Mining Patents Data through Semantic Network Analysis

Project Proposal

J. Raimbault

September 19th

Abstract

Patents are a central proxy in the study of the economy of innovation. Indeed, the information contained in relations between patents reflects the underlying structure of the socio-technological system of research and development in innovative companies. Whereas recent focus was mainly on the study of patterns in the inter-patent citation network, taking technological fields as externally fixed, we propose a novel approach based on semantic analysis of patents textual contents. Indeed, measures such as co-occurences and repetitions of keywords should contain a slightly different information than the one extracted from the citation network, as for example links between domains, or informal domains that could appear as communities in the semantic network. The aim of this project is to investigate the nature and extent of this information, by the mining of regular patterns in features that should be established. We expect to test various features crossing measures extracted from both dynamic citation and semantic network, by unsupervised, and supervised if needed, datamining techniques. Expected results are to unveil the potentialities of such approach, and in the case of significant information, to apply it to classify patents and identify innovatives

1 Introduction

The study of innovation through the lens of technological patents is not a novel idea []but the recent rise of new methods and computational abilities, including datamining and network analysis []has shed a new light on the approach. With methods relatively close to applied epistemology studies such as science mapping [] or autorships networks analysis [] , recent works have studied patents citation network to understand the processes of technological innovation.

2 Research objectives

The identification of so-called *emerging research fronts* was done in the case of scientific publications in [Shibata et al., 2008]. In the same spirit, our guiding research objective is to identify such fronts for patents, i.e. to be able to classify patents in order to distinguish the "real innovations" from

3 Proposed approach

3.1 General description

3.2 Technical details

4 Project organisation

References

[Shibata et al., 2008] Shibata, N., Kajikawa, Y., Takeda, Y., and Matsushima, K. (2008). Detecting emerging research fronts based on topological measures in citation networks of scientific publications. *Technovation*, 28(11):758–775.