Paper Summary

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Paper | Research Question | Findings | Concept | Measure | Keywords/Notes |
| Hall & Trajtenberg (2006) | Can we identify GPTs using patent data? | Chemical patents have the highest generality  Computing and communications technology saw the most growth in patenting  Citation lag captures older and slower-moving technologies than GPTs  Looking at generality and patent growth in combination, ICT patents score high in both | Generality | HHI index based on   1. US patent class 2. Hall-Jaffe-Trajtenberg (2002) technology subcategories 3. Main International Patent Class 4. Industry of manufacture based on Silverman (2002) 5. Industry of use based on Silverman(2002)   Also looked at the generality of the citing patents | GPT  Both this paper and Moser & Nicholas (2004) measures citation lags, but for different reasons: In this paper, the assumption is that GPTs will take longer time to pervade the economy and thus have longer citation lags. Moser & Nicholas (2004) uses citation lags as a measure of longevity of the technology. |
| Innovation-Spawning (Patenting Growth) | Average within-class growth for three sub-periods (1975-83, 1984-92, 1993-99)  Also looks at growth of citing patent classes |
| Citation Lags | Mean citation lags |
| Moser & Nicholas (2004) | Does electricity have GPT characteristics? | Electricity patents do not exert GPT characteristics by conventional definitions (high originality but low generality scores, fewer citations per patent, shorter citation lags) | Originality | (Binary) whether the focal patent is cited by a future patent as its earliest citation | GPT  Focuses on historical patents granted in year 1920-1928  Compares electricity patents with chemical, mechanical, and other patents |
| Longevity | Mean and maximum lag between patent grant and forward citation |
| Generality | HHI index based on three-digit USPTO classes of citing patents |
| Petralia (2020) | Develops a text-based measure of GPT patents | The new measure align well with anecdotal and historical evidence on GPTs  Having GPT patents identified by the new measure is positively correlated with firm performance (OR/employee) | Technological Improvement | Patenting growth | GPT |
| Pervasiveness | Extract bigrams from USPC class and subclass definitions  Calculate the ratio between the bigram’s frequency in its own class and in other classes.  If the ratio is over 5, the bigram is thought to be representative of the class  Count the number of technological classes in which there is a patent containing more than 3 different keywords |
| Innovation Complementarity | Number of technological classes the focal class co-occur with in a patent in a given year |
| Bhatt et al. (2023) | Identifies the technology trajectory of Blockchain using patent citation analysis | Identified five significant technology clusters based on the key-route main path analysis | Technological Trajectory | Main path analysis: construct a citation network, assign weights to each edge according to its traversal count, identify paths with highest weights | Disruptive Innovation |
| Momeni & Rost (2016) | Identifying possible disruptive technologies in the photovoltaic industry | Patent-development paths, k-core analysis and topic modeling identify thin-film technology as a possible disruptive technology | Technological Trajectory and Disruptive Technology | Main path analysis to figure out key path, apply k-core and topic modeling on the key path to identify subfield | Disruptive Innovation |
| Funk & Owen-Smith (2017) | Developing a network measure of technological change | The new index measures the extent to which a new invention consolidates or destabilizes existing technology streams | Consolidating and Destabilizing Technologies | CD index:  For all forward citations  f\_i = 1 if the patent cites the focal patent  b\_i = 1 if the patent also cites the focal patent’s predecessor (backward citation)  positive value means the patent is destabilizing, negative means consolidating | Technological Change, Disruptive Innovation  Wang et al. (2024) has a related, but more complex, measure of disruptive innovation |
| Briggs & Buehler (2018) | Is a radical innovation more likely to be a breakthrough innovation? | Inverted U-shaped relationship between radicalness and likelihood of breakthrough  Established innovator, university ownership, and joint patent ownership moderates the optimal level of radicalness | Breakthrough Innovations | (Binary) whether a patent is within top 1% in forward citation counts in a given year and technology subcategory  (Count) number of forward citations received within 5 years | Radical innovation, Breakthrough Innovations |
| Radicalness | OECD’s radicalness index (Squicciarini, Dernis, and Criscuolo 2013) number of unique four-digit IPC classes in the backward citations not allocated to the focal patent, divided by the total number of unique IPC classifications in the backward citations |
| Capponi, Martinelli, & Nuvolari (2022) | How to identify breakthrough innovations? | Develops a method that uses award-winning innovations | Breakthrough Innovation | Uses data on the Queen’s Award for Innovation in the UK and matches the innovation to US patents | Breakthrough Innovation  Doesn’t distinguish breakthroughs with radical innovations |
| Dahlin & Behrens (2005) | When is an invention radical? | Develops a measure for radicalness and identifies radical inventions in the tennis racket industry | Novelty | Citation structure dissimilarity (pairwise measure: number of overlapping citations divided by the number of unique citations of both patents) with past patents  (lower, the more novel) | Radical Innovation |
| Uniqueness | Citation structure dissimilarity with concurrent (same grant year) patents  (lower, the more unique) |
| Impact | Citation structure dissimilarity with future patents  (higher, the more impactful) |
| Kaplan & Vakili (2015) | Does recombination lead to breakthrough innovation? | Counter to theories of recombination, patents that originate new topics are more likely to be associated with local search, while economic value is the product of broader recombinations as well as novelty | Novelty | Topic-originating patents:  Topics identified through topic modeling (LDA)  Identify topic-originating patents as patents over a threshold weighting for the topic and appearing in the first 12 months of the topic formation | Breakthrough Innovation, Recombination |
| Technological Distance | Trajtenberg et al. (1997)  Overlap in three-digit classes between focal patent and backward citations |
| Technological Diversity | Same as an “originality” measure: 1 -HHI (backward citation USPC) |
| Breakthrough Innovation | (Binary) whether a patent is in the top 5 percent of cited patents (five year forward citations) |

Kelly, Papanikolaou, Seru, & Taddy (2021)

Research Question: How can we measure the degree of technological progress over time?

* Much of economic growth is attributed to improvements in productivity, which is a function of technological progress
* Measuring this progress is difficult
* Existing measures (e.g., citation-weighted patents) have disadvantages

Methodology

* Idea: a patent is important if it is different from prior patents (novel), but similar to future patents (impactful)
* Data sources
  + USPTO patent search website: records all patents beginning from 1976
  + Google Patents for pre-1976 records (missing inventor/assignee addresses, examiner, and attorney)
  + Final dataset: full text of over 9M patents over the period 1840-2010
* Clean and convert text into document term matrix (DTM)
  + Denoted
  + Columns correspond to words, rows to patents
  + : count of word in patent
* TFIDF transformation: gives more weight to “important” words
  + A word is more important if it appears frequently in the focal document
  + Term Frequency:
  + A word is less important if it appears frequently in other document
  + Inverse Document Frequency:
  + High : term appears relatively frequently in document but does not appear in most other documents
  + is especially representative of document ’s semantic content
* Standard is not suitable for the paper’s purposes
  + Ex: Nicola Tesla’s 1888 patent of an AC motor
  + One of the first patents to use the phrase “alternating current”
  + Standard for the term would be low because many future patents use it
  + The importance of the term in earlier patents is discounted
* Backward- (
  + “alternating current” is not penalized for Tesla’s patent
  + For measuring similarity between two patents, apply for the earlier patent to both (to reflect the impact of the earlier invention on future innovation)
* Constructing the similarity between patents and
  + Calculate the cosine similarity between the vectors
* 9M X 9M matrix of pairwise similarities

A comparison of graphs and diagrams

Description automatically generated with medium confidence

The more similar, the higher the likelihood of citation

Patents are highly dissimilar

Average: 0.102

Median: 0.078

P90: 0.176

P95: 0.229

* Similarity networks (nodes are connected when similarity is 0.5 or higher)

A diagram of different types of information

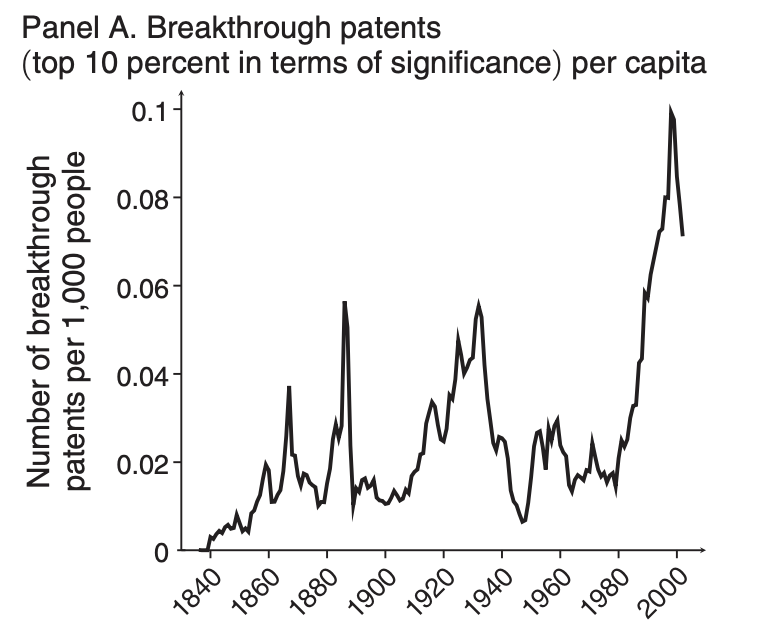
Description automatically generated with medium confidence

* From similarity to importance
  + Again, a patent is important if it is both novel and impactful
  + Backward Similarity: (Higher: less novel)
  + Forward Similarity: (Higher: more impactful)
  + Quality: (

Validation of the new measure

* The new measure captures historically established important patents better than forward citations
* Positively correlated with forward citations and Kogan et al. (2017) measure based on stock market reactions

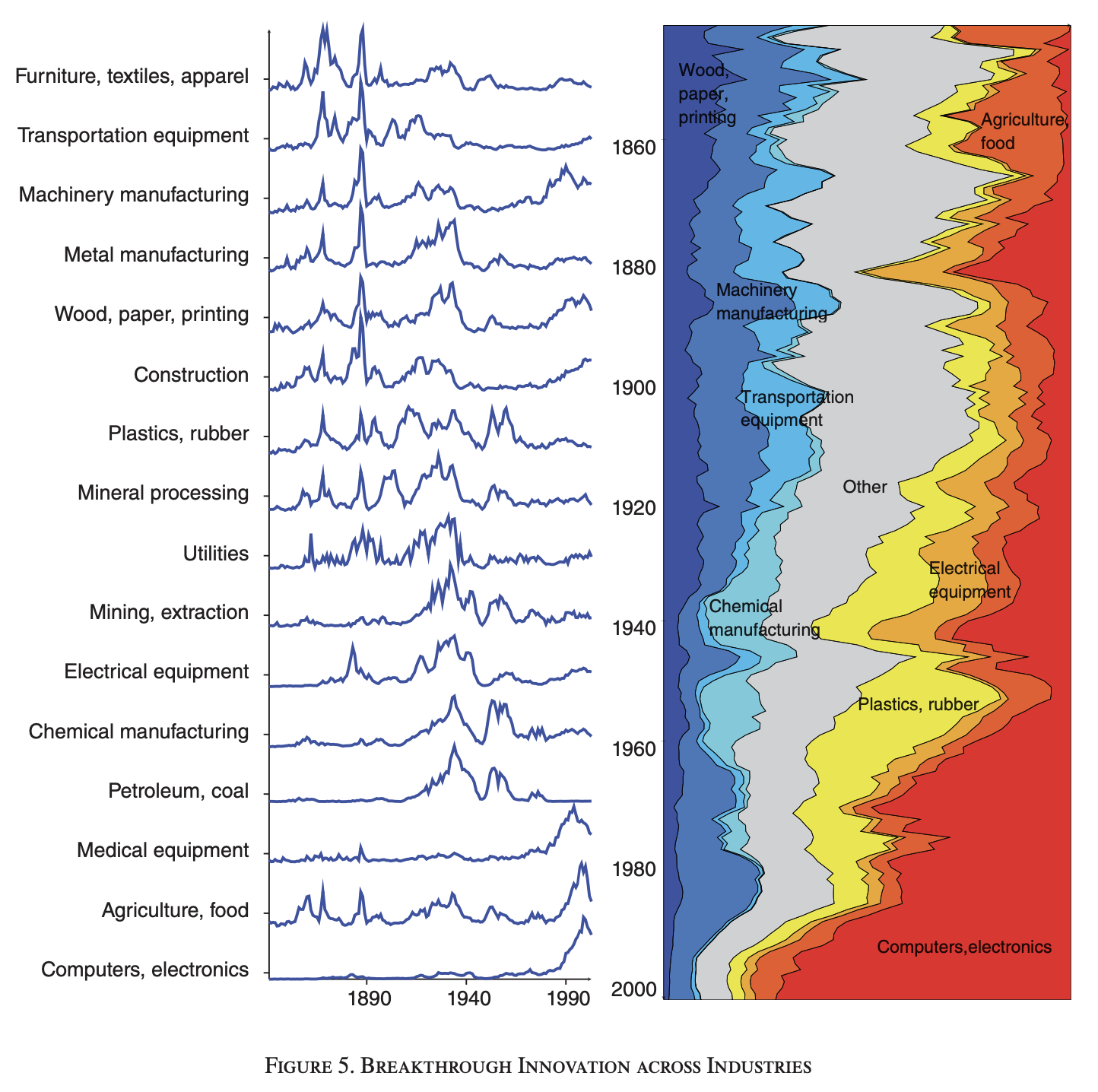
Constructing indices of technological progress



2nd industrial revolution

Advances in manufacturing

Computing, genetics, and telecommunication



Engineering and construction, consumer goods, manufacturing

(Bollman Bridge, sewing and knitting machines)

Electrical equipment

(electric light, telephone)

Chemistry, Materials

(Bakelite, PVC, Nylon, Penicillin)

Transportation

(railroads, airplanes)

Electronics

(Microchip, computers, software)

Thoughts

* Good measure of patent quality
* Could apply the basic idea (novelty + impact) to more advanced measures of text similarity (Transformer models, etc), but computation may be costly