Investigating Patterns of Technological Innovation

Communication Proposal, CCS 2016

Antonin Bergeaud[1], Yoann Potiron[2] and Juste Raimbault[3,4]

[1] Department of Economics, London School of Economics
[2] Faculty of Business and Commerce, Keio University
[3] UMR CNRS 8504 Géographie-cités
[4] UMR-T IFSTTAR 9403 LVMT

Keywords : *Technological Innovation, Text-mining, Multilayer Network*

The understanding of patterns of technological innovation is crucial for both economical growth theories and practical applications in research and development. Yet a precise characterization of breakthrough inventions has not been fully investigated. We propose to answer this issue through a large-scale data-mining and network approach on patent data.

We construct from US Patent Office raw data an open consolidated database that includes detailed patent information, technological classifications, citation links, and abstract texts, yielding a database of around $4 \cdot 10^6$ patents on a time range from 1976 to 2012. Aiming at capturing the semantic information contained in texts, that has been shown to be complementary to classification data, we extract relevant n-gram keywords and obtain for each year a semantic network based on co-occurrences. The multi-objective optimization of network modularity and size is performed on network construction parameters (filtering thresholds) through high performance computing. We obtain for each year a multi-layer network, containing semantic community relations, technological classes relations and citation relations between patents. The mining of network layers yields interesting results, such as an increase in time of patent semantic originality combined with a counter-intuitive loss of class-level interdisciplinarity, what confirms the stylized facts of both invention refinement and specialization in time. Citation-level interdisciplinarity is investigated by combining the different layers. Further work is planned towards the use of these heterogeneous features produced by multi-layer network analysis into machine learning models to predict success and breakthrough level of inventions.

Our contribution to the study of socio-technical complex systems is on the one hand thematic, with the construction of an open-access large scale consolidated patent database and insights into the temporal evolution of inventions, and on the other hand methodological with a technique that can be generated to any network whose nodes contain a textual description.