Identifying Scientific and Technical “Unicorns”

Lucy L. Xu1,2, Miao Qi1, Fred Y. Ye†

1Jiangsu Key Laboratory of Data Engineering and Knowledge Service, School of Information Management, Nanjing University, Nanjing 210023, China
2University Library, Nantong University, Nantong 226019, China

Abstract

Purpose: Using the metaphor of “unicorn,” we identify the scientific papers and technical patents characterized by the informetric feature of very high citations in the first ten years after publishing, which may provide a new pattern to understand very high impact works in science and technology.

Design/methodology/approach: When we set C as the total citations of papers or patents in the first ten years after publication, with $C \geq 5,000$ for scientific “unicorn” and $C \geq 500$ for technical “unicorn,” we have an absolute standard for identifying scientific and technical “unicorn” publications.

Findings: We identify 165 scientific “unicorns” in 14,301,875 WoS papers and 224 technical “unicorns” in 13,728,950 DII patents during 2001–2012. About 50% of “unicorns” belong to biomedicine, in which selected cases are individually discussed. The rare “unicorns” increase following linear model, the fitting data show 95% confidence with the RMSE of scientific “unicorn” is 0.2127 while the RMSE of technical “unicorn” is 0.0923.

Research limitations: A “unicorn” is a pure quantitative consideration without concerning its quality, and “potential unicorns” as $C \leq 5,000$ for papers and $C \leq 500$ for patents are left in future studies.

Practical implications: Scientific and technical “unicorns” provide a new pattern to understand high-impact works in science and technology. The “unicorn” pattern supplies a concise approach to identify very high-impact scientific papers and technical patents.

Originality/value: The “unicorn” pattern supplies a concise approach to identify very high impact scientific papers and technical patents.

Keywords Unicorn; Scientific paper; Technical patent; Citation analysis; Patent analysis

† Corresponding author: Fred Y. Ye (E-mail: yye@nju.edu.cn).
1 Introduction

The “unicorn” is a legendary creature that has been described since antiquity as a beast with a single large, pointed, spiraling horn projecting from its forehead (Wikipedia, 2020). It is imaged as a mythical. Usually, the white animal generally depicted with the body and head of a horse with long flowing mane and tail and a single often spiraled horn in the middle of the forehead. Eileen Lee (2013) published a signed article titled “Welcome to the unicorn club: learning from billion-dollar start-ups” on TechCrunch, in which she introduced this “unicorn” into the economy as a company which indicated of rarity and worth. From then on, the title of “unicorn” from the venture capitalist characterization of private start-up companies quickly become popular in Silicon Valley, which has achieved the almost mythical accomplishment of reaching 1 billion dollars or more valuation in ten years (Casanova, Cornelius, & Dutta, 2018). Gradually, venture capitalists have funded today’s “unicorn” high-tech companies such as Uber Technologies Inc., Airbnb Inc., Palantir Technologies Inc., etc. in the United States. Several start-up companies emerge (Cbinsights, 2020) and become “unicorn” in China as well, and prominent examples include Xiaomi Inc. and Didi Chuxing. By January 2019, there are 347 “unicorns,” of which 89 are in China and 174 in the United States, with a total valuation of 1.093 trillion dollars (Wikipedia, 2020). Applying to the smart city mega-developments, the “unicorn” planning (Cugurullo, Datta, & Shaban, 2016; Rebentisch et al., 2020) has been used to convey the potentially massive profits for high-tech companies, the ambition to instantly invigorate local and regional economies and the idealized expectation of overnight success.

As early as 2007, The National Human Genome Research Institute (NHGRI) endorsed a multi-taxon genome-sequencing initiative termed “unicorn” (Ruiz-Trillo et al., 2007), which generated extensive genomic data from animals to summarize the rationale guiding the choice of organisms. In addition, the term “unicorn” assisted the biocuration and validation of structure entries (Akune et al., 2016). It were chosen for its mythical connotations into cardiology (Elbarouni et al., 2017), to discuss the anticipated benefits to the broader scientific and technical community. Researchers increasingly gain insights into the medical, health, and life science sectors. Notably, 2018 has brought unprecedented success for the biotech and health-care industry, with 16 companies earning the title of “unicorn”—a valuation mark of the 1 billion dollars (Wharton et al., 2019). Moreover, accompanying the growing number of health-care “unicorns” in 2019, including Babylon Health, Doctolib, and CMR Surgical, today’s “unicorns” have focused on defining interventions and services of patients’ health and wellbeing to achieve the rare and almost mythical accomplishment (The Lancet Digital Health, 2019).
Meanwhile, in informetrics, citation analysis (Garfield, 1979) contributed an effective method for estimating the number of important scientific achievements (Bonaccorsi, 2007; González-Betancor & Dorta-González, 2017). Balancing quality and quantity, the idea of “swan” (Zeng et al., 2017) provided an interesting interpretation of key scientific contributions in science, and the “swan groups” (Zhang, Zuccala, & Ye, 2019) contributed a broader metaphor of remarkable academic achievements in science and social science. In this article, we introduce the concept of informetric “unicorn” as a useful metaphor, which is different from the idea of “swan” and “swan group.” While the “swan” has both qualitative contents and quantitative citations, the “unicorn” focuses on the publications that have a very high impact of quantitative citations only, which may be a useful concept for identifying the rare and worthy works in science and technology.

2 Literature review

Citation analysis (Garfield, 1979; Meyer, 2000) as a methodology is always one of the most essential emphases in scientometrics, via its citation-based metrics, by analyzing scientific journals (Garfield, 1972; Guerrero-Bote & Moya-Anegón, 2012; Moed et al., 2012; Silva, 2016), scientific and technical papers (Kuan & Cheng, 2014; Narin, 1994), and authors (Grimwade & Garfield, 2002; Hirsch, 2010; Zou & Peterson, 2016), which reveal intrinsic characteristics, distribution rules, and significant influences. Bornmann and Mutz (2015) analyzed modern science’s growth rates by the data for the natural sciences, the medical and health sciences. In the authors’ analysis, Wang et al. (2018) tracked the scientific fame of great scientists in physics and revealed the greatest minds had gone but not forgotten. Meanwhile, for patents, Harhoff et al. (1999) found that the higher an invention’s economic value estimate was, the more the patent was subsequently cited. Generally, scientific and technical citation research evaluation (Leydesdorff, 2004; Persson, 1986; Tijssen, 2001) provides theoretical guidance and practical experience for national scientific and technical management and policy. However, the citations to papers and patents grow at different pace and pattern. Covering 1996–2000, Glänzel and Meyer (2003) assessed the frequency and characteristics of papers citing patents. The data source for papers was SCI, and for patents was USPTO, and the result showed that only 1.7% of all papers from SCI contain patent references, most of which were from periodicals. Using 4.8 million US patents and 32 million WoS research articles, Ahmadpoor and Jones (2017) found that about 80% of cited scientific publications (i.e. cited at least once by other scientific journals) eventually link forward to a future patent. Patents directly cited only 10% of cited scientific publications.
While citation analysis penetrates different disciplines, such as economics and management (Laengle et al., 2017; Merigo & Yang, 2017), sociology (Moed, Luwei, & Nederhof, 2002; White, Boell, & Yu, 2009; White, 2015), and biomedicine (Bornmann & Daniel, 2006; Comins & Leydesdorff, 2017; Garfield, 1991), highly-cited analysis shows its unique advantages to generate interesting and meaningful exploration in different discipline areas, where we mention that Bornmann and Leydesdorff (2015) did research in the domain of top-cited papers to investigate the development of the BRICS countries and scientific excellence.

On the one hand, researchers analyzed the content of highly-cited papers in the field of biomedicine. Davis and Cunningham (1990) suggested creative thought in neurosurgical research by citations analysis of journals. Bornmann and Marx (2014) gave information about a researcher’s productivity and the impact of their publications based on 10% percentiles of citations, who worked in the natural and life sciences. Moral-Munoz et al. (2018) and Perez-Cabezas et al. (2018) identified and conceptualized microbiology and rheumatology by highly cited papers. Ye et al. (2013) provided new insight into the relationship between Nobel awards and landmark papers in physiology or medicine. In terms of highly cited papers and Nobel Prize-winners’ discoveries are reasonably similar, Rodriguez-Navarro (2016) suggested that the United States’ research success was almost three times that of Europe, which had also published in *Nature* and *Science*.

On the other hand, researchers focused on the tendency of highly-cited papers in the field of biomedicine. Garfield (1979) looked at 37 “core” primary journals from 1968–1977 to track the trends in biochemical literature. Based on information extracted from the Science Citation Index database, he found that the biochemical literature was still growing faster than the scientific literature. There is a significant variation of citations for papers in different research fields (Bornmann & Daniel, 2008), e.g. biology papers, on average, receive a larger number of citations than mathematics papers. Ponomarev et al. (2014) studied Medline annual data sets for each of 1995–2004 and found that research fields related to medicine and biology had a stable citation threshold. Boyack et al. (2018) analyzed in-text citations of more than five million articles from PubMed Central Open Access Subset and Elsevier journals. They found that the reference distributions for biomedical and health sciences were more highly cited than other types of papers. However, citation analysis in medical patents is relatively simple. Huang, Zolnoori, and Balls-Berry (2019) analyzed more than 5 million US patent documents between 1995 and 2017, which provided a deep understanding of the focuses and trends of technological innovations in biomedicine.

In summary, serving as a functional linkage between ongoing scientific efforts with prior endeavors (de Solla Price, 1965; Garfiled, Malin, & Small, 1978; Radicchi Fortuno, & Castellano, 2008), citations (especially highly-cited papers) quantify the
scholar impact of research, assess the utility of scientific and technical achievements, originate outside of traditional medicine (e.g. medical principles and clinical treatments) and become three enabling biomedical innovation forces.

Therefore, we pay attention to highly-cited scientific papers and technical patents, to explore a new pattern and way to reveal highly-impact achievements.

3 Methodology

Essential Science Indicators (ESI; Essential Science Indicators, 2020) is a publication-and-citation-based research analytic tool provided by Clarivate Analytics, which delivers the in-depth coverage you need to effectively evaluate the impact of countries, ranks significant trends and top performers, analyze and benchmark research institutes (Csajbók et al., 2007; Fu et al., 2011; Harzing, 2015). Identifying for WoS-indexed item, the period for ESI counts is ten years. Based on a 10-year rolling file, highly-cited papers (Citation Thresholds, 2020; Highly Cited Papers, 2020) have a clear overview Family form ESI in the Clarivate website, which reflects the top 1% of papers by field and publication year. With consideration of comparative 10-year data for both papers and patents and the definition of “unicorn” in ten years, we chose a 10-year time window in the study.

Although, citations have been inflating over time (Persson, Glänzel, & Danell, 2004), and later papers on average are cited more than earlier ones. Comparing the scientific and technological impact of research, the “unicorn” will have an informetric feature as rarity and highly citations just in the first ten years after its publication, as in Figure 1. Empirically, most scientific discoveries happened before technical invention. Therefore, we set $T_{p1}$ as the publishing time of scientific paper and $T_{p2}$ as the technical patent’s publishing time. As patent applications should do after paper publication, the $T_{p2}$ should be the patent’s publishing time.

Figure 1. A designed model for informetric “unicorn.”
According to the model, the total citations \( C_T \) of the first 10 years after paper or patent published can be calculated as

\[
C_T = \int_{T_p}^{T_p + 10} C_p(t) dt ,
\]

where \( T_p \) is \( T_{p1} \) or \( T_{p2} \), and \( C_p \) denotes the citation curve of a scientific paper (\( P_s \)) or technical patent (\( P_n \)) in the publishing year. The Eq. (1) extended discrete citation counting to continuum variable for analysis.

According to our design, \( C_T \) should be significantly large, with much more than average level, so that we had to set it as very high citations. As papers received more citations than patents, we also had to differentiate paper and patent. As a result, selecting from the most recent ten years of data, we also proposed an approach as less than 1% of highly-cited papers (top 1% of papers) to quantify how much rarity and worth papers can be considered as “unicorn,” which is to see where science and technology are going and who’s leading the way. With consideration of field-independent computation, we applied absolute values as \( C_T \geq 5,000 \) citations for scientific papers and \( C_T \geq 500 \) citations for technical patents in ten years empirically, replacing relative 1% highly-cited ones. Then we would like to introduce the following definitions.

**Definition 1. Scientific “unicorn”:** a scientific “unicorn” is a publication that received \( C_T \geq 5,000 \) citations in ten years after publication, with an increasing citation curve in the first two years as the start-up emerging.

**Definition 2. Technical “unicorn”:** a technical “unicorn” is a patent that received \( C_T \geq 500 \) citations in ten years after it published, with an increasing citation curve in the first two years as the start-up emerging.

The definitions are field-independent. When we use these definitions for practical computation, about 50% “unicorns” fall in biomedicine (see Results).

Since we meet an increasing citation curve, it is reasonable to discuss the models of citation curves. Price (1963) established the exponential growth model of scientific and technical publications. However, we knew that rare or important documents increased linearly. American science historian and intelligence scientist Rescher (1978) also proposed a hierarchical sliding index as the mathematical model (Zhang, Vogeley, & Chen, 2011) by describing the rule of lengthening in scientific and technical publications with introducing a valued index \( \lambda \) in the book of Scientific Progress. When \( \lambda = 1 \), it means the entire literature; \( \lambda = 3/4 \), it represents meaningful literature; \( \lambda = 1/2 \), it means high important Meaningful literature; \( \lambda = 1/4 \), it represents very high important literature.

\[
C_p(t) = (ae^{bt})^\lambda , b > 0, \lambda \in [0,1].
\]
When $\lambda=0$, it means the first class important literatures, and the law of exponential growth is broken. Eq. (2) is defined as linear relation as $C_p(t) = \ln C_0 + bt$. $C_0$ is the number of publications at the start of statistics. Therefore, we payed attention to the linear model only.

$$C_p(t) = a + bt, b > 0,$$

(3)

where $a$ is the amount at the start point, and $C_p(t)$ is the citations in the $t$ year after publishing, $b$ is the growth coefficient, and $t$ is the time.

In this paper, we processed data with fixed effects to reduce the effect of different publications and then established a linear regression model using programs by Python and R. The ordinary least squares (OLS; Bertoli-Barsotti & Tommaso, 2019) was used. For comparing the error between the linear model and the actual acquisition data, root means squared error (RMSE) was calculated according to the following formula

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (Predicted_i - Empirical_i)^2}{N}},$$

(4)

in which $N$ is the total number of data.

For measuring the ratio of “unicorn” in a field every year, we introduce the unicorn-ratio ($Ur$) as an index as follows

$$Ur = \frac{P_T}{P} \times 100\%,$$

(5)

where $P$ is the total publications and $P_T$ is the “unicorn” papers in the field. It is expected that $Ur$ is a very low ratio.

3.1 Data and data processing

Empirical data came from WoS and Derwent Innovations Index (DII). We selected and searched data by restricting citations in WoS from 2000 to 2012 until November 2, 2019. Meanwhile, we also used Python to crawl the patent with the restricted number of citations from 2000 to 2012 until November 19, 2019. The research direction was a set of classification methods used by all product databases under WoS. There were five research directions for papers in WoS: Arts & Humanities, Biochemistry & Molecular biology, Natural sciences, Social sciences, and Applied sciences (Research areas, 2020). We also found the classification basis through the International Patent Classification (IPC) in DII, A61 stands for patents in Medical OR Veterinary Science; Hygiene. Whether papers (Biochemistry & Molecular biology) or patents (Medical OR Veterinary Science; Hygiene) were directly or
indirectly related to the biomedical fields. As a result, the distribution of publications in the research areas were obtained from WoS by its classification (WC=Biochemistry & Molecular biology), and data from DII using the features of IPC=A61. The total number of citations was required from the publication year up to the first ten years. Apart from biomedical fields, other unicorns were classified by WC for papers and IPC for patents in the WoS Categories field. According to WC and IPC, we selected unicorns in different disciplines from 2001–2012. The citations were also required from the publication year up to the date of the first ten years.

In informetrics, we can provide the expression for the distribution of citations in terms of the time. Based on objective data from WoS and DII, how to evaluate the characteristics of the growth to the evolution of scientific and technical “unicorn” publications as a whole? Thus, under formula conversion, empirical values coming from WoS and DII were calculated in models. We used the Linear regression analysis for optimization, which has been used in biomedicine (Karakülah et al., 2019), finance, investing, and other disciplines. Regression analysis attempts to determine the strength of the relationship between one dependent variable (usually denoted by Y) and a series of other changing variables (known as independent variables).

4 Results

The absolute number and the relative percentage of both scientific and technical “unicorns” show rarity, as shown in Table 1.

| Year | WoS total papers | Absolute unicorn number | Ur: Relative unicorn ratio (%) | DII total patents | Absolute unicorn number | Ur: Relative unicorn ratio (%) |
|------|------------------|-------------------------|-------------------------------|------------------|-------------------------|-------------------------------|
| 2000 | 874,542          | 2                       | 0.0002                        | 672,139          | 33                      | 0.0049                        |
| 2001 | 872,370          | 5                       | 0.0006                        | 729,168          | 21                      | 0.0029                        |
| 2002 | 889,519          | 4                       | 0.0004                        | 786,869          | 12                      | 0.0015                        |
| 2003 | 927,047          | 6                       | 0.0006                        | 791,486          | 16                      | 0.0020                        |
| 2004 | 967,440          | 11                      | 0.0011                        | 824,382          | 13                      | 0.0016                        |
| 2005 | 1,016,231        | 6                       | 0.0006                        | 893,139          | 22                      | 0.0025                        |
| 2006 | 1,070,302        | 8                       | 0.0007                        | 924,045          | 17                      | 0.0018                        |
| 2007 | 1,122,363        | 13                      | 0.0012                        | 1,085,298        | 30                      | 0.0028                        |
| 2008 | 1,201,425        | 15                      | 0.0012                        | 1,163,732        | 41                      | 0.0035                        |
| 2009 | 1,251,718        | 30                      | 0.0024                        | 1,224,936        | 12                      | 0.0010                        |
| 2010 | 1,294,375        | 28                      | 0.0022                        | 1,384,960        | 3                       | 0.0002                        |
| 2011 | 1,373,399        | 16                      | 0.0012                        | 1,473,876        | 1                       | 0.0001                        |
| 2012 | 1,441,144        | 21                      | 0.0015                        | 1,774,920        | 3                       | 0.0002                        |
| Total | 14,301,875       | 165                     | 0.0012                        | 13,728,950       | 224                     | 0.0016                        |
The results show the proportion of “Biomedical unicorns” from Table 2 in the number of “Absolute unicorn number” from Table 1. There is an apparent disciplinary bias that the ratio of biomedical “unicorns” is respectively 57.58% (95/165) in WoS and 47.32% (106/224) in DII, which means that disciplinary distribution of “unicorns” is asymmetric and biomedical “unicorns” occupied almost 50%.

Table 2 shows the biomedical scientific and technical “unicorns,” where we see that the annual distribution is not average, it changes over time yearly.

Table 2. The annual distribution of biomedical “unicorns” (2000–2012).

| Year/Biomedical unicorns | Scientific papers | | Technical patents |
|--------------------------|-------------------|-------------------|-------------------|
|                          | Absolute number   | Relative ratio (%)| Absolute number   | Relative ratio (%)|
| 2000                     | 2                 | 0.0002            | 5                 | 0.0007            |
| 2001                     | 2                 | 0.0002            | 6                 | 0.0008            |
| 2002                     | 2                 | 0.0002            | 1                 | 0.0001            |
| 2003                     | 4                 | 0.0004            | 1                 | 0.0001            |
| 2004                     | 6                 | 0.0006            | 1                 | 0.0001            |
| 2005                     | 3                 | 0.0003            | 9                 | 0.0010            |
| 2006                     | 3                 | 0.0003            | 10                | 0.0011            |
| 2007                     | 7                 | 0.0006            | 24                | 0.0022            |
| 2008                     | 5                 | 0.0004            | 35                | 0.0030            |
| 2009                     | 19                | 0.0015            | 10                | 0.0008            |
| 2010                     | 18                | 0.0014            | 3                 | 0.0002            |
| 2011                     | 8                 | 0.0001            | 0                 | 0.0000            |
| 2012                     | 16                | 0.0002            | 1                 | 0.0001            |
| Total                    | 95                | 0.0007            | 106               | 0.0008            |

Except for the discipline of biomedicine, science and technology “unicorns” in other fields resemble rare. Here we concluded the “unicorns” distribution of other top 5 disciplines in Table 3, where we see that the values are lower than the biomedical “unicorns.”

Individually, most “unicorns” are important leading papers or patents in science or technology, in which we selected representative one case per year for ten years, listed in Tables 4 and 5, respectively.

In Table 4, there are some breakthrough scientific achievements which belong to biomedical “unicorns.” For example, “Initial sequencing and analysis of the human genome” (Lander et al., 2001) is a famous initial paper on the human genome, Takahashi and Yamanaka (2006) reported their discovery of induced pluripotent stem (iPS) cells, and successfully applied to human cells (Takahashi et al., 2007), which led to Shinya Yamanaka’s winning the Nobel Prize for Physiology or Medicine in 2012.
Table 3. The scientific and technical “unicorns” in other five disciplines (2000--2012).

| Year | Scientific papers (absolute number) | Technical patents (absolute number) |
|------|-------------------------------------|-------------------------------------|
|      | Chemistry | Multidisciplinary Sciences | Computer Science | Physics | Astronomy & Astrophysics | Computing & Calculating & Counting | Basic Electric Elements | Agriculture | Electric Communication Technique | Sports & Games & Amusement |
| 2000 | 0         | 0                        | 0                  | 0       | 0                      | 16                                 | 1                           | 0          | 4                          | 4                          |
| 2001 | 0         | 2                        | 1                  | 0       | 0                      | 9                                  | 0                           | 0          | 2                          | 1                          |
| 2002 | 1         | 0                        | 0                  | 1       | 0                      | 6                                  | 0                           | 0          | 0                          | 0                          |
| 2003 | 1         | 0                        | 0                  | 0       | 1                      | 5                                  | 4                           | 1          | 0                          | 0                          |
| 2004 | 2         | 1                        | 2                  | 0       | 0                      | 0                                  | 4                           | 5          | 1                          | 0                          |
| 2005 | 0         | 3                        | 0                  | 0       | 0                      | 2                                  | 3                           | 5          | 0                          | 0                          |
| 2006 | 1         | 1                        | 2                  | 1       | 0                      | 2                                  | 4                           | 0          | 0                          | 0                          |
| 2007 | 4         | 1                        | 0                  | 0       | 0                      | 2                                  | 3                           | 0          | 0                          | 0                          |
| 2008 | 6         | 2                        | 0                  | 0       | 0                      | 3                                  | 2                           | 0          | 0                          | 0                          |
| 2009 | 3         | 4                        | 2                  | 2       | 0                      | 2                                  | 0                           | 0          | 0                          | 0                          |
| 2010 | 8         | 0                        | 0                  | 1       | 0                      | 0                                  | 0                           | 0          | 0                          | 0                          |
| 2011 | 1         | 1                        | 1                  | 1       | 2                      | 1                                  | 0                           | 0          | 0                          | 0                          |
| 2012 | 0         | 3                        | 0                  | 0       | 2                      | 2                                  | 0                           | 0          | 0                          | 0                          |
| Total| 27        | 18                       | 8                  | 6       | 5                      | 50                                 | 21                          | 11         | 7                          | 5                          |
| Code  | Title                                                                 | Author(s)                                                                 | Source                                                                 |
|-------|----------------------------------------------------------------------|--------------------------------------------------------------------------|----------------------------------------------------------------------|
| Pr1   | Initial sequencing and analysis of the human genome                  | Lander, ES et al.                                                         | NATURE 2001, 409(6822):860                                          |
| Pr2   | Reduction in the incidence of type 2 diabetes with lifestyle intervention of metformin | Knowler, WC; Barrett-Connor, E; Fowler, SE et al.                         | NEW ENGLAND JOURNAL OF MEDICINE 2002, 346(6):293-303.               |
| Pr3   | Measuring inconsistency in meta-analyses                            | Higgins, JPT; Thompson, SG; Deeks, JJ et al.                             | BRITISH MEDICAL JOURNAL OF MEDICINE 2003, 327(7414):557-560.       |
| Pr4   | MicroRNAs: Genomics, biogenesis, mechanism, and function             | Bartel, DP                                                                | NATURE 2004, 116(2):281-297.                                        |
| Pr5   | Arlequin (version 3.0): an integrated software package for population genetics data analysis | Excoffier, Laurent; Laval; Guillaume; Schneider; Stefan                    | EVOLUTIONARY BIOINFORMATICS 2005, 1:47-50.                            |
| Pr6   | Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors | Takahashi, Kazutoshi; Yamakawa, Shinya; Takahashi, Koji et al.             | CELL 2006, 126(4):663-676.                                           |
| Pr7   | Induction of pluripotent stem cells from adult human fibroblasts by defined factors | Takahashi, Kazutoshi; Yamakawa, Shinya; Takahashi, Koji et al.             | CELL 2007, 131(5):861-872.                                           |
| Pr8   | Analyzing real-time PCR data by the comparative C-T method          | Schmittgen, Thomas D.; Livak, Kenneth J.                                  | NATURE PROTOCOLS 2008, 3(6):1101-1108.                               |
| Pr9   | MicroRNAs: Target Recognition and Regulatory Functions               | Bartel, David P.                                                          | CELL 2009, 136(2):215-233.                                          |
| Pr10  | PHENIX: a comprehensive Python-based system for macromolecular structure solution | Adams, Paul D.; Afonine, Pavel V.; Bunkoczi, Gabor et al.                 | ACTA CRYSTALLOGRAPHICA SECTION D-STRUCTURAL BIOLOGY 2010, 66(2):213-221. |
Table 5. A selected top 10 technical patents of biomedical “unicorns” (2001–2010).

| Code | Title | Inventor(s) | Patent No. | CT |
|------|-------|-------------|------------|----|
| Pn1  | Producing humanized immunoglobulin, involves producing a cell containing DNA segments encoding humanized heavy and light chain variable regions, and expressing the DNA segments in the cell | QUEEN C L; SELICK H E | US6180370-B1; 2001 | 542 |
| Pn2  | Analyte level monitoring device for diabetes treatment, has transmitter arranged on substrate of electrochemical sensor, for transmitting signal indicating analyte level in bodily fluid | HELLER A; DRUCKER S M; JIN R Y; FUNDERBURK J V; et al. | WO200258537-A2; US2003100821-A1; US6560471-B1; 2002 | 514 |
| Pn3  | Spinal cord stimulation system includes surgical components, which consist of insertion needle and tunneling tools to aid implantation of electrode array and lead extension | MEADOWS P; MAN MANN C M; PETERSON D K; et al. | US6516227-B1; 2003 | 674 |
| Pn4  | Surgical stapling instrument for laparoscopic and endoscopic clinical procedure has firing device that has a distally presented cutting edge longitudinally received between the elongated channel and the anvil | SHELTON IV F E; SETSER M E; HEMMELE GARN B J; et al. | EP1479349-A1; US2004232196-A1; CA2467795-A1; 2004 | 514 |
| Pn5  | Surgical instrument for endoscopically inserting end effector, e.g. endo-cutter, grasper, cutter, and staplers, includes articulation control comprising actuator, and motion conversion mechanism | WALES K S | US2005006430-A1; CA2473482-A1; JP2005028148-A; 2005 | 717 |
| Pn6  | Surgical instrument e.g. endo-cutter for use during fastening of buttress pads to tissue, comprises staple applying assembly attached to elongate shaft, which includes opposing tissue compression surfaces | SHELTON F E; SHELTON F; WALES K S; et al. | EP1621141-A2; JP2006043451-A; US2006025816-A1; 2006 | 722 |
| Pn7  | Medical device e.g. surgical stapler, for e.g. stapling tissue, has articulation joint actuator to hold passive joint and end effector in fixed articulation state during unactuated state and to release joint during actuated state | SMITH K W; PALMER M A; KLINE K R; et al. | US2007187453-A1; US7404508-B2; AU2015201382-B2; 2007 | 817 |
| Pn8  | Surgical stapling apparatus includes drive assembly that is supported in the tool assembly and which has a knife blade disposed in an elongated longitudinal slot formed by the anvil plate and the staple cartridge | TARINELLI D; ARANYI E; SIMPSON R; et al. | WO2008109125-A1; US2009134200-A1; AU2008223389-A1; 2008 | 674 |
| Pn9  | Disposable loading unit for endoscopic surgical stapling instrument for incising fastened tissue, has anvil portion that is provided with staple-deforming cavity, and cover plate is secured for supporting anvil portion | ARMSTRONG G A; BLAIR G B; BRUEWER D B; et al. | EP2090235-A2; US2009206140-A1; CN101507634-A; 2009 | 571 |
| Pn10 | Motor e.g. stepper motor, driven surgical cutting and fastening instrument i.e. endoscopic instrument, for use by e.g. physician for endoscopic application, has motor with operational modes for portions of cutting stroke cycle of instrument | LAURENT R J; SHELTON F E; SMITH B W; et al. | EP2165664-A2; JP2010075694-A; US2010076474-A1; 2010 | 675 |
In Table 5, most technical biomedical “unicorns” belong to surgical instruments and their appendages, where 91 technical patents are directly or indirectly related to ETHICON ENDO-SURGERY, INC., which indicates that its sewing products are the market leader in the world. Meanwhile, Shelton was a famous inventor in the company. He invented a surgical stapling instrument for the laparoscopic and endoscopic clinical procedure (Hemmelgarn, Setser, & Shelton, 2004), endo-cutter for use during fastening of buttress pads to tissue (Shelton et al., 2006), and a stepper motor with operational modes for surgical cutting and fastening instrument in 2010.

After 2010, the rapid development of computer and network technologies had widely affected the biomedical field. Both “unicorns” in scientific papers and technical patents become much more “technical.” For example, Molecular Evolutionary Genetics Analysis (MEGA) (Tamura et al., 2011) has used statistical methods (Maximum Likelihood, Evolutionary Distance, Maximum Parsimony, Bayesian, and so on) to analyze sequence alignment from 2004 to 2011. The ImageJ (Schneider, Rasband, & Eliceiri, 2012) website got about 7,000 visitors a day, and there were about 1,900 subscribers to the ImageJ mailing list. PHENIX can save significant time and effort, which has provided a comprehensive Python-based system for macromolecular crystallographic structure solution, emphasizing on automation of all procedures instead of traditional performing by hand. Finally, the Pearson correlation coefficient is tested for patent families and citations by SPSS 23, which find $r=0.140$ and $p=0.164$ in the confidence interval of 95%. So, the relationship between patent families and citations is almost irrelevant from 2001 to 2010.

5 Analysis and discussion

According to linear model Eq. (3), we substitute the real yearly values into the regression model for fitting, the linear equation supports more powerful in scientific (red) and technical (blue) “unicorns,” with results are showed in Figure 2.

In the regression analysis, the statistical sign is significant at $p<0.01$. All the parameters shown in Table 6.

When we take the data back to the linear mathematical model, we calculate all in the 95% confidence interval, as shown in Table 7. The first ten records of the prediction results are selected and kept in Table 6, and the RMSE of the scientific “unicorn” is 0.2127, while the RMSE of technical “unicorn” is 0.0936.
Table 6. The fitting parameters of scientific and technical “unicorns” (p<0.01).

| Code | Scientific papers | Technical patents |
|------|-------------------|-------------------|
|      | Eq. (3)           | Eq. (2)           | Eq. (3)           | Eq. (2)           |
| B    | 121.364           | 121.364           | 8.057             | 8.057             |
| a    | 19.895            | 148.005           | 8.988             | 13.196            |
| Obs  | 805               | 805               | 1008              | 1008              |
| $R^2$| 0.650             | 0.642             | 0.541             | 0.522             |
| $R^2_{adj}$ | 0.609          | 0.600             | 0.490             | 0.469             |
| RSE  | 506.307           | 506.307           | 24.372            | 24.372            |
| F    | 15.897            | 15.362            | 10.562            | 9.793             |

Table 7. A comparison of theoretical (predictive) and empirical values.

| Code | Scientific papers | Technical patents |
|------|-------------------|-------------------|
|      | theoretical       | empirical         | theoretical       | Empirical         |
| P1   | 7,205             | 8,204             | 453               | 593               |
| P2   | 6,755             | 6,746             | 444               | 554               |
| P3   | 13,028            | 8,725             | 426               | 617               |
| P4   | 11,354            | 5,813             | 417               | 511               |
| P5   | 7,619             | 7,062             | 426               | 556               |
| P6   | 7,286             | 5,151             | 462               | 832               |
| P7   | 6,035             | 8,339             | 498               | 542               |
| P8   | 7,421             | 7,661             | 543               | 751               |
| P9   | 7,367             | 8,688             | 516               | 566               |
| P10  | 6,089             | 6,463             | 444               | 507               |
| RMSE | 0.2127            | 0.2127            | 0.0923            | 0.0923            |

Figure 2. Fitting curves of the linear model for scientific and technical “unicorns.”
For any \( b > 0 \), according to Eq. (3), we can estimate theoretically

\[
C_T = \int_0^{10} (a + bt)dt = at + \frac{1}{2}bt^2 \bigg|_0^{10}.
\]  

The quadratic curve indicates conic growth, which means that the total citation curve of “unicorns” will be quickly increasing.

Also, we mention the limitations of this research. As this is a purely quantitative study, we do not know the real quality of “unicorns.” Similarly, we do not know whether the company holding “unicorn” patents will necessarily become a “unicorn” company. Comparing with coupled patents (Kuan, Chen, & Huang, 2019), we hope to learn more via patent analysis in the future.

6 Conclusion

By considering informetric quantity only, we suggest the model for finding scientific unicorn (\( C_T \geq 5,000 \) in 10 years) and technical “unicorns” (\( C_T \geq 500 \) in 10 years), which may be a useful concept for identifying rare and very high impact works in science and technology, particularly in biomedicine.

During 2000–2012, there are 165 scientific “unicorns” in 14,301,875 WoS papers, with ratio 0.0012%, and there are 224 technical “unicorns” in 13,728,950 DII patents, with ratio 0.0016%, in which the rate of biomedical “unicorns” are respectively 57.58% in WoS and 47.32% in DII. The rare “unicorns” increased following linear model, the fitting data show 95% confidence with the RMSE of scientific “unicorn” is 0.2127 in WoS while the RMSE of technical “unicorn” is 0.0923 in DII.

Finally, it would be interesting and significant to explore “potential unicorns” on \( C_T \) near 5,000 for papers and \( C_T \) near 500 for patents, which could also belong to remarkable discoveries in scientific and technical fields. The proportion of reduced \( C_T \) is less than 10%. We remain “potential unicorns” for future studies.

Competing interests

The authors declare no competing interests.

Acknowledgements

The authors acknowledge National Natural Science Foundation of China Grant No. 71673131, and Jiangsu Key Laboratory Fund, as well as support from the International Joint Informatics Laboratory operated cooperatively by the University of Illinois at Urbana-Champaign, USA and Nanjing University, China.
Author contributions

Lucy L. Xu (xululu@ntu.edu.cn) collected and processed data with modelling, Miao Qi (161070055@smail.nju.edu.cn) assisted literature review, and Fred Y. Ye (yye@nju.edu.cn) initiated the idea and wrote the paper.

References

Ahmadpoor, M., & Jones, B.F. (2017). The dual frontier: Patented inventions and prior scientific advance. Science, 357(6351), 583–587.

Akune, Y., Lin, C.H., Abrahams, J.L., Zhang, J.Y., Packer, N.H., Aoki-Kinoshita, K.F., & Campbella, M.P. (2016). Comprehensive analysis of the N-glycan biosynthetic pathway using bioinformatics to generate UniCorn: A theoretical N-glycan structure database. Carbohydrate Research, 431, 56–63.

Bartel, D.P. (2004). MicroRNAs: Genomics, biogenesis, mechanism, and function. Cell, 116(2), 281–297.

Bertoli-Barsotti, L., & Tommaso, L. (2019). How mean rank and mean size may determine the generalised Lorenz curve: With application to citation analysis. Journal of Informetrics, 13(1), 387–396.

Bonaccorsi, A. (2007). Explaining poor performance of European science: Institutions versus policies, Science and Public Policy, 34, 303–316.

Bornmann, L., & Daniel, H.D. (2006). Selecting scientific excellence through committee peer review—A citation analysis of publications previously published to approval or rejection of post-doctoral research fellowship applicants. Scientometrics, 68(3), 427–440.

Bornmann, L., & Daniel, H. (2008). What do citation counts measure? A review of studies on citing behavior. Journal of Documentation, 64, 45–80.

Bornmann, L., & Marx, W. (2014). How to evaluate individual researchers working in the natural and life sciences meaningfully? A proposal of methods based on percentiles of citations. Scientometrics, 98(1), 487–509.

Bornmann, L., & Mutz, R. (2015). Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. Journal of the Association for Information Science & Technology, 66(11), 2215–2222.

Bornmann, L., Wagner, C., & Leydesdorff, L. (2015). BRICS countries and scientific excellence: A bibliometric analysis of most frequently cited papers. Journal of the Association for Information Science & Technology, 66(7), 1507–1513.

Boyack, K.W., Eck, N.J., Colavizza, G., & Waltman, L. (2018). Characterizing in-text citations in scientific articles: A large-scale analysis. Journal of Informetrics, 12(1), 59–73.

Casanova, L., Cornelius, P.K., & Dutta, S. (2018). Financing entrepreneurship and innovation in emerging markets. Salt Lake city: Academic Press, 185–218.

Cbinsights. (2020). The Global Unicorn Club Current Private Companies Valued At $1B+ (including whisper valuations). Retrieved from https://www.cbinsights.com/research-unicorn-companies.

Citation Thresholds. (2020). archive.sciencewatch.com/ [Online]. Retrieved from: http://archive.sciencewatch.com/about/met/thresholds/
Comins, J.A., & Leydesdorff, L. (2017). Citation algorithms for identifying research milestones driving biomedical innovation. Scientometrics, 110(3), 1495–1504.

Csajbók, E., Berhidi A., Vasas, L., & Schubert, András. (2007). Hirsch-index for countries based on essential science indicators data. Scientometrics, 73(1), 91–117.

Cugurullo, F., Datta, A., & Shaban, A. (2016). Mega-urbanization in the global south: Fast cities and new urban utopias of the postcolonial state. New York: Routledge, 23(3), 66–80.

Davis, R.A., & Cunningham, P.S. (1990). Creative thought in neurosurgical research: The value of citation analysis. Neurosurgery, 26(2), 345–353.

de Solla Price, D.J. (1965). Networks of scientific papers. Science, 149(3683), 510–515.

Elbarouni, B., Ducas, J., Friesen, D., & Zhang, H. (2017). The uninterrupted anticoagulation in coronary catheterization (unicorn) registry. A single center experience. Canadian Journal of Cardiology, 33(10), S144–S145.

Essential Science Indicators. (2020). clarivate.com. Retrieved from: https://clarivate.com/products/essential-science-indicators/.

Fu, H.Z., Chuang, K.Y., Wang, M.H., & Ho, Y.A. (2011). Characteristics of research in China assessed with essential science indicators. Scientometrics, 88(5), 841–862.

Karaküläh, G., Arslan, N., Yandım, C., & Suner, A. (2019). TEffect R: An R package for studying the potential effects of transposable elements on gene expression with linear regression model. Peer J, 7, 6–20.

Garfield, E. (1972). Citation analysis as a tool in journal evaluation: Journals can be ranked by frequency and impact of citations for science policy studies. Science, 178(4060), 471.

Garfield, E. (1979). Citation Indexing: Its theory and application in science, technology and humanities. New York: Wiley.

Garfield, E. (1979). Is citation analysis a legitimate evaluation tool? Scientometrics, 1(4), 359–375.

Garfield, E. (1979). Trends in biochemical literature. Trends in biochemical literature, 4(12), N290.

Garfield, E. (1991). A citation analysis of Austrian medical-research and wiener-klinische wochenschrift. Wiener Klinische Wochenschrift, 103(11), 318–325.

Garfield, E., Malin, M.V., & Small, H. (1978). Citation data as science indicators. In Y. Elkana, J. Lederberg, R.K. Merton, A. Thackray, & H. Zuckerman (Eds.), Toward a metric of science: The advent of science indicators, 179–207. New York: Wiley.

González-Betancor, S.M., & Dorta-González, P. (2017). An indicator of the impact of journals based on the percentage of their highly cited publications. Online Information Review, 41, 398–411.

Glänzel, W., & Meyer, M. (2003). Patents cited in the scientific literature: An exploratory study of ‘reverse’ citation relations. Scientometrics, 58(2), 415–428.

Grimwade, A., & Garfield, E. (2002). The Scientist on the Web. Scientist, 16(16), 10.

Guerrero-Bote, V.P., & Moya-Anegón, F. (2012). A further step forward in measuring journals’ scientific prestige: The SJR2 indicator. Journal of Informetrics, 6(4), 674–688.

Harzing, A.W. (2015). Health warning: Might contain multiple personalities—The problem of homonyms in Thomson Reuters essential science indicators. Scientometrics, 105(3), 2259–2270.

Harhoff, D., Narin, F., Scherer, F.M., & Vopel, K. (1999). Citation frequency and the value of patented inventions. Review of Economics and Statistics, 81(3), 511–515.
Hemmelgarn, B.J., Setser, M.E., & Shelton, IV.E. (2004). Surgical stapling instrument for laparoscopic and endoscopic clinical procedure has firing device that has a distally presented cutting edge longitudinally received between the elongated channel and the anvil. EP 1479349-A1, 2004-11-24. Retrieved from https://worldwide.espacenet.com/patent/search/family/042229108/publication/EP1479349A1?q=EP1479349A1

Highly Cited Papers. (2020). archive.sciencewatch.com. Retrieved from http://archive.sciencewatch.com/about/met/core-hcp.

Hirsch, J.E. (2010). An index to quantify an individual’s scientific research output that takes into account the effect of multiple co-authorship. Scientometrics, 85(3), 741–754.

Huang, M., Zolnoori, M., & Balls-Berry, J.E. (2019). Technological innovations in disease management: Text mining us patent data from 1995 to 2017. Journal of Medical Internet Research, 21(4), 1–7.

Kuan, C.H., Chen, D.Z., & Huang, M.H. (2019). Bibliographically coupled patents: Their temporal pattern and combined relevance. Journal of Informetrics, 13(4), 100978.

Kuan, C.H., & Cheng, H.J. (2014). Do we miscount patent citations? An empirical study on the impact of overlooking the citations to a patent’s pre-grant publication. International Conference on Industrial Engineering and Engineering Management, 1034–1037.

Laengle, S., Merigo, J.M., Miranda, J., Słowiński, R., Bomze, I., Borgonovo, E., Dysone, R.G., Oliveira, J.F., & Teunterg, R. (2017). Forty years of the European Journal of Operational Research: A bibliometric overview. European Journal of Operational Research, 262(3), 803–816.

Lander, E.S., Lauren, L.M., Birren, B., Chad, N., Michael, C.Z., Baldwin, J., Devon, K., Dewar, K., Doyle, M., & Fitzhugh, W. (2001). Initial sequencing and analysis of the human genome. Nature, 409(6822), 860–921.

Leydesdorff, L. (2004). The evaluation of research and the scientometric research program: Historical evolution and redefinitions of the relationship. Studies in Science of Science, 22(3), 225–232.

Lee, E. (2013). Welcome to the unicorn club: Learning from billion-dollar startups. Retrieved from https://techcrunch.com/2013/11/02/welcome-to-the-unicorn-club/

Merigo, J.M., & Yang, J.B. (2017). A bibliometric analysis of operations research and management science. Omega-International Journal of Management Science, 73, 37–48.

Meyer, M. (2000). What is special about patent citations? Differences between scientific and patent citations. Scientometric, 49(1), 93–123.

Moed, H.F., Colledge L., Reedijk, J., Moya-Anegon, Felix., Guerrero-Bote, V., Plume, Andrew, & Amin, M. (2012). Citation-based metrics are appropriate tools in journal assessment provided that they are accurate and used in an informed way. Scientometrics, 92(2), 367–376.

Moed, H.F., Luwei, M., & Nederhof, A.J. (2002). Towards research performance in the humanities. Library Trends, 50(3), 498–520.

Moral-Munoz, J.A., Lucena-Anton, D., Perez-Cabezas, V., Carmona-Barrientos, I., González-Medina, G., & Ruiz-Molinero, C. (2018). Highly cited papers in Microbiology: Identification and conceptual analysis. Fems Microbiology Letters, 365, 20.

Narin, F. (1994). Patent bibliometrics. Scientometrics, 30(1), 147–155.
Identifying Scientific and Technical “Unicorns”

Perez-Cabézas, V., Ruiz-Molinero, C., Carmona-Barrientos, I., Herrera-Viedma, E., Cobo, M.J., & Moral-Munoz, J.A. (2018). Highly cited papers in rheumatology: Identification and conceptual analysis. Scientometrics, 116 (1), 555–568.

Persson, O. (1986). Online bibliometrics—a research tool for every man. Scientometrics, 10(1–2), 69–75.

Ponomareva, I.V., Williams, D.E., Hackett C.J., Schnell, J.D., & Haak, L.L. (2014), Predicting highly cited papers: A Method for Early Detection of Candidate Breakthroughs. Technological Forecasting and Social Change, 81, 49–55.

Price, D.S. (1963). Little science, big science. New York: Columbia University Press.

Radicchi, F., Fortuno, S., & Castellano, C. (2008). Universality of citation distributions: Toward an objective measure of scientific impact. Proceedings of the National Academy of Sciences, 105(45), 17268–17272.

Rebentisch, H., Thompson, C., Côté-Royc, L., & Moserc, S. (2020). Unicorn planning: Lessons from the rise and fall of an American ‘smart’ mega-development. Cities, 101, 102686–102692.

Research areas. (2020). images.webofknowledge.com. Retrieved from: http://images.webofknowledge.com//WOKRS535R52/help/zh_CN/WOS/hp_research_areas_easca.html.

Rodriguez-Navarro, A. (2016). Research assessment based on infrequent achievements: A comparison of the United States and Europe in terms of highly cited papers and Nobel Prizes. Journal of the Association for Information Science & Technology, 67(3), 731–740.

Ruiz-Trillo, I., Burger, G., Holland, P.W.H., King, N., Lang, B.F., Roger, A.J., & Gray, M.W. (2007). The origins of multicellularity: A multi-taxon genome initiative. Trends in Genetics, 23(3), 113–118.

Schneider, C., Rasband, W.S., & Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nature Methods, 9(7), 671–675.

Shelton, IV.F.E. (2006). Surgical instrument e.g. endo-cutter for use during fastening of buttress pads to tissue, comprises staple applying assembly attached to elongate shaft, which includes opposing tissue compression surfaces. EP 1621141-A2, 2006-02-01. Retrieved from https://worldwide.espacenet.com/patent/search/family/035285352/publication/EP1621141A2?q=EP1621141A2

Silva, M.R. (2016). Journal impact factors for the year-after the next can be objectively predicted. Medical Express, 3(5), M160506.

Takahashi, K., & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell, 126(4), 663–676.

Takahashi, K. Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., & Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell, 131(5), 861–872.

Tamura, K., Peterson, D., Peterson, N., Stecher, Glen., Nei, M., & Kumar, S. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution, 28(10), 2731–2739.

The Lancet Digital Health. (2019). Unicorns and cowboys in digital health: The importance of public perception. The Lancet Digital Health, 1(7), e319.

Tijssen, R.J.W. (2001). Global and domestic utilization of industrial relevant science: Patent citation analysis of science-technology interactions and knowledge flows. Research Policy, 30(1), 35–54.
Wang, G.Y., Hu, G.Y., Li, C.F., & Tangd, L. (2018). Long live the scientists: Tracking the scientific fame of great minds in physics. Journal of Informetrics, 12(4), 1089–1098.
White, H.D. (2015). Co-cited author retrieval and relevance theory: Examples from the humanities. Scientometrics, 102(3), 2275–2299.
Wharton, G.A., Sood, H.S., Sissons, A., & Mossialos, E. (2019). Virtual primary care: Fragmentation or integration? The Lancet Digital Health, 1(7), e330–e331.
White, H.D., Boell, S.K., & Yu, H. (2009). Libcitations: A measure for comparative assessment of Book Publications in the Humanities and Social Sciences. Journal of The American Society for Information Science & Technology, 60(6), 1083–1096.
Wikipedia. (2020). Unicorn From Wikipedia, the free encyclopedia. Retrieved from https://en.wikipedia.org/wiki/Unicorn.
Wikipedia. (2020). Unicorn (finance) From Wikipedia, the free encyclopedia. Retrieved from https://en.wikipedia.org/wiki/Unicorn_(finance).
Ye, S.Q., Xing, R., Liu, J., & Xing, F.Y. (2013). Bibliometric analysis of Nobelists’ awards and landmark papers in physiology or medicine during 1983–2012. Annals of Medicine, 45(8), 532–538.
Zeng, C.J., Qi, E. P, Li, S.S. Stanley, H.E., & Ye, F.Y. (2017). Statistical characteristics of breakthrough discoveries in science using the metaphor of black and white swans. Physica A, 487, 40–46.
Zhang, H.H., Zuccala, A.A. & Ye, F.Y. (2019). Tracing the ‘Swan-groups’ of Physics and Economics in the Key Publications of Nobel Laureates, Scientometrics, 119, 425–436.
Zhang, J.A., Vogeley, M.S., & Chen, C.M. (2011). Scientometrics of big science: A case study of research in the Sloan Digital Sky Survey. Scientometrics, 86(1), 1–14.
Zou, C., & Peterson, J.B. (2016). Quantifying the scientific output of new researchers using the zp-index. Scientometrics, 106(3), 901–916.

This is an open access article licensed under the Creative Commons Attribution-NonCommercial-NoDerivs License (http://creativecommons.org/licenses/by-nc-nd/4.0/).