The dynamics of innovations and citations

Christian Ghiglino [a,b,*], Nicole Tabasso [c]

[a] *University of Essex, Department of Economics, Colchester, CO4 3SQ, UK*
[b] *GSEM, Geneva, Switzerland*
[c] *University of Surrey, School of Economics, Guildford, GU2 7XH, UK*

HIGHLIGHTS

- Patent citations occur as new ideas are produced by combining existing ideas.
- Ideas are intervals in a variety space.
- Interval lengths determine the likelihood of citation.
- The model derives exponential aging of patents, which fits the data very well.
- Endogenous aging sets the model apart from preferential attachment models.

ARTICLE INFO

ABSTRACT

We present a model in which patent citations occur as new ideas are produced from combinations of existing ideas. An idea's usability in this process is represented as an interval in a variety space of ideas, whose length determines the likelihood of citation. This process endogenously derives exponential aging of patents, which is consistent with empirical observations. The endogeneity of aging sets our process apart from the standard preferential attachment literature.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Citation patterns between academic papers or patents have received considerable attention. They have been analyzed predominantly with network models based on preferential attachment or intrinsic fitness in which papers/patents are nodes, and citations directed links between them (Atalay, 2013; Barabási et al., 1999; Jackson and Rogers, 2007; Peterson et al., 2010; Valverde et al., 2007). In these models the probability to link to an existing node is proportional to a scalar quantity, e.g., intrinsic fitness or the number of past citations. However, without the introduction of a specific aging function they are not able to match the observed aging of patents, see Fig. 1 or Marco (2007). In particular, with pure preferential attachment citation rates are only negatively affected by the total network size which provides a weak form of aging.

We rather leave preferential attachment and model the attachment process between patents as random, guided by the applicability of patents, representing intrinsic heterogeneity. We model applicability in a way that leads to aging very naturally through a behavioral choice of innovators. In our model, patents represent ideas, which are built from combination of older ideas as in Auerswald et al. (2000); Ghiglino (2012), or Weitzman (1998). Innovations arrive to innovators, who decide which ideas to combine to realize the innovation. This choice is largely driven by technological identity of ideas. In particular, we model a patent as an interval in the variety space of ideas, which represents its applicability

* Corresponding author at: University of Essex, Department of Economics, Colchester, CO4 3SQ, UK.
E-mail addresses: cghig@essex.ac.uk (C. Ghiglino), n.tabasso@surrey.ac.uk (N. Tabasso).
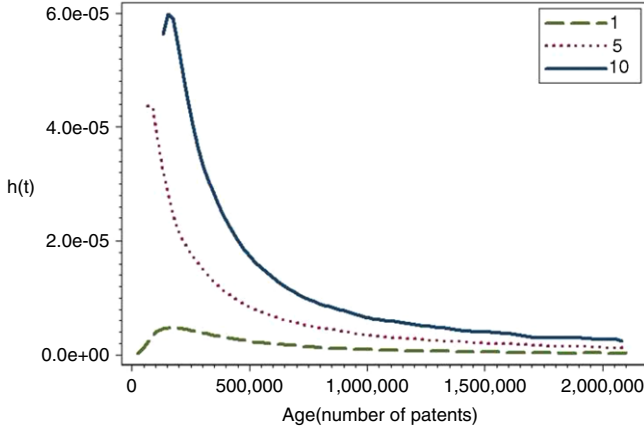
Fig. 1. Citation dynamics. Non-parametrically estimated population hazard rates of being cited for USPTO patents that have received 1, 5, and 10 citations respectively.

range. More broadly applicable patents thus are more likely to be cited. The choice of the innovator that leads to aging is simple: If there exist multiple ideas that may be used as inputs for his innovation, he chooses the youngest. Such behavior might be justified if innovators do not know perfectly which input idea is best for them to use.

2. The model

We model patents/ideas as nodes in a network, and citations among them as directed links. A link from node i to j thus signifies that j is an input idea to i. Ideas are of different varieties, and the support of the variety space is a circle of circumference 1. At time t there are $N(t)$ nodes. Time is continuous and new nodes arrive sequentially, as a Poisson process with arrival rate of 1. Each node i is characterized by an interval $I_i \subset (0, 1]$, a set of m scalars, $\mu_k^i \in (0, 1]$ with $k = 1, \ldots, m$, and its birth date t_i. We assume that each μ_k^i is extracted from a uniform distribution on $(0, 1]$, and I_i is extracted from a distribution Ψ such that the position (middle point) of the (connected) interval is extracted from a uniform distribution on $(0, 1]$, and $|I_i| \in (0, \frac{1}{m})$.

A necessary condition for the existence of a link from i to j, is that for at least one $k = 1, \ldots, m$, $\mu_k^i \in I_j$ and $t_j \leq t_i$, in which case j is a feasible input for i. However, there might be several nodes that satisfy this condition. Let the set of these nodes be $\hat{I}_k$. A sufficient and necessary condition for a link from i to j is that $j \in \hat{I}_k$ and $t_j > t_{j'}$ for all $j' \in \hat{I}_k$. As nodes are added sequentially there is, at most, only one such node. The attachment process is illustrated in Fig. 2 for $m = 3$.

We proceed to obtain the probability for a node to receive a link. First, note that a given node j has m chances to receive an additional edge from a newly entered node i. For each of these, as μ_k^i is uniformly extracted from $(0, 1]$, the probability that j fulfills the necessary condition that $j \in \hat{I}_k$ is equal to

$$Pr(\mu_k^i \in I_j) = |I_j|. \tag{1}$$

In turn, the probability that $t_j > t_{j'}$ for all $j' \in \hat{I}_k$, is the probability that between t_j and t, no other node j' has entered the system for which $\mu_k^i \in I_{j'}$. Given the Poisson arrival process of ideas, the mean arrival rate of such nodes j' is given by the average interval length, denoted $\bar{I}$,

$$Pr(t_j > t_{j'} | j, j' \in \hat{I}_k) = e^{-\bar{I}(t-t_j)}. \tag{2}$$

Let $k_j(t)$ be the number of edges that node j has received up to t (its *in-degree*). The expected change in $k_j(t)$ is computed assuming
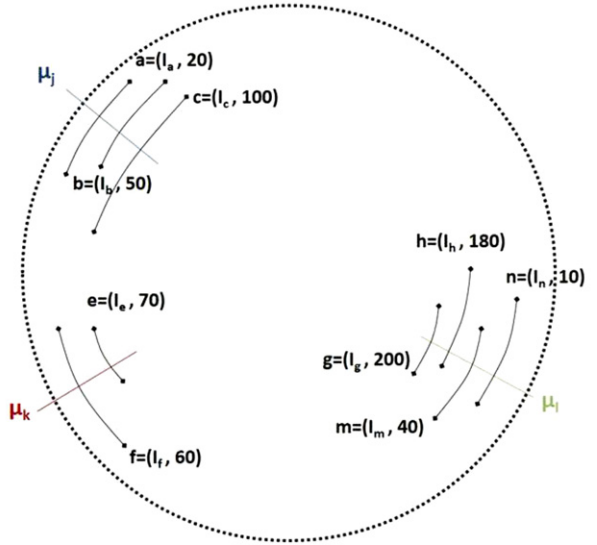

Fig. 2. Attachment process. Node i enters the system, with $\mu^i = \{\mu_j, \mu_k, \mu_l\}$. Nodes c, e, and g will receive links.

that $k_j(t)$ is continuous and that the mean-field approximation holds (see, e.g., Barabási et al. (1999) or Jackson and Rogers (2007)), which allows us to denote the expected rate of change as the actual one. Consequently, the probability that node j's in-degree will increase by at least one at t, $\Pi(I_j, t_j, t)$ (the *hazard rate* of node j),[1] can be expressed as the continuous rate of change of $k_j(t)$:

$$\Pi(I_j, t_j, t) = \frac{\partial k_j(t)}{\partial t} = m \cdot |I_j| \cdot e^{-\bar{I}(t-t_j)}. \tag{3}$$

The predictions of (3) are in line with stylized facts of patent citations: citation rates vary across patents and older patents are less likely to be cited. In contrast to preferential attachment models the likelihood to receive an edge depends on the *age* of a node, rather than on time t itself, and is independent of the current number of edges. Note that the model does not attempt to describe the increase in the likelihood of receiving an edge that is observed early in the life of patents. Instead, it predicts exponential aging of patents, which we now test against available patent citation data.

3. Results

Our dataset consists of a random sample of $N = 214,071$ patents granted by the United States Patent and Trademark Office (USPTO) between 1975 and 1999, made available by the National Bureau of Economic Research. We observe each patent whenever it gets cited and at the end of the 25 year period, which provides a total of $n = 1,059,475$ observations. We measure time as the number of patents granted, i.e., it coincides with the number of patents in the system.

We test our model by comparing the predicted hazard rates from (3) against the empirically estimated citation rates. To do so, we first integrate (3) subject to $k_j(t_j) = 0$, which yields

$$k_j(t) = m \frac{|I_j|}{\bar{I}} \left(1 - e^{-\bar{I}(t-t_j)}\right). \tag{4}$$

This allows us to calculate $|I_j|$ as

$$|I_j| = k_j(t) \cdot \bar{I} \cdot \frac{1}{m\left(1 - e^{-\bar{I}(t-t_j)}\right)}. \tag{5}$$

---

[1] While theoretically, node j can be cited multiple times by i, we find that the distribution of $|I_j|$ makes this probability negligible.