ICT-Enabled Devices for Better Living

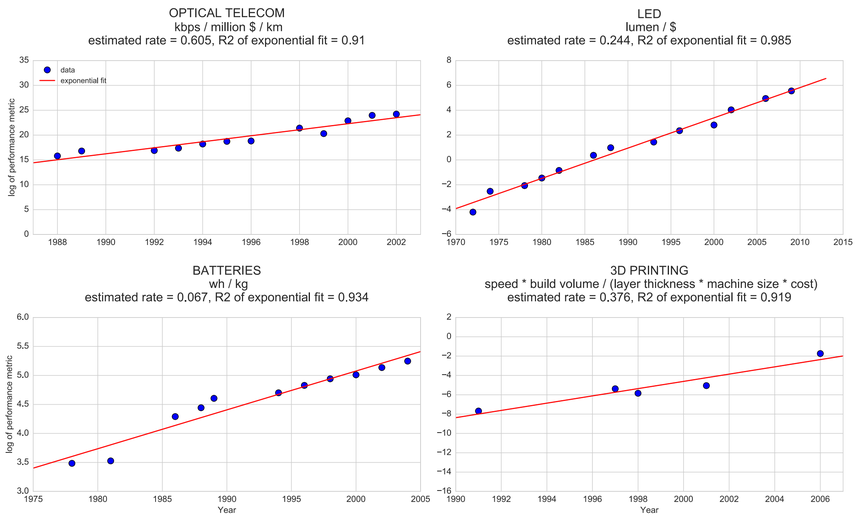
Predicting industry economic performance by dynamic mapping industry-technology linkages and mining patent data

Giorgio Triulzi
Chris Magee

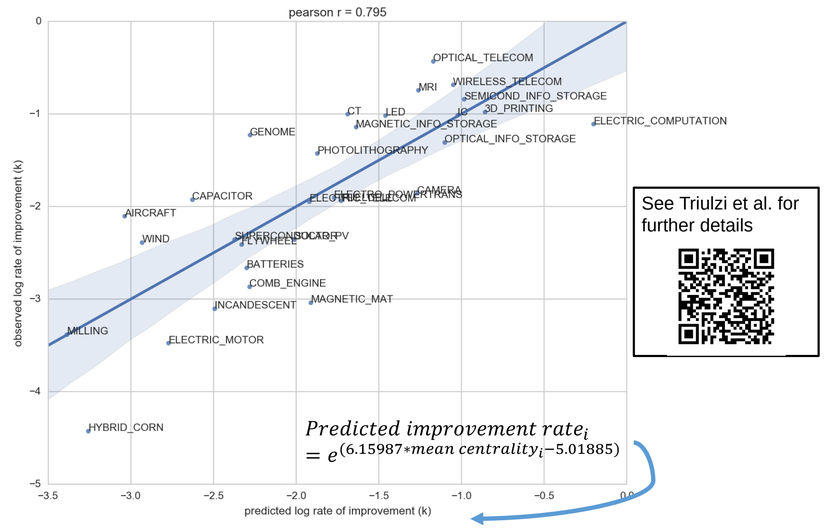
OVERALL GOALS

1. Identify set of patents that best represent an industry or product category over time using a combination of machine learning and patent network analysis
2. Use output from 1) to investigate correlation of economic and technology dynamics (e.g. technology functional performance and industry productivity and real output)
3. Identify the next fast growing industries and products that will experience rapid productivity improvements

Functional performance has been shown to improve exponentially over time for many different technologies (e.g. Koh and Magee 2006 and 2008, Magee et al., 2016), much like the famous Moore's law for integrated circuits, but with varying yearly rate of improvements.



These improvement rates have been shown to be predicted, for 30 different technology domains, by the centrality of a domain's patented inventions in the overall patent citation network (see Triulzi et al., 2017 for details). The predicted rate can be obtained using the equation below, whose coefficients have been estimated using patent and performance data for 30 domains.



See Triulzi et al. for further details

Research hypothesis

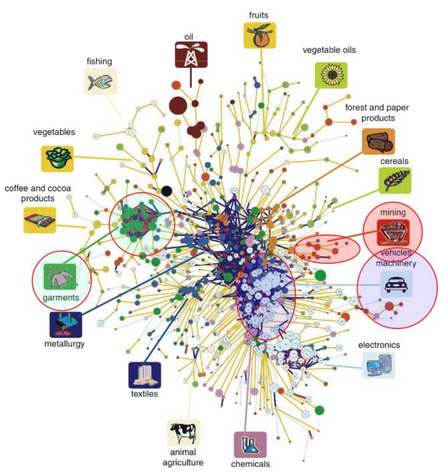
Multi-factor productivity growth should be proportional to technology performance improvement rates

Patent centrality is associated with technology improvement rates

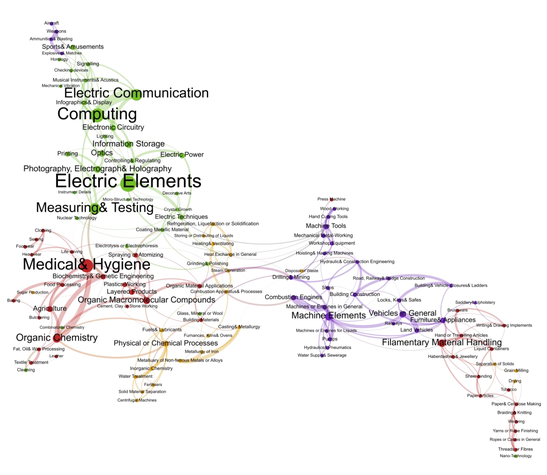
→ An industry's patent centrality should predict its productivity growth and real-output growth

A technology's functional performance improvements over time should be associated with improvements in the economic (total-factor) productivity for the products in which the technology is used. Economic theory suggests that higher productivity may cause larger real output. Therefore we expect to observe higher total-factor productivity (TFP) and real output in industries whose products incorporate faster improving technologies.

Link industry and product categories to patents



The Product Space
Fig 1.b in Hidalgo, Klinger, Barabasi and Hausmann, 2007, Science



The Technology Space
Figure from Jeff Alstott, Giorgio Triulzi, Bowen Yan and Jianxi Luo, 2017, Scientometrics

Economic data such as total factor productivity (TFP) and real output are recorded using industry and product classification systems. Therefore, to test our hypothesis, we need to identify the sets of most relevant patents for each industry or product category.

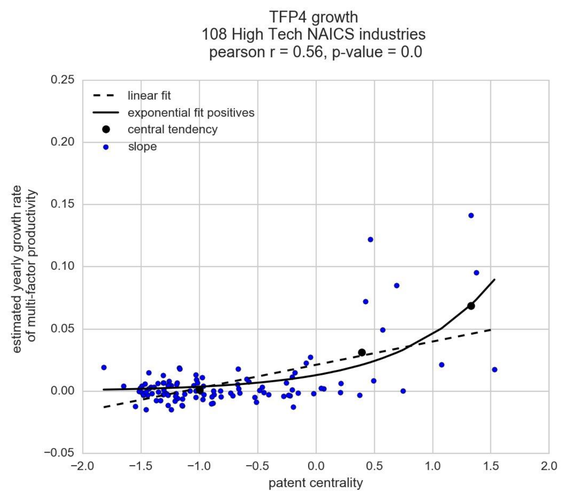
To achieve this goal we will use an unsupervised Machine Learning approach.

We will perform a hierarchical clustering of patents based on:

- Membership of community of inventors based on patent co-authorship
- Membership of community of firms based on patent co-ownership or shared inventor careers
- Membership of path of engineering improvements based on patent citations
- Membership of technology sub-class
- Membership of community of patents based on topic similarity

Then we use key words from industry categories' descriptions and retrieve the most relevant cluster of patents by key word counts.

Estimate relationship between yearly rate of total-factor productivity growth and patent centrality for High-Tech industries



As a preliminary result we test the validity of our hypothesis by computing the correlation between patent centrality and industry (total-factor) productivity growth by using an existing concordance table between high tech industries and patent classes. These are identified by using the North-America Industry Classification System (NAICS) and the United States Patent Classification System (USPC). The existing concordance table is described in Lybbert and Zolas, 2014.

We find that high tech industries that experience faster productivity growth also tend to have more central patents.

We also found that even the best existing concordance tables are affected by considerable noise, possibly because they are created by only using information on key words contained in industry descriptions and patent text corpuses.

The existence of a relationship consistent with our hypothesis despite this noise is encouraging. It provides reasons to believe that our Machine-Learning approach will produce better crosswalks between industry/product categories and patents that will improve our ability to predict the next fast growing industries and high-tech products.