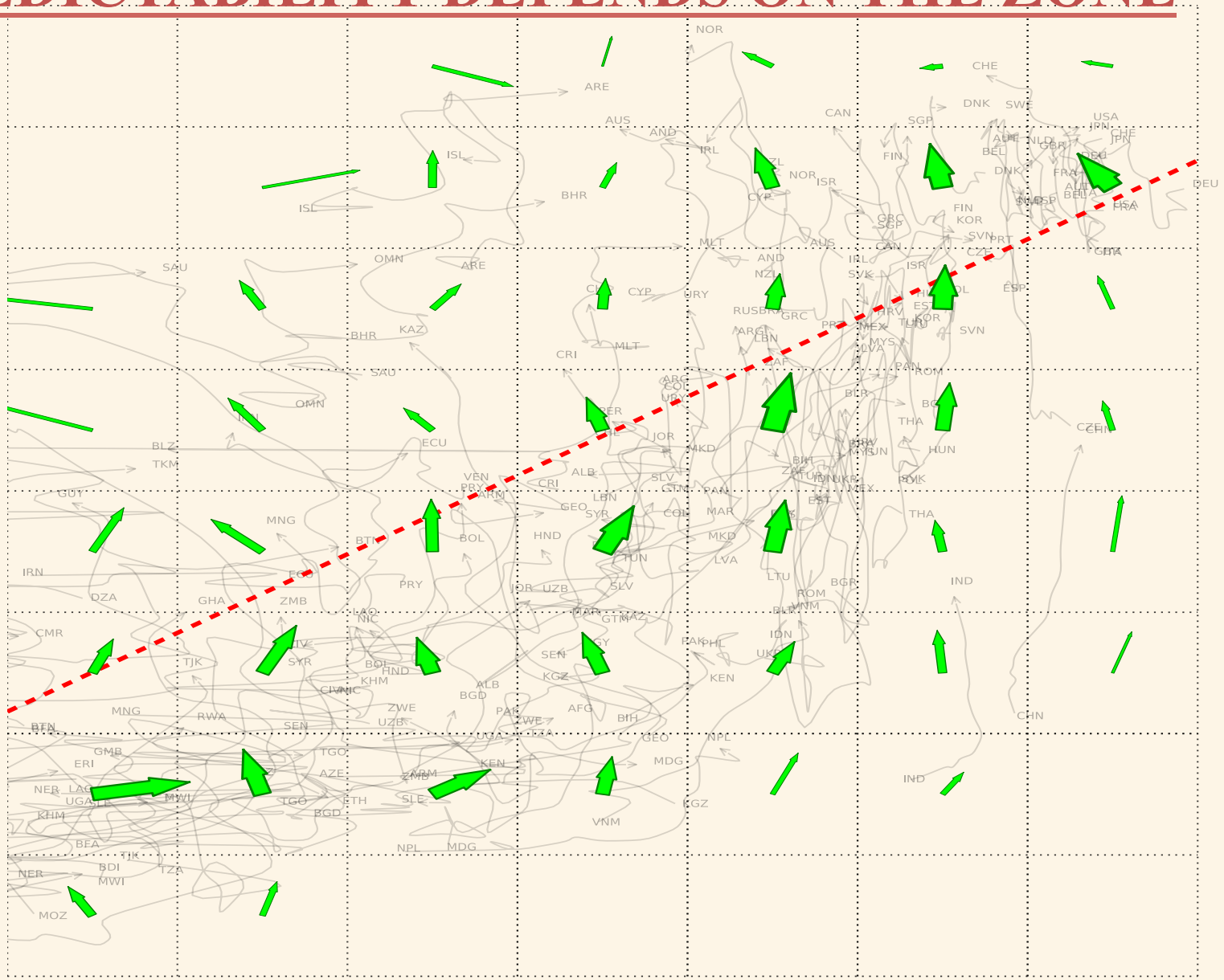
Bloomberg News November 2015
Goldman's BRIC era ends as fund
Defaulted after years of losses

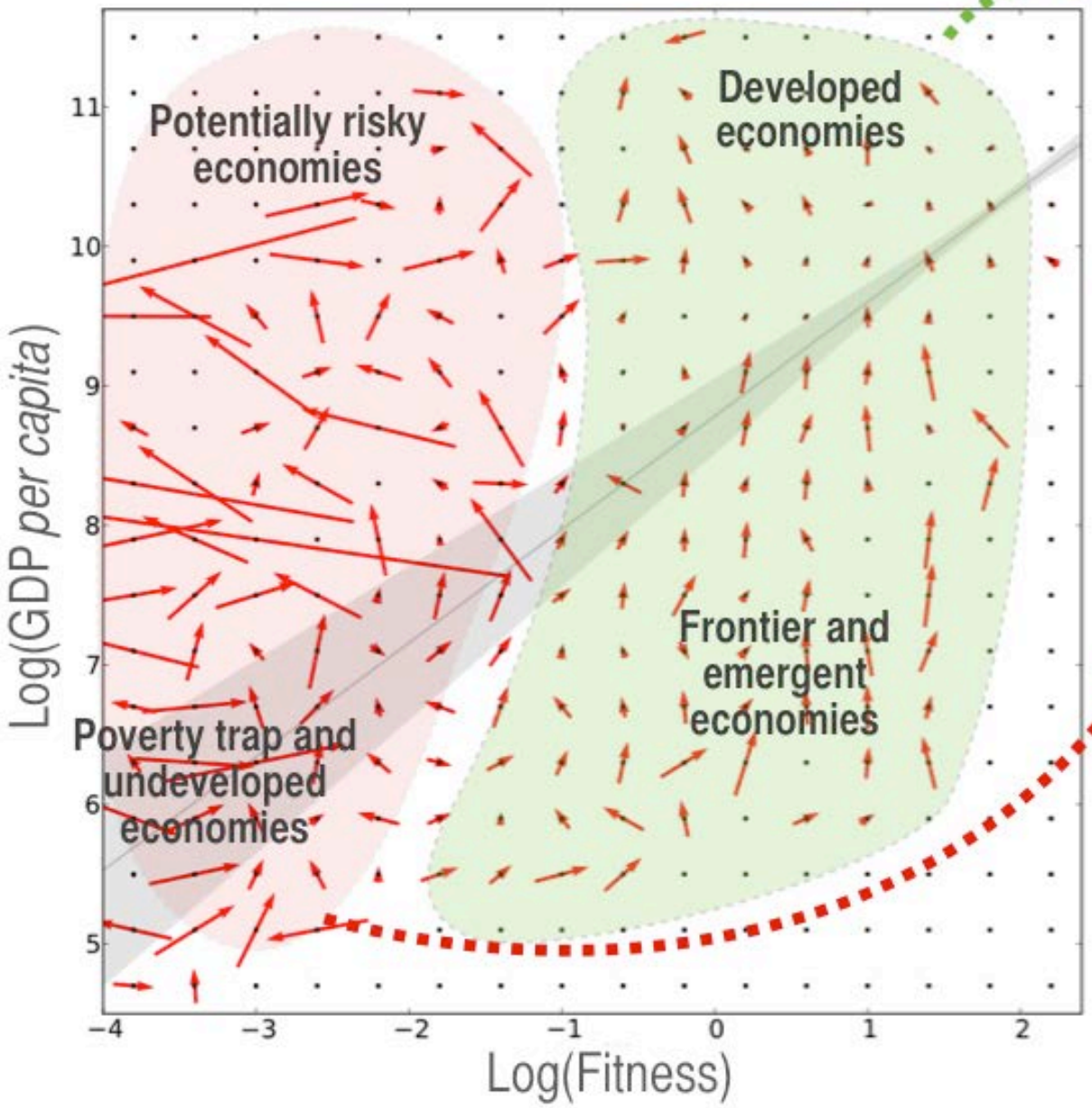
ECONOMIC DYNAMICS IS HETEROGENEOUS



PLoS ONE **10**(2): e0117174 (2015)

COARSE GRAINED DYNAMICS: PREDICTABILITY DEPENDS ON THE ZONE





Laminar regime

Fitness is the relevant and driving variable for the economic dynamics in this regime

high predictability

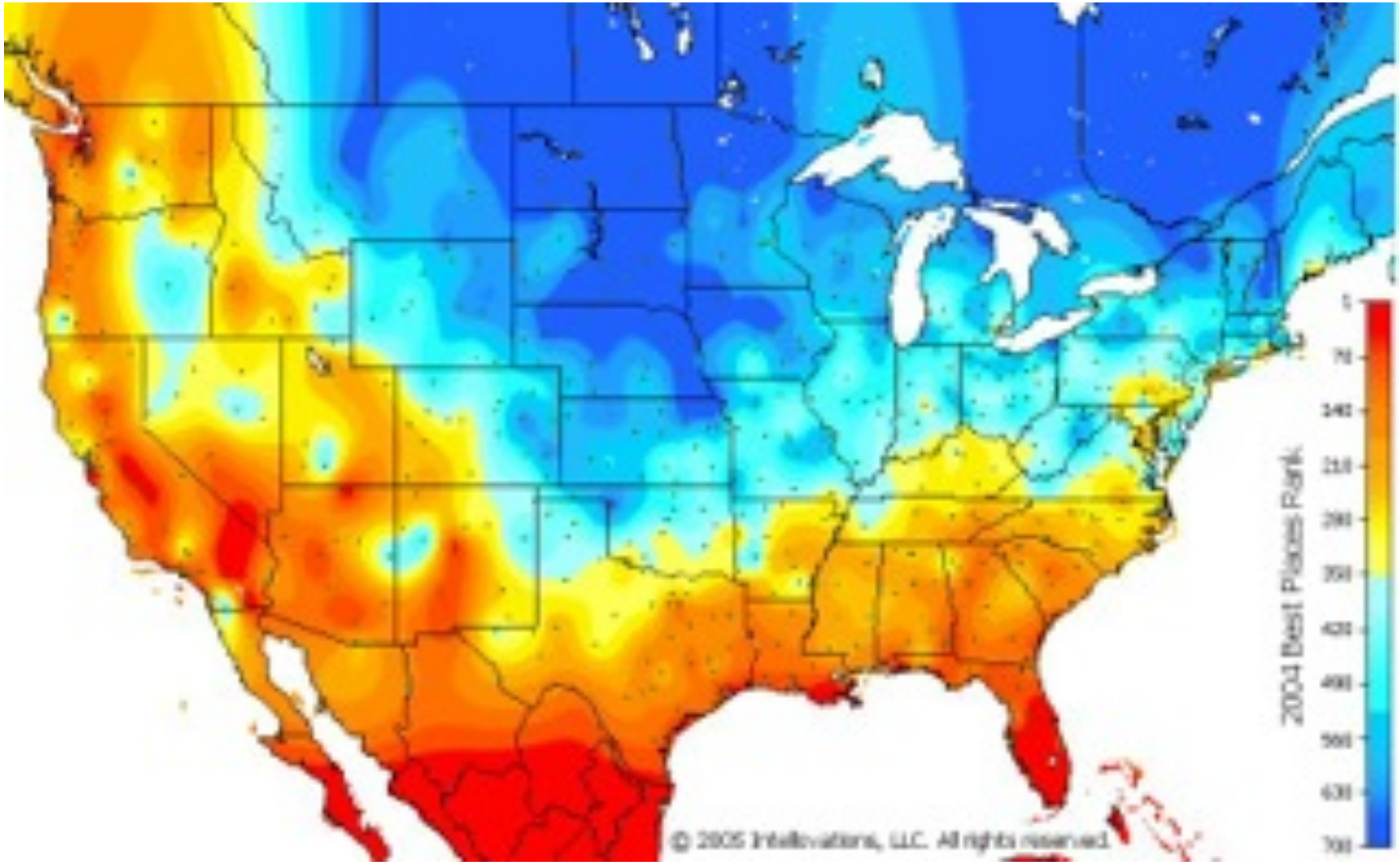
Chaotic regime

Dynamics is ruled by several other exogenous factors competing with Fitness

low predictability

Hetreogeneous Weather Forecasting:

RED: High predictability BLUE: Low predictability



Impact in the Scientific Debate

nature
International weekly journal of science

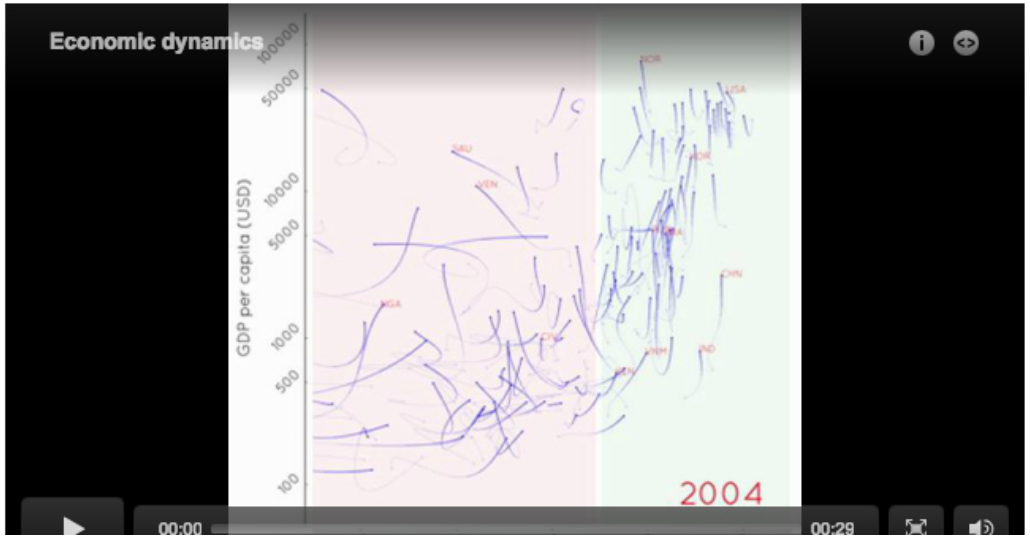
NATURE | NEWS

Physicists make 'weather forecasts' for economies

The development of some countries is as predictable as steady winds, but for others it is more chaotic, physicists find.

Richard Van Noorden

23 February 2015



What about scientific development? Data sets

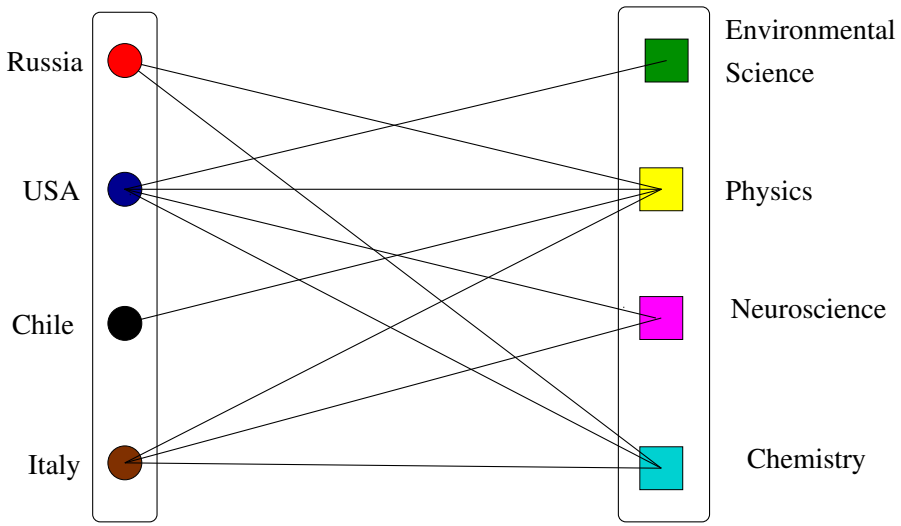
- Citation data over years 1996-2012 (source: Scimago website, based on Scopus): **N**=238 nations, **D**=27 scientific domains and **d**=307 sub-domains

- Higher Education expenditure on Research & Development (HERD, source: OECD): good index for national effort in research

Fundamental object and tool

In analogy with Economic Complexity[1] we build the binary bipartite network of nations and scientific domains

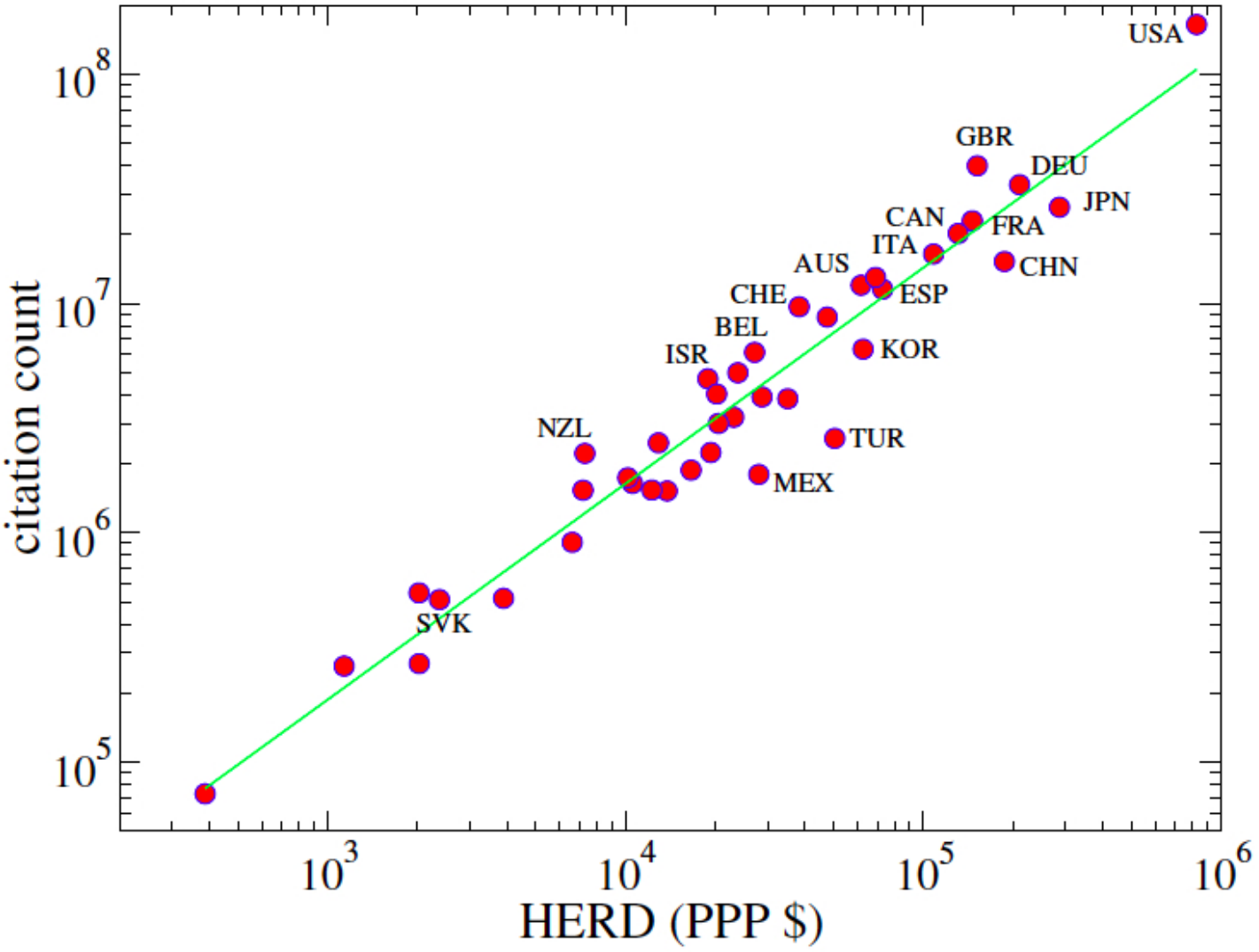
Example



[1]Cristelli M, Gabrielli A, Tacchella A, Caldarelli G, Pietronero L (2013). PLoS ONE **8**(8): e70726.

Tacchella A, Cristelli M, Caldarelli G, Gabrielli A, Pietronero L (2012). Scientific Reports **2**, 723.

Which measure of scientific production has to be used? Aggregated citations are a good indicator: e.g. string correlation with HERD (from OECD data) averaged over years



The bipartite network is described by the bi-adjacency matrix M_{cs} which says if a country *c* is "productive" ($M_{cs} = 1$) in scientific domain *s* or not ($M_{cs} = 0$)

$M_{cs} = 1$ → link, $M_{cs} = 0$ → no link

Which criterion to use to decide productiveness in a scientific sector?

Two possible choices

1. Extensive: M_{cs}= 1 if nation *c* ranks in the top-**T** for number of citations in the scientific domain *s*

2. Intensive: M_{cs} = 1 if nation *c* ranks in the top-**T** for number of citations in the scientific domain *s* per unit spent on HERD

We used different choices of T=10, 20 (weak dependence)
Results refer to T=10 (we can use also percentage)

Extensive criterion

It measures the absolute competitiveness of countries in the different sectors:

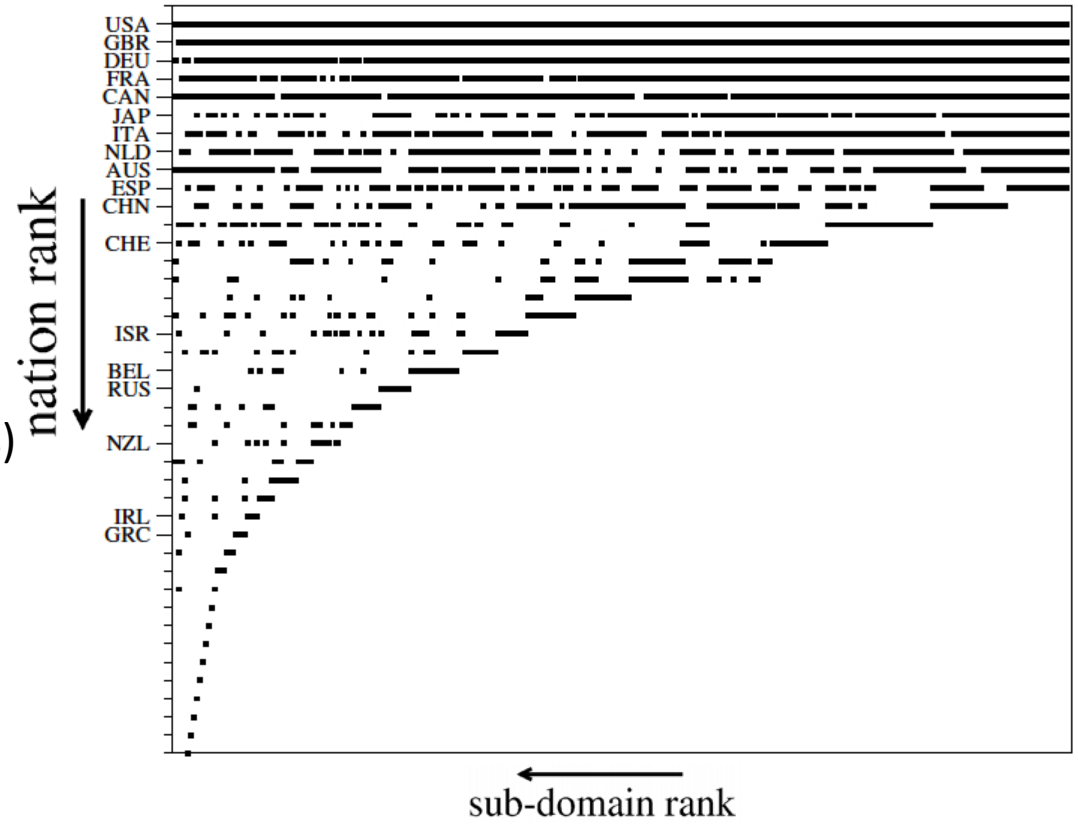
Size is important!

Successful nations do not specialize

Diversification appears as the main feature

Countries ranked in decreasing fitness (measure of competitiveness)

Domains ranked in increasing complexity index
(See below)



Intensive criterion

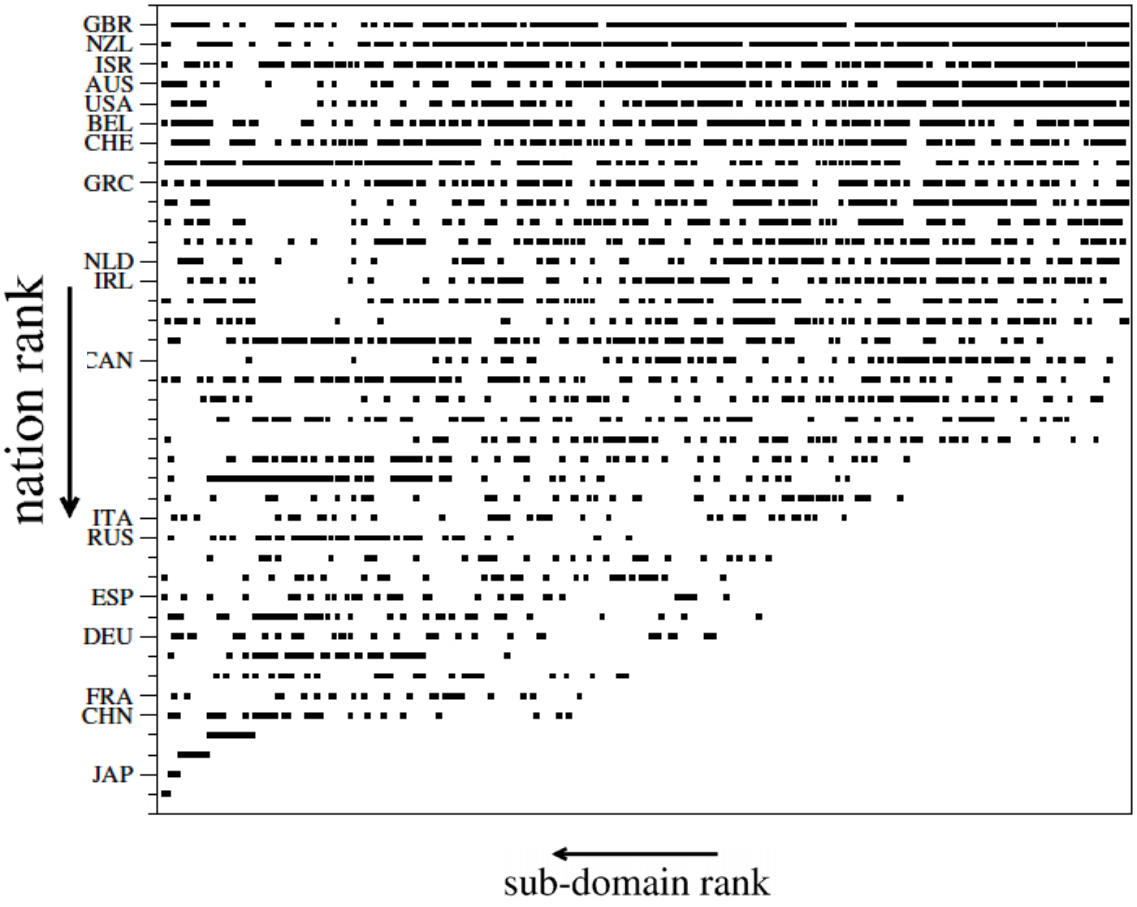
It measures the national efficiency of the expenditures in High Education and Research

Size is less important! Efficiency matters!

Again triangular shape

Diversification is the main feature of scientific development

Countries ranking changes towards more efficient countries



A strong analogy can be done for the scientific production of nations and the complexity of scientific domains

What is the simplest algorithm to link countries fitness and complexity of scientific sector?

Non-linear iterative algorithm evolved up to the fixed point

Fitness:

$$\tilde{F}_c^{(n)} = \sum_s M_{cs} Q_s^{(n-1)}$$

$$F_c^{(n)} = \frac{\tilde{F}_c^{(n)}}{\left\langle \tilde{F}_{c'}^{(n)} \right\rangle_{c'}}$$

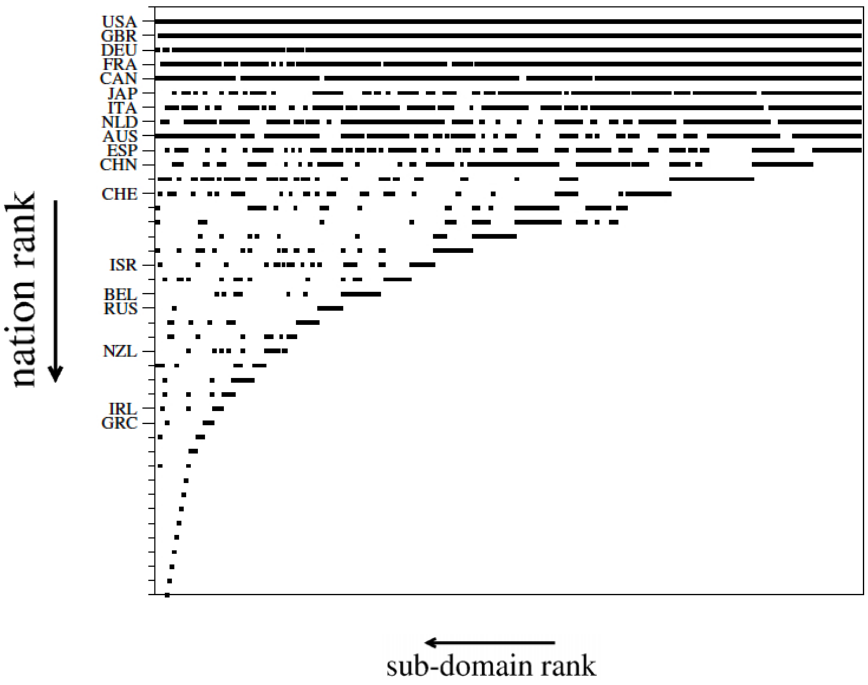
Complexity:

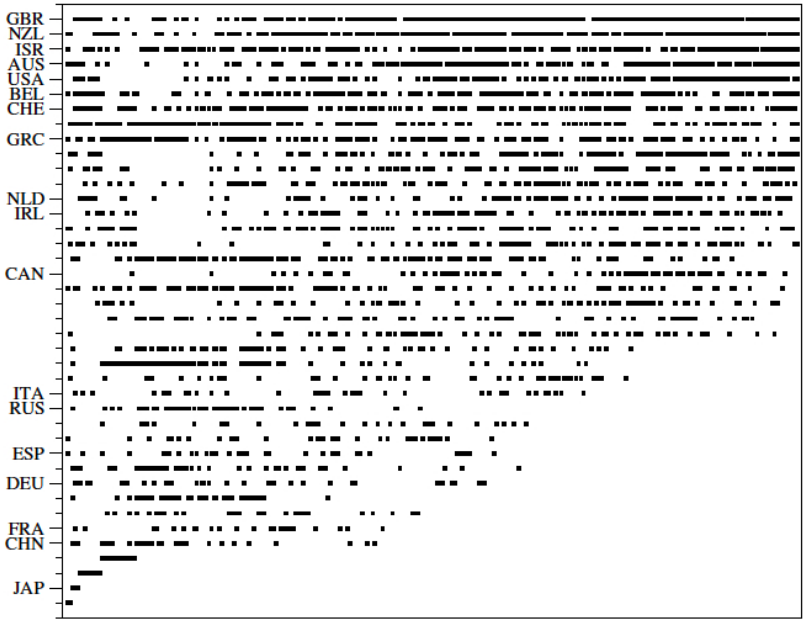
$$\tilde{Q}_s^{(n)} = \frac{1}{\sum_c M_{cs} \frac{1}{F_c^{(n-1)}}}$$

$$Q_s^{(n)} = \frac{\tilde{Q}_s^{(n)}}{\left\langle \tilde{Q}_{s'}^{(n)} \right\rangle_{s'}}$$

- Fitness of a nation measured by summing the complexities of the domains belonging to its research pool
- A domain on which scarcely competitive nations make research achieves low complexity
- To achieve high complexity, a domain must be part of the research system of only highly competitive nations

Scientific international ranking





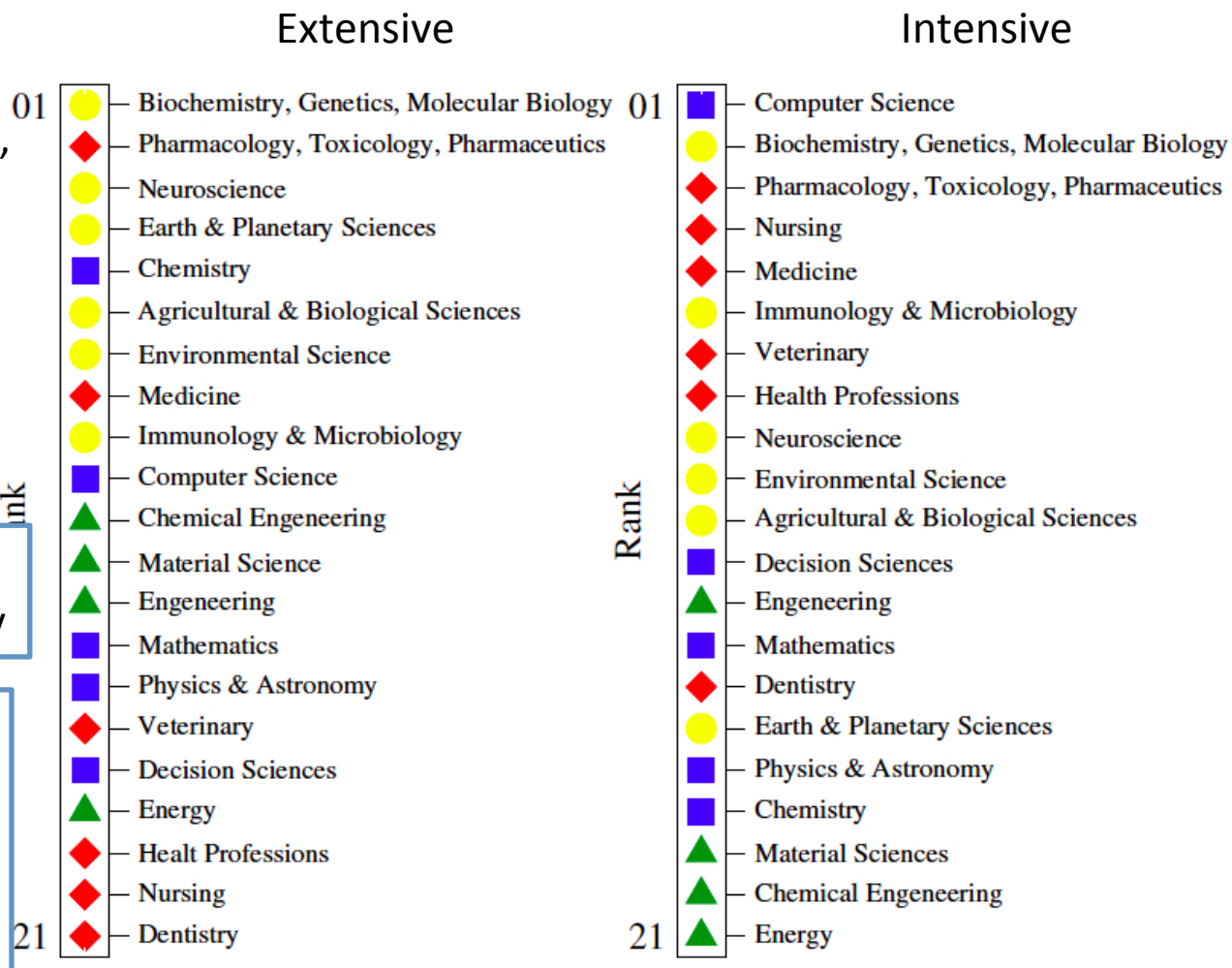
What is the most complex scientific domain/area? (Let's bet!)

Complex = "sophisticated"

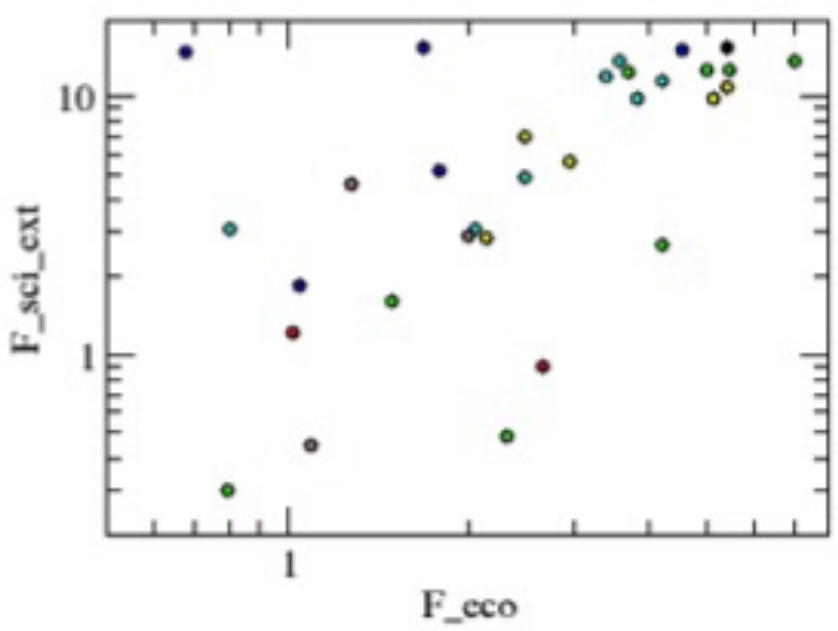
The development of complex sciences needs a development of simpler ones

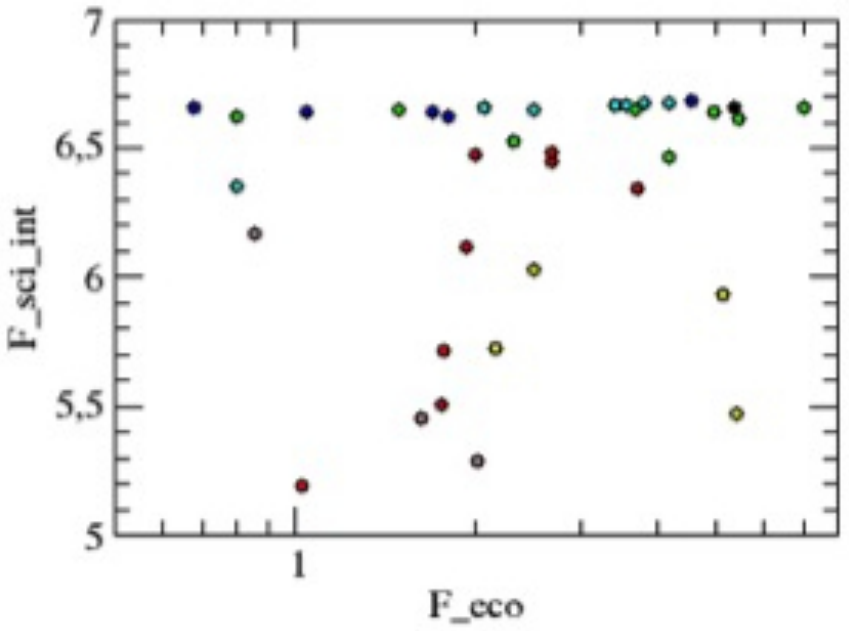
E.g. Neuroscience needs Physics, Medicine, Biology

Nursing becomes a real science only if medicine is highly developed, and is highly ranked as they belong to efficient countries



Comparison Scientific and Economic fitnesses

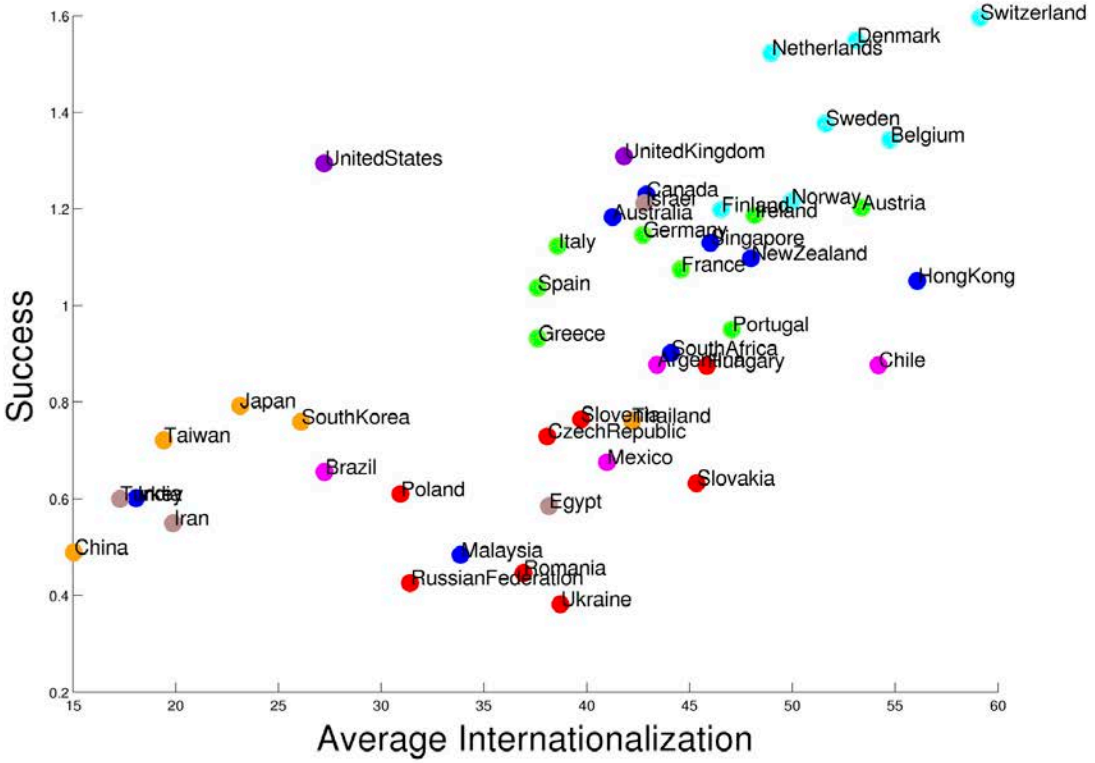




Correlation is present, but not so strong: the set of countries considered for science is much smaller than for export. North and west Europe countries are more diverse in export than science: EC grant policies homogeneisation effects?

Does Internazionalization index correlate with success parameter?

G. Cimini, A. Zaccaria, A. Gabrielli, J. of Informetrics **10**, 200 (2016)



$$S = \frac{\%\ \text{citations}}{\%\ \text{papers}}$$

Not only correlation, but

Automatic selection of geographical homogeneous regions

Basic information

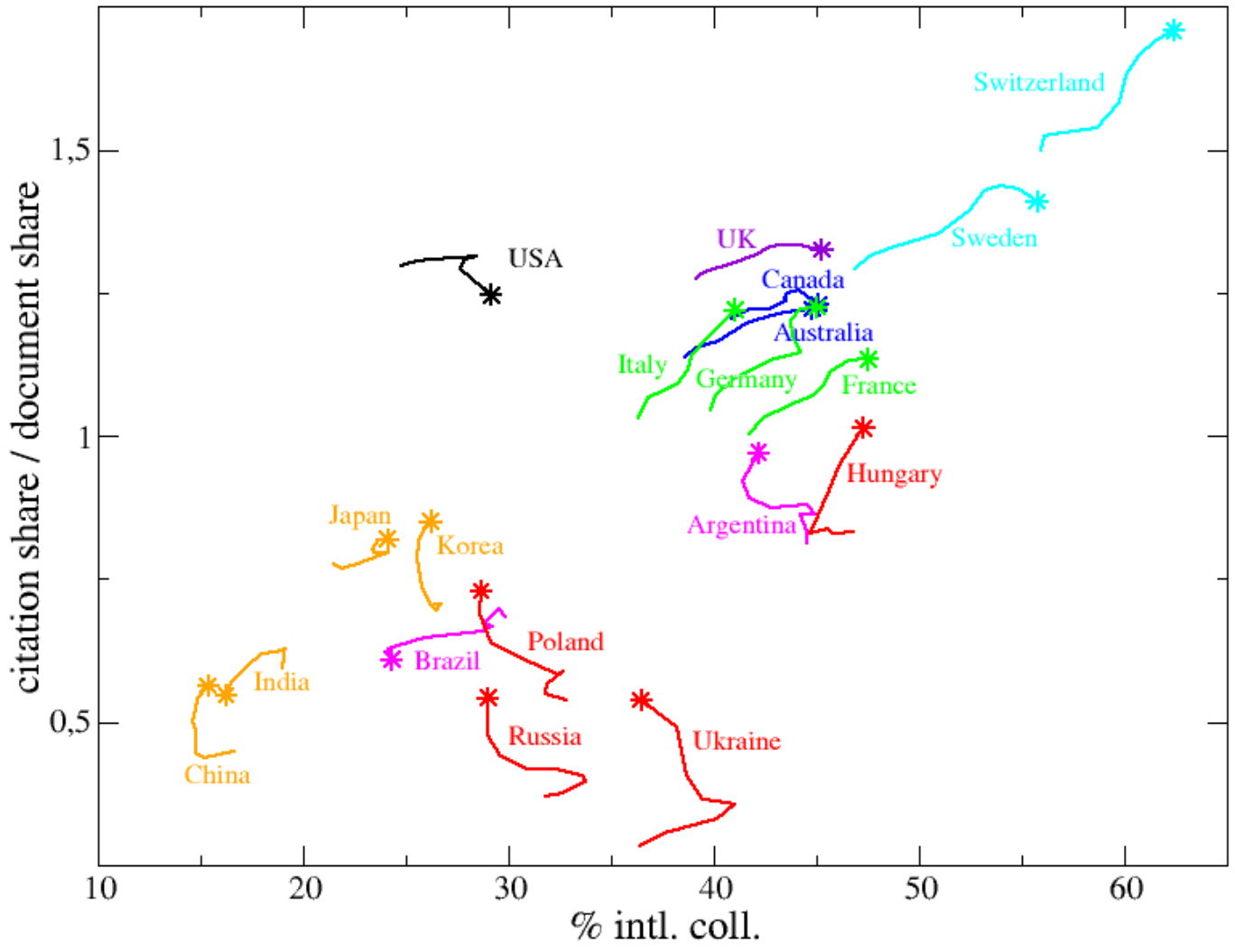
- Internationalization is stronger in Western Europe: EU integration research projects
- US evolves similarly but with a negative bias: larger self-consistency
- Is Eastern Europe yet in the “Russian” attractor?
- Asian regions are not increasing internationalization: different reasons

Total number of citations is correlated with total HERD

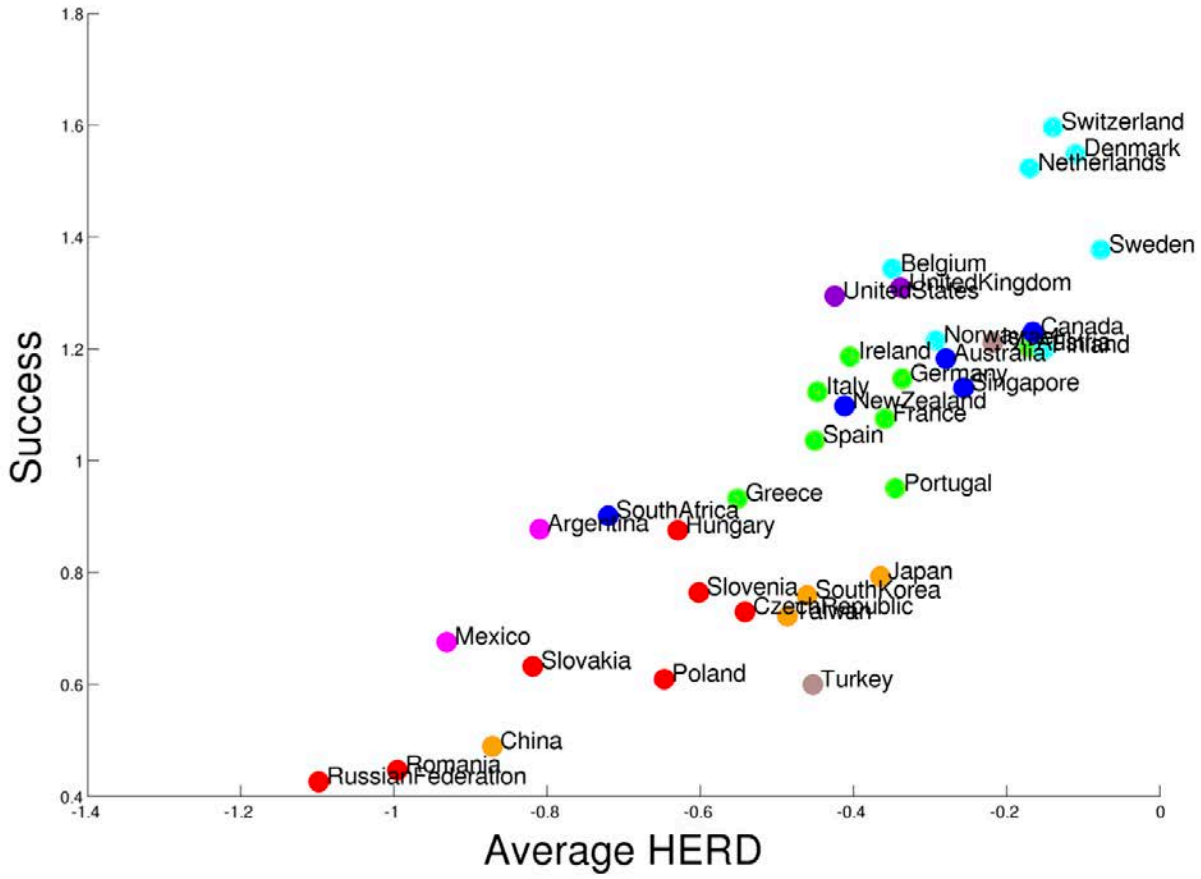
But they are extensive quantities depending on the country size

Can we define an intensive parameter for scientific impact?

How this situation evolves in time?

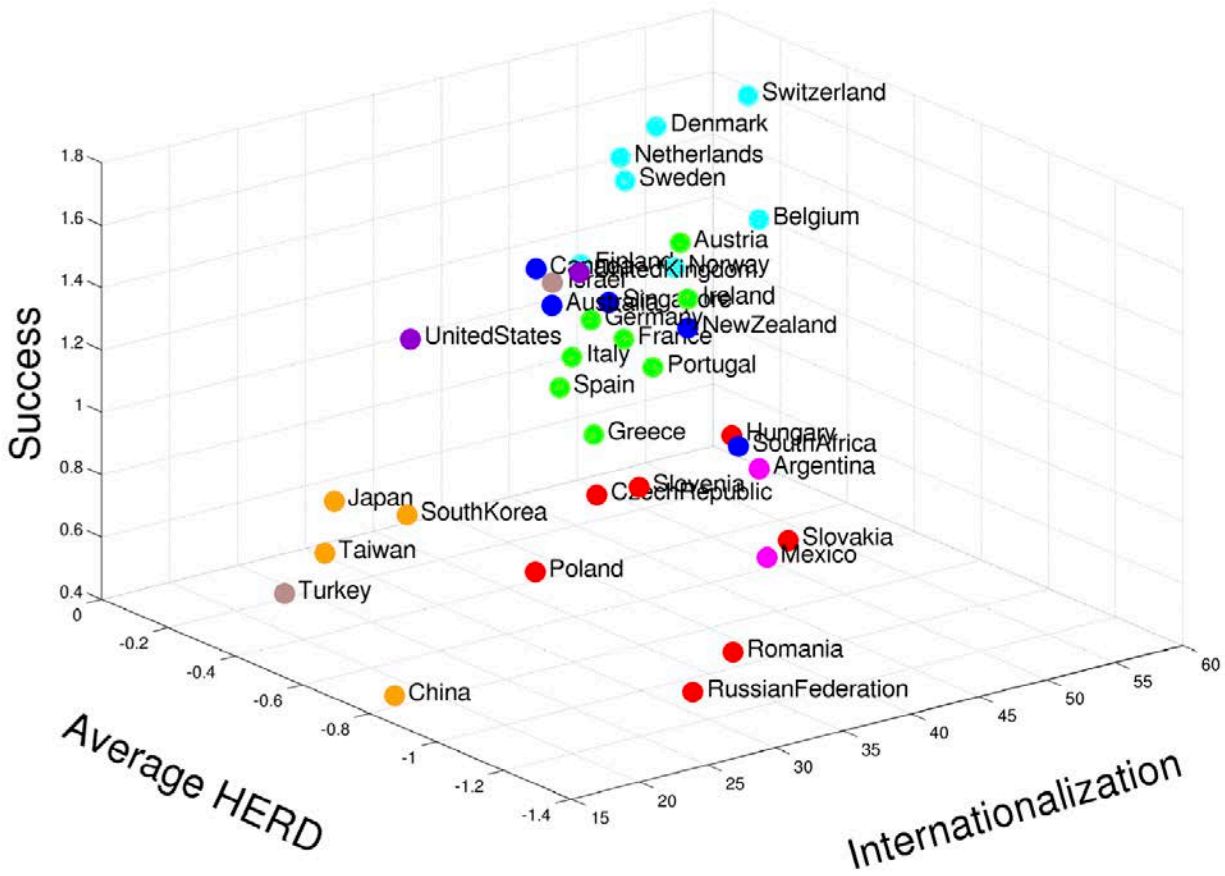


Does HERD/GDP correlate with S parameter?

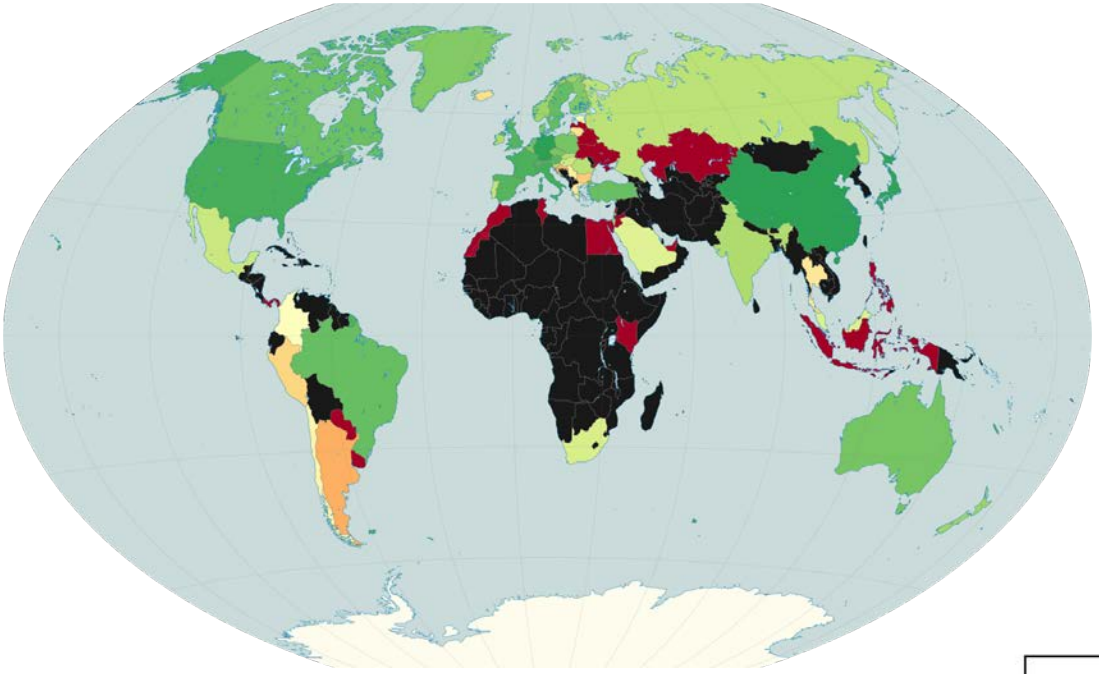


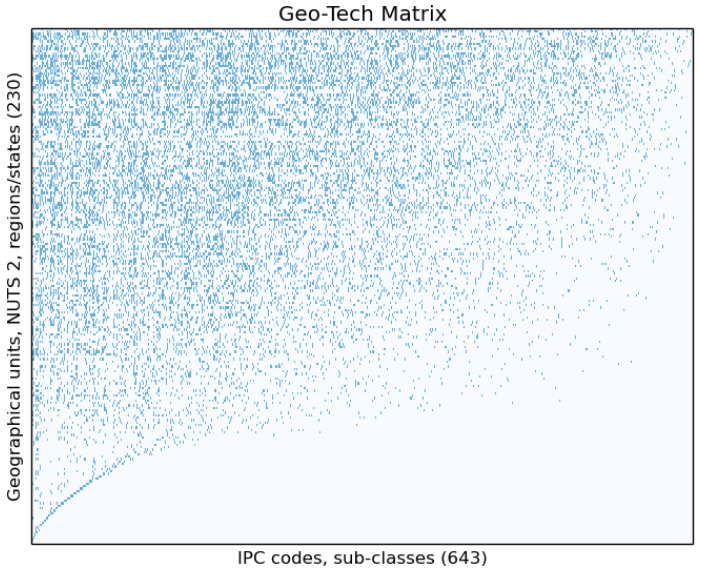
Similar to Internationalization with a shift of Asian countries

Feature selection approach

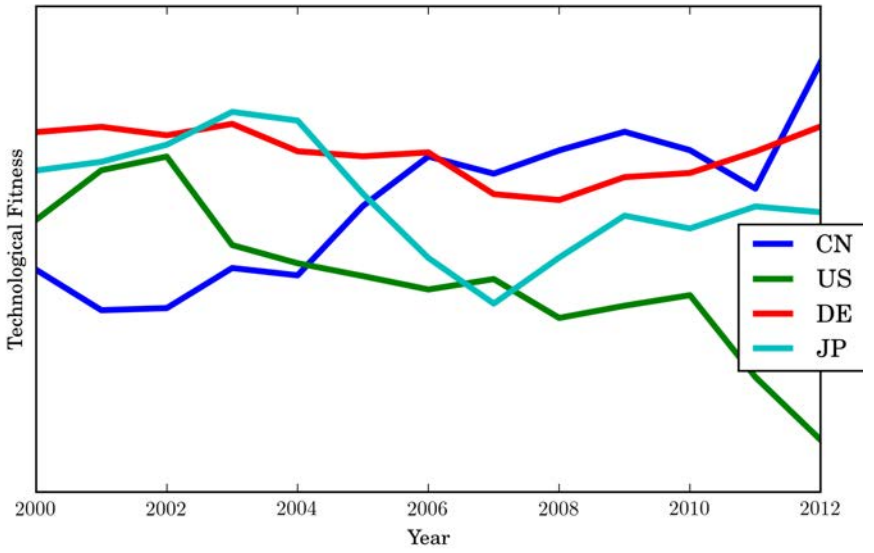


Technological fitness and complexity of technologies

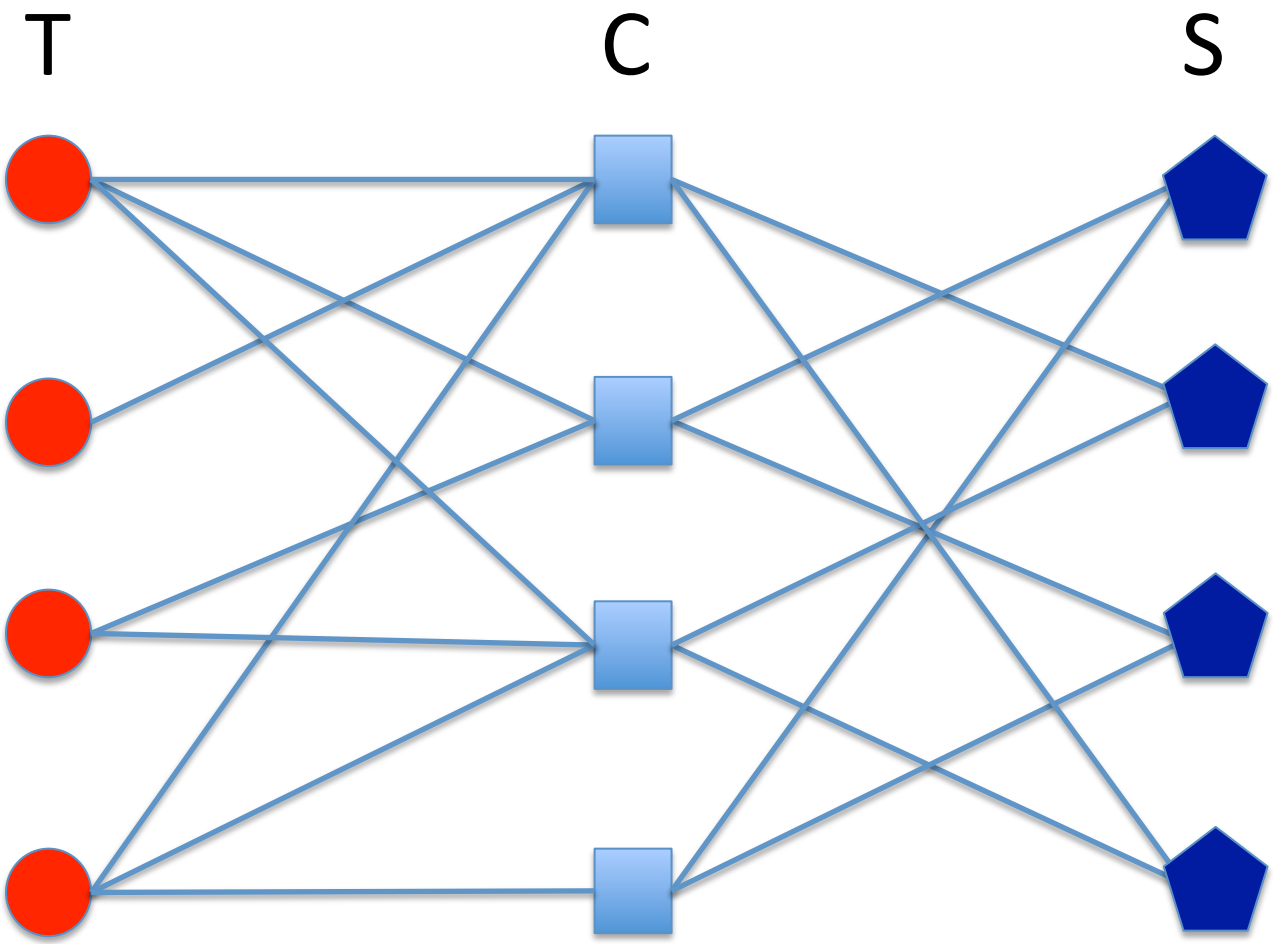




An “Economic Complexity” approach can be developed also for technologies and countries using data form patents database as PATSTAT

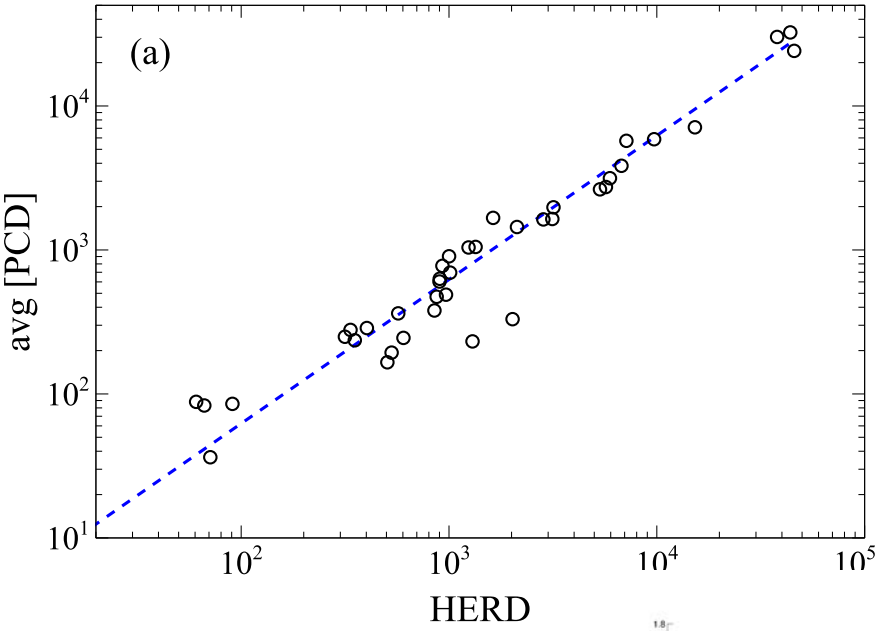


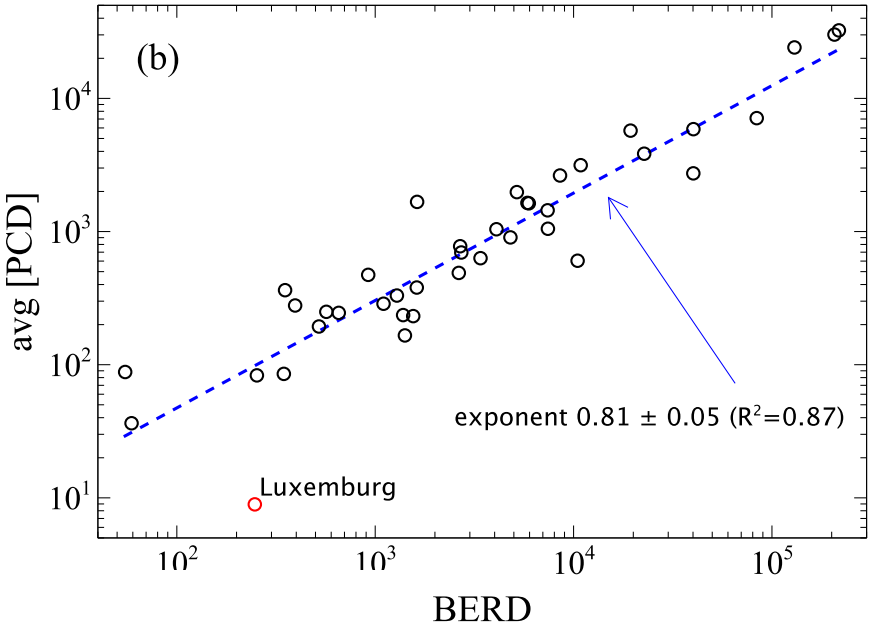
Multilayer space of development: science vs technology
Slow information flux from science to technology



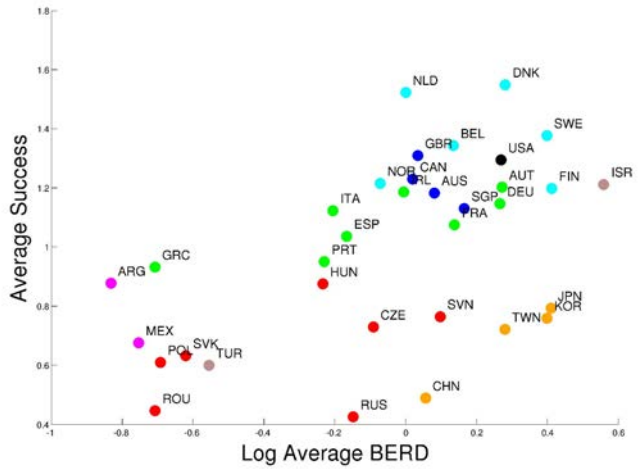
How science reflects in technology?

Scival → Citations of scientific documents in patents

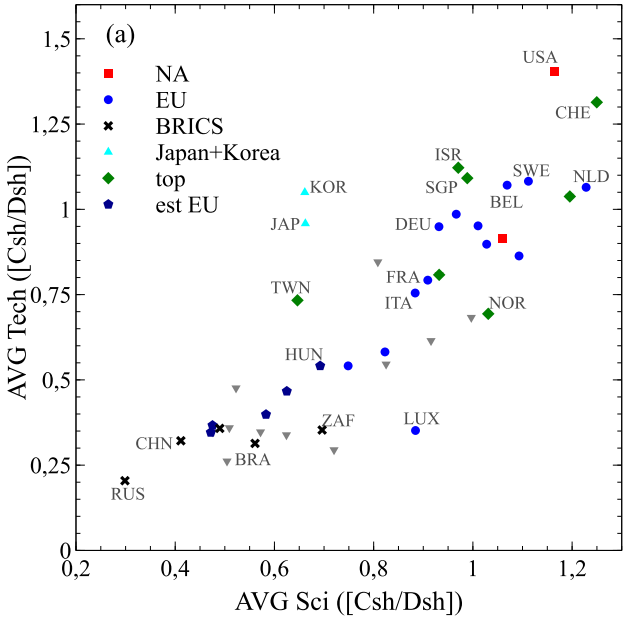


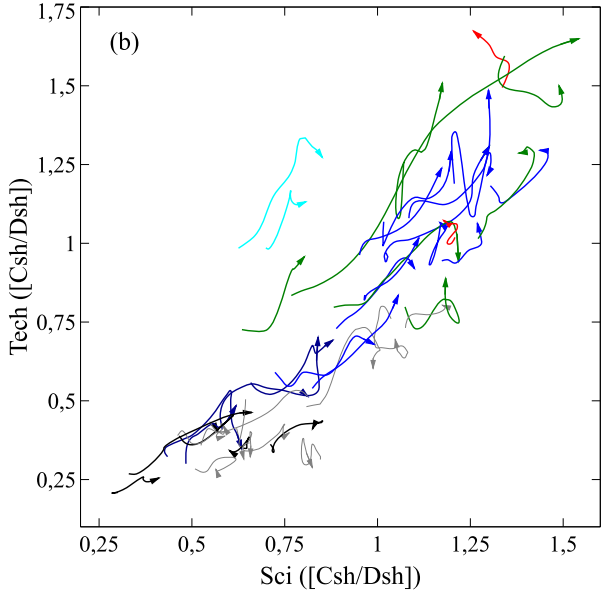


While …



A. Patelli, G. Cimini, E. Pugliese, A. Gabrielli, J. of Informetrics **11**, 1229 (2017)

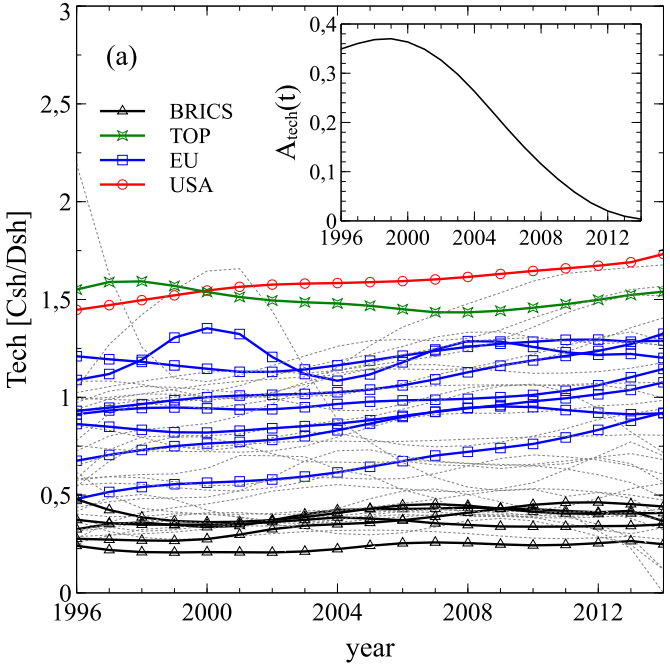




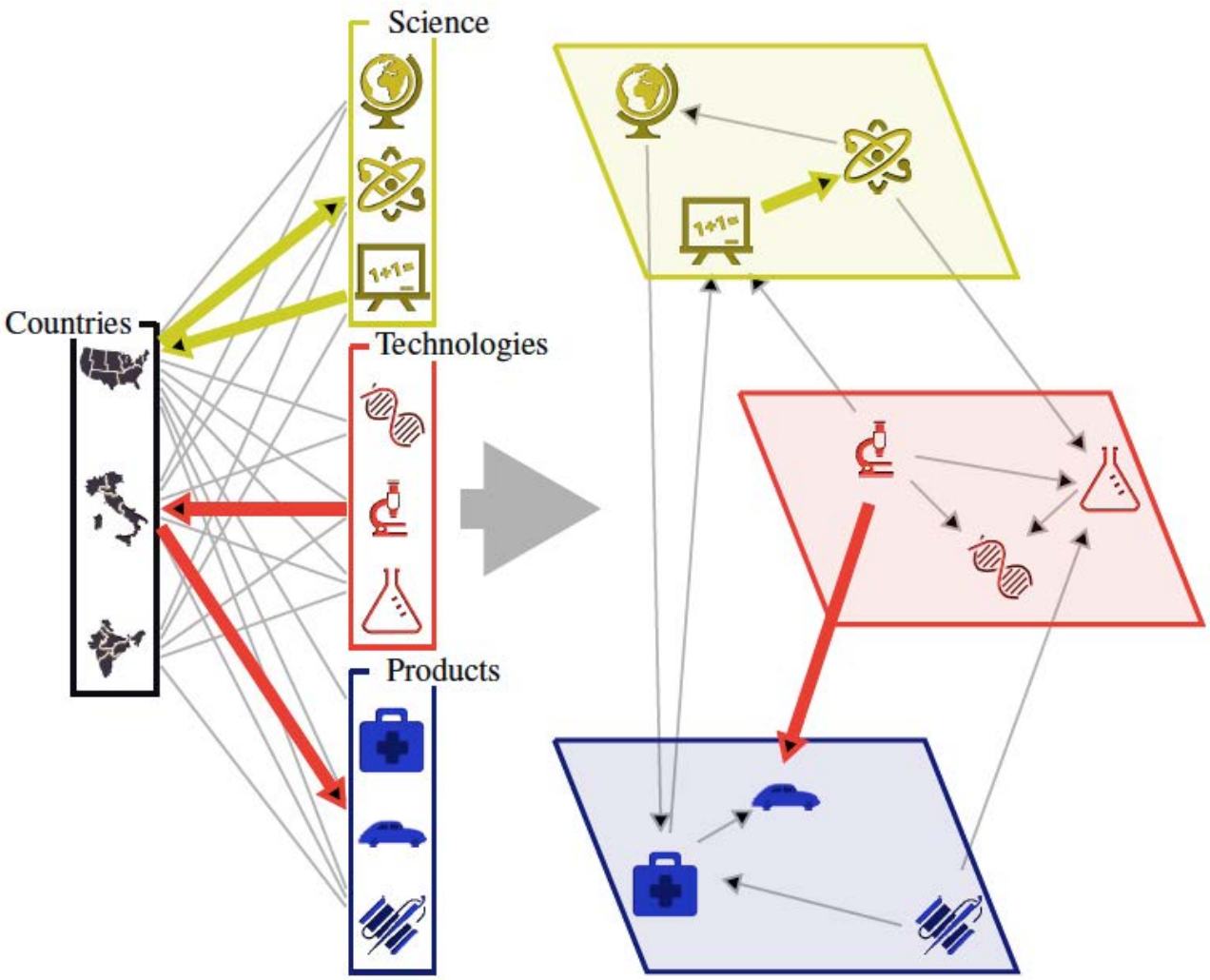
Scientifically successful papers correlate well with technologically successful ones

But someone is better than others to transfer science to technology: US, CH

And someone are worse: BRICS



Science, Technology, Industry: Multilayered Innovation Space



Left panel: schematic visualization of the triple bipartite network Countries - Sciences, Technologies, and Products;

Right panel: tri-layer representation of the resulting Assist matrix between activities. The generic element is equal to the probability that a bit of information, randomly diffusing in the triple bipartite network, travels from one activity to another either in the same or in another layer

E. Pugliese, G. Cimini, A. Patelli, A. Zaccaria, L. Pietronero, A. Gabrielli, https://arxiv.org/abs/1707.05146 (2017), submitted to Science

Probability that a bit of information starting from activity a_1 of the layer L_1 in the year y_1 arrives at activity a_2 of layer L_2 in the year y_2

Composition of Markovian random walks

$$B_{a_1 \to a_2}^{L_1 \to L_2}(y_1, y_2) = Pr(a_2; y_2 | a_1; y_1) = \sum_c Pr(a_2; y_2 | c) Pr(c | a_1; y_1) =$$

$$= \sum_c \frac{M_{c,a_2}^{L_2}(y_2)}{d_c^{L_2}(y_2)} \frac{M_{c,a_1}^{L_1}(y_1)}{u_{a_1}^{L_1}(y_1)},$$

$d_c^L(y)$ = # of activities of country c in the layer L (S,T or P) in the year y

$u_a^L(y)$ = # of countries owning activity a in the layer L (S,T,P) in the year y

Multi-layer space of development

That is,

$$B_{t,p}^{T,P}(y^T, y^P) = Pr(t|p, y^T, y^P) = \sum_c Pr(c|p, y^P) Pr(t|c, y^T) =$$

$$= \sum_c \frac{M_{c,t}(y^T)}{d_c^T} \frac{M_{c,p}}{u_p}$$

With this formula we compute the probability that an innovation that reached a product, started from a specific technological code. In the following we will study in this fashion a multi-layer space connecting products, technologies and scientific fields. We want to look how ideas and innovations scatter to products by looking at co-occurrences in countries.

Validation of results through statistical physics tools

Randomization of the network with constrained mean in and out-strengths

$$\langle C_a \rangle \equiv \sum_G C_a(G) P(G) = C_a^* \quad \forall a \quad (1)$$

G = generic graph configuration with N nodes

$P(G)$ = ensemble measure on all graph configurations

$$S(G) = -\sum_G P(G) \log P(G)$$

maximizzation with constraints $(1) \Rightarrow$

$$P(G) = \frac{1}{Z} \exp[-H(G)] \quad \text{where } H = \sum_a \theta_a C_a(G)$$

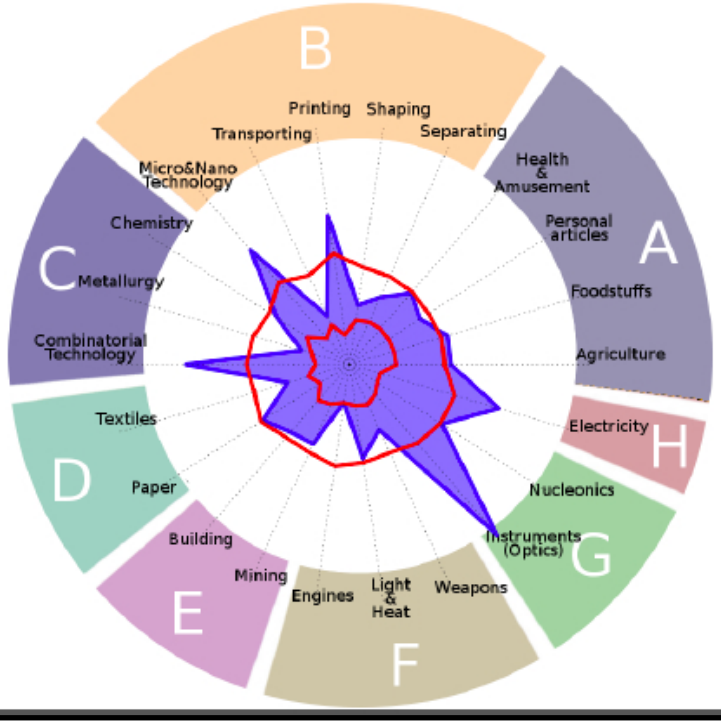
Se $\{C_a\} = \{k_i^{in}, k_i^{out}\}$ i=1,…,N → $H(G) = \sum_i (\theta_i^{in} k_i^{in} + \theta_i^{out} k_i^{out})$

$$p_{i \to j} = \frac{x_i^{out} x_j^{in}}{1 + x_i^{out} x_j^{in}}$$

$x_i^{(in,out)} = \exp[-\theta_i^{(in,out)}]$

Technology to Product: Computer desktop, 847149

From Technologies to a Product, probability that an innovation reaching a product is starting from a specific technology. Contemporary matrices. The large (small) red line is the 95% (5%) significance area. Spikes larger than the line are significant, the rest is noise.



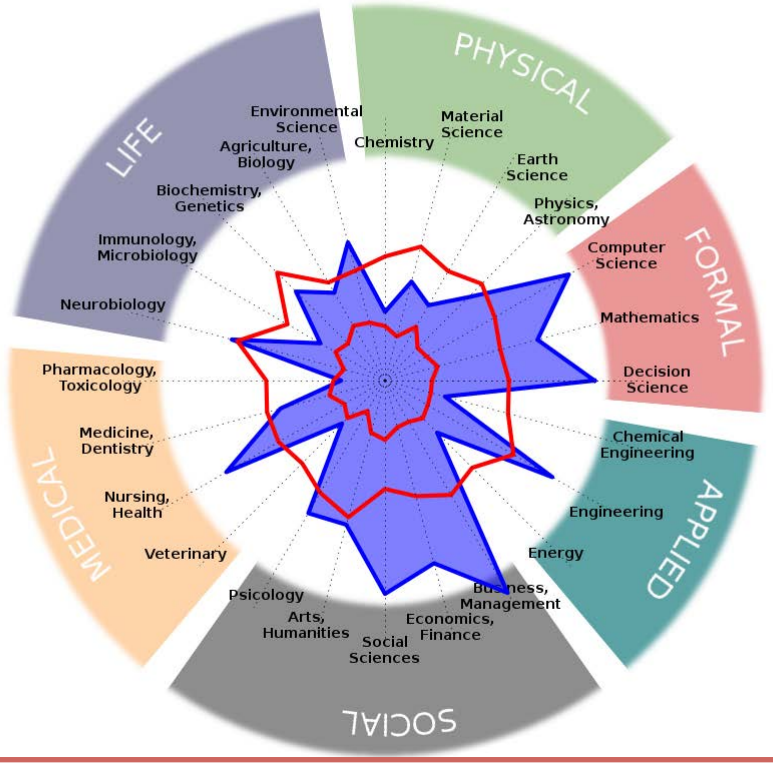
Validation: null model = bipartite configuration model

Connecting the different levels of competition, Innovations Dynamics

Science → technology → Market ??? Not really!

Market → Technology → Science

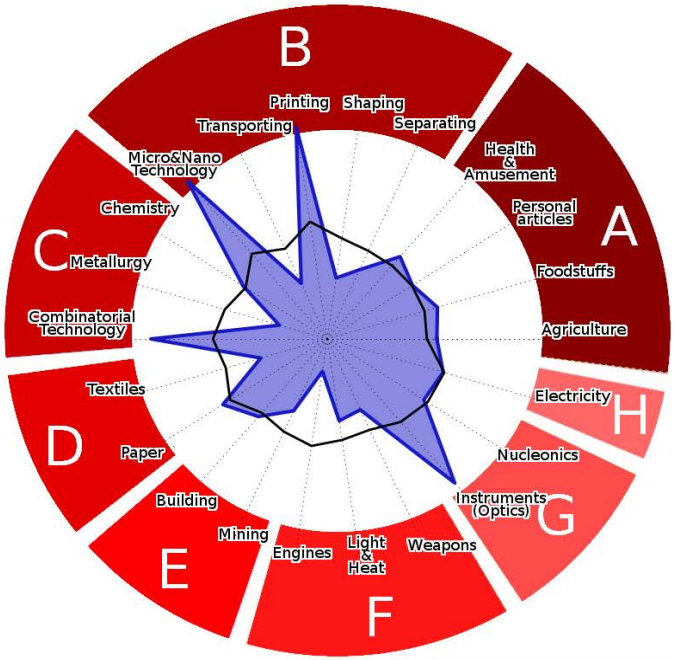
Co-occurrence (with dT=1) of scientific papers with Patents for technological code
Ex.: G06F: ELECTRICAL DIGITAL DATA PROCESSING



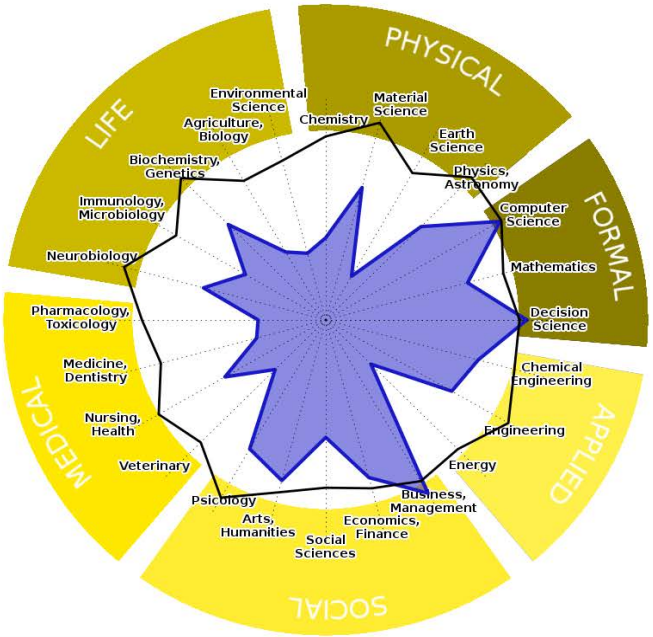
Example: Superconductivity – Criogenics; neuroscience and NMR

Products to technology and science

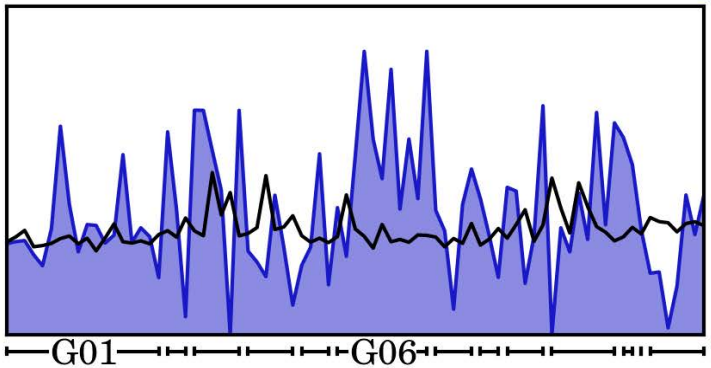
Product category: Desktop Computers



(a)



(b)

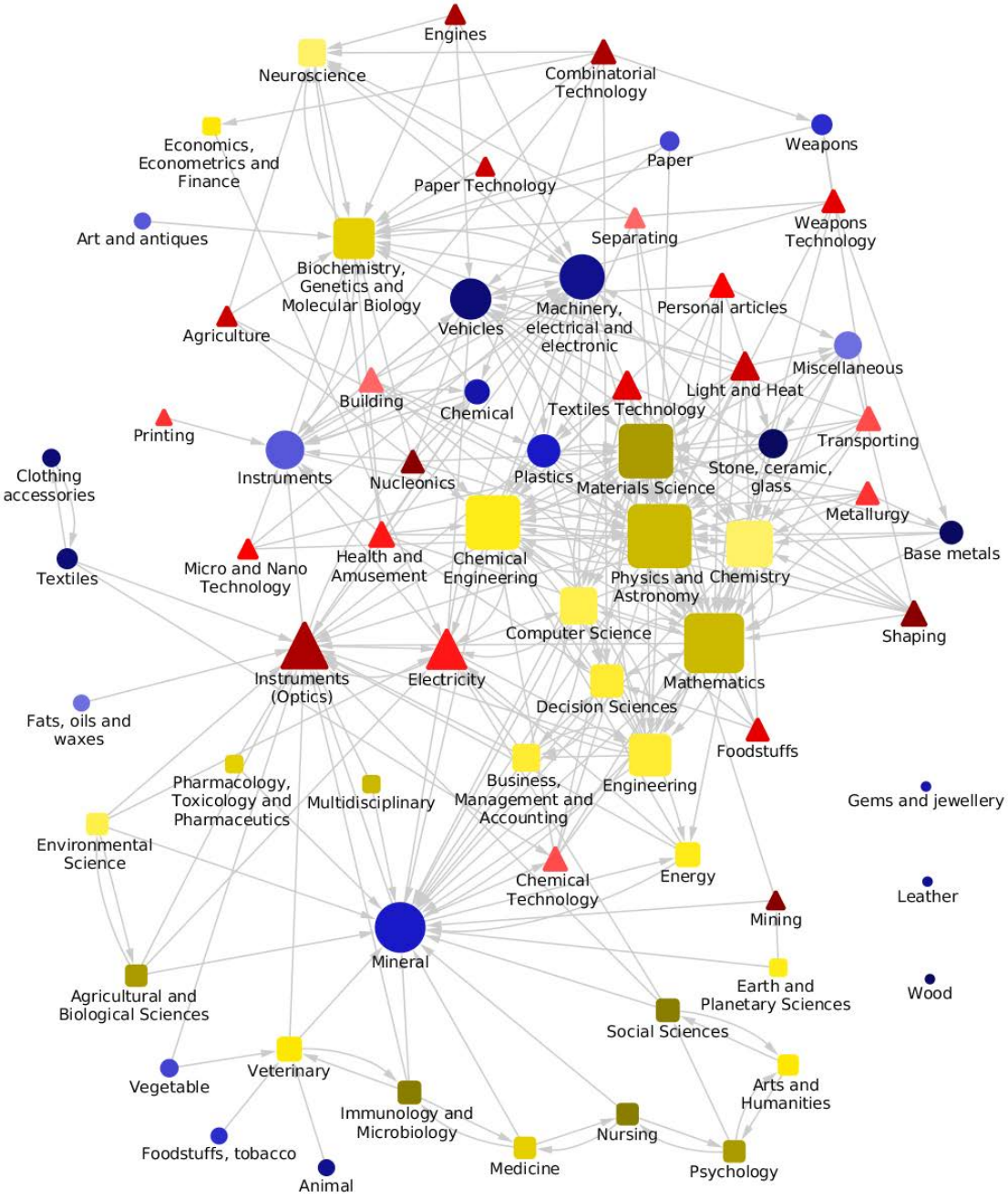


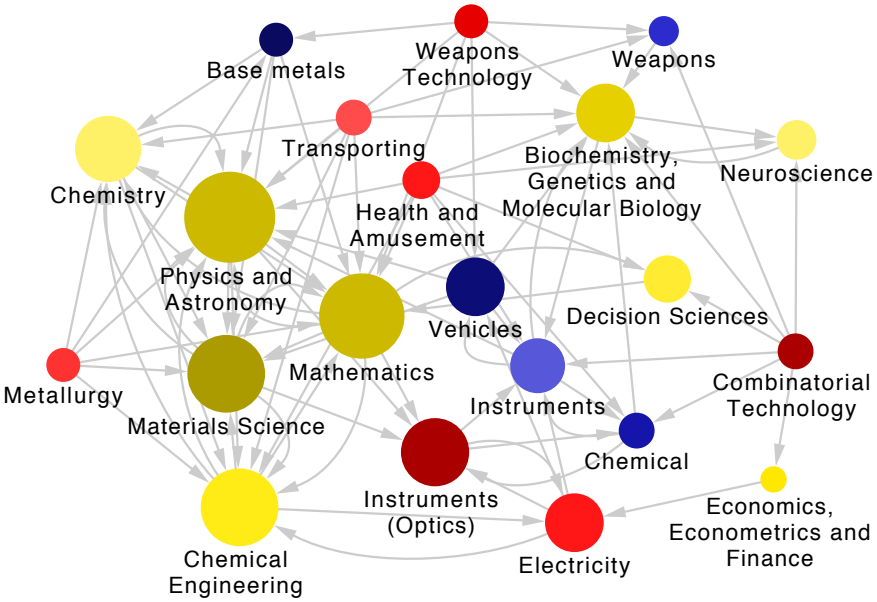
(c)

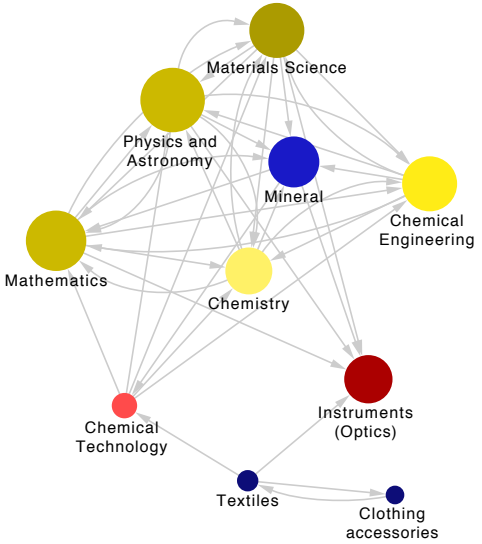
Tech. sect. “G: Physics”
G06 = Computing

Multilayer network of broadly aggregated activities (23 scientific major categories, 25 technological sub-sections, 21 product sections):

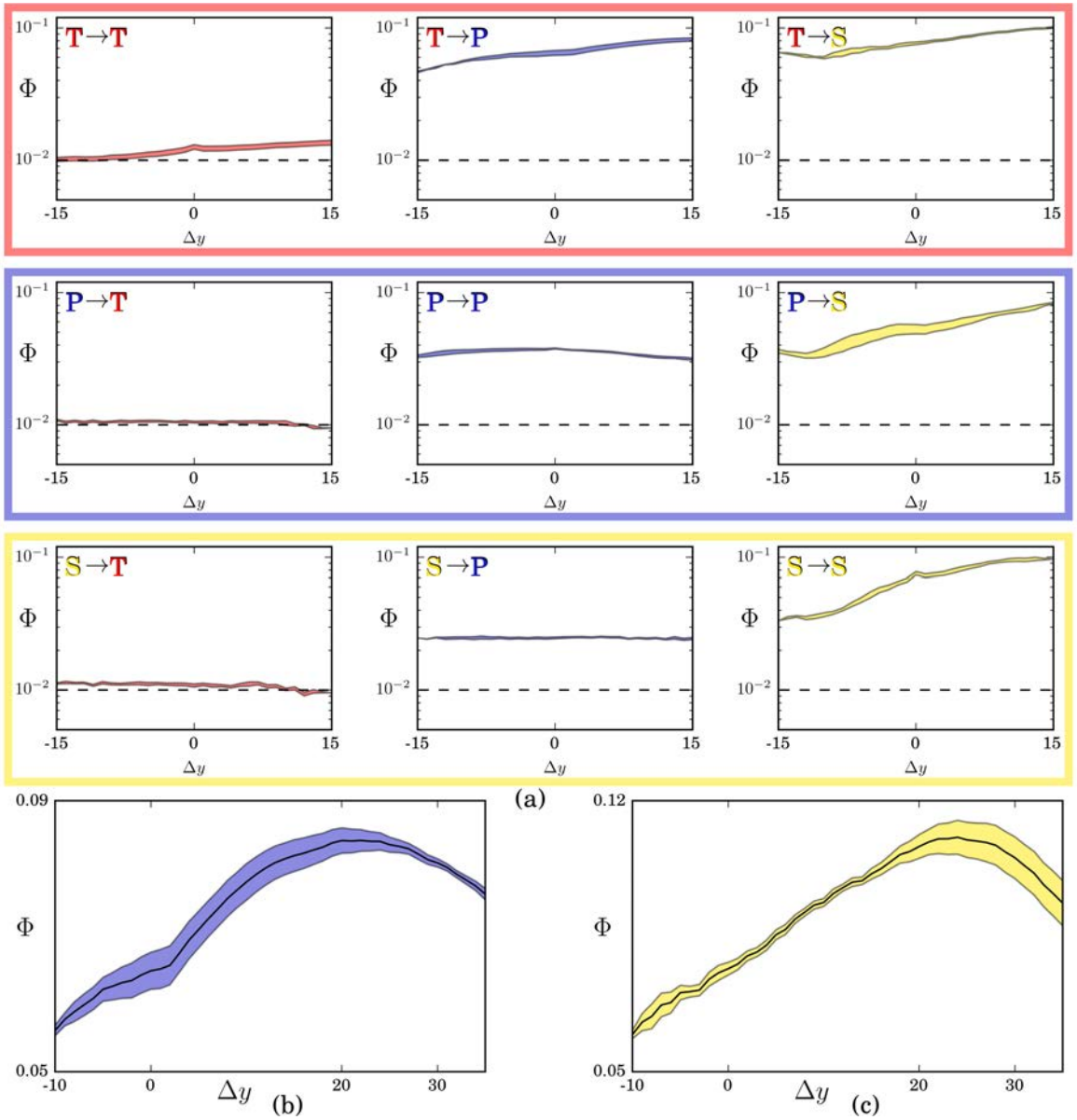
Links are obtained using a significance level of 99.999%. To increase the signal-to-noise ratio, we compute B as the average of three consecutive years in the middle of our sample (2008-2010). Red nodes represent technologies, yellow nodes represent scientific fields and blue nodes represent the export of products.




Base metals
Weapons Technology
Weapons
Transporting
Biochemistry, Genetics and Molecular Biology
Neuroscience
Chemistry
Health and Amusement
Physics and Astronomy
Vehicles
Decision Sciences
Mathematics
Metallurgy
Materials Science
Instruments
Combinatorial Technology
Chemical
Instruments (Optics)
Economics, Econometrics and Finance
Chemical Engineering
Electricity


Materials Science
Physics and Astronomy
Mineral
Chemical Engineering
Mathematics
Chemistry
Instruments (Optics)
Chemical Technology
Textiles
Clothing accessories

What triggers main innovation avalanches? Science, Technology or Products?



(a) Signal Φ of activities for varying time differences Δy. The time series are build aggregating three years and looking at all pairs of years giving the desired Δy. The shaded area denote the one sigma confidence interval. The analysis is done at a medium level of aggregation: 600 T subclasses, 300 in S and 1000 in P. Dashed black lines mark the noise level = 1%, as we consider significant links at the 99% confidence interval. The same analysis with a longer time frame is reported for T-P (b) and T-S (c) relations.

Main collaborations with policy makers

Ministero degli Affari Esteri

REPUBBLICA ITALIANA
CNEL
CONSIGLIO NAZIONALE ECONOMIA E LAVORO

WORLD BANK GROUP

Conclusions

- International competition of nations on science and technology share the fundamental features of economic competition
- Economic Complexity non-linear algorithm captures these analogies
- HERD and internationalization rate look fundamental parameter for an healthy national scientific system
- Science show a slow constant flow towards technology
- New ideas move in a multilayer science-technology-market space in a non-trivial way with maximal bursts of innovation from technology to science and industry

Funding Projects

GROWTHCOM

GROWTH AND INNOVATION POLICY-MODELLING - APPLYING NEXT GENERATION TOOLS, DATA, AND ECONOMIC COMPLEXITY IDEAS