

An Hypernetwork Approach to Accurately Measure Technological Innovation

Methods and Results

J. RAIMBAULT^{1,2} and A. BERGEAUD³

¹ UMR CNRS 8504 Géographie-cités

² UMR-T IFSTTAR 9403 LVMT

³ London School of Economics

Abstract

This working paper details technical methods used in the project and gives a first overview of results.

1 Introduction

See project Proposal

2 Methods

Originality Measures The originality measure is defined by Hall *et al.* (2001) [Hall et al., 2001] as

$$O_i = 1 - \sum_{j=1}^{n_i} c_{i,j}^2$$

where $c_{i,j}$ is the percentage of citations made by patent i to a patent in class j out of n_i technological classes to which patent i belongs. If the scope of technologies which the patent uses and cites is large, then the originality measure will be high. Radicalness is more difficult to define. It is constructed in the same way as the originality index but here we only consider the technology classes of patents cited by patent i but to which patent i does not belong. These two indicators are good proxies and great start to estimate if a patent is protecting a new product that can hardly be classified into the official technological field space.

Citation Network We define a binary relationship between each pair of patents $Cit(i, j) = 1$ if j cites i or i cites j , otherwise $Cit(i, j) = 0$.

Technological Class Network For each patent i , let B_i be the set of technological class of i . We then define a relationship between each pair i and j as 2 times the number of technological class in common divided by the total number of class of i and j .

$$Class(i, j) = 2 \frac{|B_i \cap B_j|}{|B_i| + |B_j|}$$

Thus, if two patents have no class in common, $Class(i, j) = 0$ while if the two patents are exactly identical in terms of their sets of technological class $Class(i, j) = 1$.

Semantic Network We first assign to a patent p a set of significant keywords $K(p) \in \bigcup_{n \in \mathbb{N}} \mathcal{A}^{*n}$, that are precisely extracted following a similar procedure to the one detailed in [Chavalarias and Cointet, 2013] :

- Text parsing and tokenizing.
- Part-of-speech tagging, normalization.
- Stem extraction and multi-stems constructions.
- Relevant multi-stems filtering.

Text processing operations will be implemented in `python` in order to use the `nltk` library [] which is highly ergonomic and supports most advanced state-of-the-art natural language processing operations.

Possible Features

Multilayer Network Analysis Check out the method proposed in [Iacovacci et al., 2015]. → Interesting for some kind of “between-layers correlation” ?

3 Results

References

- [Chavalarias and Cointet, 2013] Chavalarias, D. and Cointet, J.-P. (2013). Phylomemetic patterns in science evolution—the rise and fall of scientific fields. *Plos One*, 8(2):e54847.
- [Hall et al., 2001] Hall, B., Jaffe, A., and Trajtenberg, M. (2001). The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools. Papers 2001-29, Tel Aviv.
- [Iacovacci et al., 2015] Iacovacci, J., Wu, Z., and Bianconi, G. (2015). Mesoscopic structures reveal the network between the layers of multiplex datasets. *arXiv preprint arXiv:1505.03824*.