Alternative methods to measure breakthrough innovations: the case of historical patents

Benjamin Cornejo Costas, [1[1]] Raoul Schram[2[1]], Shiva Nadi Najafabadi[3[1]], Marteen Schermer[4[1]], Jelle Treep[5[1]] and Andrea Morrison[6[1,2]]

[1] Utrecht University, Utrecht, The Netherlands
[2] University of Pavia, Pavia, Italy
b.j.cornejocostas@uu.nl

Abstract. Breakthrough innovations represent significant disruptions to existing technological trajectories. They can potentially lead to paradigm shifts and regional diversification through unrelated and disruptive changes in the technological portfolio of a region (Dosi, 1982). Therefore, breakthrough innovations are significant players in a place's structural change. Continuous or incremental innovations, at the same time, improve upon existing technologies. Breakthrough innovations are rare and difficult to assess, with most inventions being of moderate value.

Most scholars agree that for a technology to be a breakthrough, it should embody a novel combination of technologies that incorporate new knowledge from existing practices (Verhoeven, Bakker, & Veugelers, 2016). In addition, breakthrough innovation should contain an impactful combination of technologies with a particular influence on subsequent inventions.

Measuring breakthrough innovations is more straightforward in qualitative studies for small-scale disruptive events. Yet, for a large sample of innovation output, such as patents, the measurements available are the key challenge to identifying and finding those radical inventions (Capponi, Martinelli, & Nuvolari, 2022). Conventionally, they are often measured by the number of citations received (Squicciarini, Dernis, & Criscuolo, 2013). However, this measure needs to be more accurate as these metrics are inconsistent over time because of the disparities between patents themselves (Kelly, Papanikolaou, Seru, & Taddy, 2018).

For historical patents, we developed a new algorithm using NLP techniques to overcome the inconsistencies in the metrics of technological classes and the inexistence of old citations. In our research, we propose a novel methodology for assessing patent quality. For this, we followed the seminal work by Arts, Hou, & Gomez (2021) and

Kelly et al. (2018), but with significant differences. First, we adopt a distinctive approach by treating all words as independent vectors instead of relying solely on unigrams, bigrams or trigrams. Furthermore, we comprehensively explored various models and algorithms to identify the best-performing one, employing the technological classifications provided by the US Patent and Trademark Office (USPTO) as a point of reference.

This innovative approach extends beyond merely comparing our breakthrough innovation findings with historically acknowledged instances, such as the inventions of the telegraph, the telephone, or submarines. Instead, we embark on an ex-ante evaluation, leveraging technological categories as a benchmark, thereby enhancing the comprehensiveness of our assessment.

We tried four different algorithms for each set of patents (all the patents for all the years). We compared them to how close they were to the technological classes as per the register of the USPTO (also known as CPC Codes, standing for Cooperative Patent Classification from USPTO, the European Patent Office, EPO). In each case, we could observe how distant the algorithms were from the technological classes, and accordingly, we chose the one to find the dot product between the focal patent and the rest of each patent. These algorithms are TFIDF, Count2Vec, word2Vec and Bert.

We found that TFIDF was the best algorithm for these historical patents to predict the similarities of patents, using the CPC codes as the benchmark.

The results of our study yield valuable insights into the assessment of breakthrough innovations, a category of inventions that represents significant disruptions in technological trajectories, often leading to paradigm shifts and regional diversification.

We also ran a bunch of robustness checks, considering different exponents for each algorithm (to penalise and account for the repetition of words), and we also ran more stringent conditions to set a patent to be novel or impactful and, therefore, breakthrough.

Our measurement of breakthrough innovations for historical patents is not the first of its kind. Still, the first one considers the importance of a patent compared to previous patents (originality) and its importance in future inventions (impact). It is also an open-access analysis available to any researcher who wants to use the algorithm to analyse

the relevance of patents no matter the year of invention, as our indicator also works with new patents.

Keywords: Breakthrough Innovations, NLP techniques, Patent quality.

1. Arts, S., Hou, J., & Gomez, J. C. (2021). Natural language processing to identify the creation and impact of new technologies in patent text: Code, data, and new measures. *Research Policy*, *50*(2), 104144. https://doi.org/10.1016/j.respol.2020.104144
2. Capponi, G., Martinelli, A., & Nuvolari, A. (2022). Breakthrough innovations and where to find them. *Research Policy*, *51*(1), 104376. https://doi.org/10.1016/j.respol.2021.104376
3. Dosi, G. (1982). Technological Paradigms and Technological Trajectories. *Innovation, Organization and Economic Dynamics*, (11), 147–162. https://doi.org/10.4337/9781782541851.00007
4. Kelly, B. T., Papanikolaou, D., Seru, A., & Taddy, M. (2018). Measuring Technological Innovation over the Long Run. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3279254
5. Kerr, W. R. (2010). Breakthrough inventions and migrating clusters of innovation. *Journal of Urban Economics*, *67*(1), 46–60. https://doi.org/10.1016/j.jue.2009.09.006
6. Squicciarini, M., Dernis, H., & Criscuolo, C. (2013). OECD Science, Technology and Industry Working Papers 2013/03 Measuring Patent Quality: Indicators of Technological and Economic Value. *OECD Science, Technology and Industry Working Papers*. Retrieved from https://dx.doi.org/10.1787/5k4522wkw1r8-en
7. Verhoeven, D., Bakker, J., & Veugelers, R. (2016). Measuring technological novelty with patent-based indicators. *Research Policy*, *45*(3), 707–723. https://doi.org/10.1016/j.respol.2015.11.010