Data MASTER Research Proposal: The Effects of Patent Filing Acceleration on the Evolution of Technological Innovation

Ryan Steed
The George Washington University

April 11, 2018

Abstract

The 2011 America Invents Act (AIA) disrupted the United States patenting process, replacing first-to-invent policy with the first-to-file system popular in Europe and sparking a furious debate over the impact of the policy on innovation and technological growth. The institution of patent protection, an integral part of intellectual property law and economic competition, is fundamentally changed by its passage, but many of its effects are unknown and unexplored. How does accelerating the process of patent filing affect the evolution of technological innovation?

To investigate these effects, I intend to 1) identify key properties of the patent citation network that are affected by the AIA, 2) employ a time series model to forecast network evolution for several specific patent classes, and 3) compare the rate of evolution before and after the implementation of the AIA.

The topological structure of the citation network alone may provide useful insight into the effects of the AIA on patent filing behavior. I hypothesize that many new prolific firms and inventors (as measured by H-indices) will become prominent in a denser citation network while already-prolific firms will not change output significantly, resulting in a more modular network. Beyond static topological analysis, I anticipate that the AIA has a catalyzing effect on network evolution, increasing rates of technology proliferation and mimicry in the citation network.

Project Description

Problem Statement

United States patent filings, often the first public release of information about cutting-edge industry technology, comprise an ever-growing complex network of documents linked by citations and rigid U.S. Patent & Trademark Office (USPTO) classifications. Unlike text-based networks, citation network edges are consistently defined and such a network is generally regarded as useful structural representations of patent documentation, especially when combined with USPTO classifications to form a multipartite network. Open source patent data can be used to construct a direct citation network that depicts a complex landscape across industries useful for a variety of technology, inventor, and firm analysis (Henrique et al., 2018). Oh et al. (2016) establish base network statistics for measuring technological innovation over time, including degree centrality and the Hirsch index (H-index), a measure of individual or firm productivity used in social and financial networks. More complex models attempt leverage citation network structure for forecasting. Chang et al. (2015) use the H-index as an indicator of the correlation between patent citation count and market value for individual firms, linking moderately productive firms to high market value. Meanwhile, Daim et al. (2006) construct a system dynamics model that uses bibliometric techniques to identify emerging technologies in three separate industries. Other studies have attempted to link patent data as such to market value, breakthrough technologies, and even GDP (Karanikic et al., 2017).

The size and complexity of the patent citation network make it ideal for network evolution analysis, from large-network generalizations to case-specific claims (Leskovec et al., 2007). As of 1999, the direct citation network for publicly released patent fillings from the USPTO database contained nearly 4 million nodes and over 16.5 million edges and is nearly connected (Leskovec et al., 2007). Martinelli and Nomaler (2014) cast concepts of genetic inheritance to patent relations, identifying interruptions in patent development over time by breaks in the "lineage" of a particular subfield. Most pertinent is the work of You et al. (2017), who apply two time series models (Bass and ARIMA) to a single USPTO patent subclass to forecast a customized indicator of "development potential."

However, little investigation has been conducted into the specific policy problem at hand. Miyagiwa (2015) conducts a simple patent race model to determine the effects of the AIA on innovation, determining by qualitative criteria that first-to-file is a slightly worse catalyst for innovation. Pierce (2012), who considers it "the most substantial change to American patent law since the Patent Act of 1952," holds that legal stipulations in the Act will inhibit collaborative research (perhaps resulting in a citation network with higher modularity). Research on the network effects of the AIA is lacking, and nearly no temporal analyses of the policy change exist. To understand the progress of innovation in competitive markets, it is essential to examine the legal institutions governing intellectual property; those legal institutions dictate the private research and development practices of technology firms. Further, if patent filing has an effect on the evolution of invention, patent policy plays an extremely important role in technological evolution. Fortunately, public patent data provide the means to track this progression of knowledge and its interaction with government policy.

Research Plan

To address these questions, I will explore the topological structure of a few contained patent classes seeking statistical differences in network properties before and after the implementation of the AIA. I will also examine the use of forecasting models for predicting trends in development indicators and attempt to determine whether forecasting is improved by the policy change. I propose the following method to achieve these specific aims:

1. Parse public patent data from the USPTO, 1976-2018, into an direct citation network with a patent classification layer for industry-specific analysis with a custom script in Python or another high-level programming language (Marco et al., 2015; You et al., 2017).

2. Analyze topological differences between pre-AIA and post-AIA using established or designed metrics for technological innovation and development.

3. Build and execute two separate time series models (see You et al. (2017)) to forecast trends in patent subclasses and citation clustering based on individual node attributes and statistics in the pre-AIA and post-AIA citation networks, cross-validating with a separate data set within each time series.

4. Attempt to establish subject matter interpretations of the topological differences between the two data sets, and draw conclusions about the change in the rate of technological evolution as a result of the AIA.

Since the data collection segment of this research is largely a parsing problem, the most crucial portion of this project is the definition of a development metric based on topological structure and node statistics. You et al. (2017) cite useful sources for development measures used in their time series forecasting which may be adapted to the problem at hand. It will be necessary to seek industry and economics expert analysis to suggest potential impacts of the AIA on innovation practice at firms and in the market at-large which will be used to evaluate identified topological trends.

Outcomes

I expect to produce an informed qualitative interpretation of changes in citation network topology (supported by statistical analysis when necessary) and a time series model that aggregates relevant network properties to describe change in innovation based on previous literature. A data-driven evaluation of the effects of the AIA is invaluable and timely, as firms have adjusted their research and development practices and new norms of patent filing procedure are in place. Not only do I hope to illuminate the effect of an accelerating policy on a large-scale, generative network, I also hope to provide a more generalized metric for the evolution of technology.

These conclusions are useful to economics and patent law researchers, network mathematicians, and industry professionals seeking to understand and predict the process and vulnerabilities of technological development.

References

Chang, K., W. Zhou, S. Zhang, and C. Yuan (2015). Threshold effects of the patent h-index in the relationship between patent citations and market value. *Journal of the Association for Information Science and Technology 66*(12), 2697–2703.

Daim, T. U., G. Rueda, H. Martin, and P. Gerdsri (2006). Forecasting emerging technologies: Use of bibliometrics and patent analysis. *Technological Forecasting and Social Change 73*(8), 981–1012.

Henrique, B. M., V. A. Sobreiro, and H. Kimura (2018). Building direct citation networks. *Scientometrics 115*(2), 817–832.

Karanikic, P., I. Mladenovic, S. Sokolov-Mladenovic, and M. Alizamir (2017). Prediction of economic growth by extreme learning approach based on science and technology transfer. *Quality & Quantity 51*(3), 1395–1401.

Leskovec, J., J. Kleinberg, and C. Faloutsos (2007). Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data (TKDD) 1*(1), 2–es.

Marco, A. C., A. F. Myers, S. Graham, P. D'Agostino, and K. Apple (2015). The USPTO Patent Assignment Dataset: Descriptions and Analysis. *USPTO Economic Working Paper* (2015-2).

Martinelli, A. and Z. Nomaler (2014). Measuring knowledge persistence : a genetic approach to patent citation networks. *Journal of Evolutionary Economics 24*(3), 623–652.

Miyagiwa, K. (2015). The 2011 america invents act: Does it undermine innovation? *Journal of Economics Management Strategy 24*(2), 211–227.

Oh, G., H. Kim, and A. Park (2017;2016;). Analysis of technological innovation based on citation information. *Quality & Quantity 51*(3), 1065–1091.

Pierce, N. S. (2012). The effect of the leahy-smith america invents act on collaborative research. *Journal of the Patent and Trademark Office Society 94*(2), 133.

You, H., M. Li, K. W. Hipel, J. Jiang, B. Ge, and H. Duan (2017). Development trend forecasting for coherent light generator technology based on patent citation network analysis. *Scientometrics 111*(1), 297–315.