A general index to quantify the academic-level of cross-domain and cross-era scholars

Yuhang Yao[1] Yang Li[2] and Chengliang Si[3]

Abstract— **The measure of how to judge and compare the academic-level of scholars has long raise the attention worldwide. A common index is the cumulative citation of scholars' publication. However, as the number of papers and scholars increase exponentially nowadays, considering on the various area and era of different papers, the simple use of citation is not a scientific way to judge the academic-level of scholars. What's more, there is no common measure to compare cross-domain and cross-era scholars. In this paper, we give an index named Turing Index to gives the reference citations should be reached for becoming top-level scholars in computer science. We then give a more common index named Citation Power Parity for measuring the development of different area in different era. We then give a weighted citation measure to describe and predict the achievement of scholars. Based on these, we finally develop a system based on Acemap to compare any cross-domain and cross-era scholars and predict the time-stamp when a scholar's academic-level surpass one another.**

I. INTRODUCTION

Science and technology are people's rational understanding of the laws of the objective world. They are the synthesis of experience, methods, techniques, techniques, and capabilities that humans have accumulated in the process of understanding the world and transforming the world. They are material sciences.

History has proven that the development of scientific thought and spirit can lead to the development of politic, society, economy and culture. For example, in the ancient world, the invention of the wheel makes the territory of the Empire expand constantly. The industrial revolution promoted by the development of Newtonian mechanics theory has greatly liberated social productivity and raised the bourgeoisie and the working class. The human society gradually moves from a feudal monarchy to a more advanced republican system. Maxwell's electromagnetic theory and the inventions of Tesla, Edison, etc. enabled electric energy to enter human life and production, and thus promoted economic development. The rapid increase in computing power predicted by Moore's Law has created a rich and varied Internet culture. Moreover, based on facts, the scientific spirit focus on experiments which are the ultimate source of scientific laws and scientific theories and the highest standards of testing. Such spirit gurantee the sustainable development of human civilization.

[1] Means this part was done by Yao
[2] Means this part was done by Li
[3] Means this part was done by Si

There is a common consensus that the development and application of scientific theories requires the joint efforts of a large number of professional technicians. Therefore, the academic community has always been the cornerstone for the development of human civilization.

Nowadays, more and more people have devoted themselves to academia and have developed a large number of scientific research fields and achieved a large number of scientific research achievements.Today, various scientific journals, magazines, and conferences all have a clear orientation around the world. It mainly includes papers in different subdivision directions in different fields. At the same time, scholars have also divided these journals, magazines and conferences into different grades to demonstrate their academic influence. For example, the list of papers recommended by the China Computer Federation (CCF). In addition, various academic research institutions such as universities, research institutes, companies, and academic organizations also have correspondingly mature evaluation systems, such as the Global University Rankings published by The Times every year and the global university rankings published by Shanghai Jiaotong University. These evaluation systems are helpful for scholars to understand the latest relatively correct and reasonable academic research directions, help establish authority and disseminate academic ideas to the public, and are conducive to the sustained and stable development of the academic community and the prosperity of the whole society. In summary, in academic circles, the evaluation of academic strength and influence is particularly important.

As mentioned above, there are already mature evaluation systems for academic journals, magazines, conferences, and academic institutions. However, the evaluation of scholars is relatively incomplete. Today, it is generally recognized that: citations in Google scholar.

Due to the deficiencies in the existing work, we have put forward our evaluation indicators for scholars' academic level and academic influence. For different fields, our evaluation system can evaluate the development of the field in different historical stages, and at the same time put forward the development status of the Turing index in the evaluation field. For scholars in the same field, based on the number of papers published by scholars and the quality of the papers, we conducted a systematic and comprehensive evaluation of scholars' academic level and academic influence in the field. Our method considers the contribution of scholars to the theory, the age of the scholars, the change of the field with time, the fieryness of the field and so on. After putting forward evaluation systems for different fields and scholars, we have also proposed a method to measure the relationship between scholars and their fields. In addition, we also propose a method to compare the influence of scholars in different fields on their field of influence.

Based on the above work, our academic evaluation system can systematically and dynamically assess the overall level of a scholar, and at the same time predict the future development of the scholar, compare the academic influence of the two scholars in the field, and predict whether young scholars can exceed it. tutor. In addition, our evaluation system also considers the situation of uneven development between different fields and can judge whether the level of a scholar can keep up with the development of the field.

Based on a large number of academic scholars' information and paper information, we conducted detailed experiments to verify the validity and rationality of our evaluation system. In the experiment, our team conducted processing, statistics, modeling, fitting, and forecasting operations from the 1895 to 2015 data. The results obtained are very good proof that we have validated our proposed method. We also conducted meticulous analysis based on our experimental results, and obtained some heuristic academic information.

In the later part of the report, we first introduce related work, and then introduce the mathematical definition and methods of our evaluation system. Then introduce our experiments and show them, analyze and explain. In the end we came to our conclusion.

II. Related Work

A. Ace-Map

AceMap[1] is an academic system to analyze the academic big data, present the results through a novel map approach, and thus help the researchers better do their work.

Unlike existing academic systems which mainly adopt textbased methods, AceMap displays the information in a clear and intuitive way. Currently, AceMap contains the following functions: dynamic citation network display, paper clustering, academic genealogy and academic path finding. Ace-map uses distributed network analysis algorithms, perform the algorithms in a Spark system and utilize modern visualization tools to present the results.

B. Science Citation Index

For a long time, the quotation number is an important indicator of the scholars academic level.

The Science Citation Index (SCI) is an important citation index, which is originally produced by the Institute for Scientific Information (ISI) and created by Eugene Garfield[2]. The index is made available online through different platforms, such as the Web of Science and SciSearch. (There are also CD and printed editions, covering a smaller number of journals). This database allows a researcher to identify which later articles have cited any particular earlier article, or have cited the articles of any particular author, or have been cited most frequently. Thomson Reuters also markets several subsets of this database, termed ”Specialty Citation Indexes”, such as the Neuroscience Citation Index and the Chemistry Citation Index.

C. H-index

The h-index[3] is an author-level metric that attempts to measure both the productivity and citation impact of the publications of a scientist or scholar. The index is based on the set of the scientist's most cited papers and the number of citations that they have received in other publications. The index can also be applied to the productivity and impact of a scholarly journal as well as a group of scientists, such as a department or university or country.

H-index received great attention once it was proposed. The h-index has been proven a effective index by comparison of WoS, Scopus and Google Scholar[4]. Moreover, Degree, H-index and coreness are widely used metrics, but previously treated as unrelated. H-index was extended to the general framework of Information Production Processes (IPPs), using a source-item terminology by L Egghe etc.[5]. Linyuan L etc.[6] showed their relation by constructing an operator , in terms of which degree, H-index and coreness are the initial, intermediate and steady states of the sequences, respectively.

Based on the h-index, some researchers raised other index to evaluate the scholar level. F Franceschini etc. introduced the success-index[7], aimed at reducing the NSP -indexs limitations, although requiring more computing effort. It was much more versatile for different types of analysis. And scientific index (SI)[8] was raised to solve the dilemma that the h-index does not indicate the entire scientific labour of a person, including the number of publications and the total amount of citations and someone with fewer publications but more citations for each paper would have the same h-index as a person with multiple higher numbers of publications and total amount of citations.

D. Other scholar level evaluation approaches

Many approaches are raised to evaluate the scholar level. The Becker Model[9] which was raised by Cathy C etc. can be use by both researchers and librarians to document research impact to supplement citation analysis. CL Giles etc.[10] developed automated methods for acknowledgment extraction and analysis and showed that combining acknowledgment analysis with citation indexing yields a measurable impact of the efficacy of various individuals as well as government, corporate, and university sponsors of scientific work.

III. Method

In this section, we describe the academic-level compare problem and the measures to derive an general academic-level index.

A. Problem Description[1]

As we need to compare the cross-domain and cross-era scholars, we need to give a accurate description of the judgement when a scholar outperform another scholar. A common view of the Academic-level Compare Problem is:

Suppose we have two scholar p and q. For scholar p in field f and year t, the academic-level index is $I_{p,f,t}$. Similarly, for scholar q in field g and year s, the academic-level index is $I_{q,h,s}$. If $I_{p,f,t} > I_{q,h,s}$, then we say the academic-level of scholar p in field f and year t is greater than that of scholar q in field g and year s.

To compare the lifetime academic-level, we set the lifetime academic-level of scholar p in field f and year t ($t > t_0$, start year of scholars' academic career) as

$$S_{p,f,t} = \sum_{t=t_0}^{t} I_{p,f,t} \tag{1}$$

For time-stamp t_s and $\forall$ years $t > t_s$, if $S_{p,f,t} > S_{q,h,t}$, we say the lifetime academic-level of scholar p in field f is greater than that of scholar q in field g since year t_s.

The key problem is how to derive the academic-level index $I_{p,f,t}$. As the index considering on the personal level of scholars and global level of fields in different years, the index can be describe as

$$I_{p,f,t} = \frac{A_{p,f,t}}{C_{f,t}} \tag{2}$$

where $A_{p,f,t}$ denotes the achievement of scholar p in field p and year t and $C_{f,t}$ denotes the development level of field f in year t. Then our work is to use scientific way to measure $I_{p,f,t}$ and $C_{f,t}$

TABLE I

Notations and Definitions

Notation	Notation
p, q	Scholars
f, g	Academic Fields
t, s	Years
a, b	papers
c, d	citations
t_0	start year of scholars' academic career
$I_{p,f,t}$	Academic-level index of author p in field f and year t
$A_{p,f,t}$	Achievement of author p in field f and year t
$Ac_{p,f,t}$	Active index of author p in field f and year t
Ac_f	Active index of field f
$C_{f,t}$	Development level of field f in year t.
$Co_{p,f}$	Contribution index of scholar p in field f.
$G_{p,f,t}$	Growth rate of scholar p in field f and year t
$G_{f,t}$	Growth rate of field f in year t
$S_{p,f,t}$	Lifetime Academic-level of author p in field f and year t
$R_{p,f,t}$	Cumulative citations of scholar p in field f and year t
$T_{f,t}$	Turing Index at year t

B. Field Index[1]

To get $C_{f,t}$, we give two different way: Turing Index and Citation Power Parity. Turing Index is more specific and give meaningful reference for top-level scholars. Citation Power Parity is more general and give global view of the development of fields.

1) Turing Index: The Turing Index is mainly considering on the cumulative citation of top-level scholars when they won the top awards. We specifically considering on the use in Computer Science and ACM A.M. Turing Award(an annual prize given by the Association for Computing Machinery to an individual selected for contributions of lasting and major technical importance to the computer field). The Turing Award is generally recognized as the highest distinction in computer science and the "Nobel Prize of computing".

In Computer Science area where f='Computer Science', We calculate scholars' cumulative citation when they won the Turing Award. We set the cumulative citations(references) of scholar p at year t to be $R_{p,f,t}$. If scholar p won the Turing Award in year t, the Turing index $T_{f,t} = R_{p,f,t}$. If there are more than one winner in year t, we just calculate the average number of $R_{p,f,t}$. In this case, $C_{f,t} = T_{f,t} - T_{f,t-1}$. We consider the Turing Index as reference citations should be reached for becoming top-level scholars in computer science. If $R_{p,f,t} \geq T_{f,t}$, we consider scholar p has the potential to become the top-level scholar and win the Turing Award.

2) Citation Power Parity: Citation Power Parity is a more general way as it mainly focus on the additional citations. In Econometrics, Purchasing Power Parity[11] is the price ratio of goods in different countries. It denotes the number of goods in different country can be purchased by one unit of base currency. By similar idea, Citation Power Parity denotes the output ratio by one unit of citation. For convenience, We just set Citation Power Parity $C_{f,t}$ in field f and year t to be the increased citations in field f and year t. The increased citations in field f and year t is the sum of increased citations papers of field f in year t. Increased citations is general and convenient without manual selection like Turing Index.

For field f and year t, field g and year s, if $C_{f,t} > C_{g,s}$, we know that the development of field f in year t is greater and faster than that of field g in year s. It also means that same academic-level scholars can achieve more citations in field f and year t than that in field g and year s.

C. Scholar index[3]

To get $I_{p,f,t}$, we proposed three Gradually perfect way to measure the achievement of author p in field f and year t. Paper numbers is a simple and raw evaluation index. Citation power is the a common evaluation index for a scholar. Academic-level is more reasonable and consider the real contribution of a scholar to the field.

1) Paper numbers: The paper numbers is relatively simple and crude. Suppose scholar p have published m papers until year t in field f. Then we define the paper numbers of scholar p in field f in year t as

$$I^{p}_{p,f,t} = m \tag{3}$$

2) Citation power: The paper numbers is relatively crude because it cannot distinguish the papers impact and quality. Consider a situation that a scholar p only publish one paper, which is a big breakthrough in field f in year t, while another scholar q publish three Supplementary work of scholar p, Obviously, scholar p have higher impact and contribution.Then we proposed citation power to handle such unreasonable situation.

The citation power is a evaluation index for scholar in specific field f and year t. Suppose scholar p have published m papers until year t in field f, note these papers as $a_{1,f,t}, a_{2,f,t}, ..., a_{i,f,t}, ..., a_{n,f,t}$. Each papers published have citations c in year t, note citations as $c_{1,f,t}, c_{2,f,t}, ..., c_{i,f,t}, .., c_{n,f,t}$. Then we define the citation power of scholar p in field f in year t as

$$I^c_{p,f,t} = \sum_{i=1}^{m} c_{i,f,t} \tag{4}$$

Note that the citation power is represent the whole citations of the papers scholar published before in year t, instead of the papers published in year t.

3) Academic level: Although citation power have considered the difference of paper quality, the contribution of paper authors are different. In most fields, the order of author represent the contribution of the papers. Consider a situation that the first order author contribute almost all work of a paper, and the last order author just do some Auxiliary work, then if we use citation power is unfair and unreasonable. Based on above consideration, we propose our academic level index to evaluate the real contribution and impact of a scholar p reasonably.

Similar to citation power, suppose scholar p have published m papers, $a_{i,f,t}$, until year t in field f. Each papers published have citations c, $c_{i,f,t}$, in year t. Suppose scholar p is the j order author of paper $a_{i,f,t}$, we set a parameter λ_f represent the attenuation parameter. Then we define the academic level as

$$I_{p,f,t} = \sum_{i=1}^{m} \lambda_f^j * c_{i,f,t} \tag{5}$$

The scholars' academic level comprehensively considers the quality of the thesis, the author's actual contribution and other factors, and is more convincing than the evaluation of the number of papers and citations, and can better reflect the comprehensive influence and actual contribution of a scholar in a certain field. In the following texts, the academic achievements of the scholars mentioned by us and the evaluation of bachelor's degree all refer to academic level.

D. Scholar-field relationship[3]

In the current era of subdivision of this discipline, different scholars have different abilities in different fields. For example, Einstein's achievements in physics are very deep, but he may not have achieved much in other fields. Based on the above analysis and evaluation of fields and scholars, we propose three different indicators to measure the different relationships between scholars and fields.

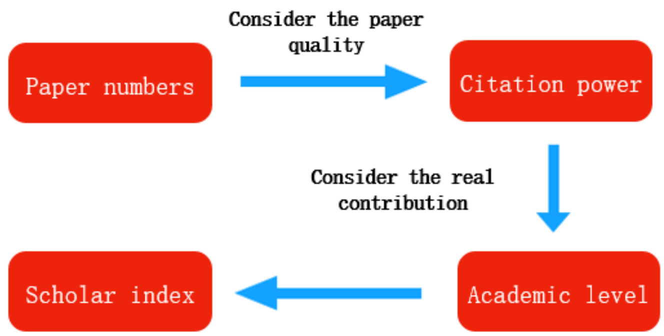

Fig. 1. Scholar index process

1) Citation power: Based on above work, we have proposed the citation power for field and scholar respectively. As we mentioned before, citation reflect the paper quality, in this part, we use the citation power in both field and scholar to evaluate the active index of a scholar in field p at year t.

$$Ac_{p,f,t} = \frac{I^c_{p,f,t}}{C_{f,t}} \tag{6}$$

The active coefficient of scholars indicates the academic activity of the scholar in a certain field in the year. The quota of the citation of the paper in the field in the whole year can effectively reflect the contribution of the scholars work in that field to the field, thus reflecting the scholars The level of activity in this field.

2) Lifetime impact: As mentioned in problem description, we have set the lifetime academic level of scholar p in field f until year t ($t > t_0, startyearofscholars'academiccareer$) as

$$S_{p,f,t} = \sum_{t=t_0}^{t} I^c_{p,f,t} \tag{7}$$

In order to found the lifetime impact of a scholar in the whole field, we set the field active index Ac_f of field f as

$$Ac_f = \sum_{t=t_0}^{t} C_{f,t} \tag{8}$$

Where t_0 is the start year of field f, and t is the current time.

Then based on the lifetime academic level and field active index, we define the lifetime contribution index of scholar p in field f as

$$Co_{p,f} = \frac{S_{p,f,t}}{Ac_f} \tag{9}$$

Obviously, with the constant development of a certain field, the field active index will gradually increase, and most scholars will begin to become less and less as the age increases. Our field contribution index reflects this phenomenon, and most scholars have The contribution coefficient will gradually decrease, but for some breakthrough work will continue to be cited, reflecting the author's continued influence.

3) Scholar increment: In the above work, we proposed some static relations between scholars and fields. We propose a method to measure the dynamic relationship between scholars and fields. At the same time, we can use this method to judge whether scholars can keep up with the development trend of the field.

We count the increment of a scholar in field f in year t, note the increment as $\Delta I_{p,f,t}$

$$\Delta I^c_{p,f,t} = I^c_{p,f,t} - I^c_{p,f,t-1} \tag{10}$$

Then the growth rate of the scholar in field f in year t define as

$$G_{p,f,t} = \frac{\Delta I^c_{p,f,t}}{I^c_{p,f,t-1}} \tag{11}$$

Similarly, we count the increment field f in year t, note the increment as $\Delta C_{f,t}$

$$\Delta C_{f,t} = C_{f,t} - C_{f,t-1} \tag{12}$$

Then the growth rate of the scholar in field f in year t define as

$$G_{f,t} = \frac{\Delta C_{f,t}}{C_{f,t-1}} \tag{13}$$

To judge whether scholars can keep up with the development trend of the field. we set compare the growth rate of scholar and field. For time-stamp t_s, if $G_{p,f,t} > G_{f,t}$, we say the scholar can keep up with field f in year t. Because as we know, in students' paper, the professor usually is the last order author, but the professor is the guide of the paper, so we choose citation power to handle this issue instead of academic level.

TABLE II

Scholar-field relationship

Relation	Property
Citation power	The activity level
Lifetime impact	The lifetime contribution
Scholar increment	The development trend

E. *Curve fitting and prediction*[2]

According to the data we get from dataset, the citation number of some young scholars shows a J-shaped growth trend. Using this function to fit the data and predict the development of a scholar is unreasonable because a researcher can't keep such a fast speed on his field, not to mention faster and faster. If we use the J-shape function to predict the trend, the citation number of this scholar will reach infinity in a short time, which is contradict to our common sense.

In fact, a more proper explanation is that the growth trend is a S-shaped function. At the beginning of the scholar's research career, the development are usually slow because of the weak accumulation on knowledge and the lack of experience. The time which a researcher makes a quick progress and gets a lots of research results in is in the middle of the career. Because during this time, the scholars have a certain foundation on knowledge and their brains are in the most flexible phase. At the later stage of research career, the thinkings are not as agile as before. Accordingly, their research results grow slowly. This trend can be proved through the data of elder scholars. The growth trend we get from young scholars shows J-shaped curve just because they are in their meridian research time.

In our paper, we use logistic function to fit the trend. Standard logistic sigmoid function. A logistic function or logistic curve is a common "S" shape defined by following equation:

$$f(x) = \frac{a}{1 + e^{-b(x+c)}}$$

where a, b, c are pending fitting parameters and x is the scholar number. In some situations, the data of a scholar is scarce and it can't be fitted. Then we use linear function or quadratic function to fit them:

$$f(x) = ax + b$$

$$f(x) = ax^2 + bx + c$$

Admittedly, the linear function and quadratic function are not as precise as logistic function. But the function can be improved by the increase of data. When the data is enough, the function can be fitted by logistic function.

When fitting the curve, the scholar index calculated through our method is proportional expanded to make the fitting process effective. And some data will be noisy points and we remove their influence. This process will be explained clearly in experiment part. The figure2 shows the process of curve fitting.

IV. EXPERIMENT

A. *Dataset*[1]

We use the dataset from Acemap supported by Microsoft Academic[1] which is extremely large and contains nearly all papers, scholars and related relationships in academic. Table III shows the detail of dataset.

B. *Initial data processing & generation of Turing Index*[2]

The initial data processing can be divided into four main parts. Through these four parts, we get the fundamental reference data to support our following research and building up scholar index. This process is shown in figure3.

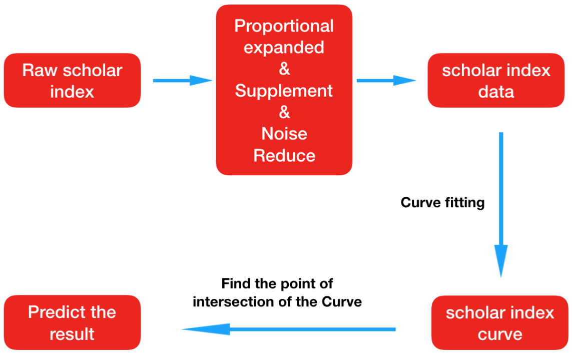

Fig. 2. Curve fitting process

TABLE III
Dataset description of Acemap

Items	number
Scholars	114796077
Papers	127324384
References	536583464
Fields	53834
begin year	1895
end year	2016

- **Massive data**: The data consists of 2 main parts. First, the pair of paper and author, which includes virtual all areas such as computer science, math and physics. Second, the reference relationship in papers. Some relationships are also contained in the dataset such as author name - author number, paper - author. They are tools for establishing the relation networks. The datas coverage is so broad that its size achieves 40G.
- **Construct paper-author-year network through massive data** :By using the reference relationship and affiliation relationship between papers and authors, we constructed the information net. It connected the isolate messages, such as the publication year of paper, reference papers, and the authors of reference papers. It exactly raise the possibility of finding the inner link in massive data.
- **Count the reference data of author per year**:Through author-paper net we constructed before, we counted the reference statistics of author per year. The work made the inner link in the net readable for human being and uncovered the increase rule of reference number, which is a vital factor to measuring the achievement of a professor.
- **Analysis the data of Turing Award Winner**:How to choose the proper samples to analysis is key to our work. The data of Turing award winners is typical in Computer Science area. The reference number tendency of Turing award winners shows a good result corresponding to our prospection. The result of analysis on Turing Award Winner have been shown in previous part clearly.

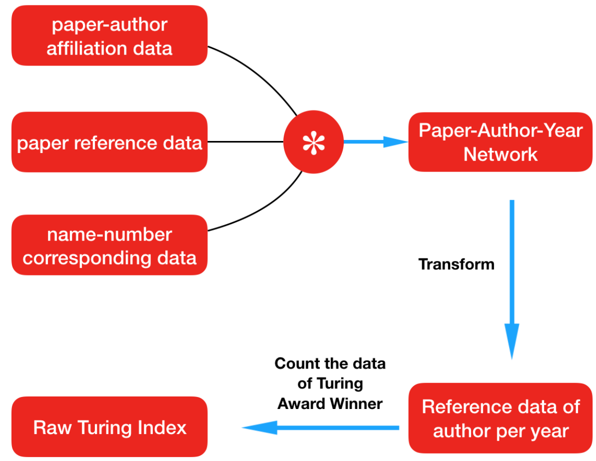

Fig. 3. Initial data processing

C. Turing Index[1]

After the generation, we get the Turing Index in Computer Science. Figure4 shows the curve of Turing Index. As we can seen, at the 1970s, as Computer Science just started up and was a small field with a few scholars. There are not many paper published and to win the Turing Award doesn't need more than one hundred citations. However, as the general application of computer and the Internet, the development of Computer Science increases exponentially. Up to now, a Turing Award winner need nearly 10,000 citations. A scholar who had 100 citations in Computer Science at the 1970s may be considered has the same academic-level as a scholar who has 10,000 citations at the 2010s. This index can be the evaluation of top-level scholars in computer science. At now, a scholar who has one unit of Turing Index may have the potential to win the Turing Award.

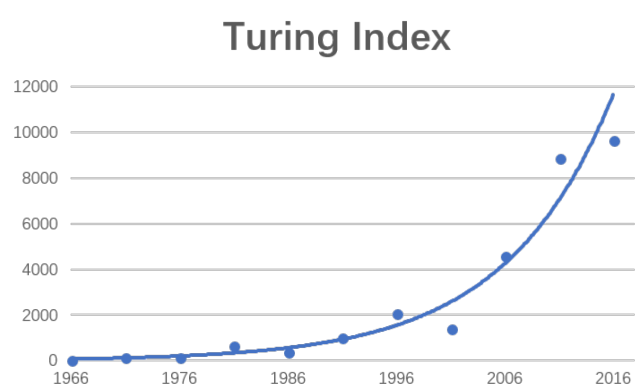

Fig. 4. Turing Index in Computer Science

As shown in Figure5, we also found that the total citation increment each year of papers follow the Moore's Law and doubles every 5 years. It shows the prosperous development of academia and make it necessary to eliminate the inflation of citation for the evaluation of academic-level.

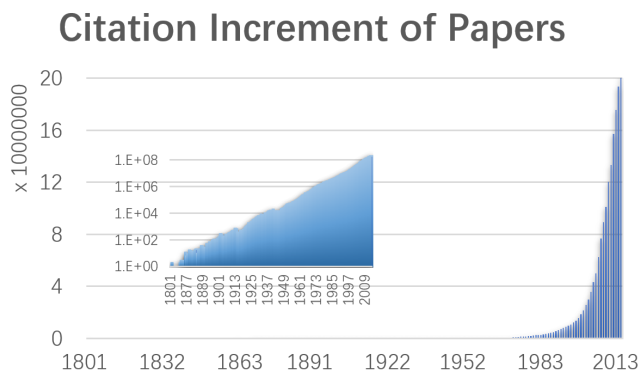

Fig. 5. Citation Increment of Papers

D. Generate the scholar index

1) Field fitting[3]*:* The larger the amount of data, the prediction results, and the effectiveness of our experiments, we have screened the fields. In our experiments, we only selected fields that historically have a total citation of more than 500,000.

2) The preparation of scholar index[3]*:* We do mass data processing, in order to get the useful data and network for counting each scholar's scholar index, we do the preparation by following steps:

- **Construct paper-author network**:Obtain all papers of a scholar through the previously completed paper-author-year network and calculate the author's contribution to the paper $\lambda_{a_i,f}$.
- **Count the reference data of author per year**:Through author-paper net we constructed before, we counted the reference statistics of the papers published by scholar p before in year t.
- **Construct the paper-field network**:Through mass data, we obtain the field each papers belongs to. Note that one paper may belongs to several fields, since a scholar usually publish lots of papers in his life, so a scholar may have many field.

3) Generation of scholar index[3]*:* On the basis of the above-mentioned full preparation, we use **Algorithm 1** to calculate the scholar index in respective fields year by year.

4) The preparation of data for curve fitting[2]*:* The preparation of data is to make the curve fitting more precise. The process can be divided into three parts: Supplement, Accumulation and Noise Reduce. After the preparation, we will get incremental and continuous data, which is fit the real situation of a researcher's scholar level. The process of preparation is shown in Figure6.

- **Supplement**: The raw data of the scholar index are always discrete. For example, if a research lasts a long time, the authors of this paper may don't have a data during these years, which creates the "fault" of data. In order to make the data cover all of the years, we define the scholar index as 0 in this "fault". Through supplement process, the discrete data is converted into continuous one.
  In our experiment, we find the researcher's earliest and latest year he had nonzero scholar index in the dataset, and define them as beginning year and ending year. The data is continuous between beginning and end years.
- **Accumulation**:Scholar index represents the contribution of a researcher to his career. Hence, when we fit the curve and predict the trend, the data of a year should be accumulate-scholar index. That's to say, the data at the certain year contains all the contribution before this year. Through accumulation process, the data is converted into incremental one.
  In our experiment, we accumulate the data of a scholar between beginning and end years. Because the dataset only contains part of paper data after 2016, the accumulation stops before 2016 and the data after 2016 is removed. The data is incremental between beginning and end years.
- **Noise Reduce**:The scholar index is usually small. In order to make the S-shape obvious, we have a proportional expansion on scholar index. In this experiment, the data is magnified 1000000 times. But some data points are disproportionate because of the imperfection of dataset. Luckily, these data points only have a tiny proportion in the dataset. And we delete the data points which are 1000 times more than average number. The noise reduce process makes the curve fitting more precise.

Algorithm 1 Academic level algorithm

Require: P_a:Paper-author; P_f:Paper-field; P_r:Paper-reference;
Ensure: S:Scholar-index

$S \leftarrow \emptyset$
for Author p in P_a **do**
  for Papers a of Author p **do**
    for Field f of paper a in P_a **do**
      for Year t in P_r **do** Count $I_{p,f,t}$
      end for
    end for
  end for
end for

E. Prediction[2]

In this part, the growth function is fitted by current data points and used to predict when a scholar can achieve even exceed another. The year is the intersect point of two function curves. The algorithm is shown in Algorithm2.

1) Curve fitting[2]*:* The modified data we get from last step is used to curve fitting process. The logistic function is used to fit the curve mainly. The logistic function have following form:

$$f(x) = \frac{a}{1 + e^{-b(x+c)}}$$

This S-shaped curve can fit the data points well.

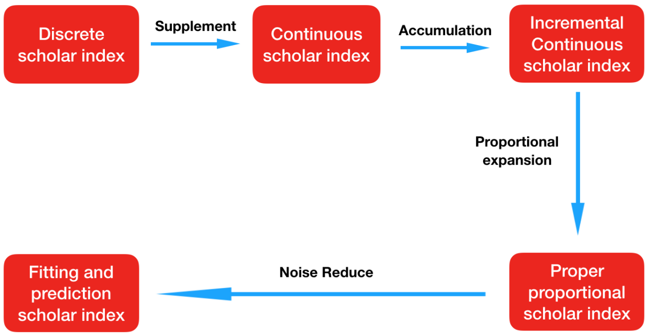

Fig. 6. Data preparation for curve fitting

Algorithm 2 Prediction algorithm

procedure PREDICTION(a, b) ▷ Data of two scholars
  $f_a \leftarrow Get\_curve\_fit\_function(a)$
  $f_b \leftarrow Get\_curve\_fit\_function(b)$
  while $i \in (beginning, end\ prediction\ year)$ **do**
    if $f_a(i) > f_b(i) \& f_a(i-1) < f_b(i-1)$ **then**
      $year \leftarrow i$
    end if
  end while
  return $year$ ▷ Scholar a exceed scholar b
end procedure

2) Field year removal[2]*:* As is explained before, in some years, the field index data is so small that it is unconvincing and it is removed. Correspondingly, the scholar index of this year has to be deleted in order to make the prediction precise. In this experiment, the data after 2016 is removed from dataset.

3) Prediction Result and Analysis[2,3]*:* In this experiment, we use some pairs of scholars to test out prediction method. The results are surprisingly interpretable. In the comparison between Xiaoou Tang and Jian Sun, two famous scholars in computer vision, their scholar index curves show their scholar level obviously. Xiaoou Tang (H-index 95) exceeded Jian Sun (H-index 71) in this field. Google scholar h-index supports our prediction. This is shown in Figure7.

Another comparison is Xinbing Wang and his teacher Eun. The curve shows the great potential of professor Wang and it predicts that Wang will exceed his teacher in 2016. This result is quite close to reality. This is shown in Figure8.

Another example is the comparison between Andrew Y Ng and Michael I Jordan, two famous scholars in Machine Learning, their scholar index curves show their scholar level obviously. Michael I Jordan always have higher scholar index than Andrew Y Ng in this field. It implies that Andrew Y Ng can never exceed Michael I Jordan.

In comparison of Edward J Coyle and Ness B Shroff, the curve shows the great potential of professor Shroff and Coyle's impact stop at 1990s, which is quite close to reality.

As for different fields, our method shows amazing precision. In the comparison between Kaiming He(H-index 37),

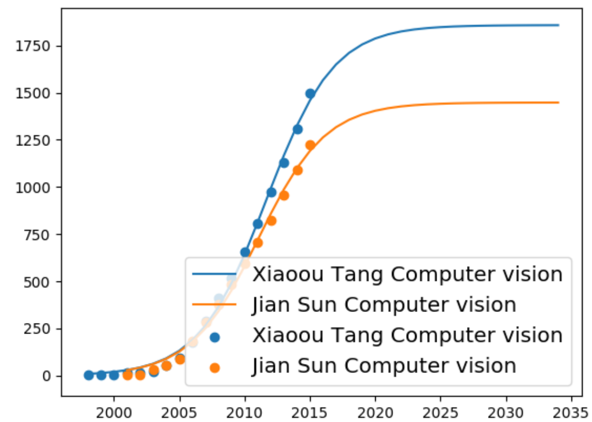

Fig. 7. Comparison between Xiaoou Tang and Jian Sun

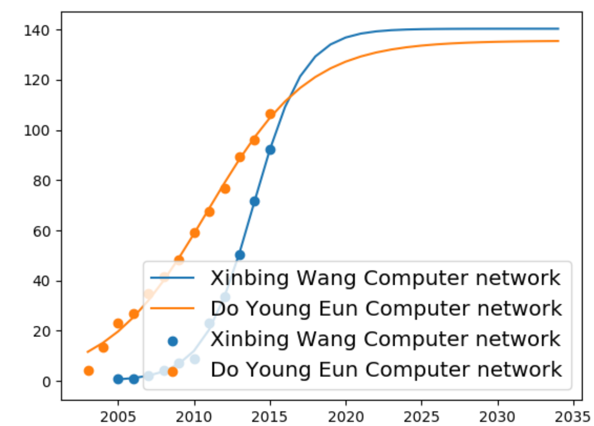

Fig. 8. Comparison between Xinbing Wang and Eun

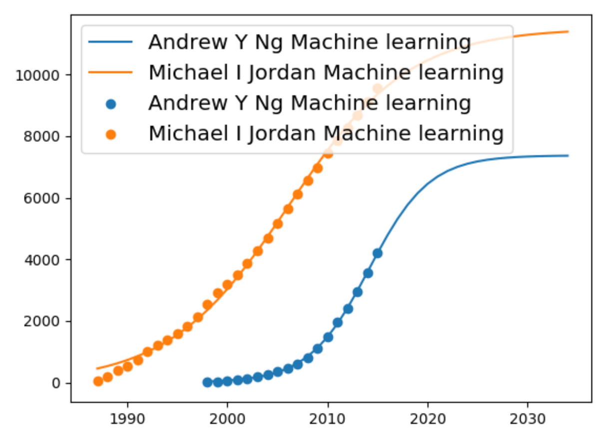

Fig. 9. Comparison between Andrew Y Ng and Michael I Jordan

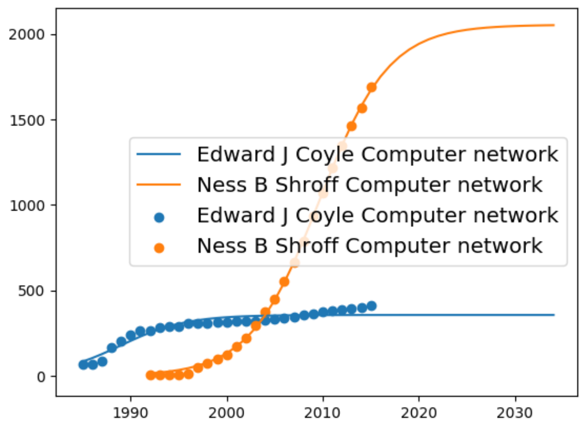

Fig. 10. Comparison between Edward J Coyle and Ness B Shroff

who major in computer vision and Kenny Q Zhu(H-index 16) , who mainly studies in Data Analysis, it shows that in 2013 He will exceed Zhu in their own field. Considering the paper they published, this result is convincing, and it prove that our method is reasonable. This is shown in Figure11.

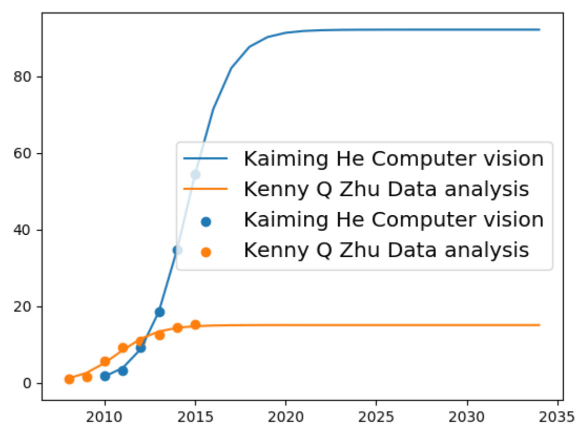

Fig. 11. Comparison between Kaiming He and Kenny Q Zhu

Another example for different fields, our method shows amazing precision. In the comparison between Kaiming He, who major in computer vision and Xinbing Wang , who mainly studies in Computer Network, it shows that the academic level of professor Wang in his major is higher than Kaiming He.

V. CONCLUSIONS

In this paper, we propose Turing Index and Citation Power Parity to describe the development level of fields. We then provide a weighted citation measure to describe and predict the achievement of scholars. Based on these, we finally develop a system based on Acemap to compare any cross-domain and cross-era scholars and predict the time-stamp

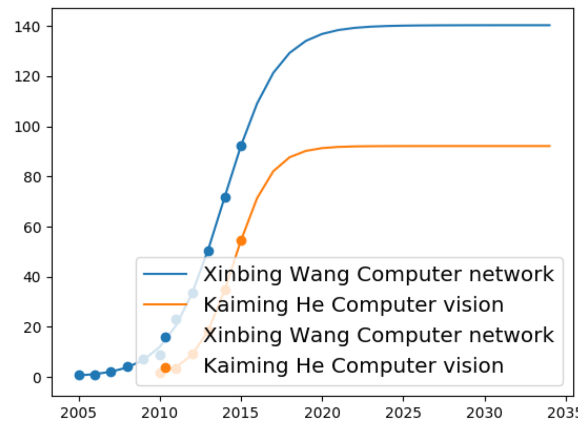

Fig. 12. Comparison between Xinbing Wang and Kaiming He

when a scholar's academic-level surpass one another. This work is considered meaningful reference of academic-level evaluation and prediction. In the future, the work will focus on maintain the structure of model and build a stable website to compare any scholars just by user-defined.

ACKNOWLEDGMENT

Thanks for the guidance of Professor Xinbing Wang and Luoyi Fu.

REFERENCES

[1] Z. Tan, C. Liu, Y. Mao, Y. Guo, J. Shen, and X. Wang, "Acemap: A novel approach towards displaying relationship among academic literatures," pp. 437–442, 2016. [Online]. Available: https://doi.org/10.1145/2872518.2890514

[2] E. Garfield, ""science citation index"-a new dimension in indexing," *Science*, vol. 144, no. 3619, pp. 649–654, 1964.

[3] J. E. Hirsch, "An index to quantify an individual's scientific research output," *Proceedings of the National Academy of Science*, vol. 102, pp. 16 569–16 572, Nov. 2005.

[4] J. Barilan, "Which h-index? a comparison of wos, scopus and google scholar," *Scientometrics*, vol. 74, no. 2, pp. 257–271, 2008.

[5] L. Egghe and R. Rousseau, "An informetric model for the hirsch-index," *Scientometrics*, vol. 69, no. 1, pp. 121–129, 2006.

[6] L. Lü, T. Zhou, Q.-M. Zhang, and H. E. Stanley, "The h-index of a network node and its relation to degree and coreness," in *Nature communications*, 2016.

[7] F. Franceschini, M. Galetto, D. Maisano, and L. Mastrogiacomo, "The success -index: an alternative approach to the h -index for evaluating an individuals research output," *Scientometrics*, vol. 92, no. 3, pp. 621–641, 2012.

[8] Z. H. Khan, M. Nashibi, and S. A. Javadi, "Scientific index: a complementary scale for the h-index," *BMJ Evidence-Based Medicine*, vol. 23, no. 3, pp. 118–118, 2018. [Online]. Available: http://ebm.bmj.com/content/23/3/118

[9] C. C. Sarli, E. K. Dubinsky, and K. L. Holmes, "Beyond citation analysis: a model for assessment of research impact," *Journal of the Medical Library Association Jmla*, vol. 98, no. 1, p. 17, 2010.

[10] C. L. Giles, I. G. Councill, and J. N. Gray, "Who gets acknowledged: Measuring scientific contributions through automatic acknowledgment indexing," *Proc Natl Acad Sci U S A*, vol. 101, no. 51, pp. 17 599–17 604, 2004.

[11] M. P. Taylor, "Purchasing power parity," *Review of International Economics*, vol. 11, no. 3, pp. 436–452, 2003.