Available online at www.sciencedirect.com
ScienceDirect
Physica A 374 (2007) 783–793
PHYSICA A
www.elsevier.com/locate/physa

# Modeling innovation by a kinetic description of the patent citation system

Gábor Csárdi[a,b,*], Katherine J. Strandburg[c], László Zalányi[a,b], Jan Tobochnik[d], Péter Érdi[a,b]

[a]*Department of Biophysics, KFKI Research Institute for Particle and Nuclear Physics of the Hungarian Academy of Sciences, 29-33 Konkoly-Thege Road, Budapest, H-1121, Hungary*
[b]*Center for Complex Systems Studies, Kalamazoo College, 1200 Academy Street, Kalamazoo, MI 49006, USA*
[c]*DePaul University College of Law, 25 East Jackson Boulevard, Chicago, IL 60604, USA*
[d]*Department of Physics and Center for Complex Systems Studies, Kalamazoo College, Kalamazoo, MI 49006, USA*

Received 4 April 2006; received in revised form 9 August 2006
Available online 5 September 2006

**Abstract**

This paper reports results of a network theory approach to the study of the United States patent system. We model the patent citation network as a discrete time, discrete space stochastic dynamic system. From patent data we extract an attractiveness function, $A(k,l)$, which determines the likelihood that a patent will be cited. $A(k,l)$ shows power law aging and preferential attachment. The exponent of the latter is increasing since 1993, suggesting that patent citations are increasingly concentrated on a relatively small number of patents. In particular, our results appear consistent with an increasing patent ''thicket'', in which more and more patents are issued on minor technical advances.


*Keywords:* Innovation; Patents; Citation network; Preferential attachment; Aging

## 1. Introduction

Innovation plays a key role in economic development and the patent system is intended (and required by the United States Constitution) to promote innovation. The patent system promotes innovation by giving inventors the power to exclude others from using their inventions during the patent term. The power to exclude is a double-edged sword, however, because it benefits the original inventor, but imposes costs on later innovators seeking to build on past inventions. Thus, the proper design of the patent system is an important matter—and a matter of considerable current debate. See, e.g., Refs. [1–3]. Advances in computer technology and the availability of large patent databases have recently made it possible to study aspects of the patent system quantitatively. To date the empirical analysis of the patent system has been undertaken primarily by economists, sociologists, and legal scholars. See, e.g., Refs. [4–8]. Because patents and the citations between

*Corresponding author.
*E-mail address:* csardi@rmki.kfki.hu (G. Csárdi).

0378-4371/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2006.08.022

them can be conceptualized as a growing network, techniques from statistical physics that have been used in the study of complex networks can be usefully applied to the patent citation network [9,10]. In this paper we present what we believe to be the first results of such a network theory approach to the patent system. We explore the kinetics of patent citation network growth and discuss some possible implications for understanding the patent system.

The paper is organized as follows: In Section 2 we provide background on the United States patent system and describe the citation data that is used in this study. In Section 3 we describe a general framework for modeling the kinetics of citation networks, define an "attractiveness function" for the evolving network and introduce a method for extracting this function from the history of the network. In Section 4 we use this approach to analyze the US patent citation network and explore the changes in the kinetics from 1976 to 2000. In Section 5 we discuss some possible implications of our results, and mention directions for future research.

## 2. Patentological background

While a similar approach could be applied to many patent systems, including the very important European and Japanese patent systems, we begin our analysis with the United States patent system for which an extensive database of citations has been made available through the work of economists Hall et al. [11].

An application for a US Patent is filed in the US Patent and Trademark Office (USPTO). A patent examiner at the USPTO determines whether to grant a patent based on a number of criteria, most important of which for present purposes are the requirements of novelty and non-obviousness with respect to existing technology. Once a patent is issued by the USPTO, it is assigned a unique patent identification number. These numbers are sequential in the order in which the patents were granted.

Novelty and non-obviousness are evaluated by comparing the claimed invention to statutorily defined categories of "prior art", consisting in most cases primarily of prior patents. Patents are legally effective only for a limited term (currently 20 years from the date of application), but remain effective as "prior art" indefinitely. Inventors are required to provide citations to known references that are "material" to patentability, but are not required to search for relevant references (though they or their patent attorneys often do so). During consideration of the application, patent examiners search for additional relevant references.

Patent citations include potential prior art that was considered by the examiner. They thus reflect the judgment of patentees, their attorneys, and the USPTO patent examiners as to the prior patents that are most closely related to the invention claimed in an application. Patent citations thus provide, to some approximation, a "map" of the technical relationships between patents in the US patent system. This "map" can be represented by a directed network, where the nodes are the patents and the directed edges the citations. Our research uses a statistical physics approach inspired by studies of other complex networks to attempt to gain insight from this "map".

The patent database we use for the analysis in this paper was created by Hall, Jaffe and Trajtenberg based on data available from the US Patent Office [11]. It is available online at http://www.nber.org/patents/. The database contains data from over 6 million patents granted between July 13, 1836 and December 31, 1999 but only reflects the citations made by patents after January 1, 1975: more than 2 million patents and over 16 million citations. Citations made by earlier patents are also available from the Patent Office, but not in an electronic format. The Hall, Jaffe and Trajtenberg database also contains additional data about the included patents, which is described in detail in Ref. [11].

## 3. Modeling patent citation networks

### 3.1. Defining the model framework

In this section we define the mathematical model framework we will use for studying patent citations. The raw citation data gives us a complete history of citations made and received by each patent. Our goal is to determine whether the evolution of the patent network may be consistently described in terms of variables commonly used in understanding the evolution of complex networks and then to extract the time dependence of the network growth from the detailed history. We assume as an initial matter (an assumption which turns