Technological Forecasting & Social Change 201 (2024) 123208

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

journal homepage: www.elsevier.com/locate/techfore

Citations or dollars? Early signals of a firm's research success ☆

Shuqi Xu [a], Manuel Sebastian Mariani [b,c,*], Linyuan Lü [d,**], Lorenzo Napolitano [e], Emanuele Pugliese [f], Andrea Zaccaria [g]

[a] Institute of Dataspace, Hefei Comprehensive National Science Center, Hefei, 230088, PR China
[b] Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu, 611731, PR China
[c] URPP Social Network, University of Zurich, Zurich, CH-8050, Switzerland
[d] School of Cyber Science and Technology, University of Science and Technology Of China, Hefei, 230026, PR China
[e] European Commission, Joint Research Centre (JRC), Seville, 41092, Spain
[f] United Nations University - MERIT, AX Maastricht, 6211, Netherlands
[g] Istituto dei Sistemi Complessi (ISC-CNR), UoS Sapienza, Rome, I-00185, Italy

ARTICLE INFO

Keywords:
Data-driven forecasting
Patent analysis
Technological innovation
Research success

ABSTRACT

Scientific and technological progress is largely driven by firms in many domains, including artificial intelligence and vaccine development. The early identification of the future performance of innovation players is a relevant goal for policymakers and practitioners. In this work, we investigate how the future trajectory of a firm can be predicted by the economic or technological value of its early patents. By inspecting the patenting life cycles of 7440 publicly listed firms, we find that the economic value of a firm's early patents is an accurate predictor of various dimensions of a firm's future research success. At the same time, a smaller set of future top-performers do not generate early patents of high economic value, but they are detectable via the technological value of their early patents. Importantly, the observed heterogeneity of the firms' temporal success patterns markedly differs from the patterns previously observed for individuals' research careers.

1. Introduction

Understanding the technological impact of an invention or of an economic actor actively engaging in innovation is a burning topic in economics and management due to its implications for various sectors and decision-making processes. When looking at the impact of past innovations, patents are one of the best-known data sources, and the citations they have received are a particularly effective indicator in capturing the relevance of a single invention because they can be counted by patent document, thus allowing a straightforward comparison. This makes citations particularly useful to estimate how valuable a patent is commercially or establish how successful it is from a research and development standpoint. However, it takes time for a patent to accrue citations. As effective as citations are as *ex-post* measures of inventive value, researchers must look elsewhere when they aim to predict in advance where the next breakthrough invention will come from. In the absence of reliable invention-level predictors, firm-level characteristics are natural candidates for this task. However, which firm-level indicators best forecast important innovators remains largely unexplored.

This work ventures in this direction by focusing on firm-level indicators that predict the technological and market success of R&D investing firms, of the technologies they develop, and the technological trajectories that are fueled as a consequence. Identifying a class of effective indicators in this sense has the potential to improve investors' and policymakers' decision-making as well as economic research by advancing our ability to better identify in advance future key players and, through them, future avenues of technological development (Yang et al., 2023; Kim et al., 2023). To this aim, we build and compare different classes of firm-level indicators based on their predictive ability in two respects.

☆ **Funding**: This work is supported by the National Natural Science Foundation of China (Grant No. 62306191, T2293771), STI 2030—Major Projects (Grant No. 2022ZD0211400), and the New Cornerstone Science Foundation through the XPLORER PRIZE. MSM acknowledges financial support from the Swiss National Science Foundation (Grant No. 100013-207888) and the URPP Social Networks at the University of Zurich. AZ acknowledges the European Union - Next Generation EU Italian PRIN project (Grant No. 20223W2JKJ "WECARE").
Disclaimer by LN: The views expressed in this paper belong solely to the authors and do not necessarily reflect an official position of the European Commission.
* Corresponding author at: URPP Social Network, University of Zurich, Zurich, CH-8050, Switzerland.
** Corresponding author at: School of Cyber Science and Technology, University of Science and Technology Of China, Hefei, 230026, PR China.
E-mail addresses: manuel.mariani@business.uzh.ch (M.S. Mariani), linyuan.lv@ustc.edu.cn (L. Lü).

https://doi.org/10.1016/j.techfore.2024.123208
Received 10 June 2022; Received in revised form 29 December 2023; Accepted 2 January 2024
Available online 19 January 2024

The first aspect concerns the predictability of a firm's research success, which we define in terms of the technological and economic value of its patents. This is inspired by the technological forecasting literature as well as the economics and management literature. In the former stream of literature, several works have looked for early signals of significant technologies and patents, which would be of interest to companies as well as investors (Altuntas et al., 2015; Mariani et al., 2019; Chung and Sohn, 2020; Jiang et al., 2023a; Yang et al., 2023). At the same time, the latter stream has shown that, at the firm level, the quantity and quality of a firm's patents correlate with several measures of future performance (Ernst, 2001; Andries and Faems, 2013; Wu et al., 2022; Lu and Wang, 2024). Our empirical analysis examines whether citation-based or economic indicators of the value of firms' early patents are more predictive of firms' future research success.

The second aspect concerns the timing of a firm's most valuable patents. This is inspired by recent cross-disciplinary studies that generated insights on the dynamics of research success of scientists and their scientific works (Sinatra et al., 2016; Liu et al., 2018; Wang et al., 2019; Yin et al., 2019; Wang and Barabási, 2021). In addition to the common use of bibliometric indicators (e.g. citations) to measure success in scientific research and technological R&D, there are also common patterns across the two fields in how novelty emerges and how the trajectories of the actors responsible for it evolve. However, if the analogy is also extended to the timing of creative success over the life cycle of scientific and economic actors, one would expect a firm's most valuable inventions to appear at random in its sequence of patents (Simonton, 1997), while the management literature exploring the relation between firm age and research performance is not conclusive on this (see Literature review section below). Our empirical analysis seeks to uncover whether early or late research hits are more likely for the firms, or whether research hits occur with the same likelihood at any life cycle stage, similar to creative careers.

It can be argued that to have sufficient data to construct an actor-level bibliometric analysis, one needs to focus on actors that have recorded a large enough production. In the present case, it implies focusing on firms with an established patent portfolio, which are typically large and mature companies. Though this could seem to be a limitation in the applicability of the analysis, it is worth noting that large R&D investing firms account for over 80% of privately funded R&D (Grassano et al., 2022) and over 60% of global patent filings (Amoroso et al., 2021), making them a relevant object of study. Here, we focus on 7440 publicly listed firms in the United States Patent and Trademark Office (USPTO) from 1926 to 2017 (Woeppel, 2019; Kogan et al., 2017). We represent a firm's research life cycle as the time-ordered sequence of its issued patents (see Fig. 1 which exemplifies the life cycle of four firms highlighting the timing of each one's hit patents). Based on this representation, we ask the previously unanswered questions: Is firms' future research success predictable from their earliest outputs? If yes, which dimensions of early research success are the most predictive ones? And do firms exhibit similar research success patterns as academic actors?

An obstacle toward answering these questions is the ambivalence of the success of patents from the applicant firm's standpoint. As mentioned above, quantitative studies of technological innovation often define the success of a given patent based on its number of received citations or some other citation-based indicator (Aristodemou and Tietze, 2018; Jaffe and De Rassenfosse, 2019), in line with the definition of scientific papers' impact in the growing literature on the science of science (Wang et al., 2013; Fortunato et al., 2018; Wang and Barabási, 2021). However, defining patent success in terms of citation counts would only capture the patents' technological value, but not their economic value, which drives firms' investment decisions (Kogan et al., 2017; Stoffman et al., 2020).

The recent dataset by Kogan et al. (2017) offers us the unique opportunity to simultaneously quantify the technological value of firms' patents via citation-based metrics (Hall et al., 2005) and their economic value based on the firms' stock-price movements following the patent's announcement (Kogan et al., 2017). By comparing different classes of value metrics over the life cycle of the firms in our data, we aim to quantify the predictability of firms' research success, and understand the different implications of patents' economic and technological value for a firm's research success. We find that the technological and economic value of a firm's early patents hold different predictive signals for the firm's later research success. The economic value of a firm's early patents is predictive of both the economic and technological value of its later patents. On the other hand, the technological value of a firm's early patents is only predictive of the technological value of the firm's subsequent patents, but not of their future economic value. We also identify a minority of firms without top-economic value patents in the early stage, which are "sleeping beauties" achieving high-economic value patents in the later stages of their life cycle. These firms markedly differ from those of "predictable" firms that are among the top ones by economic value in both the early and late stages. In particular, in the case of sleeping beauties, the technological value of patents is a much stronger predictor of the future innovative success of the firm. Overall, our findings suggest that indicators of the early economic value of firms' innovations are very relevant in the prediction of the research success of R&D investing companies. Furthermore, we find empirical support for the claim that, contrary to the timing of scientists' highest-impact papers (Sinatra et al., 2016), the timing of a firm's best research over its life cycle is non-random.

Our results have important implications both for stakeholders in the innovation system – policymakers, investors, and managers – and academic scholars. From a managerial perspective, our findings show indeed that a firm's future technological trajectory can be early predicted through a patent-level indicator which can be estimated within few days after the patent's issuance. This underscores the crucial role of picking up market signals on firms' patents for investment and policy decisions. On the research side, our findings contribute by demonstrating that different early signals on a recent firm's patents are not equally predictive. This is evident in the differences between the predictive power of the economic and technological value of a firm's early patents. At the same time, as our findings are correlational and predictive, more research is needed to analyze empirically and theoretically the subject, and identify possible confounding effects.

The rest of the paper is organized as follows. Section 2 reviews the literature and the open questions addressed by this work; Section 3 describes the research methods; Section 4 presents the research results; Section 5 discusses our results as well as their implications for scholars and stakeholders in the innovation system.

2. Literature review

A defining characteristic of the literature on Technology Forecasting (TF) is the systematic attempt of its contributions to predict and understand the potential direction, speed, characteristics, and effects of technological change. A testament to this is the multitude of studies and reviews regularly published on methodological advances in the field (Martino, 2003; Firat et al., 2008; Cho and Daim, 2013; Park et al., 2020). TF initially emerged during the 1950s and 1960s as a tool to anticipate military technology needs and aid in R&D planning. However, over time, the significance of TF has expanded beyond military applications to encompass business and policy considerations (Porter, 1999; Coates et al., 2001). This evolution reflects the increasing need for investors, companies (both large and small), and countries to have access to reliable tools that can support their decision-making processes, including setting research and development priorities and enhancing technological competitiveness (Porter, 1999). Various forms of forecasting technology developments and their impacts have emerged from this prolific field of research, such as technology intelligence, forecasting, roadmapping, assessment, and foresight (Technology Futures Analysis Methods Working Group, 2004). These methods serve a common goal: to provide insights into growth forecasts, interrelationships between

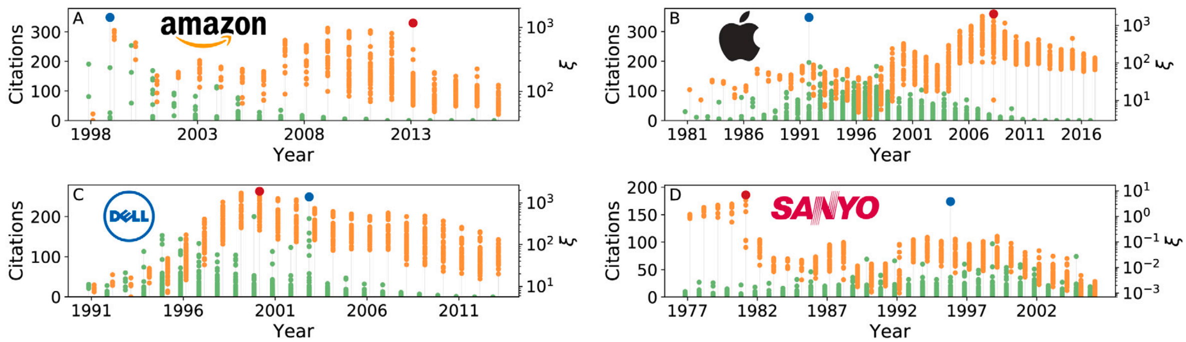


Fig. 1. The research life cycles of firms extracted from their patents. We characterize a firm using the temporal sequence of its issued patents, which we refer to as the firm's *life cycle*. The technological value of a patent (green dots, blue for the most cited patent) is a function of the number of its citations. The economic value ξ of a patent (orange dots, red for the highest value patent) is a function of the stock-price movement of the applicant firm related to the patent's announcement. The four panels show the patenting life cycles of four major firms – *Amazon*, *Apple*, *Dell*, and *Sanyo* – that achieve the most cited patent and the patent with the highest economic value at different stages of their life cycles. *Amazon*'s highest tech-value patent came early in the firm's life cycle, while its highest economic value patent came late; the opposite is true for *Sanyo*. For *Apple*, the most cited patent and highest economic value patent emerged at the early-middle and late-middle stages, respectively. As for *Dell*, both kinds of highest-value patents emerged in succession. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

technology areas, influential researchers or research groups, and the underlying factors that influence technological advancements (Firat et al., 2008).

Of the many qualitative and quantitative methods developed over the years to pursue the arduous task of gaining insights into (or, better still, anticipating) the evolution of technology, bibliometric and patent data-based approaches have become increasingly popular in recent years. This has happened as a result of the availability of comprehensive data collections (Hall et al., 2001; Maraut et al., 2008; Woeppel, 2019; De Rassenfosse et al., 2019; EPO, 2023b) bringing together extensive data about the contents of patent documents and other aspects related to the inventions that they cover (e.g. citations, technology codes, geographical information, patent text, links to economic or firm-level data). This trend in data availability has been accompanied by the rise of computational resources and the emergence of techniques allowing to extract valuable information from large collections of data about technological trends (e.g. Rotolo et al., 2015), innovation activity, and the competitive landscape within specific industries.

Two popular methodological alternatives in the literature focus, respectively, on the analysis of patent codes (sometimes in combination with keyword searches) or patent citations as predictive tools. Technology codes are standard symbols used to classify the technological domains in which a given patent document innovates with respect to the prior art. A very desirable characteristic that technology codes bring to the table for large-scale data analysis is that the most important classifications[1] (International Patent Classification (WIPO, 2023) and Cooperative Patent Classification (EPO, 2023a)) are shared by most patent offices worldwide, which allows a language-independent comparison of documents submitted to patent authorities around the world. Code-based studies of technologies consider the diffusion and evolution of codes over time a measure of how dynamic innovation is in general over a certain period (Strumsky et al., 2012; Youn et al., 2015) as well as to predict how successful or promising specific technological applications are (Daim et al., 2012; Altuntas et al., 2015; Baumann et al., 2021; Yuan and Cai, 2021; Ghaffari et al., 2023; Metzger et al., 2023).

Similarly to technology codes, also the citations received by a technology or an invention[2] are an effective instrument to measure or predict innovative success or value (Hall et al., 2001; Chang et al., 2009; Lee, 2009; Cho and Shih, 2011; Jaffe and De Rassenfosse, 2019). Contrary to codes, which characterize the scope of an invention very well, and hence can be used to create measures of similarity or relatedness between inventions (Yan and Luo, 2017; Whittle and Kogler, 2020; Sbardella et al., 2018), citations are particularly effective in capturing the relevance of a single invention because they can be counted by document and allow comparison over a single dimension. This makes citations particularly useful in applications that aim to estimate how valuable a patent is commercially or establish how successful it is from a research and development standpoint.

However, as mentioned in Section 1, citations have inherent limitations as early predictors of technological success. Finding better metrics in this respect would not only be useful from the viewpoint of TF, but would also be informative for the economics and management literature, which have tackled the issue at the firm level through many relevant studies. For instance, studies have found a significant relation between the quantity and quality of firms' patents and future sales (Ernst, 2001), financial performance (Chen and Chang, 2010; Andries and Faems, 2013; Hsu et al., 2013), export performance (Wu et al., 2022), domestic innovative sales (Thompson and Woerter, 2020), and start-up growth (Guzman and Stern, 2020; Farre-Mensa et al., 2020). Besides, researchers have used patent data to correlate various dimensions of firm performance with different innovation strategies (Michelino et al., 2019; Pugliese et al., 2019b; Cammarano et al., 2020; Xie et al., 2021; Cammarano et al., 2022). These works made extremely valuable contributions to the literature by pointing to the prominent role of patent analysis in understanding firms' research activities (Katila, 2000; Hall et al., 2005; Ponta et al., 2021), but they did not attempt to predict a given firm's future research success based on the value of the firm's early patents. To our best knowledge, there are no previous attempts in the literature to determine whether firms' patent outputs in the early stage can predict firms' future research success and if yes, which early indicator is the most predictive one. We aim to address this gap with our analysis.

Understanding the timing of innovation within the life cycle of an economic actor is a further relevant aspect in improving our ability to

[1] The International Patent Classification (IPC) is a standardized system used for the classification of patents and patent applications on a global scale. It categorizes patents into defined sections, classes, subclasses, and groups. The World Intellectual Property Organization (WIPO) is responsible for maintaining and updating the IPC. The Cooperative Patent Classification (CPC) is a patent classification system jointly developed by the United States Patent and Trademark Office (USPTO) and the European Patent Office (EPO). Compared with IPC, CPC refines classifications further, incorporating additional subgroups and symbols to provide a more granular representation of technological features.

[2] Generally referred to as forward citations to distinguish them from the citations that a patent document contains to acknowledge prior art; these are called backward citations. Since in this paper we are interested in forward citations only, we will henceforth use the term citation as a synonym of forward citation.

predict, which has been under-appreciated in the economics literature, but has instead been extensively explored in recent cross-disciplinary studies about the dynamics of research success of scientists and their scientific works (Sinatra et al., 2016; Liu et al., 2018; Wang et al., 2019; Yin et al., 2019; Wang and Barabási, 2021). This literature has emphasized that despite substantive differences between the two processes, scientific research and technological R&D share several commonalities. For instance, previous studies showed that both build their success on prior knowledge (Uzzi et al., 2013; Mukherjee et al., 2017; Shi and Evans, 2019; Pugliese et al., 2019a,b); that the impact of papers and patents has a predictable evolution over time (Wang et al., 2013; Higham et al., 2017); and that team size predicts research impact and disruptiveness (Wu et al., 2019).

If the analogy between scientific and economic actors engaging in innovation also holds for the timing of creative success over their life cycle, one should expect a firm's most valuable inventions to appear at random points in its sequence of patents (Simonton, 1997). However, the management literature exploring the relation between firm age and research performance has provided arguments that could potentially support both the predominance of early and late research hits along a firm's life cycle (see Table S1 in Supplementary Information for a summary). Some works claimed that newcomer firms are more likely to produce innovations of high technological quality (Huergo and Jaumandreu, 2004; Balasubramanian and Lee, 2008). This is because as firms age, they might gradually refine their innovation competence and organizational routines (Sørensen and Stuart, 2000); in this phase, benefits from new technological advances might reduce (Thornhill, 2006; Balasubramanian and Lee, 2008). Hence, inventions by experienced firms are more likely to be the extension and improvement of their established innovative domains and technologies (Sørensen and Stuart, 2000). At the same time, the innovation literature has identified factors that might favor research success at the late stage of firms' life cycles, including time-consuming knowledge acquisition (Jones, 2009; Withers et al., 2011), experience (Withers et al., 2011), reputation and centrality in relevant collaboration networks (Höflinger et al., 2018; Withers et al., 2011), the hiring of new CEOs (Cucculelli, 2018) and so on (see Table S1 in SI for more details). In sum, some of these works introduce mechanisms that could suggest that a firm's research hit might occur early on along the firm's life cycle (Sørensen and Stuart, 2000; Huergo and Jaumandreu, 2004; Thornhill, 2006; Balasubramanian and Lee, 2008), while other works introduce mechanisms that suggest that late hits might be more likely (Withers et al., 2011; Cucculelli, 2018). Our work provides the first empirical measurement of the distribution of the timing of a firm's most valuable hit, and our results contribute to this long-standing debate.

3. Methods

3.1. Data and general approach

To examine the predictability and dynamics of firms' research success, we analyze the 2,458,402 patents granted to 7440 publicly listed firms by the USPTO from 1926–2017 (the dataset are available at (Woeppel, 2019)). We leverage a recent dataset (Kogan et al., 2017), based on which we can quantify simultaneously both the technological and the economic value of firms' patents. To build the dataset, the authors downloaded from Google patents the entire history of US patent documents and processed the data to disambiguate applicant names. Further, they matched all patents in the data to corporations whose returns are in the Center for Research in Security Prices (CRSP) database using a name-matching algorithm and filtered out mismatched patents (Kogan et al., 2017). The result is a dataset in which each patent applicant is a firm publicly listed on the US stock market. The average number of issued patents per firm is 330.6, and the largest number is 123,220 (granted to *IBM*).

To measure the technological value of a patent at year t, we compute its cumulative citation count until year t based on the citation information in the dataset. As for the economic value metric, we rely on the ready-made measure ξ, which is defined as the present value of the monopoly rents associated with the patent. A patent's ξ is estimated by focusing on a short time window around the patent announcement and filtering the stock price reaction to the patent from the total stock return over the window. To compare patents issued in different years, we normalize both metrics by requiring that the score of a patent is not biased by its issuing year (Waltman, 2016; Mariani et al., 2019). Since we are interested in firms with a sufficiently productive research activity, in the main text we limit the firm-level analysis to the 2819 firms that have at least 15 patents. In the Supplementary Information, we show that our main results are qualitatively robust when filtering the firms based on their number of years of research activities, see Fig. S15 in SI.

The crucial advantage afforded by the data we employ is that, besides patent citation information allowing us to compute the technological value of patents, it also provides an estimate of their economic value based on firms' stock price movements. The possibility to assess the predictability of firms' patent value along two distinct dimensions pushed us to focus uniquely on this dataset, even though other well-known sources of data about firms' patents exist (e.g. the NBER Patent Data Project (NBER, 2012), the Global Corporate Patent Dataset (Darden, 2017)) and the Chinese patent database of listed firms (Zhang, 2017; CNRDS, 2020).

3.2. Quantifying the value of patents

In this work, we consider two dimensions of research value: *technological value* and *economic value*. To assess the predictive power that these metrics have on the research success of firms, we need to construct firm-level indicators accounting for the innovative output of each actor over a given time window. To this aim, in this Section, we first need to define technological and economic value at the patent level. We then move on to define the aggregate metrics at the firm level in the next Section.

To quantify the technological value of a patent, we measure its number of received citations (Hall et al., 2005). A potential shortcoming of the citation count, even when restricted to a fixed temporal window (see e.g. Sinatra et al., 2016), is its strong bias by patent age (see Fig. 2A and Fig. S1 in SI), which makes it unsuited to perform a fair comparison between patents issued in different years (Mariani et al., 2019). To eliminate this bias, we define the *age-normalized citation value* (*NCV*) of a patent as its relative ranking position by citation count among all patents issued in the same year. In formulas, patent i's technological value is $NCV_i = 1 - r_i/N(t_i)$, where $N(t_i)$ is the number of patents issued in the same year t_i as patent i, and r_i denotes the ranking of i by citation count among the $N(t_i)$ patents of the same age. $r_i = 1$ if i is the top patent; $r_i = N(t_i)$ if i is the last one, which corresponds to $NCV_i = 1 - N(t_i)^{-1}$ and $NCV_i = 0$, respectively. We assign the average rankings to all tied values. Therefore, the resulting score $NCV_i \in [0, 1)$ is close to one (zero) for high-value (low-value) patents. Crucially, differently from the rankings by citation count (see Fig. 2A) and C_{10} (i.e., citation count restricted to the first 10 years after the patent issuance, see SI, Fig. S1) (Sinatra et al., 2016), the ranking by *NCV* is consistent with an age-unbiased ranking (see Fig. 2B).

To quantify the economic value of a patent, we rely on a recent measure based on the granted firm's stock price movement upon the patent's grant event, which is denoted as ξ (Kogan et al., 2017). The rationale behind the ξ metric is that the market learns that a USPTO patent application has been successful on the patent's issue date, when the patent is listed in the USPTO's official journal *Official Gazette*. Therefore, the patent's economic value can be estimated based on the firm's stock-market reaction around the patent issue date. Stock market reactions to specific announcements have been also

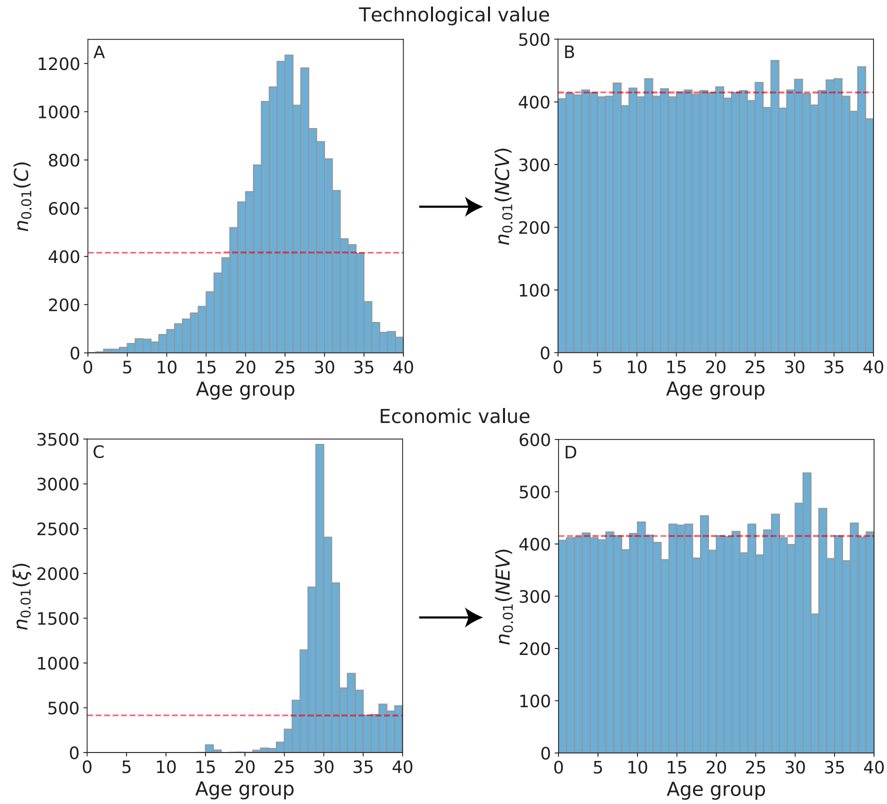


Fig. 2. Normalizing patent-level technological and economic value metrics. Age distribution of the top-1% patents ranked by the citation count (panel A), economic value ξ (panel C), and the proposed normalized value metrics (panels B and D). We divide all patents into 40 equally-sized groups by age and show the number of top-1% patents from each age group (Mariani et al., 2019). The dashed red line denotes the expected value for an age-unbiased ranking, $0.01\,N/40$, where N denotes the total number of patents of all firms. The distributions for the raw value metrics (C and ξ) are strongly biased by patent age (panels A and C), which makes them unsuitable for studying the temporal patterns of firms. By contrast, the two normalized value metrics (NCV and NEV) exhibit a flat temporal profile (panels B and D).

used in event studies that were able to measure the impact of the announcements of crowdsourcing activities (Cappa et al., 2019) and M&A operations (Cappa et al., 2022; Pinelli et al., 2022), among others. To grasp the main idea behind (Kogan et al., 2017)'s ξ, assume that, prior to the issue event, the market can observe the market value ξ_i of a given patent application i, and estimate that the probability that the patent application would be successful is π_i. The firm's stock market reaction to the patent's issuance, on the issue day, is reflected in the change of the firm's stock price, given by the equation (Kogan et al., 2017)

$$\Delta V = (1-\pi_i)\,\xi_i. \tag{1}$$

By making additional assumptions on the distributions of the components of the firm's stock market returns linked and not linked to the patent announcement, Kogan et al. (2017) provided empirical estimates of ξ for all patents in the collected dataset, which are available at Woeppel (2019).

Kogan et al. (2017)'s ξ metric is admittedly not the only method to measure patents' economic value. Previous works in economics considered alternative definitions based on surveys of patent holders (Harhoff et al., 1999; Gambardella et al., 2008), licensing incomes (i.e., the revenues that the patent holder earns from patent licensing) (Sampat and Ziedonis, 2004) and the transfer of patent rights (Serrano, 2006), among others. The major problem with these metrics is that they are backward-oriented (i.e., they can only capture past information), which could lead to imprecise estimates (Hall et al., 2007). Meanwhile, the patent holders have no interest in disclosing the trade details of the patent (Frietsch et al., 2014). Previous studies have found that a firm's patenting activities are linked with its stock market valuations (Hall et al., 2005; Nicholas, 2008), which motivated (Kogan et al., 2017) to propose the ξ metric used here. It is a forward-looking evaluation because of the nature of the stock market (Sandner, 2009; Hall et al., 2005), which can get rid of the above limitations. The authors revealed that compared with patent citations, this economic measure of patents is more correlated with firm growth (Kogan et al., 2017). The ξ measure has been increasingly used in the innovation literature (Hsu et al., 2021; Cascaldi-Garcia and Vukotić, 2022; Kim and Valentine, 2021), which motivates its use in our paper.

Differently from the cumulative number of patent citations, a patent's ξ is determined shortly after the patent's issuance and does not change over time. Similarly to citation count, the economic value metric ξ also exhibits strong bias by patent age, as shown in Fig. 2C. Again, to prevent this bias from influencing our firm-level results, we define the *age-normalized economic value* (NEV) of a patent as its ranking position by ξ among all patents issued in the same year (see Fig. 2D). Specifically, patent i will obtain a score $NEV_i = 1 - r_i/N(t_i)$. Likewise, NEV ranges in $[0,1)$.

3.3. Quantifying the value of firms

To quantify the research success of a firm, one could average or sum the value of all its patents. However, it is well-known that patents' quality is highly heterogeneous (Silverberg and Verspagen, 2007), and prior works placed a greater emphasis on a firm's most prominent innovation than on ordinary innovations (Ahuja and Morris Lampert, 2001; Fleming and Sorenson, 2003; Dunlap-Hinkler et al., 2010). For this reason, we focus on a firm's patents with the highest technological and economic value, which we refer to, respectively, as its technological and economic hits (Ahuja and Morris Lampert, 2001; Srivastava and Gyawali, 2011). The two hits coincide for only about 2% of the analyzed firms (see Fig. S2 in SI for the correlation details). As we

Table 1
The different technological value (*NCV*) and economic value (*NEV*) of five historically significant patents. Among the historically significant patents identified by Strumsky and Lobo (2015), we show a sample of five patents whose *NCV* differs from the *NEV*. We refer to Table S2 in SI for the value metrics of all 31 significant patents.

Patent #	Issue year	Applicant firm	*NCV*	*NEV*	Title/description
2895584	1959	INTERNATIONAL BUSINESS MACHS COR	0.77	0.96	Selectric typewriter printing head
3728480	1973	SANDERS ASSOCIATES INC	0.98	0.17	First video game
3821715	1974	GENERAL ELECTRIC CO	1.00	0.83	Intel 4004 microprocessor
4504982	1985	OPTICAL RADIATION CORP	0.99	0.41	An intraocular lens for permanent implantation into a human eye
6469012	2002	PFIZER INC	0.66	0.99	Viagra

show below, the value of early economic and technological hits have substantially different implications for firms' future research success.

Based on the hits, we define the two dimensions of the research value of a given firm α: its technological value (TV) and its economic value (EV). We define the TV and EV of a firm as the technological value of the firm's technological hit and the economic value of its economic hit, respectively. In formulas, $TV_\alpha = max_{i\in\mathcal{P}_\alpha}\{NCV_i\}$ and $EV_\alpha = max_{i\in\mathcal{P}_\alpha}\{NEV_i\}$, where $\mathcal{P}_\alpha$ denotes the set of patents that were granted to firm α. Note that to simplify exposition, in the following, we refer to a "firm's value" as a shorthand for its research value, i.e. the value of its patents. This should not be confused with the firm's stock price or other measures of the firm's performance, which are not considered here.

4. Results

4.1. Patent-level economic and technological value correlate weakly

A patent may represent a major technological advance, but its announcement might fail to restrict competition or attract the attention of investors, thereby generating a modest impact on the firm's stock price. For example, patent US3728480 from *Sanders Associates* (see Table 1) reported the invention of the first video game that could be played on a home television. This can be considered as a substantial technological advance compared to computer games, and the patent was highly cited, resulting in a high technological value ($NCV = 0.98$). At the same time, the patent failed to capture market interest shortly after its issuance because of the recession in the cable TV industry at that time (Winter, 2019), which resulted in a low economic value ($NEV = 0.17$). We refer to Tables 1 and S2 in SI for the NCV and NEV of a set of expert-selected historically significant patents (Strumsky and Lobo, 2015), and to Tables S3–S4 for a list of top patents by *NCV* and *NEV*, respectively.

Overall, the Pearson correlation between patents' technological and economic value is as low as $r(NCV, NEV) = 0.09$, and the correlation between the two non-normalized variables is also low ($r(c, \xi) = 0.09$, see Fig. 3). To explain the discrepancy between our finding and previous claims of a high positive correlation between technological and economic value (Cremers et al., 1999; Hall et al., 2005; Kogan et al., 2017), we show that such correlation increases as patents are grouped into increasingly large sets of patents with a similar citation value (see Fig. 3). Therefore, whereas previous works demonstrated that groups of patents with higher citation impact exhibit higher economic value (Kogan et al., 2017), the low correlation reported here indicates that there is little predictability of economic value from citation value at the individual patent level, and suggests that the two dimensions of value may hold different forecasting implications.

4.2. Early economic value predicts future research success

To uncover regularities in the dynamics behind firms' research success, we divide firms into three groups according to their technological value and economic value. Specifically, we consider the top-5% firms as high-value firms, the bottom-35% as low-value firms, and the intermediate 60% as medium-value firms. All our results do not strongly depend on the exact choice of these separation thresholds (see Figs. S13 and S14 in SI). These three groups of firms exhibit markedly different productivity (in terms of the number of issued patents) and value dynamics (see SI Fig. S3). High-value firms exhibit a sustained advantage over medium and low-value firms in terms of both productivity and value. This gap is evident even in the very early stage. Motivated by this finding, in the following, we examine whether early activity data can be used to predict firms' future value.

To this end, we split each firm's research life cycle into a 5-year *initial window* of early activities and a *later window* composed of all its remaining years. Our results are qualitatively unchanged for different choices of the initial period's duration and the later period's duration (see SI S4.3). A firm's *early technological value* is defined as the technological value of its early technological hit (i.e., the highest-value patent among the patents issued within the initial 5-year window as measured by the *NCV*). Similarly, a firm's *early economic value* is defined as its early economic hit measured in terms of the *NEV*. We can define firms' subsequent technological and economic value in the same way by referring to their patent output in the later window.

Analyzing the early and late value of firms, we find strong predictability: firms among the top-5% by early economic value are 21.9 times more likely to be among the top-5% by subsequent economic value than the other firms; firms among the top-5% by early technological value are 5.1 times more likely to be among the top-5% by subsequent technological value than the other firms (see Fig. 4). These initial findings motivate the question: Is early technological or economic value more predictive of firms' subsequent research success?

To quantify the predictive power of firms' early technological and economic value, we study a set of classification problems where we use information on firms' early patents to predict whether firms will subsequently be among the top-5% by two dimensions of future research success: the technological value of the firm's late technological hit (i.e., the highest-value patent among the patents issued in the late window), and the economic value of its future economic hit. We consider various metrics of firms' early performance that could be predictive of future success: not only the firms' early technological and economic value, but also their early productivity (in terms of the total number of issued patents in the early window) (Ahuja and Katila, 2001; Zhang et al., 2020), total citations of early patents (Trajtenberg, 1990; Turkina et al., 2019), and other aggregate measures of early patent value (in terms of cumulative ξ, *NCV*, and *NEV*). For each of these early performance metrics, we measure various predictive accuracy metrics, including precision, recall, area under the precision–recall curve (*AuPRC*) (Powers, 2011), for a Naïve Bayes Classifier that classifies a firm as successful in research if and only if it is among the top-z% by the metric in the early window, where z is a parameter that can be tuned to achieve a desired value of recall (see SI S4.1). Fig. 5 contains a simple example to convey the intuition behind the prediction of late *EV* and *TV* based on early *EV* and *TV* in the case of two firms, *Microsoft* and *Hitachi*. The left panel shows that *Microsoft*, which has a high economic hit in the early window (patent US5021974) goes on to obtain a high economic hit (patent US5893915) and a high

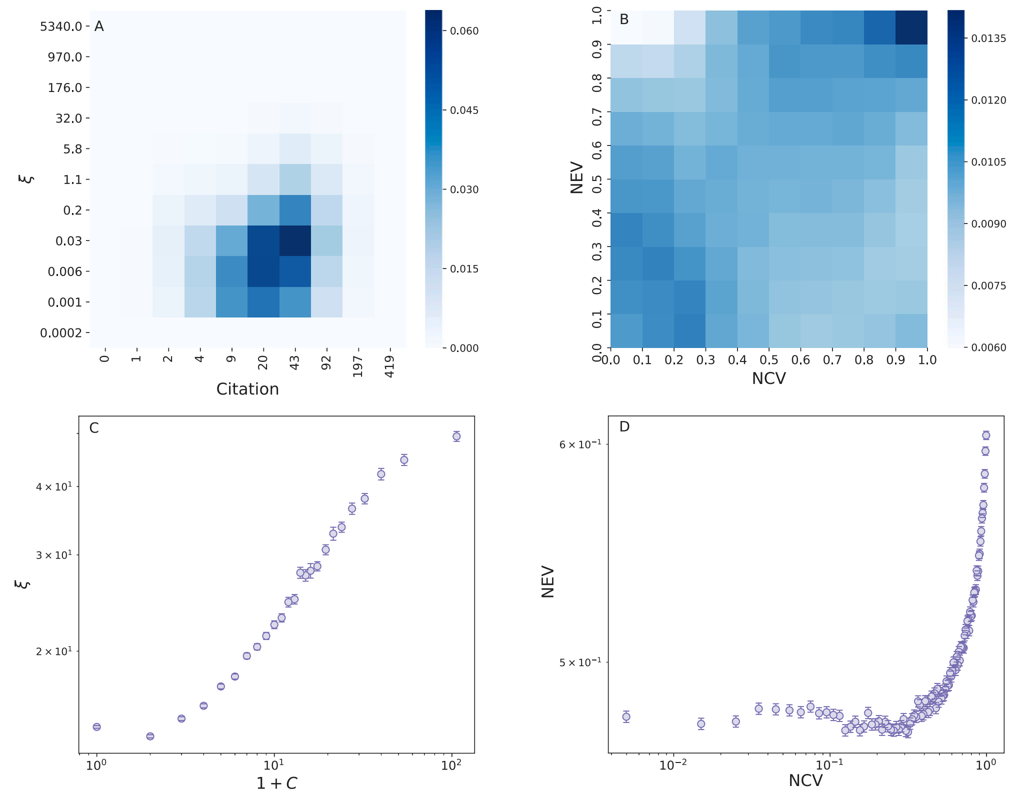


Fig. 3. The correlation between patent's technological value and economic value. (A) The joint probability of the number of citations and ξ for all patents. The Pearson correlation between ξ and the number of citations is 0.09. (B) The joint probability of the *NCV* and *NEV* for all patents. The Pearson correlation between the *NCV* and *NEV* is 0.09. (C) All patents are grouped into 100 groups based on their citations. The Pearson correlation is computed based on the average patent citations and average patent ξ in each group. We obtain a high correlation: $r = 0.88$. This is consistent with the result in Kogan et al. (2017), and it indicates that the correlation increases as patents are grouped into growing sets of patents with a similar citation value. (D) is similar to (C), but reporting the correlation between the group-level *NCV* and *NEV* ($r = 0.86$).

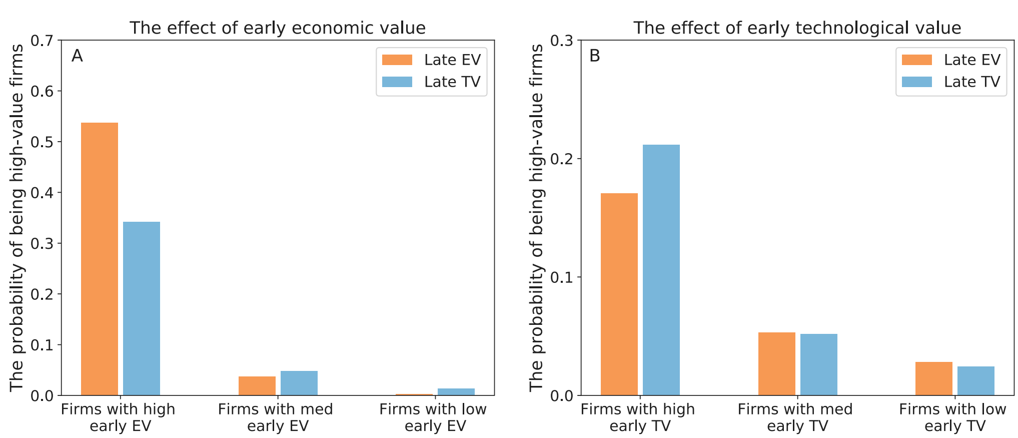
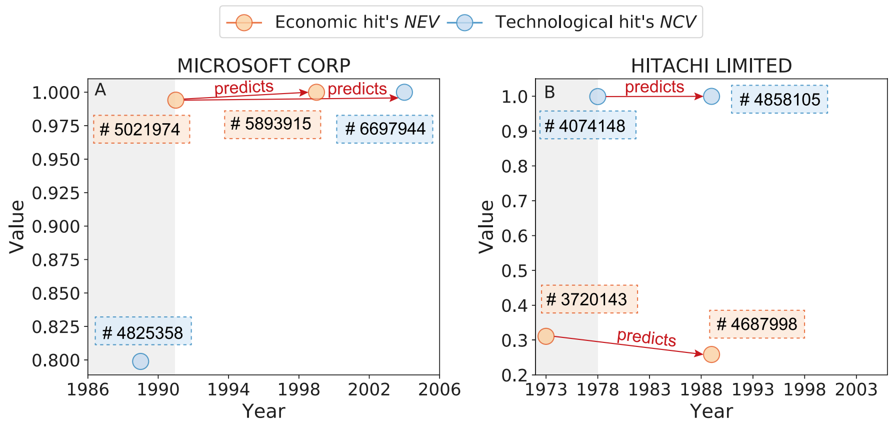


Fig. 4. Success probability for firms in different groups classified by early patent value. Early value is measured based on firms' earliest 5 years after the first patent issuance. We exclude the activity during the earliest 5 years when evaluating late value. (A) The probability of being ranked among the top 5% by late economic/technological value for firms in different groups of early economic value. Firms with high early economic value (top-5%) are 21.9 (9.8) times more likely to be among the top-5% by late economic (technological) value than other firms. (B) The probability of being ranked among the top 5% by late economic (technological) value for firms with different groups of early technological value. Firms among the top-5% by early technological value are 3.9 (5.1) times more likely to be among the top-5% by late economic (technological) value than other firms.

technological hit (patent US6697944) in the late window. On the other hand, *Hitachi* achieves high *TV* early on (patent US4074148), but low *EV*. The data shows that, in the late window, *Hitachi* achieves another high technological hit (patent US4858105), but ends up with low *EV*. This provides hints of the discrepancy between the predictive ability of early *EV* and *TV*.

Our predictive analysis that early economic value is the strongest predictor of both high-economic value firms and, more surprisingly, high-technological value firms in the future (see Fig. 6 for results). By considering classifiers with $z\% = 5\%$ which predicts a firm will belong to the top 5% of late *EV*/*TV* if it is among top $z\%$ based on early metrics, the precision of the classifier based on early economic value reaches 51.6% and 31.9% for the prediction of high economic and technological value firms in the future (10.3-fold and 6.4-fold increase compared with a random classifier, respectively), as opposed to the smaller precision of the classifier based on early technological value (2.8-fold and 3.4-fold increase compared to a random classifier, respectively), as shown by Figs. 6A and B. The results based on raw



Patent #	Issue year	Applicant firm	Title	Value
5021974	1991	MICROSOFT CORP	Method for updating a display bitmap with a character string	0.99
5893915	1999	MICROSOFT CORP	Local font face selection for remote electronic document browsing	1.00
4825358	1989	MICROSOFT CORP	Method and operating system for executing programs in a multi-mode microprocessor	0.80
6697944	2004	MICROSOFT CORP	Digital content distribution, transmission and protection system and method, and portable device	1.00
4074148	1978	HITACHI LIMITED	Address buffer circuit in semiconductor memory	1.00
4858105	1989	HITACHI LIMITED	Pipelined data processor capable of decoding and executing plural instructions in parallel	1.00
3720143	1973	HITACHI LIMITED	Mask for selectively exposing photo-resist to light	0.31
4687998	1989	HITACHI LIMITED	Pulse width generating circuit synchronized with clock signal and corresponding to a reference voltage	0.26

Fig. 5. An illustration of the predictive signals from the early patents of two firms. (A) *Microsoft* exhibits a high-value economic hit in both the early and late stages. During the early period (gray area), *Microsoft*'s economic hit (panel A) was at a high level and its technological hit was at a low level; in the subsequent years (white area), the company produced both high-economic-value patents and high-technological value patents. (B) *Hitachi Limited* exhibits high-value technological hits, but low-value economic hits, in both the early and late stages. The table provides the details of these hit patents.

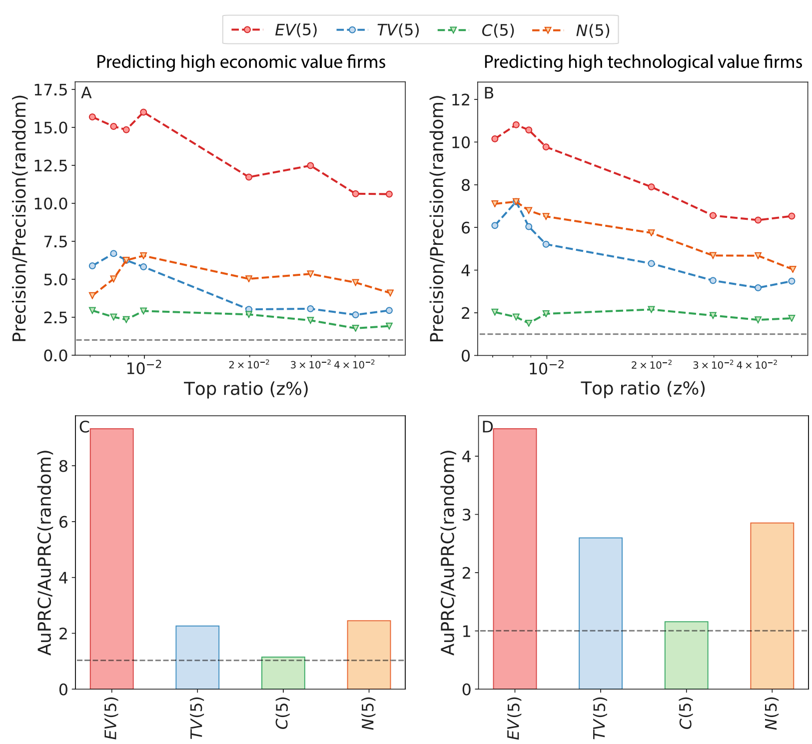


Fig. 6. Predicting top firms in the late window from their early patents. We evaluate the predictive power of classifiers based on the top-z% firms by various metrics of their early patents (within the earliest five years after their first patent issuance): economic value (EV), technological value (TV), total number of citations (C) and total number of patents (N). The EV-based classifier outperforms the others in predicting top-5% firms by late economic value (A, C), and late technological value (B, D). Panels A, B refer to the classifiers' precision normalized by the precision of a random guess as a function of z%; panels C, D refer to the area under the precision–recall curve ($AuPRC$).

accuracy metrics are shown in Fig. S4 in SI. By summing over all possible values of $z\%$, the *AuPRC* of the classifier based on early economic value is 4.12 times and 1.72 times larger than that of the classifier based on early technological value in the prediction of future high economic value firms and high technological firms, respectively, as shown by Figs. 6C and D.

The predictive power of firms' early economic value is substantially stronger than that of other predictors from the literature (such as early productivity and total citations), and significantly larger than that of a random classifier (see the dashed black lines in Fig. 6). These conclusions are robust to alternative choices of the prediction evaluation metric (see Fig. S4 in SI) and variations in the duration of the early window (see SI Fig. S5) and subsequent window (see SI Fig. S6). Combining all the early performance metrics via a binary logistic regression model can moderately improve the predictive accuracy only for the prediction of high-technological-value firms, at the cost of increasing model complexity (see Fig. S4 in SI). For this reason, in the main text, we only show the results of single performance metrics.

Importantly, the stronger predictive power of early economic value holds as well when restricting the analysis to individual industrial sectors: by considering 10 macrosectors based on the first two digits of firms' Standard Industrial Classification (SIC) code,[3] we find that the early economic value is the strongest predictor of future research success for all 10 industries except for the *Transportation & Public Utilities* sector (see SI, Fig. S10 for details). This exception might occur because in this sector, the economic value of generated research is a weaker determinant of governments' and agencies' investment decisions.

The observed predictive power motivates us to better characterize the association between the firms' early economic value and their late productivity, technological value and economic value. Through propensity score matching (Rosenbaum and Rubin, 1983), we detect two groups of firms whose early economic value significantly differs, but whose early productivity and technological value do not (see Fig. S8A in SI). In line with the predictive results above, we find that the two groups of firms significantly differ in their late productivity, technological value, and economic value (see Fig. S8A in SI). To understand whether the late success might be driven by the increased productivity of the high early economic value firms, we perform an additional matching analysis where we add the firms' late productivity among the covariates. We still find a significant success advantage for the group with higher early economic value (see Fig. S8B in SI), which suggests that the late research success of firms with higher early economic value is not due to their higher late productivity. While this analysis better characterizes the relation between early and late value, we do not aim to provide a detailed identification of the mechanisms responsible for the strong connection between early economic value and late research success, as it is beyond the scope of this paper to investigate causality. Nevertheless, the exercise points in a potentially interesting direction for future research aiming to uncover the causal mechanism linking early and late value, and to identify the reasons why only the early economic value – and not the early technological value – effectively predicts the late research success.

4.3. Early patent value predicts sleeping beauties

The observed predictability indicates that early top firms are more likely to be among the high-value ones in the future. We refer to firms with early high economic value that maintain high economic value in the subsequent years as *early bloomers*. Despite the high precision of the resulting classifiers, there exist 2.3% of firms that are not initially among the top-performing ones (i.e., top-5% by early economic value) and later end up among the top-5% (see Fig. 7A). These late-blooming firms, which we refer to as *sleeping beauties*, are reminiscent of sleeping beauty papers in science (Ke et al., 2015): They achieve a high-value economic hit only after a relatively long time after their first patent issuance. Here, we aim to quantify the early detectability of the set of sleeping beauties that transition from medium or low value in the early window to high value in the late window.

Table 2
A simple classification of firms.

	Top late EV	Non-top late EV
Top early EV (top 5%)	early bloomers	submersibles
Non-top early EV (bottom 95%)	sleeping beauties	bottom feeders

Both *early bloomers* and *sleeping beauties* exhibit high late economic value, as opposed to the 92.7% of remaining firms that never reach high economic value, which we refer to as *bottom feeders*; finally, we refer to the 2.3% of firms that achieve high economic value in the early window and descend to a lower value level in the late window as *submersibles*. Table 2 summarizes this classification of firms. We show the heterogeneous economic-value trajectories of 14 well-known firms in Fig. 7A. Among them, *Microsoft*, *General Electric*, *AT&T*, *eBay* and *Apple* maintained a high value (and are therefore classified as early bloomers according to our definition), while *Amazon* fell from high to medium value (therefore, we tag it as a submersible). By contrast, *Intel*, *IBM*, and *HP* went up from medium to high value, and *Applied Materials* rose from low to high value; these four firms qualify as sleeping beauties according to our definition.

Applied Materials is a telling example of sleeping beauty. The firm was unable to produce high economic-value patents within its earliest 5 years of patenting activity, although it was granted high-technological value patents in the early window. After 1982, its economic value exhibited steady growth, and subsequently, the firm became able to produce high economic-value patents (see Fig. 7B and Fig. S7 in SI for more examples of notable sleeping beauties). This transition is reflected in the company's history. *Applied Materials* went public in 1972. In the subsequent few years, the company followed a diversified business strategy. During this period, its technological value was high, while its economic value was low. In 1976, it changed its CEO and refocused on its core business of semiconductor manufacturing equipment (Turner, 2005). After that, its economic value rapidly increased, whereas its technological value stayed at a high level. At the time of writing, the company is a global leader in its core industry.

The existence of sleeping beauties raises the question: Are they predictable? The *Applied Materials* example suggests that high early technological value might predict transitions from low or medium early economic value to high late economic value. We confirm this conjecture in two ways. First, we compare the early technological value for four groups of firms with distinct economic value dynamics (see SI, Fig. S9A). We find the average $TV(5)$ for sleeping beauties is 0.957 (s.e.m. 0.009), which is markedly larger than that for submersibles (0.939 (s.e.m. 0.022)), bottom feeders (0.876 (s.e.m. 0.003)), and even slightly larger than early bloomers (0.954 (s.e.m. 0.016)). Subsequently, we perform a propensity score matching analysis in which we only consider firms with non-top early economic value, and the early technological value is used to split pairs of firms among a "high-T" (high-early technological value) and "low-T" (low-early technological value) group. Among same-industry firms with similar non-top early economic value and early productivity, those with high early technological value exhibit 10% higher late economic and technological value than low-T ones (see Fig. 7C and SI, Table S8). These findings indicate that among firms with non-top early economic value, an early advantage in early technological value translates into a late advantage in terms of economic value.

We further evaluate our ability to early detect the sleeping beauties via their early economic and technological value. To this end, we measure firms' economic and technological value within the earliest 5 years, and we evaluate the predictive performance of a Naïve Bayes

[3] https://siccode.com/

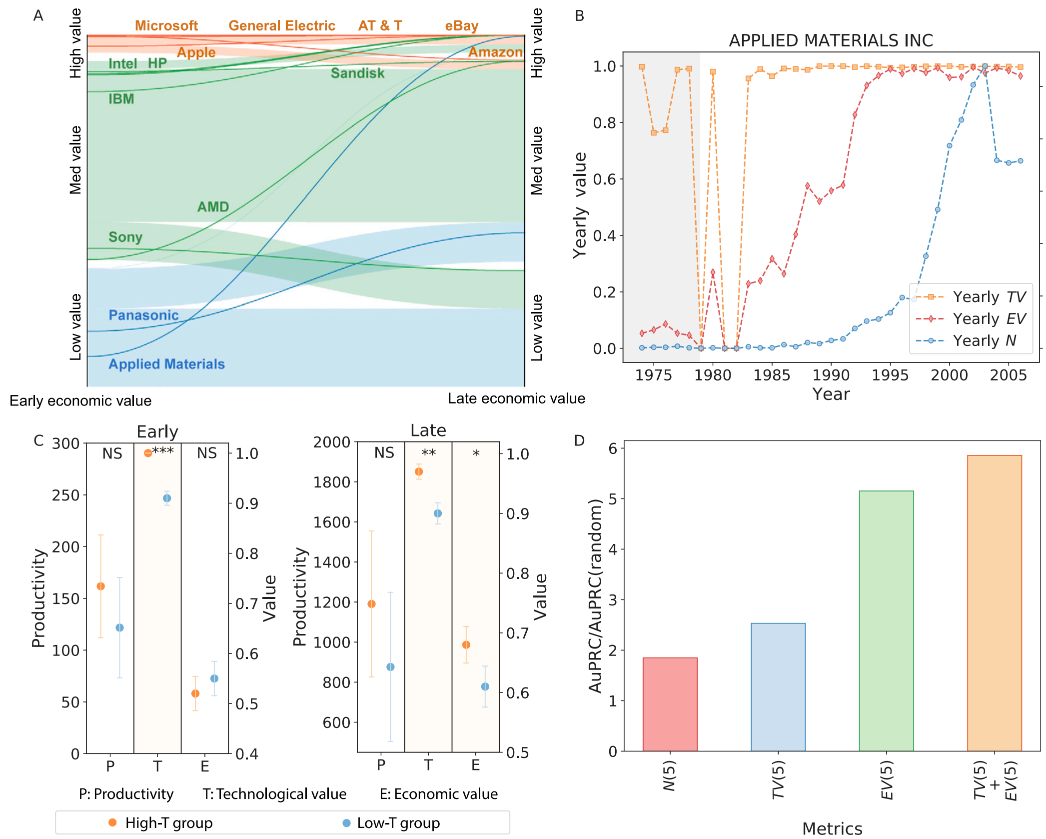


Fig. 7. Predicting sleeping beauties. (A) Illustration of the transition of firms' economic value level from the early stage (first 5 years) to the late stage (from the sixth year on). (B) An example of a sleeping beauty: *Applied Materials Inc*. This panel shows the yearly economic (red dashed line) and technological value (orange dashed line) and yearly number of issued patents (blue dashed line). Over the earliest 5 years (gray shadow), *Applied Materials* was granted high technological-value patents, but stayed at low economic value. After the early period, the economic value of its patents steadily grew, until the firm's patents reached a high economic value. See the main text for a discussion and more examples in SI, Fig. S7. (C) Matched pair analysis restricted to firms with medium or low economic value during the early phase. Firms in the high-T and low-T groups significantly differ in terms of their early technological value. Firms in the high-T group exhibit a significant advantage in terms of future economic and technological value (see Table S8 in SI for the complete results). Error bars stand for standard errors of the mean. (D) Accuracy of simple classifiers for the early detection of sleeping beauties, in terms of their *AuPRC* normalized by the *AuPRC* for a random classifier. The best predictive performance is achieved by the simple combination of early economic and technological value.

classifier that classifies a firm as a sleeping beauty if and only if it is among the top-$z\%$ by a given metric, where z is a parameter that can be tuned to achieve a desired value of recall. We consider various performance metrics, including early productivity, $N(5)$, early economic value, $EV(5)$, early technological value, $TV(5)$, and the sum of early economic and technological value, $EV(5)+TV(5)$. We find that the $EV(5)$ alone achieves a 5.1 fold increase in *AuPRC* compared to a random classifier (see Fig. 7D). This signals that, unsurprisingly, firms that are nearer the top threshold in the early stage are more likely to transition to high value. More interestingly, the $TV(5)$ alone achieves a 2.5 fold increase in *AuPRC* compared to a random classifier, and a combination of the $EV(5)$ and $TV(5)$ achieves the most accurate prediction, leading to a 5.8 fold increase in *AuPRC* compared to a random classifier (see Fig. 7D and Fig. S9B in SI for shortening the duration of early window), which confirms the key role of early technological value in the transition to high economic value.

4.4. The timing of firms' hit patents is not random

The observed predictability of firms' future hits from early patents motivates us to study the temporal dynamics of firms' patent value. Do firms tend to be granted their hits at the beginning of their research life cycles? Or are firms' highest-value patents randomly distributed along firms' life cycles, similarly to the highest-impact works for scientists, artists, movie directors, and musicians (Sinatra et al., 2016; Liu et al., 2018; Janosov et al., 2020)? How do these patterns differ for early bloomers and sleeping beauties? We find that differently from results for individuals' creative works (Sinatra et al., 2016; Liu et al., 2018; Janosov et al., 2020), the temporal position of a firm's hits is markedly non-random.

Specifically, we study the distributions $P(N_C^*/N)$ and $P(N_E^*/N)$ of the relative position of a firm's technological hit (N_C^*) and economic hit (N_E^*), respectively, compared to the firm's total number of issued patents, N (Sinatra et al., 2016; Janosov et al., 2020). Both types of hits are significantly more likely to occur among earliest patents than expected by chance, which is demonstrated by the left peaks of the two distributions (Figs. 8A and B). The observed peaks cannot be explained by randomized patenting histories where for each firm, patents' value scores are randomized, while the total number of patents is preserved (see the shadowed area in Figs. 8A and B) (Sinatra et al., 2016). At the same time, while the probability of achieving the technological hit steadily decreases as a firm is granted more patents (Fig. 8A), the probability of achieving the economic hit exhibits a second peak around the end of the life cycle (at $N_E^*/N \sim 1$, see Fig. 8B). These results hold not only when considering all firms together, but also when considering separately high-value (top 5% by their hits' value), medium-value (middle 60% by their hits' value), and low-value firms (bottom 35% by their hits' value), and when considering firms from different industries, see Figs. S11 and S12 in SI.

The heterogeneity of firms' hit position is well-illustrated by a few key case studies in Fig. 8C (see Table S5 in SI for details). *IBM* achieved its economic hit (about an integrated circuit with dielectric insulation) in 1976, whereas it achieved its technological hit significantly later (in 2002) with a patent on controlling access to shared storage devices.

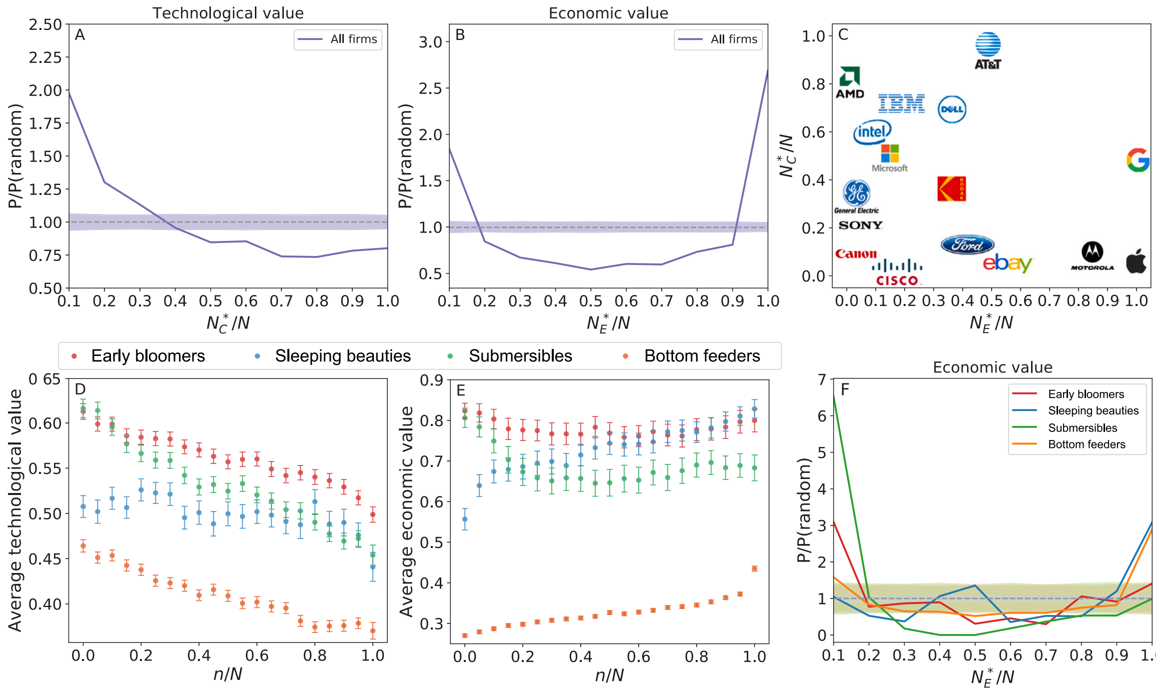


Fig. 8. Heterogeneous patterns of research success over a firm's life cycle. (A) Probability distribution of N_C^*/N for all analyzed firms, where $N_C^*/N \in [1/N, 1]$ denotes the relative temporal position of the firms' technological hits (equal to $1/N$ or 1 if the hit is the first or last issued patent of the firm, respectively). The overall decreasing trend significantly deviates from the expectation $P^{(random)}(N_C^*/N) = 1$ for a randomized life cycle. The shadow area shows the standard error of the results for 200 times randomized firms' life cycles. (B) Probability distribution of N_E^*/N for all analyzed firms, where $N_E^*/N \in [1/N, 1]$ denotes the relative temporal position of the firms' economic hits. We observe a two-peaked distribution that significantly deviates from the expectation $P^{(random)}(N_E^*/N) = 1$ for a randomized life cycle. (C) Illustration of N_C^* *vs.* N_E^* for a sample of 16 famous firms, it shows firms like *IBM*, *AMD* and *Intel* have an early economic hit but a relatively late technological hit, while for *Apple* and *Motorola*, the opposite is true, see details in SI, Table S5. (D) Average *NCV* of firms' patents as a function of relative patent order for four groups of firms: early bloomers (2.7%), sleeping beauties (2.3%), bottom feeders (92.7%), and submersibles (2.3%) firms. All groups of firms exhibit a declining trend. (E) Average *NEV* of firms' patents as a function of relative patent order for the same four groups of firms. Whereas the average *NEV* of patents by early bloomers and submersibles decline with patent order, the patents by bottom feeders and sleeping beauties exhibit a clearly increasing trend. (F) $P(N_E^*/N)$ for the same four groups of firms. The two peaks are manifestations of the heterogeneity of firms' value dynamics: early bloomers and submersibles only contribute to the early peak of the distribution, whereas sleeping beauty and bottom feeders only contribute to the late peak.

On the other hand, the *Apple*'s technological hit appeared in 1992 (on a powered manager for a portable laptop computer), whereas its economic hit was issued substantially later (in 2006, on an improved method for generating multimedia non-linear effects).

The different behavior of $P(N_C^*/N)$ and $P(N_E^*/N)$ is a reflection of firms' heterogeneous value dynamics, which is linked to the predictive problem studied above. To demonstrate this point, we consider the previously-defined four groups of firms: early bloomers, sleeping beauties, submersibles, and bottom feeders. For the four groups of firms, we find that the average technological value of their patents tends to steadily decrease over time (Fig. 8D), which matches the higher probability of early appearance of technological hits. The only exception is the group of sleeping beauties, which exhibits a more stable trend. This suggests that the sleeping beauties' innovation ability does not diminish as they mature, which could be the key to their later transition. By contrast, the dynamics of average economic value exhibits heterogeneous patterns. Whereas the average economic value of the patents of early bloomers and submersibles tends to remain stable or decrease over the firms' life cycles, the economic value of sleeping beauty and bottom feeders sharply increases over time (Fig. 8E). This different behavior is reflected in the behavior of $P(N_E^*/N)$: early bloomers and submersibles only contribute to the early peak, whereas sleeping beauty and bottom feeders only contribute to the late peak (Fig. 8F).

The observed early peak of $P(N_C^*/N)$ supports the claim that newcomer firms are more likely to produce innovations of high technological quality (Huergo and Jaumandreu, 2004; Balasubramanian and Lee, 2008). Based on our previous results, one could conjecture that the second economic peak of sleeping beauties might be due to the increasing ability of technologically competitive firms to attract interest from the market. In some cases, like *Applied Materials*, this might be due to organizational transformations. In other cases, the late peak might be due to factors that have been associated with late success in innovation research, including time-consuming knowledge acquisition (Jones, 2009), experience (Withers et al., 2011), and reputation (Höflinger et al., 2018). Uncovering the mechanisms underlying the observed patterns of research success goes beyond the scope of this contribution, but remains an important challenge for future research.

Taken together, our findings indicate that the firms' hits are not uniformly distributed along the firms' research life cycles, which markedly differs from previous findings on the timing of success for scientists (Sinatra et al., 2016), artists, movie directors (Janosov et al., 2020), and musicians (Janosov et al., 2020). Relatedly, our analysis shows that it is possible to construct a class of prediction indicators of the research success of firms based on the early economic value of patents, which outperforms purely citation-based indicators.

5. Discussion

Our work aims to identify early warnings of firms' research success to be able to forecast their technological trajectories. By viewing each firm as a collection of its granted patents, we quantify firms' research success according to the economic and technological value of their patents in two periods (an early and a late stage). We demonstrate that the economic value of a firm's early patents is highly predictive of both the economic value and technological value of the firm's late patents. By contrast, we find that, surprisingly, the early technological value of a firm's patents is only predictive of the technological value of the firm's late patents. Among firms with late patents of high economic value, we distinguish between " early bloomer" and "sleeping beauty" firms (namely, firms with and without high early economic value, respectively). We identify early signals that enable the early detection of sleeping beauties. Specifically, for firms with relatively low economic value in the early stage, high early technological value can promote late economic value. Moreover, we find that early bloomers and sleeping beauties exhibit considerably different patterns of research success over time: The patents by early bloomers exhibit an approximately stable average economic value, whereas sleeping beauties' patents exhibit a sharply increasing average economic value. Similarly, the economic hit patents by early bloomers tend to be among the earliest patents, but the opposite is true for sleeping beauties. Our results have value both for stakeholders in the innovation system – policymakers, investors, and managers – and academic scholars, as outlined below.

5.1. Implications for stakeholders in the innovation systems

From the point of view of policymakers and managers, the early identification of patent value and top innovation players is an important task (Yang et al., 2023). Our results show how the economic value of firms' early patents, a measure that can be recovered *just days* after the issuance of the patent, is highly informative of the firm's future technological trajectory. It is an important task both *per se*, as the early evaluation of a patent allows for prompt managerial decisions in a context of uncertainty (Fleming, 2001), and because the evaluation of patent value can inform patent-based measures. For instance, many institutional bodies use patents as a proxy for R&D investments when the latter is unavailable, or as a way of splitting R&D expenses among corporate subsidiaries or branches proportionally to the number of patents assigned to each one (Grassano et al., 2020), relying on the strong assumption that all patents have equal value. In this respect, early measures of the value of patents allow for more refined metrics, and therefore better aimed and more prompt policies.

Furthermore, the framework we propose opens the possibility of identifying future relevant players in the innovation system, as early technological and monetary success is informative of firms' future research success. The reaction of the market to the firms' early patents is an early signal of the future market and technological potential of the firm; picking up this signal can have a crucial relevance for investment and policy decisions. Indeed, the early identification of the future trajectory of firms is a crucial aspect to build effective policies (Brown et al., 2017).

5.2. Contributions and implications for researchers

Our work makes three main contributions to the technological forecasting literature. First, it demonstrates that not all early signals on a firm's early patents are the same: The patents' economic value is a substantially more accurate predictor of the firm's future research success compared to the patents' technological value. Further analysis is thus called for the mechanisms behind this predictability. Second, our work defines different classes of firms according to their economic value patterns, and it demonstrates the predictability of sleeping beauty firms. Finally, it identifies key differences in the success patterns of firms compared to innovative actors in academia (Sinatra et al., 2016; Janosov et al., 2020; Wang and Barabási, 2021). This is a strong indication that existing models for the dynamics of human innovative achievements do not apply to firms' life cycles, which calls for new models. A scientist is one individual, while a firm is a collection of individuals whose composition, internal network structures, leadership, etc affect its performance. Incorporating all this ‘‘internal complexity’’ into the predictive and timing analyses can help better understand the mechanisms behind the different patterns observed for firms.

As for the study of industrial organization and economics models, the present work questions our understanding of firms' innovation and the impact of early innovation on future trajectories. Many models of industrial organization involving innovation at the firm level assume uncorrelated innovation shocks (Belleflamme and Peitz, 2015). Our work hints at the role of the economic success of early innovations to characterize the whole innovation trajectory of the firm. This is consistent with a process of cumulative causation, where past success leads to more resources for innovation and a higher future potential. On the other hand, if the future potential of firms were shaped by idiosyncratic shocks, one would not expect to observe a significant difference between the predictive performance of indicators based on the economic and technological value of patents. Future research is surely needed.

5.3. Limitations and future research

The main limitations of our work open key directions for future research. First, as mentioned, the presence (absence) of the predictive power of the economic (technological) value of firms' early patents and the different success timing observed for early bloomers and sleeping beauties raise the question of which are the main mechanisms responsible for the observed difference. While our analysis is purely correlational and predictive, future works may attempt to identify specific mechanisms and test them on empirical data via, e.g., matching procedures, instrumental variables, and natural experiments. Such advanced methodologies can help identify confounding effects, and reveal possible causal pathways through which the success of a firm's early patents influences the firm's future success.

Second, by focusing on publicly listed firms, our analysis does not include the many ventures that never experienced growth and failed, and the companies that performed research that was never significant enough or necessary to produce a patent. To extend the implications of our insights, including failed companies is a key direction for future research, and it will likely involve the collection of new data that include research activities beyond those that lead to patents. Initial studies have revealed how similar endeavours can uncover fundamental mechanisms behind success and failure in science, entrepreneurship, and public security (Wang et al., 2019; Yin et al., 2019; Jiang et al., 2023b). Furthermore, the approach we propose has some obvious limitations in terms of scope. In fact, in addition to the well-known shortcomings of patents in capturing parts of the innovation landscape, the additional requirement that firms be listed restricts the set of economic actors to which the method can be applied. Nevertheless, firms that have not gone public or have just produced their first patent (e.g. startups) are already the focus of much research employing techniques such as machine learning and natural language processing to retrieve useful information. Even though mature companies with established patent portfolios are important players in the development of new technologies, extending the predictive analysis beyond patents and improve the predictions of the research success of small or young firms would be a very valuable advancement.

Third, while we purposely focused on two specific dimensions of firms' performance (the technological and economic values of firms' patents), future research could extend the observed predictive insights to additional dimensions of firm performance, such as the firm's revenue growth (Thornhill, 2006), labor productivity (Pugliese et al., 2019b), export competitiveness (Laudati et al., 2023; Wu et al., 2022), and green innovation performance (Cui et al., 2022), among others.

Finally, while recent strides in the science of science have deepened our understanding of the success trajectories of academic researchers (Sinatra et al., 2016; Liu et al., 2018; Wu et al., 2019; Li et al., 2019; Wang et al., 2019; Wang and Barabási, 2021), our results provide the first step toward a quantitative understanding of the evolution of firms' research success from a complexity science standpoint. Beyond firms, the research approach developed here might find applications to the prediction of the research success of other players, such as cities, regions, and nations. This could help forecast promising regions and companies, identify bottlenecks in research and innovation activities, and inform resource allocation strategies.

CRediT authorship contribution statement

Shuqi Xu: Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Visualization, Writing – review & editing, Writing – original draft, Conceptualization, Validation. **Manuel Sebastian Mariani:** Conceptualization, Funding acquisition, Methodology, Supervision, Writing – original draft, Writing – review & editing. **Linyuan Lü:** Conceptualization, Funding acquisition, Project administration, Supervision, Investigation. **Lorenzo Napolitano:** Conceptualization, Investigation, Supervision, Writing – original draft, Writing – review & editing. **Emanuele Pugliese:** Conceptualization, Methodology, Supervision, Writing – original draft, Writing – review & editing. **Andrea Zaccaria:** Conceptualization, Investigation, Project administration, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

We thank Dashun Wang for giving us the helpful suggestions. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors and do not necessarily reflect the views of the European Commission.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.techfore.2024.123208.

References

Ahuja, G., Katila, R., 2001. Technological acquisitions and the innovation performance of acquiring firms: A longitudinal study. Strateg. Manag. J. 22, 197–220.
Ahuja, G., Morris Lampert, C., 2001. Entrepreneurship in the large corporation: A longitudinal study of how established firms create breakthrough inventions. Strateg. Manag. J. 22, 521–543.
Altuntas, S., Dereli, T., Kusiak, A., 2015. Forecasting technology success based on patent data. Technol. Forecast. Soc. Change 96, 202–214.
Amoroso, S., Aristodemou, L., Criscuolo, C., Dechezleprete, A., Dernis, H., Grassano, N., Moussiegt, L., Napolitano, L., Nawa, D., Squicciarini, M., et al., 2021. World Corporate Top R & D Investors: Paving the Way To Carbon Neutrality. Technical Report, Joint Research Centre, Seville site.
Andries, P., Faems, D., 2013. Patenting activities and firm performance: Does firm size matter? J. Prod. Innov. Manage. 30, 1089–1098.
Aristodemou, L., Tietze, F., 2018. Citations as a measure of technological impact: A review of forward citation-based measures. World Pat. Inf. 53, 39–44.
Balasubramanian, N., Lee, J., 2008. Firm age and innovation. Ind. Corp. Chang. 17, 1019–1047.
Baumann, M., Domnik, T., Haase, M., Wulf, C., Emmerich, P., Rösch, C., Zapp, P., Naegler, T., Weil, M., 2021. Comparative patent analysis for the identification of global research trends for the case of battery storage, hydrogen and bioenergy. Technol. Forecast. Soc. Change 165, 120505.
Belleflamme, P., Peitz, M., 2015. Industrial Organization: Markets and Strategies. Cambridge University Press.
Brown, R., Mawson, S., Mason, C., 2017. Myth-busting and entrepreneurship policy: The case of high growth firms. Entrepreneurship Reg. Dev. 29, 414–443.
Cammarano, A., Michelino, F., Caputo, M., 2022. Extracting firms' R & D processes from patent data to study inbound and coupled open innovation. Creativity Innov. Manag. 31, 322–339.
Cammarano, A., Michelino, F., Vitale, M.P., La Rocca, M., Caputo, M., 2020. Technological strategies and quality of invention: The role of knowledge base and technical applications. IEEE Trans. Eng. Manage. 69, 1050–1066.
Cappa, F., Collevecchio, F., Oriani, R., Peruffo, E., 2022. Banks responding to the digital surge through open innovation: Stock market performance effects of M & As with fintech firms. J. Econ. Bus. 121, 106079.
Cappa, F., Oriani, R., Pinelli, M., De Massis, A., 2019. When does crowdsourcing benefit firm stock market performance? Res. Policy 48, 103825.
Cascaldi-Garcia, D., Vukotić, M., 2022. Patent-based news shocks. Rev. Econ. Stat. 104, 51–66.
Chang, S.B., Lai, K.K., Chang, S.M., 2009. Exploring technology diffusion and classification of business methods: Using the patent citation network. Technol. Forecast. Soc. Change 76, 107–117.
Chen, Y.S., Chang, K.C., 2010. The relationship between a firm's patent quality and its market value—the case of US pharmaceutical industry. Technol. Forecast. Soc. Change 77, 20–33.
Cho, Y., Daim, T., 2013. Technology forecasting methods. In: Research and Technology Management in the Electricity Industry: Methods, Tools and Case Studies. Springer, pp. 67–112.
Cho, T.S., Shih, H.Y., 2011. Patent citation network analysis of core and emerging technologies in Taiwan: 1997–2008. Scientometrics 89, 795–811.
Chung, P., Sohn, S.Y., 2020. Early detection of valuable patents using a deep learning model: Case of semiconductor industry. Technol. Forecast. Soc. Change 158, 120146.
CNRDS, C.R.D.S.P., 2020. Chinese innovation research database. Available at https://vcpe.cnrds.com/Home/Index#/FeaturedDatabase/DB/CIRD/.
Coates, V., Farooque, M., Klavans, R., Lapid, K., Linstone, H.A., Pistorius, C., Porter, A.L., 2001. On the future of technological forecasting. Technol. Forecast. Soc. Change 67, 1–17.
Cremers, K., Harhoff, D., Narin, F., Scherer, F., Vopel, K., 1999. Citation frequency and the value of patented inventions. Rev. Econ. Stat. 81, 511–515.
Cucculelli, M., 2018. Firm age and the probability of product innovation. Do CEO tenure and product tenure matter. J. Evol. Econ. 28, 153–179.
Cui, J., Dai, J., Wang, Z., Zhao, X., 2022. Does environmental regulation induce green innovation? A panel study of Chinese listed firms. Technol. Forecast. Soc. Change 176, 121492.
Daim, T., Harell, G., Hogaboam, L., 2012. Forecasting renewable energy production in the us. Foresight 14, 225–241.
Darden, U., 2017. Global corporate patent dataset. Available at https://patents.darden.virginia.edu/.
De Rassenfosse, G., Kozak, J., Seliger, F., 2019. Geocoding of worldwide patent data. Sci. Data 6, 260.
Dunlap-Hinkler, D., Kotabe, M., Mudambi, R., 2010. A story of breakthrough versus incremental innovation: Corporate entrepreneurship in the global pharmaceutical industry. Strateg. Entrepreneurship J. 4, 106–127.
EPO, E.P.O., 2023a. Cooperative patent classification (cpc). https://www.epo.org/en/searching-for-patents/helpful-resources/first-time-here/classification/cpc.
EPO, 2023b. EPO worldwide patent statistical database data catalog - 2023 spring edition.
Ernst, H., 2001. Patent applications and subsequent changes of performance: Evidence from time-series cross-section analyses on the firm level. Res. Policy 30, 143–157.
Farre-Mensa, J., Hegde, D., Ljungqvist, A., 2020. What is a patent worth? Evidence from the US patent "lottery". J. Finance 75, 639–682.
Firat, A.K., Woon, W.L., Madnick, S., 2008. Technological forecasting–a review. In: Composite Information Systems Laboratory. CISL, Massachusetts Institute of Technology, pp. 1–19.
Fleming, L., 2001. Recombinant uncertainty in technological search. Manage. Sci. 47, 117–132.
Fleming, L., Sorenson, O., 2003. Navigating the technology landscape of innovation. MIT Sloan Manag. Rev. 44, 15.
Fortunato, S., Bergstrom, C.T., Börner, K., Evans, J.A., Helbing, D., Milojević, S., Petersen, A.M., Radicchi, F., Sinatra, R., Uzzi, B., et al., 2018. Science of science. Science 359, eaao0185.
Frietsch, R., Neuhäusler, P., Jung, T., Van Looy, B., 2014. Patent indicators for macroeconomic growth—the value of patents estimated by export volume. Technovation 34, 546–558.
Gambardella, A., Harhoff, D., Verspagen, B., 2008. The value of European patents. Eur. Manag. Rev. 5, 69–84.
Ghaffari, M., Aliahmadi, A., Khalkhali, A., Zakeri, A., Daim, T.U., Yalcin, H., 2023. Topic-based technology mapping using patent data analysis: A case study of vehicle tires. Technol. Forecast. Soc. Change 193, 122576.
Grassano, N., Hernández, H., Tübke, A., Amoroso, S., Dosso, M., Georgakaki, A., Pasimeni, F., 2020. The 2020 EU Industrial R & D Investment Scoreboard. Publications Office of the European Union, http://dx.doi.org/10.2760/203793.
Grassano, N., Hernandez Guevara, H., Fako, P., Nindl, E., Georgakaki, A., Ince, E., Napolitano, L., Rentocchini, F., Tuebke, A., 2022. The 2022 EU Survey on Industrial R & D Investment Trends. Technical Report, Joint Research Centre, Seville site.
Guzman, J., Stern, S., 2020. The state of American entrepreneurship: New estimates of the quantity and quality of entrepreneurship for 32 US states, 1988–2014. Am. Econ. J. Econ. Policy 12, 212–243.
Hall, B.H., Jaffe, A.B., Trajtenberg, M., 2001. The nber patent citation data file: Lessons, insights and methodological tools.

Hall, B.H., Jaffe, A., Trajtenberg, M., 2005. Market value and patent citations. Rand J. Econ. 1, 6–38.
Hall, B.H., Thoma, G., Torrisi, S., 2007. The market value of patents and R & D: Evidence from European firms. In: Academy of Management Proceedings. New York, pp. 1–6.
Harhoff, D., Narin, F., Scherer, F.M., Vopel, K., 1999. Citation frequency and the value of patented inventions. Rev. Econ. Stat. 81, 511–515.
Higham, K.W., Governale, M., Jaffe, A., Zülicke, U., 2017. Fame and obsolescence: Disentangling growth and aging dynamics of patent citations. Phys. Rev. E 95, 042309.
Höflinger, P.J., Nagel, C., Sandner, P., 2018. Reputation for technological innovation: Does it actually cohere with innovative activity? J. Innov. Knowl. 3, 26–39.
Hsu, F.J., Chen, M.Y., Chen, Y.C., Wang, W.C., 2013. An empirical study on the relationship between R & D and financial performance. J. Appl. Financ. Bank. 3, 107.
Hsu, D.H., Hsu, P.H., Zhou, T., Ziedonis, A.A., 2021. Benchmarking us university patent value and commercialization efforts: A new approach. Res. Policy 50, 104076.
Huergo, E., Jaumandreu, J., 2004. How does probability of innovation change with firm age? Small Bus. Econ. 22, 193–207.
Jaffe, A.B., De Rassenfosse, G., 2019. Patent citation data in social science research: Overview and best practices. In: Research Handbook on the Economics of Intellectual Property Law. Edward Elgar Publishing, Cheltenham, pp. 1360–1374.
Janosov, M., Battiston, F., Sinatra, R., 2020. Success and luck in creative careers. EPJ Data Sci. 9, 1–12.
Jiang, H., Fan, S., Zhang, N., Zhu, B., 2023a. Deep learning for predicting patent application outcome: The fusion of text and network embeddings. J. Informetr. 17, 101402.
Jiang, C., Zhou, Y., Chen, B., 2023b. Mining semantic features in patent text for financial distress prediction. Technol. Forecast. Soc. Change 190, 122450.
Jones, B.F., 2009. The burden of knowledge and the "death of the renaissance man": Is innovation getting harder? Rev. Econom. Stud. 76, 283–317.
Katila, R., 2000. Using patent data to measure innovation performance. Int. J. Bus. Perform. Manag. 2, 180–193.
Ke, Q., Ferrara, E., Radicchi, F., Flammini, A., 2015. Defining and identifying sleeping beauties in science. Proc. Natl. Acad. Sci. 112, 7426–7431.
Kim, J., Kim, H., Geum, Y., 2023. How to succeed in the market? predicting startup success using a machine learning approach. Technol. Forecast. Soc. Change 193, 122614.
Kim, J., Valentine, K., 2021. The innovation consequences of mandatory patent disclosures. J. Account. Econ. 71, 101381.
Kogan, L., Papanikolaou, D., Seru, A., Stoffman, N., 2017. Technological innovation, resource allocation, and growth. Q. J. Econ. 132, 665–712. http://dx.doi.org/10.1093/qje/qjw040.
Laudati, D., Mariani, M.S., Pietronero, L., Zaccaria, A., 2023. The different structure of economic ecosystems at the scales of companies and countries. J. Phys. Complexity 4 (2), 025011.
Lee, Y.G., 2009. What affects a patent's value? an analysis of variables that affect technological, direct economic, and indirect economic value: An exploratory conceptual approach. Scientometrics 79, 623–633.
Li, W., Aste, T., Caccioli, F., Livan, G., 2019. Early coauthorship with top scientists predicts success in academic careers. Nature Commun. 10, 1–9.
Liu, L., Wang, Y., Sinatra, R., Giles, C.L., Song, C., Wang, D., 2018. Hot streaks in artistic, cultural, and scientific careers. Nature 559, 396.
Lu, X., Wang, J., 2024. Is innovation strategy a catalyst to solve social problems? the impact of r & d and non-r & d innovation strategies on the performance of social innovation-oriented firms. Technol. Forecast. Soc. Change 199, 123020.
Maraut, S., Dernis, H., Webb, C., Spiezia, V., Guellec, D., 2008. The OECD REGPAT database.
Mariani, M.S., Medo, M., Lafond, F., 2019. Early identification of important patents: Design and validation of citation network metrics. Technol. Forecast. Soc. Change 146, 644–654.
Martino, J.P., 2003. A review of selected recent advances in technological forecasting. Technol. Forecast. Soc. Change 70, 719–733.
Metzger, P., Mendonça, S., Silva, J.A., Damásio, B., 2023. Battery innovation and the circular economy: What are patents revealing? Renew. Energy 209, 516–532.
Michelino, F., Cammarano, A., Celone, A., Caputo, M., 2019. The linkage between sustainability and innovation performance in it hardware sector. Sustainability 11, 4275.
Mukherjee, S., Romero, D.M., Jones, B., Uzzi, B., 2017. The nearly universal link between the age of past knowledge and tomorrow's breakthroughs in science and technology: The hotspot. Sci. Adv. 3, e1601315.
NBER, 2012. Nber patent data project. Available at https://sites.google.com/site/patentdataproject/Home/.
Nicholas, T., 2008. Does innovation cause stock market runups? Evidence from the great crash. Amer. Econ. Rev. 98, 1370–1396.
Park, H., Phaal, R., Ho, J.Y., O'Sullivan, E., 2020. Twenty years of technology and strategic roadmapping research: A school of thought perspective. Technol. Forecast. Soc. Change 154, 119965.
Pinelli, M., Cappa, F., Peruffo, E., Oriani, R., 2022. Acquisitions of non-controlling equity stakes: Agency conflicts and profitability. Strateg. Organ. 20, 341–367.
Ponta, L., Puliga, G., Manzini, R., 2021. A measure of innovation performance: The Innovation Patent Index. Manage. Decis. 59, 73–98.
Porter, A.L., 1999. Tech forecasting an empirical perspective. Technol. Forecast. Soc. Change 62, 19–28.
Powers, D.M.W., 2011. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. Int. J. Mach. Learn. Technol. 2, 37–63.
Pugliese, E., Cimini, G., Patelli, A., Zaccaria, A., Pietronero, L., Gabrielli, A., 2019a. Unfolding the innovation system for the development of countries: co-evolution of science, technology and production. Sci. Rep. 9, 16440.
Pugliese, E., Napolitano, L., Zaccaria, A., Pietronero, L., 2019b. Coherent diversification in corporate technological portfolios. PLoS One 14, e0223403.
Rosenbaum, P.R., Rubin, D.B., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70, 41–55.
Rotolo, D., Hicks, D., Martin, B.R., 2015. What is an emerging technology? Res. Policy 44, 1827–1843.
Sampat, B.N., Ziedonis, A.A., 2004. Patent citations and the economic value of patents. In: Handbook of Quantitative Science and Technology Research. Springer, pp. 277–298.
Sandner, P., 2009. The market value of R & D, patents, and trademarks. In: The Valuation of Intangible Assets. Springer, pp. 35–72.
Sbardella, A., Perruchas, F., Napolitano, L., Barbieri, N., Consoli, D., 2018. Green technology fitness. Entropy 20, 776.
Serrano, C.J., 2006. The Market for Intellectual Property: Evidence from the Transfer of Patents (Ph.D. thesis). The University of Minnesota, Minneapolis.
Shi, F., Evans, J., 2019. Science and technology advance through surprise. arXiv preprint arXiv:1910.09370.
Silverberg, G., Verspagen, B., 2007. The size distribution of innovations revisited: An application of extreme value statistics to citation and value measures of patent significance. J. Econometrics 139, 318–339.
Simonton, D.K., 1997. Creative productivity: A predictive and explanatory model of career trajectories and landmarks. Psychol. Rev. 104, 66.
Sinatra, R., Wang, D., Deville, P., Song, C., Barabási, A.L., 2016. Quantifying the evolution of individual scientific impact. Science 354, aaf5239.
Sørensen, J.B., Stuart, T.E., 2000. Aging, obsolescence, and organizational innovation. Adm. Sci. Q. 45, 81–112.
Srivastava, M.K., Gnyawali, D.R., 2011. When do relational resources matter? Leveraging portfolio technological resources for breakthrough innovation. Acad. Manag. J. 54, 797–810.
Stoffman, N., Woeppel, M., Yavuz, M.D., 2020. Small innovators: No risk, no return. In: Kelley School of Business Research Paper.
Strumsky, D., Lobo, J., 2015. Identifying the sources of technological novelty in the process of invention. Res. Policy 44, 1445–1461.
Strumsky, D., Lobo, J., Van der Leeuw, S., 2012. Using patent technology codes to study technological change. Econ. Innov. New Technol. 21, 267–286.
Technology Futures Analysis Methods Working Group, 2004. Technology futures analysis: Toward integration of the field and new methods. Technol. Forecast. Soc. Change 71, 287–303.
Thompson, M.J., Woerter, M., 2020. Competition and invention quality: Evidence from Swiss firms. Technol. Forecast. Soc. Change 156, 120023.
Thornhill, S., 2006. Knowledge, innovation and firm performance in high-and low-technology regimes. J. Bus. Ventur. 21, 687–703.
Trajtenberg, M., 1990. A penny for your quotes: Patent citations and the value of innovations. Rand J. Econ. 17, 2–187.
Turkina, E., Oreshkin, B., Kali, R., 2019. Regional innovation clusters and firm innovation performance: An interactionist approach. Reg. Stud. 53, 1193–1206.
Turner, T.N., 2005. Vault Guide to the Top Energy Industry Employers. Vault Inc.
Uzzi, B., Mukherjee, S., Stringer, M., Jones, B., 2013. Atypical combinations and scientific impact. Science 342, 468–472.
Waltman, L., 2016. A review of the literature on citation impact indicators. J. Informetr. 10, 365–391.
Wang, D., Barabási, A.L., 2021. The Science of Science. Cambridge University Press, Cambridge.
Wang, Y., Jones, B.F., Wang, D., 2019. Early-career setback and future career impact. Nature Commun. 10, 1–10.
Wang, D., Song, C., Barabási, A.L., 2013. Quantifying long-term scientific impact. Science 342, 127–132.
Whittle, A., Kogler, D.F., 2020. Related to what? reviewing the literature on technological relatedness: Where we are now and where can we go? Pap. Reg. Sci. 99, 97–113.
Winter, D., 2019. Once upon a time the video game at sanders associates. http://www.pong-story.com/sanders.htm.
WIPO, W.I.P.O., 2023. International patent classification (ipc). https://www.wipo.int/classifications/ipc/en/.
Withers, M.C., Drnevich, P.L., Marino, L., 2011. Doing more with less: The disordinal implications of firm age for leveraging capabilities for innovation activity. J. Small Bus. Manag. 49, 515–536.
Woeppel, M., 2019. Patent–CRSP match, 1926–2017. Dropdox. Available at https://paper.dropbox.com/doc/Patent-CRSP-match-1926-2017-W3aHAj0Ce4CzKZayqCASj. Deposited 30 November 2019.

Wu, L., Wang, D., Evans, J.A., 2019. Large teams develop and small teams disrupt science and technology. Nature 566, 378–382.
Wu, L., Wei, Y., Wang, C., McDonald, F., Han, X., 2022. The importance of institutional and financial resources for export performance associated with technological innovation. Technol. Forecast. Soc. Change 185, 122040.
Xie, Z., Wang, J., Miao, L., 2021. Big data and emerging market firms' innovation in an open economy: The diversification strategy perspective. Technol. Forecast. Soc. Change 173, 121091.
Yan, B., Luo, J., 2017. Measuring technological distance for patent mapping. J. Assoc. Inf. Sci. Technol. 68, 423–437.
Yang, G., Lu, G., Xu, S., Chen, L., Wen, Y., 2023. Which type of dynamic indicators should be preferred to predict patent commercial potential? Technol. Forecast. Soc. Change 193, 122637.
Yin, Y., Wang, Y., Evans, J.A., Wang, D., 2019. Quantifying the dynamics of failure across science, startups and security. Nature 575, 190–194.
Youn, H., Strumsky, D., Bettencourt, L.M., Lobo, J., 2015. Invention as a combinatorial process: evidence from us patents. J. R. Soc. Interface 12, 20150272.
Yuan, X., Cai, Y., 2021. Forecasting the development trend of low emission vehicle technologies: Based on patent data. Technol. Forecast. Soc. Change 166, 120651.
Zhang, Y., 2017. Matching SIPO patents to Chinese listed firms. Available at https://sites.google.com/site/sipopdb/cpdp-home/sipo-chinese-listed-firms.
Zhang, S., Zhang, N., Zhu, S., Liu, F., 2020. A foot in two camps or your undivided attention? The impact of intra-and inter-community collaboration on firm innovation performance. Technol. Anal. Strategic Manag. 32, 753–768.

Dr. Shuqi Xu is a researcher at the Institute of Dataspace at Hefei Comprehensive National Science Center. She received her Ph.D. degree in computer science and technology at the University of Electronic Science and Technology of China, with a dissertation on the quantification and forecasts of technological impact driven by patent data. Her main research interests include patent and innovation analysis, socioeconomic complexity and citation networks.

Dr. Manuel Sebastian Mariani is a research group leader at the Department of Business Administration of the University of Zurich, and a visiting associate professor at the Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China. He holds a Ph.D. in Physics from the University of Fribourg, in Switzerland. His research focuses on collective behavior in socioeconomic systems, and its managerial applications for technological innovation, policies for social change, and the design of digital social and algorithmic systems.

Prof. Linyuan Lü received a Ph.D. degree from the University of Fribourg, Switzerland. She is a Professor at the School of Cyber Science and Technology, University of Science and Technology of China. Her current research interests include complex systems and networks, network information mining. She serves as a board member of the International Network Science Society. Her research has been published in Nature Communications, Physics Reports, PNAS, National Science Review, and other top journals.

Dr. Lorenzo Napolitano is a senior researcher at the European Commission's Joint Research Centre (JRC). He holds a Ph.D. in Economics from the Sant'Anna School of Advanced Studies (Pisa, Italy). His main research interests are in the area of innovation, with a focus on patent data analysis and the role that complexity plays in shaping economic systems. He has been researcher for the Italian National Council for Research and the Sant'Anna School of Advanced Studies.

Dr. Emanuele Pugliese is a senior researcher at the European Commission's Joint Research Centre (JRC). After his Master Thesis in Theoretical Physics, he obtained a Ph.D. in Economics from Sant'Anna School for Advanced Studies of Pisa, with a thesis on the effects of innovation on society in general and firms in particular. His research interests include innovation and development, with a focus on the role of complexity in economic systems. He has been consultant for the World Bank Group and researcher for the Italian National Council for Research.

Dr. Andrea Zaccaria is a senior researcher at the Institute for Complex Systems (ISC)-CNR and a consultant at the International Finance Corporation (IFC), World Bank Group. He received his Ph.D. in Physics at the "Sapienza" University of Rome, where he applied concepts and methods borrowed from Statistical Physics and the Physics of Complex Systems to the study of financial markets. His present work focuses on the Economic Complexity field, and on the application of complex networks methodologies, algorithmic tools, and machine learning to development economics.