The determinants of science-based green patents

Federico Nutarelli [1], Corrocher N.[1], Morrison A. [2]

[1] Bocconi University,
[2] University of Pavia

Motivation

- There is growing attention on the role of science to feed technological innovations tackling global challenges (Poege et al., 2019)
- **This is particularly relevant for green technologies** addressing climate change, because of their novelty, complexity, and uncertainty (De Marchi, 2012; Ghisetti et al., 2017; Popp, 2017; Marzucchi et al 2020).
- Climate change has indeed tangible and catastrophic effects, demanding the **rapid deployment of green technologies**.
- **Empirical evidence** showcasing the effectiveness of public spending on science in generating deployable green innovations is therefore highly valuable.

Contribution

- This paper aims at: i) defining a distance measure to delineate the scientific foundations of green patents; ii) investigating to which extent green patents are based on science and which are the most science-based fields; iii) formally assessing the determinants of science-based green patents (via Machine-Learning)

- Our results show that:
  1. Green patents grounded in scientific principles occupy a central position within the citation network;
  2. The percentage of citations of green patents is a key determinant of the scientific basis of patents.

Why is focusing on science-based patents so important?

Based on Ahmadpoor and Jones (2017), science-based patents should be the most cited (i.e. most valuable) ones. This finding is confirmed also via our novel definition of distance in ▸ Distance def.:

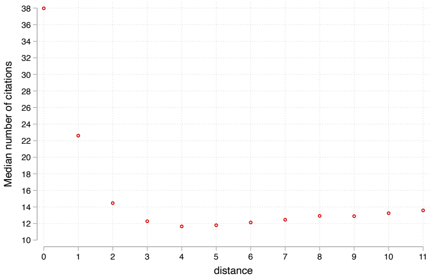


Figure 1: **Citation pattern by disatnce.**

Data

PATSTAT + EMAKG databases containing:

- Information on patent publications (e.g., CPC technology class, patent title, and claims)
- Patent citations, patent family members, legal status events (e.g. patent grants, abandonments, and expirations)
- Patent applicant variables (e.g. applicant and inventor names, countries, corporations and institutions)
- Patent to papers link (frontier)

In what follows we will focus on green patents, i.e. CPC classes Y02 and Y04 belonging to the USPTO.

Methodology (Part I)

In order to measure the scientific-base of a patent we should first construct a distance metric. Our distance metric is build via a Breadth First Search (BFS) algorithm and is inspired by Ahmadpoor and Jones (2017). Overview in the following:

① **Given graph $G = (V, E)$:**

- BFS starts at source vertex $s \in V$
- Explores neighboring vertices at current depth level
- Uses queue Q for vertices to be visited
- Uses set S of visited vertices to avoid revisiting

③ **Termination:**

- BFS stops when all reachable vertices are visited or when Q becomes empty

② **Iteration Steps:**

- Initialize Q with s and S as an empty set
- Repeat until Q is empty:
  - i. Dequeue vertex v from Q
  - ii. Mark v as visited, add it to S
  - iii. Explore unvisited neighboring vertices of v
  - iv. Enqueue them in Q and mark their distance as $d(w) = d(v) + 1$

Distance from the frontier

Idea: a patent i with $D_i = n + 1$ is one that cites a patent j with $D_j = n$ and does not cite any patent k with $D_k < n$.
$D_i \in \{0, \infty\}$.

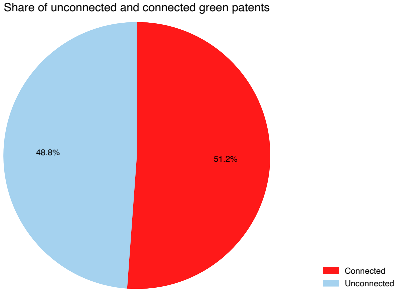


Figure 2: **Share of connected components.**

Average distance of 2 from the science frontier.

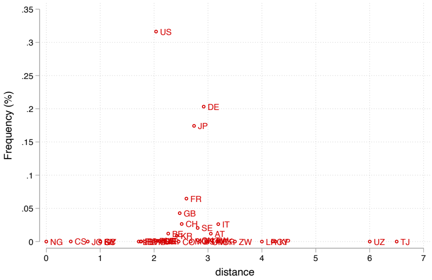
Descriptive Results on the determinants

Inventor's country

ic.i Presence of leaders: US, Japan; Germany ad Korea (confirmed in Corrocher et al. 2021);

ic.ii Presence of followers (relying on the green patents of these leaders?): Italy, Belgium, Austria (see also Lee et al. 2021);

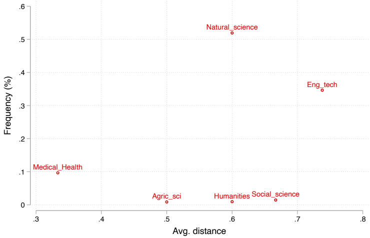
ic.iii Catching up of China

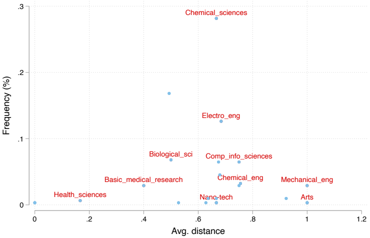


Field(s) Of Study (FOS)

fos.i The FOS (sub-)fields refer to the paper cited at the frontier by the patent;

fos.ii The results indicate that Engineering and Natural Sciences are the predominant fields of science-based green patenting. Specifically the more science-based sub-fields are: Chemical Sciences, Electronic Engineer and Materials Engineer.



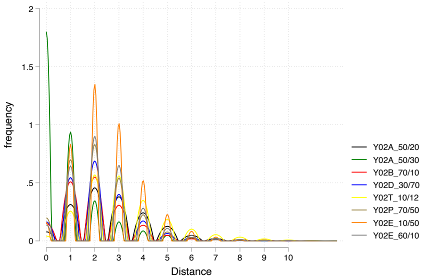


Technological classes

tc.i The figure reports the density of distances by CPC classes for a selected range of technologies;

tc.ii The central CPC classes in green patenting are Y02P, Y02T and Y02E which are, on average, closer to the frontier and have patents that are more connected to other green patents;

tc.iii Contrasting conclusions are reached in the literature (Barbieri et al. 2022, Higham et al. 2022).

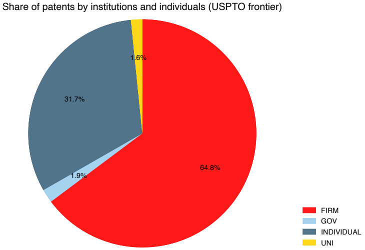


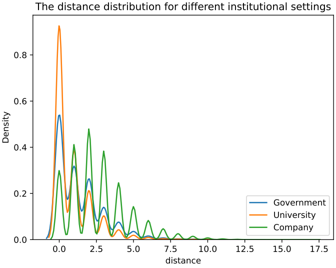
Institutions

ins.i Hospital, university, and university-affiliated company patents have more science-based patents compared to green patents from individuals and companies (in line with Popp, 2019);

ins.ii Firms hold the most green patents, followed by individuals, universities, and government institutions;

ins.iii Despite producing fewer green patents, universities generate more science-based patents, as expected. Firms adopt green technologies based on existing knowledge.





Assessing the determinants of science-based green patents via ML

Overview

Objective: Formally examine adequacy of identified characteristics for science-based green patents. To do so:

- Database structured for regularization techniques using machine learning;
- Selected features (X) including fixed and time effects, with dependent variable Y;
- Y assigned value of 1 if majority of green patents within DOCDB family have distance metric ≤ 2;
- Conducted comparative analysis of machine learning algorithms;
- **Random Forest** (RF) identified as the most effective predictor.

Methodology (Part II)

How to perform feature selection within RF? Through the Boruta algorithm which builds a measure of importance.

Boruta algorithm (informal)

- Duplicate the dataset and shuffle the values in each column to create a combined database of true features (TF) and shadow features (SF).
- Train a random forest (RF) model and compute the Mean Decrease Impurity (MDI) for each TF and SF.
- Compare the Z-scores of TF and SF for each iteration. If the Z-score of TF is higher than that of SF (i.e., TF provides a greater decrease in impurity), label TF as important for that iteration.
- If TF is determined to be important for the majority of iterations, consider it as relevant.

Variables ranked by the RF classifier based on: 1) **average importance**[1] and 2) **frequency** of variable selection across different models. (Top 10 variables shown for brevity)

Ranking	Feature	Number of models	Average Importance	I_p
1^{st}	% of green cited	5	21.39	106.95
2^{nd}	Number citing	5	20.54	102.70
3^{rd}	US	5	17.44	87.20
4^{th}	University	5	16.73	83.65
5^{th}	JP	5	14.92	74.60
6^{th}	Y02E 70	5	14.40	72.00
7^{th}	DE	5	13.10	65.50
8^{th}	GB	5	11.72	58.60
9^{th}	Y02E 10	4	13.26	53.04
10^{th}	Y02B 10	4	12.20	48.80

[1] based on decrease in impurity

Results

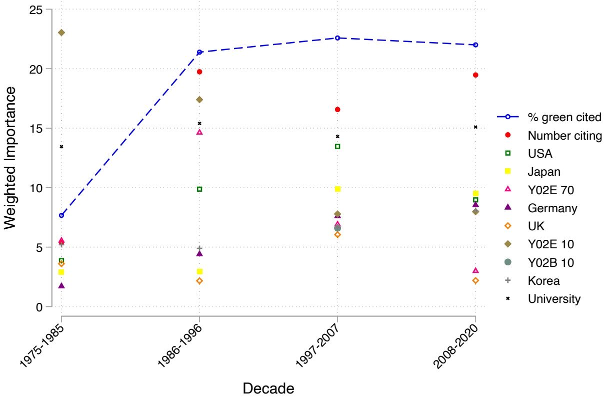


Figure 3: "Green citations" (blue) consistently exhibit higher importance compared to other features;

Discussion

1. The percentage of green patents cited matters in defining the science grounds of a patent;
2. Findings from Table above support this observation, even when considering the temporal dimension;
3. This observation highlights the importance of research comparing brown and green technologies;
4. Works by Heal (2007) and Skinner and Valentine (2023) emphasize the expanding literature on this topic;
5. Our results indicate that being classified as "green" does not guarantee scientific validity or citation centrality;
6. However, being part of the "green network" strongly correlates with scientific validity for patents.

Conclusions

Novel Definition and Feature Identification:

- First empirical attempt at defining science-based green patents;
- Identification of key features characterizing such patents;

Relevance for Policymakers and Industry Practitioners:

- Informing policy decisions related to green innovation;
- Providing insights for industry practitioners in patent examination and R&D;

Contribution to Green Innovation Literature:

- Advances understanding of science-based green patents;
- Supports the importance of pure green innovations (as opposed to brown) for sustainability and scientific validity.

THANK YOU FOR THE ATTENTION