SAPIENZA
UNIVERSITÀ DI ROMA

Progetto Schematico Tesi

The Economic Complexity Framework

Andrea Napoletano

Supervisor: Prof. Luciano Pietronero

Sapienza Università di Roma

Economic Complexity Origins

Specialization VS **Diversification**

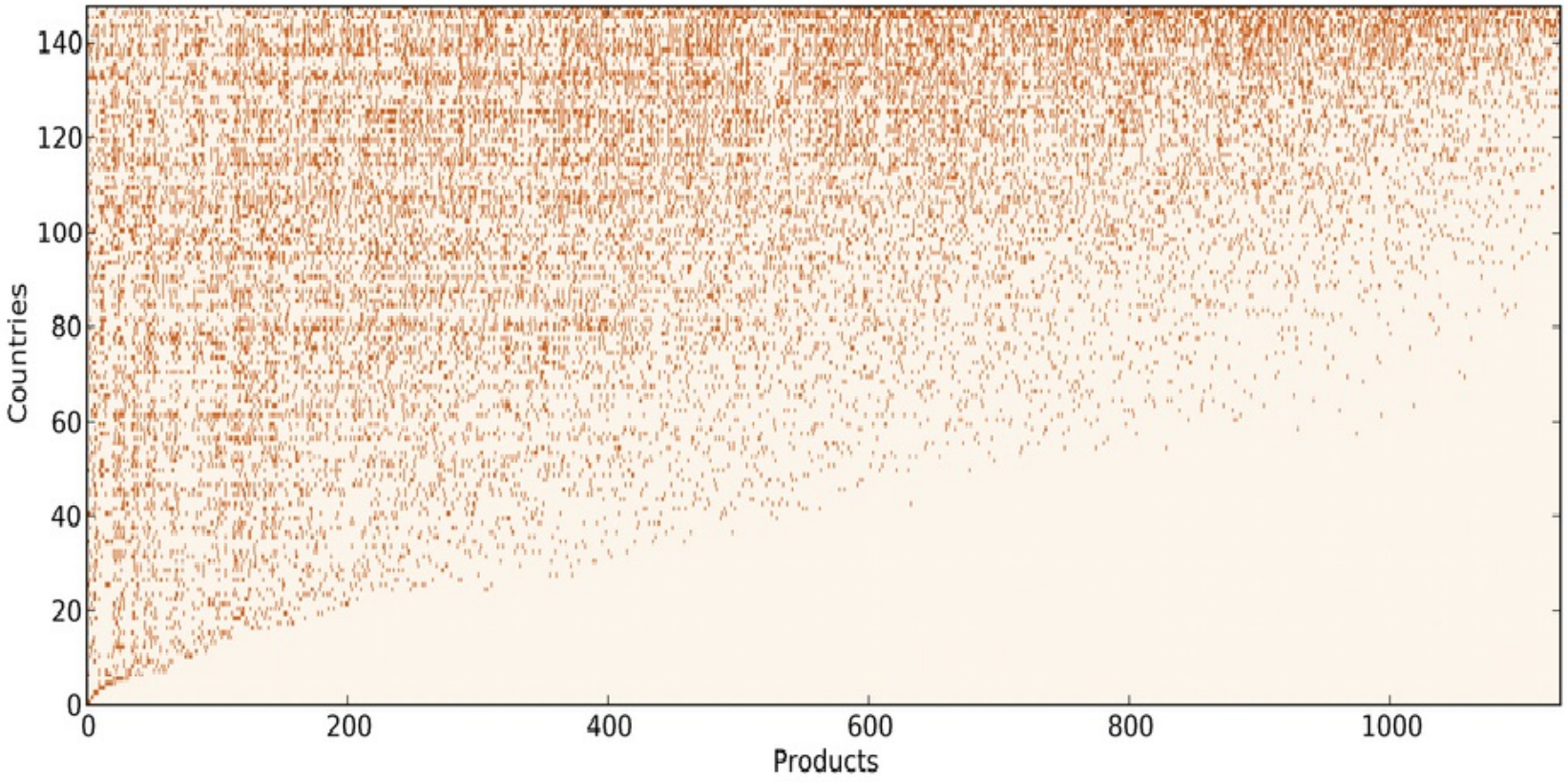
Do wealthiest countries specialize in few high tech products or produce all the available products?

Classic Economy: Smith and Ricardo

Wealthiest countries specialize their production focusing on the most advanced products that can be produced.

Economic Complexity Origins

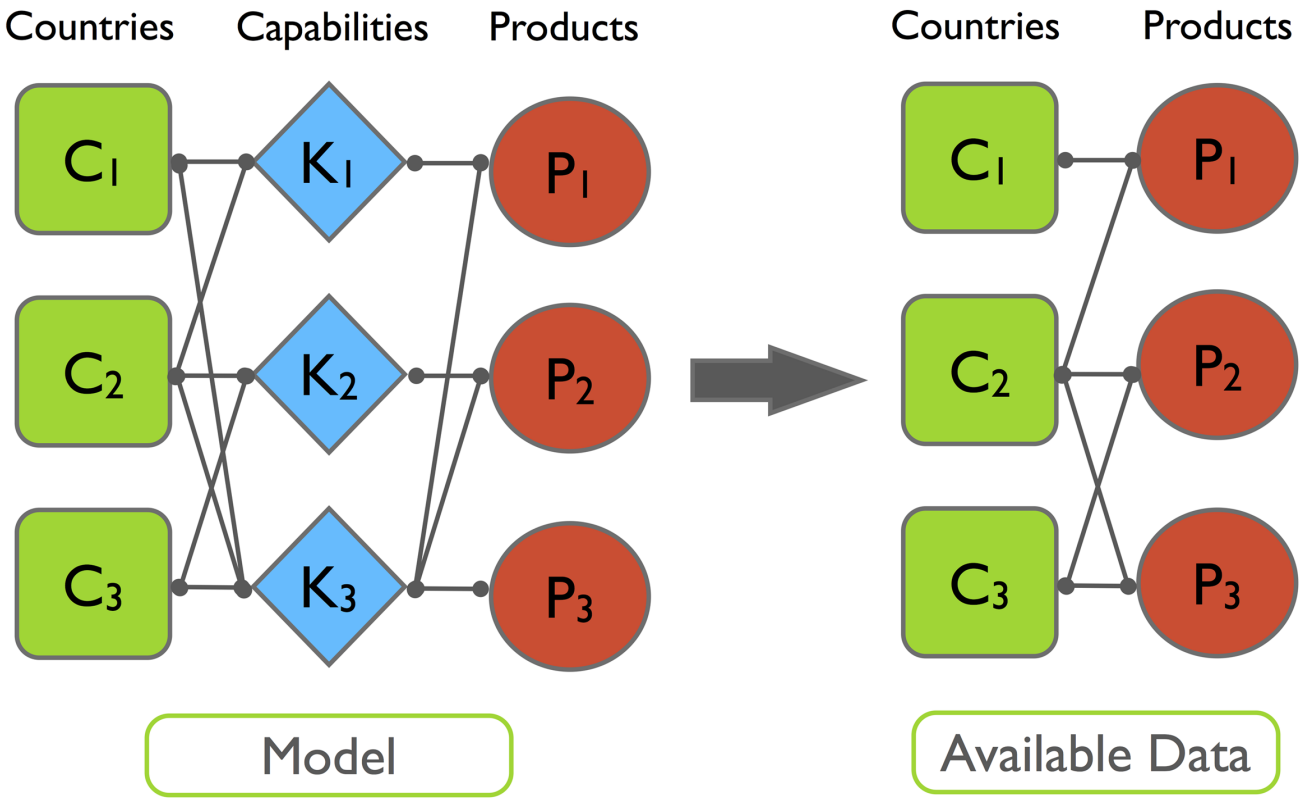
WTO Data



Wealthiest countries produce all the available products

Economic Complexity Origins

Capabilities



A product can be manufactured only by those countries who posses all the required capabilities.
Sadly there is no direct way to measure such abstract things, but we can make deductions.

Economic Complexity

More products a country produces, more capabilities it has.
More countries produce a product, less capabilities it requires.

Fitness and Complexity

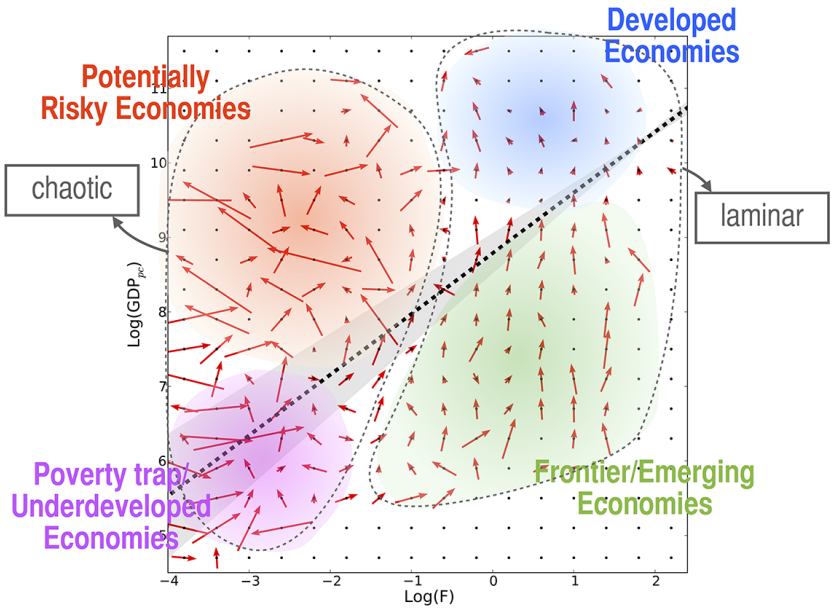
- **Fitness:** The total diversity of products produced by a country weighted by their Complexity
- **Complexity:** Determined by the Fitness of the least fit country that can produce a product

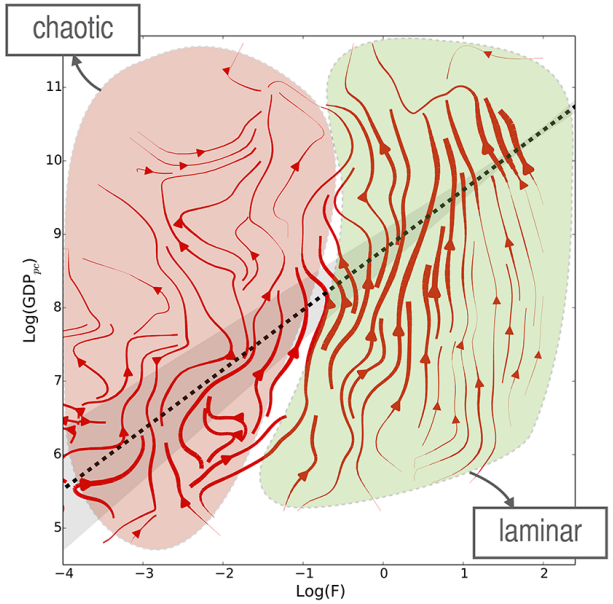
Algorithm

1. $\tilde{F}_c^{(n)} = \sum_p M_{cp} Q_p^{(n-1)}, \quad \tilde{Q}_c^{(n)} = \frac{1}{\sum_c M_{cp} \frac{1}{F_c^{(n-1)}}}.$
2. $F_c^{(n)} = \frac{\tilde{F}_c^{(n)}}{\langle \tilde{F}_c^{(n)} \rangle}, \quad Q_p^{(n)} = \frac{\tilde{Q}_p^{(n)}}{\langle \tilde{Q}_p^{(n)} \rangle}.$

Economic Complexity

GDP VS Fitness





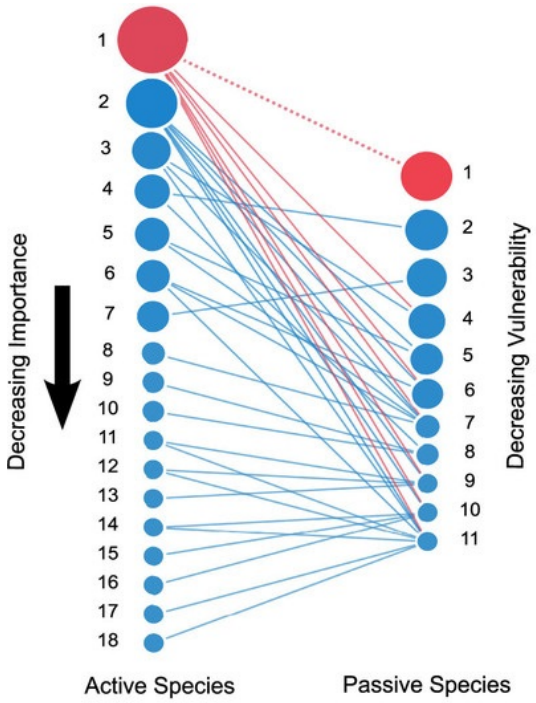
Fitness complements the GDP indicator and gives a richer description of countries evolution.

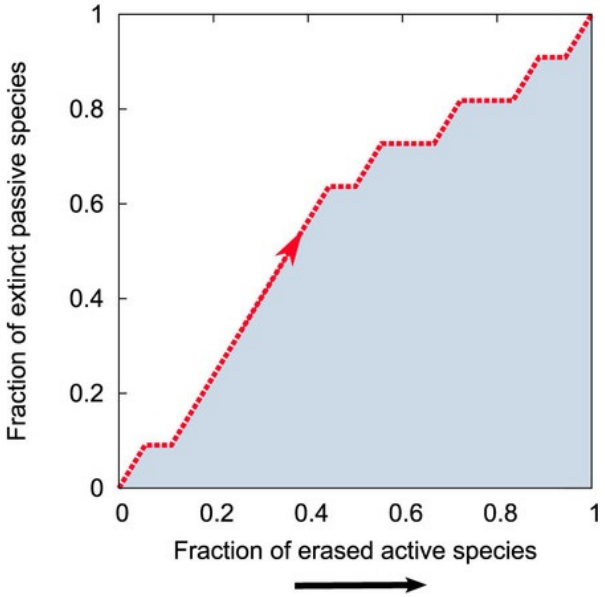
Not only Economy

Nested bipartite network in mutualistic ecosystems

Fitness $\Longrightarrow$ Importance

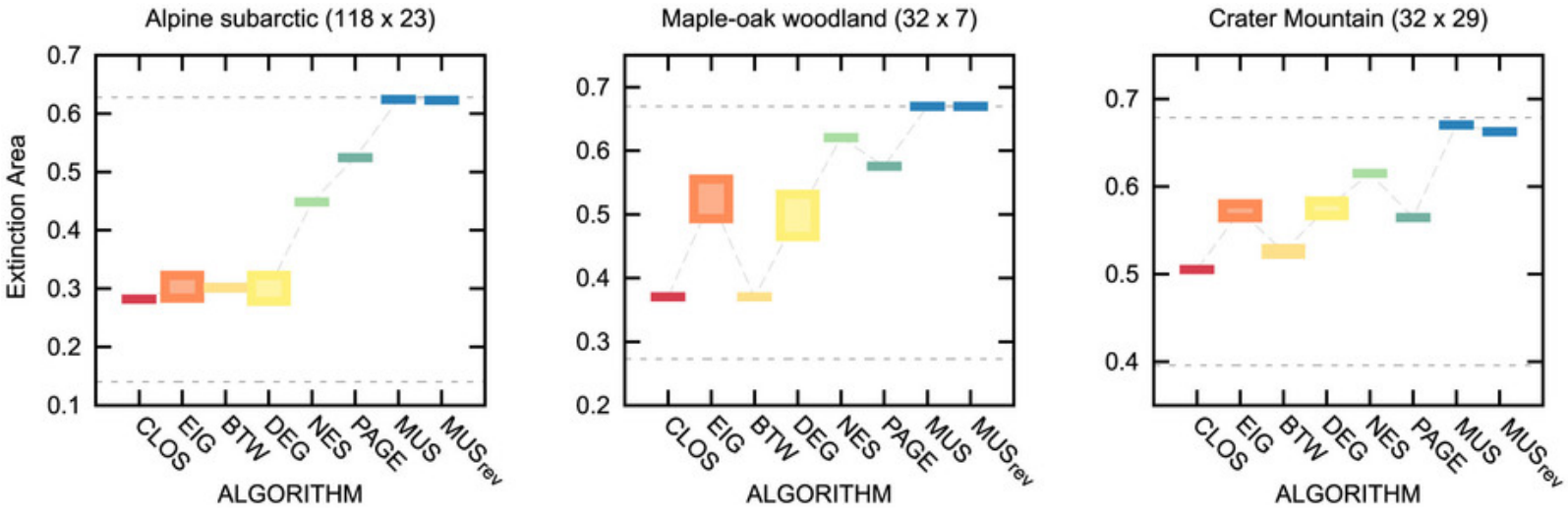
Complexity $\Longrightarrow$ Vulnerability





Not only Economy

Fitness and Complexity algorithm gives the best ranking



Among all, even Google Page Rank algorithm is outperformed. The highest extinction is given by the Fitness and Complexity algorithm.

Not only Economy

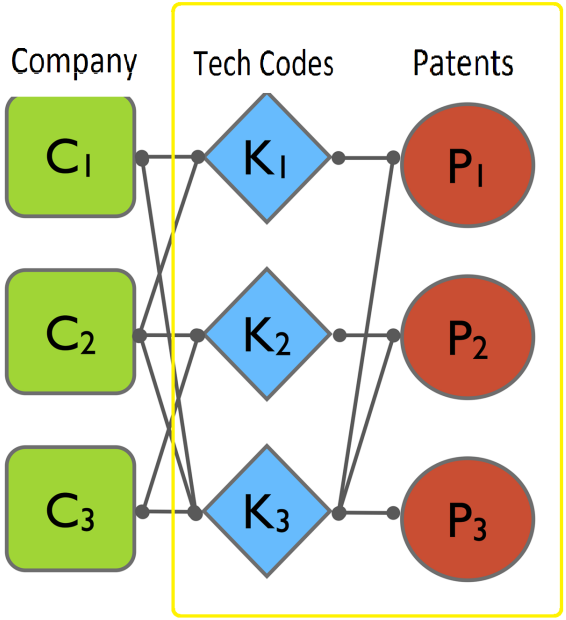
When capabilities are available

- **Patents and Tech Codes** The PATSTAT database contains year by year all the data of the patents registered world wide and it's particularly rich of data from the United States.
- **Statistical Physics of Cancer Evolution** With the advancement of technology, there is an increasing amount of genetic data on cancer available for researchers.

In both examples a direct access on the capabilities is possible, be them the technological codes or suitable pieces of Genome.

Patents and Tech Codes

Model

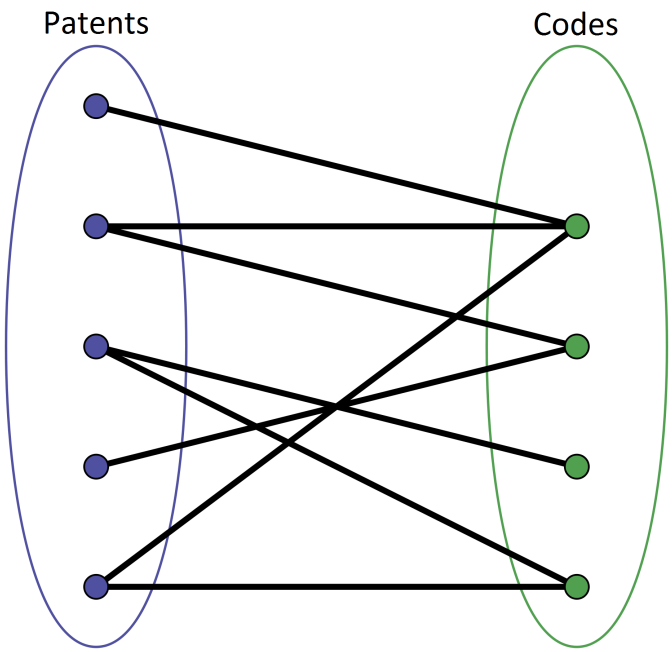


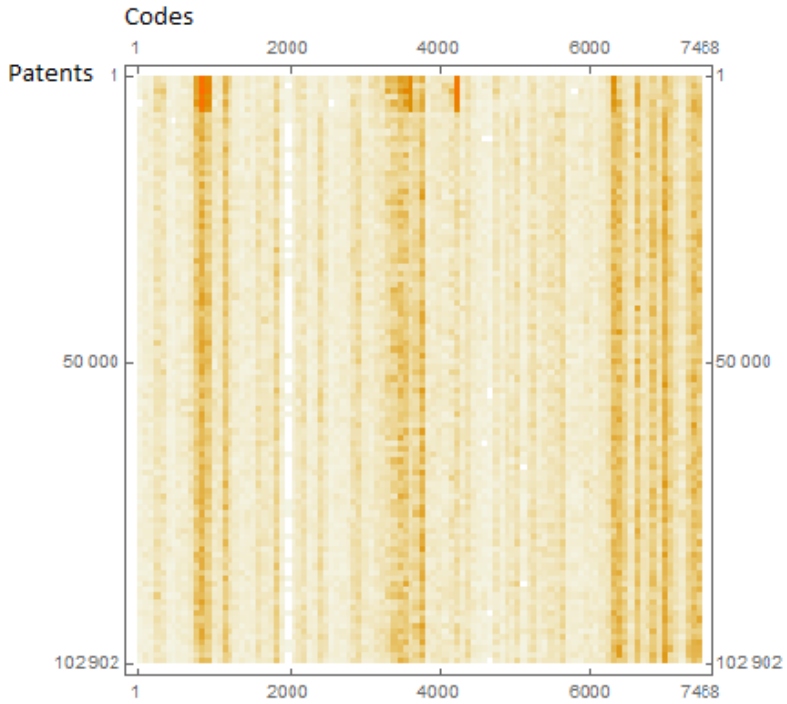
- Countries ⇒ Companies.
- Capabilities ⇒ Tech Codes.
- Products ⇒ Patents.

We focus on Patents and Tech codes, disregarding for now the companies who hold the patents.

Patents and Tech Codes

Raw data



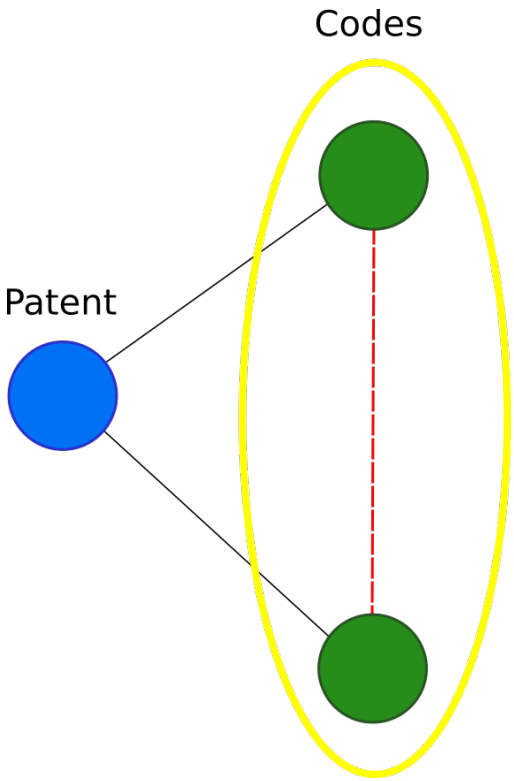


- Patents and Codes are not a nested network and codes by themselves do not provide a lot of information.
- Couples of codes however might.

Patents and Tech Codes

V Motives: Couple of Codes

How capabilities combine together



Two codes are linked if they appear in the same patent.

The strength of the link is a function of the degree of codes and patents. All couples of codes together form the network of codes.

Patents and Tech Codes

Randomizing the networks

Be d_p the degree of the patent p and u_c the degree of the code c, respectively *diversity* and *ubiquity*.

- The randomized patents-codes network is:

$$P_{pc} = \frac{d_p \, u_c}{\sum_p d_p + \sum_c u_c}$$

.

- The randomized codes network is:

$$P^p_{cc'} = P_{pc} \, P_{pc'}$$

Patents and Tech Codes

Null Model and Pvalues

For each patent p, the codes c and c' have a probability $P^p_{cc'}$ to be together in it.

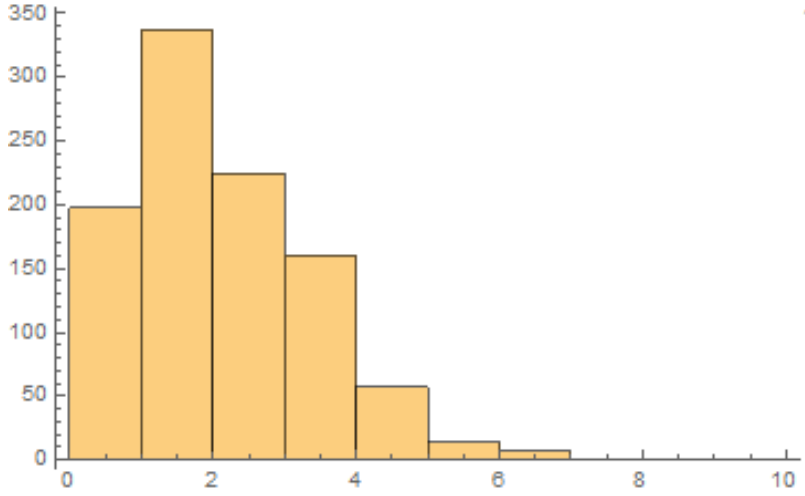
- This distribution is the Poisson Binomial distribution
- For each observed cooccurency, there are approximate methods to calculate the CDF and thus its Pvalue.

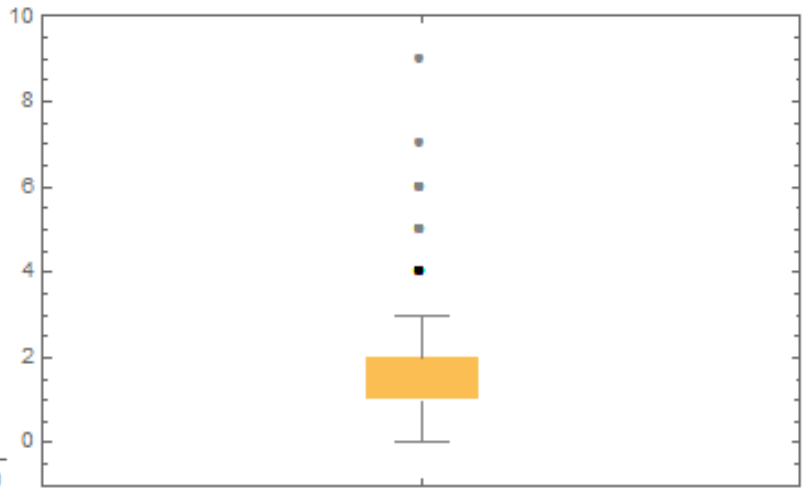
<u>The majority of the Pvalues calculated this way are 0</u>

Patents and Tech Codes

Null Model and Pvalues

Observed Cooccurencies: 100





Predicted Coccurrencies $\sim$ 1.6

The random model is too trivial for the codes network.
Nevertheless some informations can be still extracted.

Patents and Tech Codes

Weighted Network of Codes

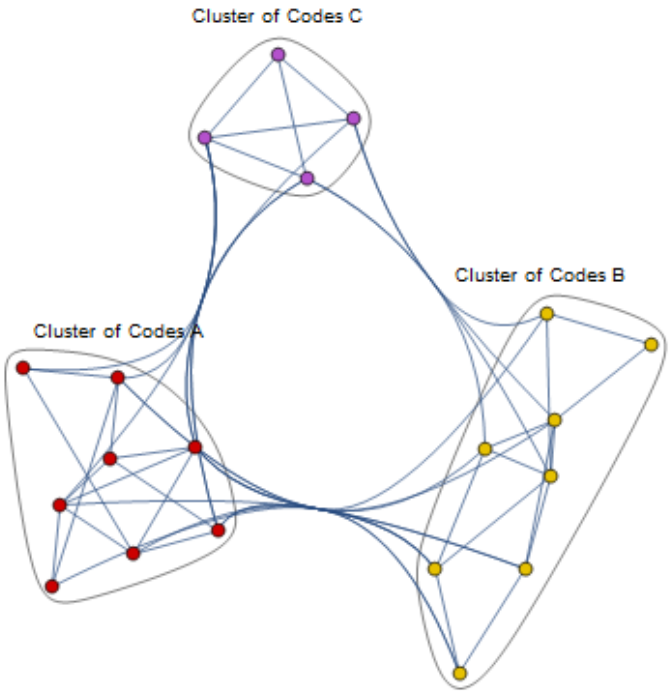
Be M_{pc} the binary matrix which yields 1 if the patent p has the code c or 0 otherwise.

$$\boxed{B_{cc'} = \frac{1}{\max\left(u_c, u_{c'}\right)} \sum_p \frac{M_{pc} M_{pc'}}{d_p}}$$

Given the *proximity matrix* $B_{cc'}$, we can construct the weighted network of codes and look for communities via community detection algorithms.

Patents and Tech Codes

Search for innovations



- Distribution of the Pvalues within each community
- Distribution of the Pvalues between different communities

Innovation manifests itself as the emergence of new links among communities not related before.

Patents and Tech Codes

What's next

- Study the evolution of communities year by year for all available years.
- Construct a refined null model and recalculate the Pvalues
- Generalize the analysis looking for more than couple of codes, triples quadruples and in general nples.

Statistical Physics of Cancer Evolution

A sketch of the project[1]

- There is an increasing amount of genetic data available on individual mutations, gene expression and epigenetic factors all about cancer.
- What is missing is a coherent analysis that combines all this informations together.
- The leading idea of our analysis will be to try to interpret cancer as an emerging collective complex phenomenon instead of the result of local events.
- My work within the Roman group will be on suitable networks of genes and microRNAs identified by our colleagues.

[1] Three research groups led respectively by Luciano Pietronero (Sapienza), Stefano Zapperi and Caterina A.M. La porta (Università di Milano)

Thank you!

References

- A. Tacchella, M. Cristelli, G. Caldarelli, A. Gabrielli, L. Pietronero; *A New Metrics for Countries' Fitness and Products' Complexity*, Scientific Reports 2, Article number: 723 doi:10.1038/srep00723 (2012)
- A. Zaccaria, M. Cristelli, A. Tacchella, L. Pietronero; *How the Taxonomy of Products Drives the Economic Development of Countries*, PLoS ONE 9(12): e113770. doi: 10.1371/journal.pone.0113770 (2014)
- M. Cristelli, A. Tacchella, L. Pietronero; *The Heterogeneous Dynamics of Economic Complexity*, PLoS ONE 10(2): e0117174. doi:10.1371/journal.pone.0117174 (2015)
- V. Domnguez-García, M. A. Munoz; *Ranking species in mutualistic networks*, Scientific Reports 5, Article number: 8182 doi:10.1038/srep08182 (2015)
- F. Chung, L. Lu; *Connected Components in Random Graphs with Given Expected Degree Sequences*, Annals of Combinatorics 6(2):125-145, doi: 10.1007/PL00012580 (2002)
- Y. Hong *On computing the distribution function for the Poisson binomial distribution*, Computation Statistic and Data Analysis, Volume 59, Pages 41-51, (2013)