

## FedUni ResearchOnline

<https://researchonline.federation.edu.au>

Copyright Notice

This is the peer-reviewed version of the following article:

Kong, X., Shi, Y., Wang, W., Ma, K., Wan, L., & Xia, F. (2019). The Evolution of Turing Award Collaboration Network: Bibliometric-Level and Network-Level Metrics. *IEEE Transactions on Computational Social Systems*, 6(6), 1318–1328.

Which has been published in final form at:

<https://doi.org/10.1109/TCSS.2019.2950445>

Copyright © 2019 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

# The Evolution of Turing Award Collaboration Network: Bibliometric-level and Network-level Metrics

Xiangjie Kong, *Senior Member, IEEE*, Yajie Shi, Wei Wang, Kai Ma, Liangtian Wan, Feng Xia, *Senior Member, IEEE*

**Abstract**—The year of 2017 for the fiftieth anniversary of the Turing Award, which represents the top level award in the computer science field, is a milestone. We study the long-term evolution of the Turing Award Collaboration Network and it can be considered as a microcosm of the computer science field from 1974 to 2016. Firstly, scholars tend to publish papers by themselves at the early stages and they began to focus on tight collaboration since the late 1980s. Secondly, compared with the same scale random network, although the Turing Award Collaboration Network has small-world properties, it is not a scale-free network. The reason may be the limited number of collaborators for each scholar. It is impossible for scholars to connect to others freely (preferential attachment) as the scale-free network. Thirdly, to measure how far a scholar is from the Turing Award, we propose a metric called the Turing Number and find that the Turing Number decreases gradually over time. Meanwhile, we discover the phenomenon that scholars prefer to gather into groups to do research with the development of computer science. Our work presents a new way to explore the evolution of academic collaboration network in the field of computer science by building and analyzing the Turing Award Collaboration Network for decades.

**Index Terms**—Turing Award Collaboration Network, network dynamics, bibliometric-level metrics, network-level metrics, Turing Number

## I. INTRODUCTION

COMPUTER science is a diverse field which is full of academic activities, including plenty of partitions. There are many prizes to commemorate computer scientists who have made outstanding contributions to the computer science field. The ACM A. M Turing Award which is established by the Association for Computing Machinery (ACM) and gains the title of “computer field’s Nobel Prize” [1], has a far-reaching influence so far. It is meaningful to study computer science from the perspective of the Turing Award. However, compared

to the Nobel Prize, the research of the Turing Award is relatively incomprehensive. There exist several works devoted to studying relevant attributes of the Nobel Prize [2] and the Nobel Prize laureates [3] while the research of the Turing Award is particularly rare. As the highest prize in the field of computer science, we believe that the analysis of the Turing Award can highlight the important contributions of the Turing Award laureates. It is also beneficial to motivate the younger generation of computer scientists to fulfill their values [4].

In order to analyze the Turing Award, we propose to study the collaboration network related to the Turing Award laureates. The study of scientific collaboration networks helps us to further understand knowledge production and innovation. Scientific collaboration networks have received growing attention in recent years [5]. Scientific collaboration is a key approach to promote the progress of computer science because it can gather data and resources to boost the collaborative development of knowledge production. The collaboration network is built from the list of published papers by treating the authors as connecting nodes if scholars write one or more papers jointly. In the view of a collaborative point, these networks have revealed patterns of collaboration and research behaviors in different areas [5].

There is an important way to analyze collaboration network besides the computer science field. In the field of mathematics, scholars can adopt the Erdős Number ( $EN$ ) to measure the distance from any mathematician to the far-reaching mathematician through a series of coauthors. It signifies mathematicians’ nearness to the great scientist Erdős. Afterward, some authors analyze the pattern of the Erdős collaboration graph [6]. Inspirationally, we intend to form a collaboration network of the top-level authors of the computer science to analyze the evolution process of the collaboration network. However, in the era of Turing, scientists are not as collaborative as current scholars. Based on the existing digital library, it is found that Turing had no collaborators on publishing papers so that no one can connect to him directly in the collaboration network. Therefore, we attempt to find a similar and alternative scientist. It is difficult to measure and find the most prestigious scientist in computer science. To this end, we can consider the Turing Award laureates as an alternative approach. We present the Turing Number ( $TN$ ) to measure the distance of a given scholar to the Turing Award laureates in the network. Unlike the first two metrics, the apparent difference of our proposed metric is that we measure the distance from a certain scholar

This work was partially supported by the National Natural Science Foundation of China (61872054, 61801076), and the Fundamental Research Funds for the Central Universities (DUT18JC09, DUT19LAB23).

X. Kong, Y. Shi, W. Wang, K. Ma and L. Wan are with the Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province, School of Software, Dalian University of Technology, Dalian 116620, China. (E-mail: xjkong@ieee.org; irene94.shi@outlook.com; ehome.wang@gmail.com; makai1163801756@outlook.com; wan.liangtian.2015@ieee.org).

F. Xia is with the School of Science, Engineering and Information Technology, Federation University Australia, Mount Helen, VIC 3353, Australia, and also with the School of Software, Dalian University of Technology, Dalian 116620, China. (E-mail: f.xia@ieee.org).

The corresponding author of this paper is Liangtian Wan, (e-mail: wan.liangtian.2015@ieee.org).

to the group of the Turing Award laureates, while the first two metrics are the distances to a specific person.

One of the most popular methods for analyzing network evolution is the bibliometric-level approach, which focuses on the quantitative and qualitative results of scientific research activities. The analysis of the bibliometric-level approach is usually based on measurable descriptions of scientific outcomes, including authorship, publications, and citations [7]. Subsequently, it has collected collaborative data to explore complex structures of contact in various fields.

Another widely used method for studying the network is the network-level analysis which has been extensively applied in collaboration network analysis [8]. Many studies propose several network-level metrics [9], such as the measure of degree, variety of centrality, and diameter. Some studies focus on describing structural properties through metrics, e.g., clustering coefficient [10], while others explore more complex problems to analyze, such as the detection of community hubs [11] and the determination of the priority attachment mechanism [12], [13]. Some authors attempt to identify the structure of small-world [14] or the property of scale-free [15]. Others propose new measures to assess the scientific collaboration [16]. Some papers study scientific collaborations focus on network changing over time [17], [18]. Comparing with them, our paper mainly focuses on collaboration network of the top-level authors of the computer science rather than general computer science collaboration network. There are plenty of studies focusing on computer science [19]–[21]. Recently, specific group (top active) collaboration [21] is investigated in the field of computer science. However, there are few studies focusing on Turing Award laureates. Our results show the same conclusion in some respects and the details are introduced in results and discussion part.

In this paper, we characterize the network evolution in computer science. Firstly, we regard all the Turing Award laureates as a group to establish the network called the Turing Award Collaboration Network. It is our aims to comprehend the structure of the entire computer science collaboration network. Then, we analyze the dynamics over a long-time period (42 years). Moreover, we intend to discover the correlation of the distance to Turing Award (Turing Number) and other considered metrics. To achieve this goal, we adopt the data from the open source of the DBLP to incorporate the paper information into the across-the-board collection of evolved network data. We also conduct many in-depth statistical analysis.

Based on the bibliometric-level metrics the development of computer science is abstracted as the collaboration evolution network centering on the Turing Award laureates. The network evolution is described both at the bibliometric level and network level. In addition, we compare calculated metrics with the same-scale random network to eliminate the impact of network dimensions.

**Contribution.** To the best of our knowledge, this work is the early study of computer science Network focused on the Turing Award. The main contributions of this study are as follows:

- We present a new approach to establishing the collaboration network in the field of computer science centering

on the Turing Award laureates.

- We propose a metric called Turing Number to measure the distance between scholars and the Turing Award.
- We provide a comprehensive analysis of the network in conjunction with the bibliometric-level and network-level metrics.
- Finally, we analyze the correlation of the calculated metrics and Turing Number to explore the scholars' properties related to the Turing Award.

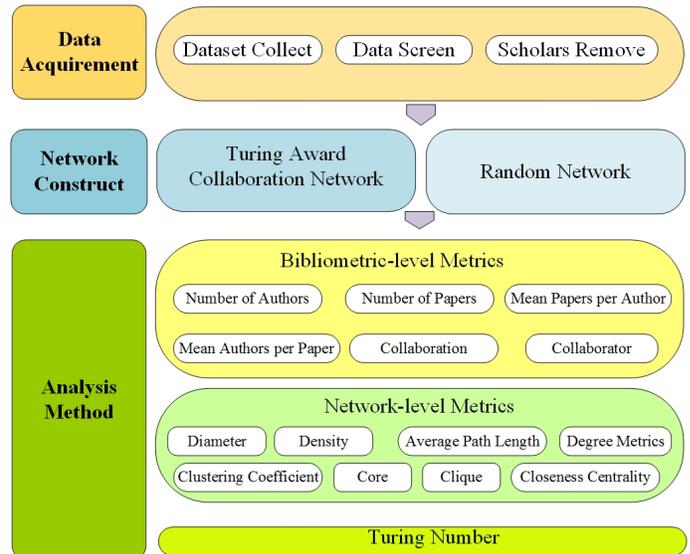


Fig. 1: The process of the evolution research.

## II. METHOD

In this section, we will introduce the methods used in the research process. These methods include data acquirement, collaboration network construct and some analysis metrics. The overall structure diagram of the study is shown in the Fig. 1.

### A. Data Acquirement

To study the evolution of the Turing Award Collaboration Network comprehensively, we collect available DBLP dataset in public. DBLP is a literature dataset of computer science that records information in regard to global computer science research. The personal bibliographic records the scientific achievements of a researcher in one's career. The entire DBLP dataset is stored in an XML form, and it can be downloaded from the website (<http://dblp.uni-trier.de/xml/>). The dataset contains 3,297,544 bibliographies of 1,735,884 authors from 1971 to 2016. The data involves a variety of categories, but we only focus on journal and conference papers whose labels are "article", "inproceeding" and "proceeding". These three categories of papers have been verified to reflect the progress of computer science prominently [22].

To exclude those scholars who leave the academia at the early career, for instance, the newly graduated students, we screen the data. We remove scholars with less than 5,8,10 papers respectively and we find that they do not affect the final

result. Finally, we select to remove scholars who published less than 10 papers, gaining 2,796,297 papers and 192,650 authors. In addition, we remove papers with more than 100 co-authors in view of the fact that such collaboration with so many authors show a weak social relationship [23]. If these authors are not excluded, it would be very unreasonable for these authors to form a fully connected graph when building a collaboration network.

### B. Turing Award Collaboration Network

When the data is obtained, we establish the collaboration network associated with the Turing Award laureates. We regard an author as a node and a co-published article as an edge. The number of common publications can be measured by the weights of links. It can cause changes in the network, mainly because their growth involves the dynamic interaction of links and weights, as well as some possible elements of accelerated growth [24]. The network allows new links to appear among existing nodes. However, we focus more on the increasing scale of the collaboration network and the distances of authors to the Turing Award in this research, so we choose the unweighted network, i.e., if two authors publish an article together, there will be an edge between them.

We treat all the Turing Award laureates as a group and then we set their number to 0. The rest of the scholars are assigned a sequence number starting at 1. If an author collaborates with any Turing Award laureate, he/she will be considered to be connected to the Turing Award. Intuitively, a Turing Award laureate is added to the Turing Award group in the year he/she won the Turing Award, while a scholar joins the network when the scholar began publishing articles. There are one or two laureates to join the Turing Award group each year. In addition, we explore the evolution of the network and investigate the rapid development of the computer science field.

Due to the marginal number of nodes in the early years, we regard 1974 as the starting year for exploration. Since some authors never collaborate with other authors, we ignore the isolated nodes of the network. Due to the tiny number of isolated nodes, it hardly affects the experimental results by deleting isolated nodes. In other words, we only consider the largest connected subgraph of the network. Eventually, we obtain the maximum connected graph related to the Turing Award group, which is named as the Turing Award Collaboration Network. The network is shortened as the Turing Network in the following paragraphs. Furthermore, we put each year as a unit to establish the time-series cumulative collaboration network separately. We also document the annual changes in the authors' properties and their collaborations in the Turing Award Collaboration Network.

### C. Random Network

Intuitively, many of calculated metrics for the Turing Award Collaboration Network are correlated with the size of the network during the year. Therefore, the annual change in metrics is not surprising. A more essential trend is to eliminate the correlation of metrics and network size so as to deepen our understanding for the internal changes of the Turing Network.

The method we choose is to compare the Turing Network to the random network with the changes of the network's size. Therefore, in order to better reflect the network characteristics and to eliminate the scale effects concurrently, we construct a random network with the same number of nodes and edges as the Turing Network.

We select the most classic Erdős-Rényi (ER) Random Network [25], which is built via the number of nodes  $n$  and edges  $m$ ,  $G(n,m)$ . The network is constructed by the following steps: (a) Initialize a given number of nodes  $n$  and edges  $m$ , (b) Select a pair of different nodes randomly without edges and add an edge, (c) Repeat step (b) until  $m$  edges in the network. Through the method above, we construct a random network that retains the number of nodes and edges of the Turing Award Collaboration Network. We expect to exclude the correlation of the Turing network scale through its equivalent random network.

In the sections of the later part of the paper, ER Random Network is used as the experimental control group for our evaluation. The "Real Value" mentioned in all the result diagrams in this paper is the analysis result of the Turing Award Collaboration Network constructed by us. Meanwhile, "Random Value" refers to the result obtained through the analysis of ER Random Network.

### D. Analysis Method

In this paper, we analyze the Turing Award Collaboration Network by the bibliometric-level and network-level methods.

1) *Bibliometric-level Metrics*: To comprehensively understand the authors' outcomes and the evolution of the collaboration pattern, we have selected some typical metrics which are widely applied in related works [26], [27]. Common bibliometric-level metrics are as follows:

- **Number of authors**: The metric describes the numbers of authors who publish papers every year. We remove scholars who published less than 10 papers, eventually obtaining 2,796,297 papers and 192,650 authors.
- **Number of papers**: This metric shows the number of papers published every year. We have utilized the papers from 1974 to 2016.
- **Mean papers per author**: This metric explores the author's average productivity.
- **Mean authors per paper**: For this measure, we explore the average number of authors in each article.
- **Collaboration**: Here, we consider the percentage of the different types of papers (one-, two-, three- and multi-authored papers).
- **Collaborator**: This measure is related to the percentage of authors who tend to collaborate with others in publishing papers.

2) *Network-level Metrics*: In the network-level analysis, the Turing Award Collaboration Network is investigated based on the macro and micro metrics. The former is the overall profile of the social network's characteristics to display the network while the latter emphasizes on the assessment of the nodes to capture the characteristics of each node [28]. In addition, to eliminate the influence of the network scale, we compare

the metrics of the Turing Award Collaboration Network with metrics of the random network. The network-level metrics which are widely used to standardize or the metrics infer the structural aspects of the network are as follows:

- **Diameter:** The given distance  $d_{ij}$  refers to the length of the shortest path connecting the two nodes  $i$  and  $j$ . The diameter  $D$  of the network is used to measure the maximum eccentricity, which is the maximum distance between any two nodes:

$$D = \max_{i,j} d_{ij}. \quad (1)$$

- **Density:** Density is used to measure connectivity across the network and it is calculated by dividing the total number of edges by the total number of possible edges in the network with the same number of nodes. The formula for density  $p$  is:

$$p = \frac{E}{\frac{1}{2}N(N-1)}, \quad (2)$$

where  $N$  is the total number of nodes and  $E$  is the total number of edges in the network. We can adopt network density to characterize the extent of coherency and linkage between nodes in the network [29].

- **Average path length:** The average shortest path length  $L$  is the average length of the shortest path between every two nodes in the network [30]:

$$L = \frac{1}{N(N-1)} \sum_{i \neq j} d_{ij}, \quad (3)$$

In the collaboration network, the distance between coauthors of a paper is 1, while the distance among the authors that they do not collaborate directly but have the same coauthor is 2.

- **Degree:** The adjacency matrix  $A = (a_{ij})_{N \times N}$  of a given graph  $G$  is an  $N$ -th order square matrix, and the element  $a_{ij}$  on the  $i$ -th row and the  $j$ -th column is defined as follows:

$$a_{ij} = \begin{cases} 1, & \text{if there is an edge between node } i \text{ and } j; \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

The degree  $k_i$  of the node  $i$  refers to the number of nodes linked to the node  $i$  which is expressed as:

$$k_i = \sum_{j=1}^N a_{ij}. \quad (5)$$

- **Average neighbor degree:** It measures the average degree of neighbors for each node. The average degree of the node  $i$  is:

$$k_{nn,i} = \frac{1}{Num(i)} \sum_{j \in N(i)} k_j, \quad (6)$$

where  $N(i)$  is the set of neighbors of the node  $i$ ,  $Num(i)$  is the number of nodes in  $N(i)$  and  $k_j$  is the degree of the node  $j$  that belongs to  $N(i)$ .

- **Degree assortativity:** In order to identify the relevance of the centrality of a node to its neighbor nodes, we

take a degree assortativity or a mean nearest neighbor connectivity as a metric of connection similarity [31]. It is a measure of the degree to either end of the edge by calculating the Pearson Correlation Coefficient  $r$ :

$$r = \frac{\sum_i k_i j_i - M^{-1} \sum_i k_i \sum_i j_i}{\sqrt{[\sum_i k_i^2 - M^{-1} (\sum_i k_i)^2][\sum_i j_i^2 - M^{-1} (\sum_i j_i)^2]}}, \quad (7)$$

where  $k_i, j_i$  are the degrees of the node for the ends of the  $i$ th edge in the network, along with  $i = 1 \dots M$ . The range of the coefficient  $r$  is between -1 and 1.

- **Clustering coefficient:** Clustering coefficient represents the ratio of neighbor nodes that are connected to each other of a node. By analyzing this metric, a highly aggregate coefficient means that the local network centering on this node is gathered together densely in a collaboration network. Suppose the degree of node  $i$  in the network is  $k_i$ , i.e., it has  $k_i$  neighbor nodes. If the  $k_i$  neighbor nodes are also neighbors, there are  $\frac{k_i(k_i-1)}{2}$  edges between these neighbors, and this is the case with the largest number of edges. Clustering coefficient  $C_i$  of the node  $i$  can be expressed as:

$$C_i = \frac{2E_i}{k_i(k_i-1)}, \quad (8)$$

where  $E_i$  actually represents the number of edges which exist between node  $i$  and  $k_i$  neighbor nodes.

- **Core:** The measure of the core can identify groups that are closely interconnected in the network [32]. Let  $G = (V, E)$  be an undirected graph and let  $H = (W, R)$  be a subgraph of  $G$ , i.e.,  $H \subseteq G$ . If subgraph  $H$  is a maximal connected subgraph in which all nodes have the degree at least  $K$ , it is defined to be a  $K$ -Core subgraph of  $G$ . It is calculated by Eq. (9), in the subgraph  $H$ , for all  $i \in W$ :

$$k_{i \in W} \geq K. \quad (9)$$

Every node  $i \in V$  has a core number of  $K$ , if it belongs to a  $K$ -core but not to a  $(K+1)$ -Core. If the nodes of the  $K$ -Core subgraph of  $G$  that corresponds to the maximum value of  $K_{max}$ , we denote the main core as the maximum core number  $K_{max}$ . In our network, connected nodes are independent of the other nodes connected to nodes existing outside the group.

- **Clique:** The clique is defined as the largest set of nodes that all nodes are directly adjacent to each other. Under the node removal operation, a clique has an invariant attribute: if a node is deleted from a clique, the rest of the subgraph is still a clique. Each node contributes a  $q$ -clique (a completely connected subgraph  $Q$  consisting of  $q$  nodes) onto  $G$ . It indicates that each node in  $Q$  is no more than a distance  $q$  away from others. Thus, we can know that a clique is a connected component which guarantees each author of this clique writes at least one paper with all other authors.
- **Degree centrality:** The node  $v$ 's degree centrality is to measure the number of other nodes which are directly connected to the node [33]. For normalization, it is

divided by the maximum possible degree  $n - 1$  of the whole network. So it is expressed as:

$$C_D(i) = \frac{k_i}{N - 1}. \quad (10)$$

- **Closeness centrality:** Throughout our established network, closeness centrality tends to give high values for nodes near the network center and high-closeness centrality nodes are generally important influencers. In order to calculate this metric, we have [33]:

$$C_C(i) = \frac{N - 1}{\sum_{j \neq i} d_{ij}}. \quad (11)$$

3) *Turing Number (TN)*: To find hierarchical relationships of the network, we explore relationships of the distance to the Turing Award and the related metrics (bibliometric-level and network-level). Enlightened by the previous study on the Erdős Number, we define the metric of the Turing Number, first proposed, to measure the distance to the Turing Award. The explanation is as follows:

**Turing Number (TN):** The Turing Number depicts the distance between the author and the Turing Award, which is similar to the concept of the shortest path distance. The *TN* of a Turing Award laureate is zero. For assigning *TN*, someone must be the coauthor of the research paper while another person needs to have a limited *TN*. Under the circumstance of treating the Turing Award laureates as a group, if one person's *TN* is  $T + 1$ , the one's distance to the Turing Award is  $T + 1$ , where  $T$  is the lowest *TN* of any coauthor. A Turing Number refers to the minimum value of the shortest path of every author to all the Turing Award laureates. In other words, we calculate the shortest path length of each scholar to all the Turing Award laureates, and we take the smallest number as the *TN*. From the perspective of complex networks, the smaller the value of *TN* is, the smaller the distance between scholars and the Turing Award laureates is. Based on the definition of *TN*, we extend the concept of the Erdős Number and measure the shortest path length of a given author to any Turing Award laureate.

We anticipate to discover correlations of authors' productivity, position in the network and distances to the Turing Award.

### III. RESULTS

An outstanding contribution of our work is the establishment of the Turing Award Collaboration Network and the exploration of related metrics to analyze this network. We explore the evolution of the network from 1974 to 2016. The results fall into two categories: bibliometric-level and network-level analysis.

#### A. Bibliometric-level Analysis

We first focus on the most intuitive metrics: the number of authors and papers. Fig. 2(a) shows the number of annual authors. The embedded graph displaces a log-linear distribution plotted with the same data that matches the exponential fitting  $a * \exp(bx)$ , where  $b = 0.063$  and  $R^2 = 0.94$ . Fig. 2(b) shows the number of papers each year and the embedded

graph displaces a log-linear distribution plotted with the same data that matches the exponential fitting  $a * \exp(bx)$ , where  $b = 0.080$  and  $R^2 = 0.97$ .  $R^2$  is used to describe the extent fitting of the linear function, where the higher the value is (i.e., closer to 1), the better the fitting is. We can observe that the two metrics in each year increase exponentially.

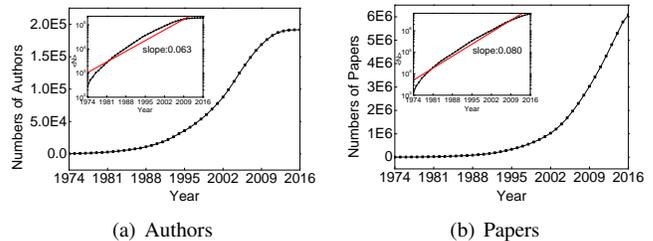
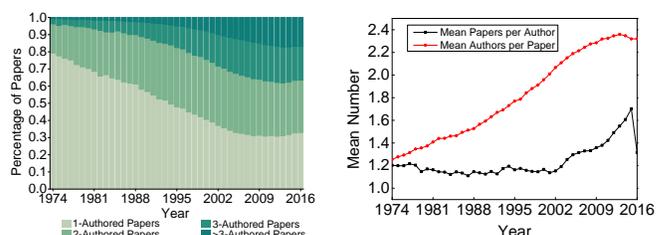


Fig. 2: The number of authors and papers each year. (a) The evolution of the number of authors versus year. (b) The evolution of the number of papers versus year. Each embedded graph displaces a log-linear distribution plotted with the same data. It indicates that the number of author and papers generated in each year increases exponentially.



(a) Distribution of authors per paper (b) Mean authors per paper and mean papers per author

Fig. 3: Percentage of papers and mean authors per paper, papers per author each year. (a) Annual percentage of papers published by one, two, three, and more than three authors. (b) The mean papers per author versus year (black line) and the mean authors per paper versus year (the red line).

As shown in Fig. 3(a), we can observe the changes in the collaboration pattern. In 1974, 79.3% authors publish papers by themselves and the collaboration strength among scholars is still weak since only 0.8% papers have more than three authors. Since then, scholars have tended to collaborate with others to publish papers. We can observe that the average number of authors per paper has increased from 1.25 to 2.28 throughout the whole period in Fig. 3(b). Subsequently, the number of papers written by individual authors continues to increase, but its percentage has dropped from 79.3% to less than 42.2%. In contrast, the number and the percentage of papers with multiple authors have increased significantly. The number of papers coauthored by two authors is the highest among those papers. The reason is probably that with the rapid development of computer science, it is increasingly difficult to publish papers alone. We can realize that it is necessary to focus on team research.

The mean number of papers per author and authors per paper are plotted in Fig. 3(b). From 1974 to 2002, the mean

papers per author is hovering between 1.0 and 1.2 yearly. After 2002, it shows an upward trend, till 2015 to about 1.70, therewith following by some decline. Since the 1970s, papers' average authors have increased quickly. At present, each paper has an average of two authors.

### B. Network-level Analysis

The analysis of these network-level metrics provides insights into the evolution of the Turing Award Collaboration Network. Meanwhile, we compare some calculated metrics with those in the random network to better reflect the realities.

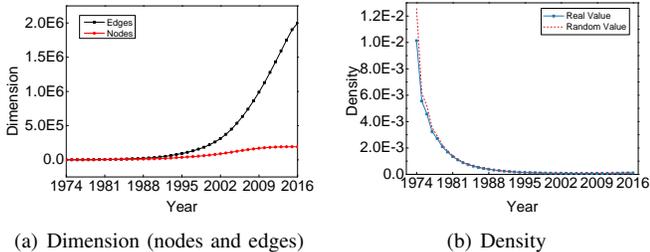


Fig. 4: Evolution of dimension and density. (a) It describes the growth of nodes and edges each year. The red line represents the trend of node, and the black line represents the trend of edge. (b) It shows density of the Turing Award Collaboration Network compared to the random network. The blue and red lines represent trends in the Turing Network and the random network, respectively.

The evolution of nodes and edges in the entire collaboration network is shown in Fig. 4(a). It is demonstrated clearly that due to the growing number of authors and their collaborators, these metrics have increased notably each year.

In Fig. 4(b), we can observe the phenomenon that with the same trend for the random network, the density is also declining. It can be understood that the number of published papers or collaborations with other authors is limited due to the increase number of new scholarships, resulting in decreased density in both networks.

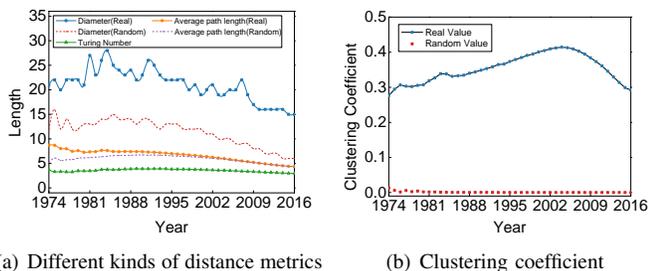


Fig. 5: Evolution of distance metrics and clustering coefficient. (a) It depicts trends of network diameter, average shortest path and  $TN$ . (b) It shows the evolution of clustering coefficient. The solid line indicates the evolution of the Turing Network, and the dotted line is the evolution of the random network.

Fig. 5(a) shows the changes of diameter, which starts at 21, and then reaches the maximum 28 in 1984. Next, it drops to

around 15 in 2016. The initial increasing of distance indicates that the network is gradually expanding, but the subsequent declining of the metric indicates that as the network continues to expand, the extent of collaboration is also increasing. It has the same tendency as the diameter of the random network, but the values are larger, and they indicate that relationships in this network are more complicated and dispersed than those in the random network.

The average shortest path length is plotted in Fig. 5(a). It can be seen that the initial network of the shortest path length is about 8.7 in 1974, and the path length steadily decreases from 1974 to 2016. Furthermore, the shortest path converges to about 4.3 at a later stage, and it indicates that a scholar in the network requires only four or five steps to achieve another scholar, which means that scientific information can be easily obtained by the need of the researcher [34]. Compared with the random network, the length is larger initially, but then it basically coincides with the random network. In accordance with the ‘‘Six Degree Separation’’ theory, the network is stabilized gradually. In addition,  $TN$  has the same trend as the average path length. But the value of  $TN$  is smaller, which shows that the distance for authors to the Turing Award is shorter than the average distance in the network. We can notice that the value of  $TN$  differs greatly from the diameter. We deem that the definition of  $TN$  and network diameter is different. The former refers to the distance of any scholars to the group of the Turing Award laureates, and the latter measures the distance between any two scholars in the network. Due to the difference in definition, the average of the two metrics differs by about 20. Therefore, the diameter of the network fluctuates relatively large.

In the Turing Award Collaboration Network, clustering coefficient increases from 0.28 in 1974 to 0.41 in Fig. 5(b). But it starts to decrease in 2004. Although the number of scholars has been increasing, they are gathered to a certain extent, not an unlimited association. This phenomenon shows specifically that in the twenty-first century, the collaboration pattern becomes increasingly unified over time. In addition, this metric is far greater than that in the random network, which indicates that the collaboration network tends to converge to form a high-density aggregation group.

The above analysis shows that scholars are gradually involved in tight collaboration and a large number of scholars join the network through the collaboration. In the dynamics of degree (Fig. 6), we can observe that the value is basically year-on-year rising with exactly the same trend in this metric of the random network. This metric shows a basic trend of linear growth and it distributes between 2 and 20. So far, the average node degree in the network has reached 20. The embedded graph in Fig. 6 displaces the logarithmic linearity that matches the exponential fitting  $a * \exp(bx)$ , where  $b = 0.020$  and  $R^2 = 0.93$ .

The Barabási-Albert model [35] of the scale-free network was first proposed in 1999. The significant feature of a scale-free network is that the degree distribution follows a power-law distribution  $p(x) = cx^{-\alpha}$ , where the scale-free coefficient ( $-\alpha$ ) is generally negative. Therefore, to explore whether the Turing Award Collaboration Network is a scale-free network,

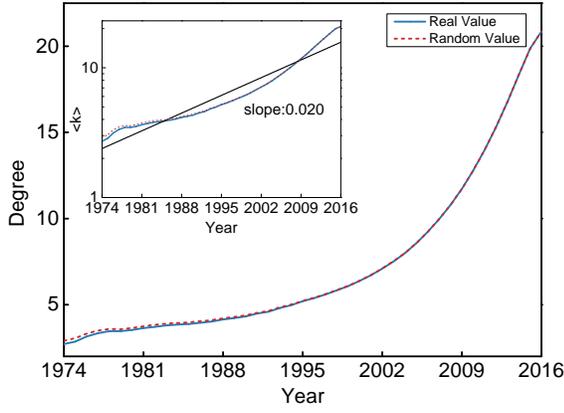


Fig. 6: Evolution of average degree. The embedded graph shows the log-linear graph. The blue line is the trend of the Turing Network and the red line is the evolution of the random network.

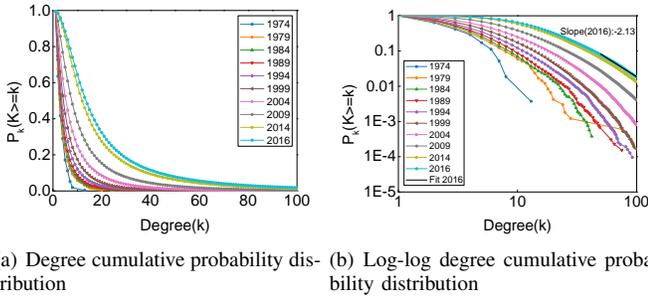


Fig. 7: Evolution of degree distribution. (a) It describes the degree cumulative probability distribution each year. (b) It shows the log-log degree cumulative probability distribution of the same network. Each line in both graphs represents the degree distribution of a certain year.

we need to analyze the degree distribution. We plot the degree distribution of the Turing Award Collaboration Network at several intervals of time points (i.e. 1974, 1979, 1984, 1989, 1994, 1999, 2004, 2009, 2014, 2016) in Fig. 7. As a result of the degree distribution, several nodes have a high degree. However, a large number of nodes are low. From the point, degree distribution seems to follow the power-law distribution essentially, especially when the dimensions of the network (nodes and edges) are large.

Subsequently, we plot the log-log distribution of degree distributions for different years in Fig. 7(b), which corresponds to the time points in Fig. 7(a). It can be concluded that these distributions are not purely power-law, otherwise, these points roughly stay in a straight line. In contrast, the tail of the distribution conforms to an exponential decay in 2016. In other words, the entire network cannot be used to fit the power-law distribution. From the analysis mentioned above, the Turing Award Collaboration Network cannot be considered as a scale-free network.

The relationship among scholars depends on the number of neighbors and their locations. In some networks, it has been noticed that the degree of their neighbors is related to

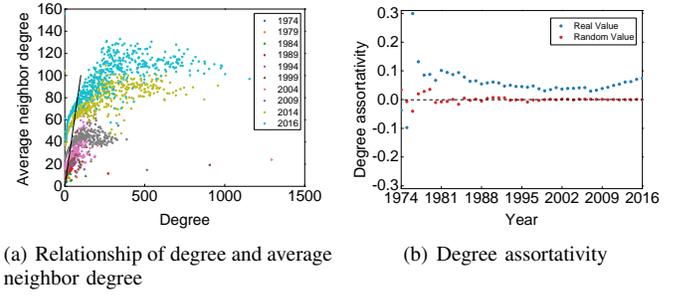


Fig. 8: Evolution of degree correlation. (a) Correlation of degree and average neighbor degree. Each color point refers to a certain year. The black line is the identity line  $x = y$ . (b) Degree assortativity of the Turing Award Collaboration Network (blue point), compared to the random network (red point).

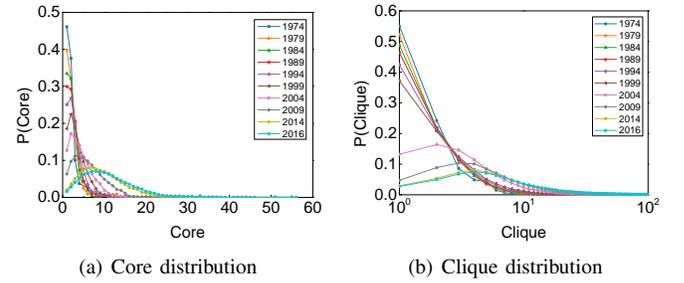


Fig. 9: Evolution of core and clique distribution. (a) This figure shows the  $k$ -core distribution of the Turing Network. (b) It shows the clique distribution of the Turing Network. Each line refers to one specific year.

their own degree. The degree for their neighbors is apt to be low when their degree is low. Conversely, the neighbor's degree is high when their own degree is high. In collaboration pattern, the prevalent authors are highly associated with other popular authors and the less popular authors are likely to be associated with the popular authors. We can quantify the relevance of these degrees by analyzing the correlation of the average neighbor degree and its own degree. The result is shown in Fig. 8(a).

The solid line represents the same degree as its average neighbor degree. From Fig. 8(a), we can discover that the average degree of the small-degree nodes' neighbors is significantly larger. But with increase of degree, the average neighbor degree of nodes is significantly higher than their degree. Another more accurate way is to calculate the assortativity coefficient, which needs to measure the preferences of nodes attached to other nodes in any way. The positive value of the degree assortativity indicates the correlation among similar-degree nodes (assortativity), while the negative value represents the relationships among different-degree nodes (disassortativity). In Fig. 8(b), the current degree assortativity of the Turing Award Collaboration Network is 0.074, positive, this condition indicates that the author tends to collaborate with other authors of a similar number of collaborators. Compared with the values in the random network, it can be seen that

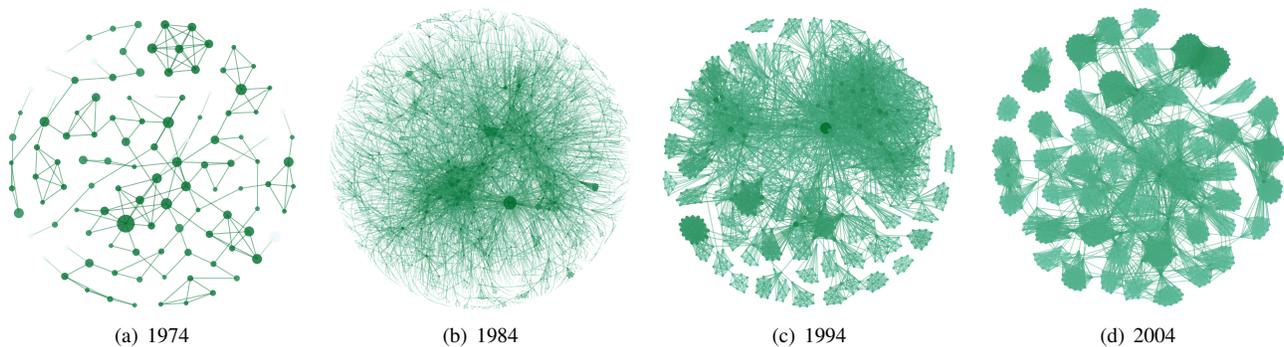


Fig. 11: Evolution of the Turing Award Collaboration Network. (a) It shows the Turing Network of all authors in 1974. (b) Due to the large number of nodes in Turing Network in 1984, it's not easy to observe network structure visually. Therefore, we select the 5-Core subnetwork of the Turing Network in 1984. (c) It shows the 8-Core subnetwork of the Turing Network in 1994. (d) It shows the 15-Core subnetwork of the Turing Network in 2004. Network graphs were produced in Gephi, using the Fruchterman Reingold layout with scaling set to 5000 and gravity to 10. Node size is proportional to the authors degree.

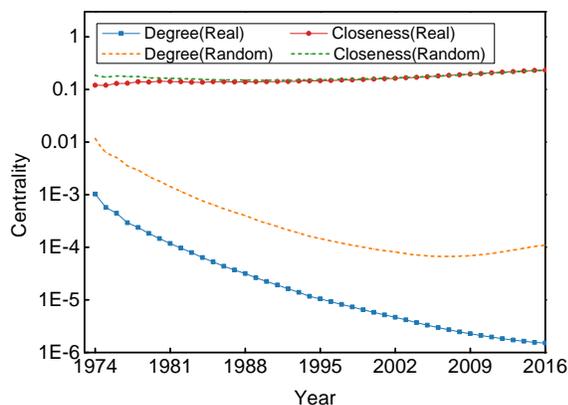


Fig. 10: Evolution of degree centrality and closeness centrality. The solid line refers to the evolution of the Turing Network, and the dotted line refers to the evolution of the random network.

the degree assortativity has been increasing steadily, which is related to the increasing number of authors.

By analyzing the core and the clique, we can acknowledge the evolution of the network group. The sizes of the core and the clique are both rising each year, which also conforms to the rule of network evolution. The value of the clique is similar to the core's at the early stages, while the clique's growth trend is significantly higher than the core's later. This phenomenon shows that extensive collaboration in the network is more prevalent than collaborating with the same person or group.

Then, in order to inquire the trend more clearly, we can observe the distribution in Fig. 9. The K-Core distribution of the Turing Award Collaboration Network is shown in Fig. 9(a). With the increase of K, the distribution of K-Core tends to be gentler. The majority of authors belong to small K-Cores (less than 9) and 20-Core contains most of the authors in 2016. The largest K-Core is 56 in 2016, which is not the group of the Turing Award laureates (K-Core is 51). The distribution of the clique is shown in Fig. 9(b), which is similar to the

K-Core. However, the value of the clique is much greater than K-Core's via the magnitude of the two pictures' abscissa (100 vs 10). This also verifies our thought: The Turing Award Collaboration Network tends to be more team collaboration (also called "baotuan" in Chinese).

Degree centrality quantifies the number of nodes connected to other nodes. Closeness centrality evaluates whether nodes are connected to other prominent nodes or not.

From Fig. 10, we can observe that degree centrality of the Turing Award Collaboration Network has been decreasing over time and it is smaller than the values in the random network. As the number of nodes in the network increases exponentially, it still leads to decreased degree centrality. This means that the links among scientists are almost similar across the network. It brings about the simultaneous development of opportunities for all scientists.

Closeness centrality shows a growth trend in Fig. 10. In contrast to degree centrality, the nodes have an increased pattern. The slope of the curve is larger at the later stages of the network dynamics, while the smaller slope of earlier stages can be attributed to the preferential attachment compared with the random network.

Because the number of nodes in the early network is small, the new added nodes are more likely to be attached to the new introduced nodes in each period. But the added nodes are more likely to be attached to higher nodes, resulting in incense of the slope of closeness centrality. This makes the newly added nodes hard to perform on the shortest path of the node pair during network evolution.

## IV. DISCUSSIONS

### A. Evolution of the Turing Award Collaboration Network

After analyzing the structure of the Turing Award Collaboration Network over years, we proceed to observe the evolution of the network intuitively. Therefore, we plot the evolution of the Turing Award Collaboration Network, as shown in Fig. 11. Because of the plentiful years involved, we take 10 years as a span to show the changes from 1974 to 1984, 1994, 2004. As there are overmuch nodes and edges, we only take

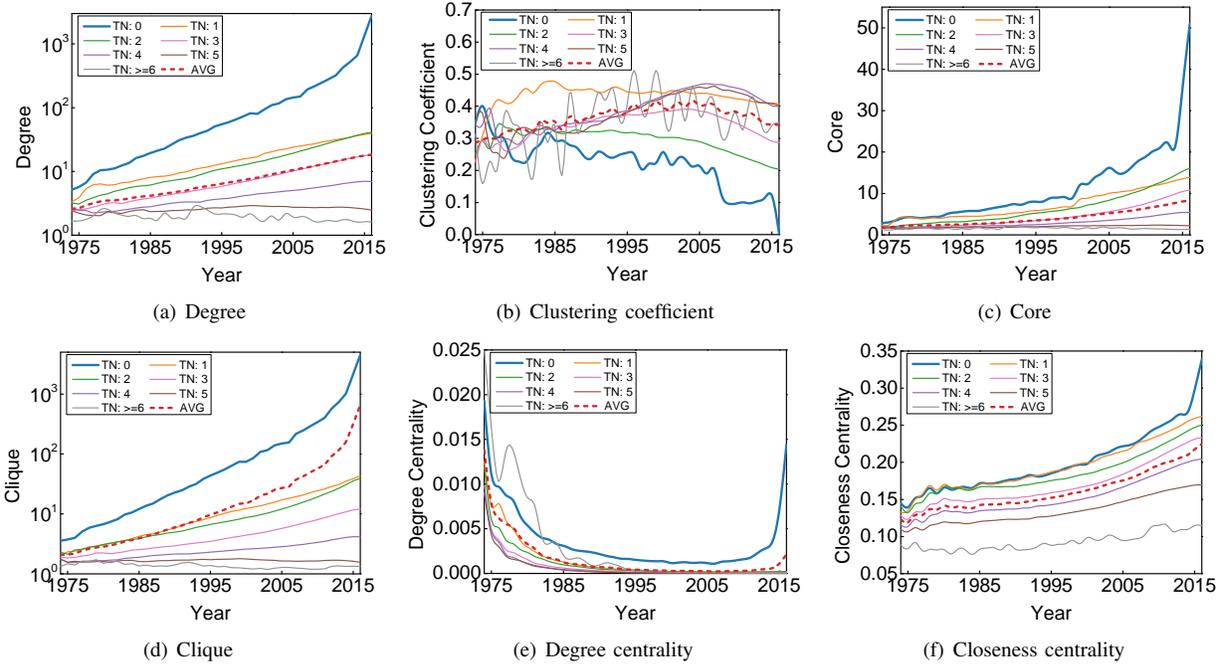


Fig. 13: Annual relationship of  $TN$  and specific network-level metrics. Each figure shows changes of a certain network-level metric against  $TN$ . The dotted line indicates average change of each metric.

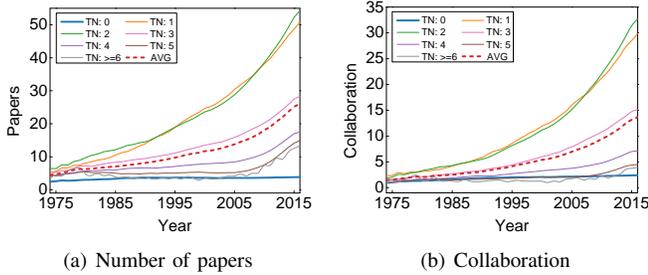


Fig. 12: Annual relationship of  $TN$  and specific bibliometric-level metrics. (a) It shows changes in the numbers of papers against the value of  $TN$ . (b) It shows changes in the numbers of collaborations with varying the value of  $TN$ . The dotted line indicates the average change of each metric.

the center of the network with the high K-Core. From Fig. 11, we can notice that with the development of the network, a growing number of nodes enter the network and the new relationships of the current nodes increase. As the number of nodes in the network increases, the connection among nodes becomes closer. In addition, the density of the network rises as the connections of nodes increase.

As the Turing Award laureates join the network gradually, they not only bring mounting nodes to join but also make the Turing Award for the network more central. Although we only take the high K-Core of the network, we can still observe that the network's group aggregation and contact closely. Moreover, the central group is also gradually expanding. This result could be used to predict those most likely to win Turing Award and the new join should fit the variation trend of the

network.

### B. Correlation of Turing Number ( $TN$ ) and Distinct Metrics

In order to further explore the relationship of the various metrics and the Turing Award, we begin to analyze the impact of different  $TN$  on the relevant metrics. Table I shows the evolution of the  $TN$ . The time phase represents the change of the  $TN$  from 1974 to 1979, 1984, 1994, 2004, 2009, 2014, 2016.

1) *Bibliometric-level Metrics Relationship*: We first analyze the relationship of the bibliometric-level metrics and  $TN$ . The results are shown in Fig. 12. When  $TN$  are 1, 2, 3, the author's number of papers and collaborations are above the average. Then, both of graphs demonstrate that the smaller the  $TN$  is, the larger the bibliometric-level metric is. The  $TN$  0 in Fig. 12 is the average of all the Turing Award laureates. It can be seen that the number of papers and collaborations of the Turing Award laureates are no more than that of the general scholars. Besides, scholars with the  $TN$  2 have a higher number of collaborations and papers than scholars' with the  $TN$  1 in the later period. There is an upward trend in the growth of the network for all authors.

2) *Network-level Metrics Relationship*: We analyze the impact of  $TN$  on the network-level metrics, and the results are shown in Fig. 13. The  $TN$  0 in Fig. 13 is the total value of all the Turing Award laureates, while The  $TN$  0 in Fig. 12 is the average. It is not appropriate for bibliometric-level metrics to use total value. For example, if we use total value, the amount of Turing Award laureates paper is tiny to the rest authors paper. Because the total Turing Award laureates number is tiny. Network-level metrics is based on collaboration network structure. The influence of Turing Award laureates

TABLE I: Evolution of  $TN$  from 1974 to 1979, 1984, 1994, 2004, 2009, 2014, 2016.

Metrics	Time phase								
	1979	1984	1989	1994	1999	2004	2009	2014	2016
Turing Number	3.469	3.767	3.899	3.866	3.760	3.571	3.305	3.039	2.964

can be reflected through the network structure. So it is better to consider the total value of  $TN$ . We can realize that the relationships of  $TN$  and the network-level metrics are not exactly the same as the bibliometric-level metrics.

The degree (Fig. 13(a)), the core (Fig. 13(c)) and the clique (Fig. 13(d)) show that the smaller the  $TN$  is, the greater the values are. However, for clustering coefficient depicted in Fig. 13(b), the correlation with  $TN$  cannot be seen and there is no growth trend towards the year, which is similar to the annual clustering coefficient in Fig. 5(b).

In terms of centrality, closeness centrality is increasing yearly. The smaller  $TN$  is better for closeness centrality. However, degree centrality does not have such a strong correlation. It is noteworthy that, as time goes by, the correlations of  $TN$  and degree centrality are not large at later stages, despite some fluctuations exist. But the changes of the closeness centrality are more obvious over time.

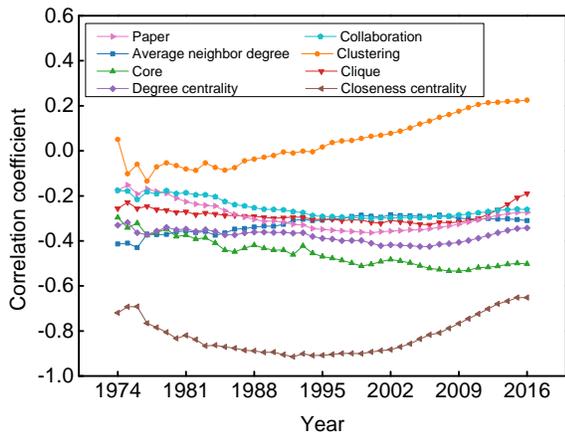


Fig. 14: Annual correlation coefficient of  $TN$  and specific metrics.

3) *Correlation Coefficient*: For the purpose of checking whether existing measures are associated with  $TN$  and changes in the subsequent years, we adopt the Spearman Correlation Coefficient to calculate the relationship of the relevant metrics and  $TN$  from 1974 to 2016, because many metrics do not conform to the normal distribution and the non-parametric correlations are more appropriate and robust than the Pearson correlations. Spearman correlation coefficient is defined as Pearson correlation coefficient between rank variables. For the samples with size  $n$ , all original data are converted into rank data. And the correlation coefficient  $\rho$  is expressed as:

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}. \quad (12)$$

where the raw data is assigned a corresponding rank based on their average descending position in the overall data. For example, given three values: 33,21,44, their rank would be 2,1,3.  $x_i$  is the  $TN$  rank of  $i$ th sample.  $\bar{x}$  is the average rank of all  $x$ .  $y_i$  is the metric rank of  $i$ th sample.  $\bar{y}$  is the average rank of all  $y$ . If  $y$  tends to increase as  $x$  increases, the Spearman Correlation is positive. If  $y$  tends to decrease as  $x$  increases, the Spearman Correlation is negative. If Spearman Correlation is 0, it indicates that  $y$  has no tendency as  $x$  increases. As  $x$  and  $y$  get close to a complete monotonic correlation, the Spearman Correlation increases in absolute value. When  $x$  and  $y$  are completely monotonic, the absolute value of the Spearman Correlation Coefficient is 1.

Fig. 14 shows that the relevant metrics are negatively correlated with  $TN$  except for the clustering coefficient. When the absolute value of correlation coefficient is closer to 1, the metric are more correlated to  $TN$ . The trends of the vast majority of metrics are the same with some slight increase over years or some fluctuations. Of all the metrics, closeness centrality is of great relevant metric (the absolute value of Correlation Coefficient is about 0.6 in 2016). There are some declines in the middle of the periods. The second is the core.

From the exploration mentioned above, we can understand that the correlation of the relevant metrics and  $TN$  is always significant over time. From the perspective of the centrality, the correlation of the closeness centrality and  $TN$  is the highest. In other words, the author with a high closeness centrality is often closer to the Turing Award. In addition, the coefficient has an increasing trend yearly for degree centrality but has a stable trend for closeness centrality.

## V. CONCLUSION AND FUTURE WORK

The ACM A.M. Turing Award is recognized as the highest honor prize in the realm of computer science. Half a century has passed since the first laureate was awarded. However, there is scarce research to explore the potential mechanism of the Turing Award and collaboration pattern of its laureates. Our work is the first research to study the evolution of the Turing Award Collaboration Network. In this paper, we analyze the evolution of the Turing Award Collaboration Network, which is extracted from the journal and conference papers recorded in DBLP from 1974 to 2016. We combine the metrics of the bibliometric-level method and the network-level approach to characterize the multiple attributes of the researchers.

The observations from the Turing Award Collaboration Network show that the collaboration of different periods has a different intensity as well as the type of collaborative behavior. The number of papers and authors of the network has both increased exponentially which indicates the field of computer science has developed rapidly over the past 42 years. In the early stage, scholars did not pay attention to collaboration, so

they were more inclined to publish papers alone. Gradually, they began to focus on collaboration so that multiple authors' papers account for a larger proportion than the single authors' papers. But the average number of multiple authors' papers is limited to two authors. This consequence is consistent with other related disciplines [36], [37].

Scholars can connect to each other in four path lengths, characterizing the small world properties. However, the Turing Award Collaboration Network is not the scale-free network, and then clustering coefficient began to decline in 2004. The reason may be that each author can collaborate with a limited number of scholars, so the scholars in the Turing Award Collaboration Network cannot be "preferential attachment" freely as the scale-free network. In addition, scholars tend to form a group to study the academic. Moreover, we further explore the relationship of the Turing Number ( $TN$ ) and the measured metrics. As a result, we draw a conclusion that the closer the authors are to the Turing Award, the better the metrics are. It shows these authors are more superior. Compared with the random network, this network is more closely linked to the group, but less connected because of the community.

As the possible future research direction, we believe that it would be necessary to expand the analysis of the performance metrics, including citations and h-index, [5], [38], to perceive whether the collaboration has an impact on them.

## REFERENCES

- [1] H. Zou, X. Zhang, and J. Yang, "ACM Turing Award: History and enlightenment," in *Education Technology and Computer Science, 2009. ETCS'09. First International Workshop on*, vol. 2. IEEE, 2009, pp. 105–111.
- [2] S. Tong and P. Ahlgren, "Evolution of three Nobel Prize themes and a Nobel snub theme in chemistry: A bibliometric study with focus on international collaboration," *Scientometrics*, vol. 112, no. 1, pp. 75–90, 2017.
- [3] E. M. Schlagberger, L. Bornmann, and J. Bauer, "At what institutions did Nobel Laureates do their prize-winning work? an analysis of biographical information on Nobel Laureates from 1994 to 2014," *Scientometrics*, vol. 109, no. 2, pp. 723–767, 2016.
- [4] V. L. Hanson, "Celebrating 50 years of the turing award," *Communications of the ACM*, vol. 60, no. 2, pp. 5–5, 2017.
- [5] J. Zhang, Z. Ning, X. Bai, X. Kong, J. Zhou, and F. Xia, "Exploring time factors in measuring the scientific impact of scholars," *Scientometrics*, vol. 112, no. 3, pp. 1301–1321, 2017.
- [6] Y. Asahiro, T. Kubo, and E. Miyano, "Experimental evaluation of approximation and heuristic algorithms for maximum distance-bounded subgraph problems," *The Review of Socionetwork Strategies*, pp. 1–19, 2019.
- [7] F. de Moya-Anegón, V. P. Guerrero-Bote, C. Lopez-Illescas, and H. F. Moed, "Statistical relationships between corresponding authorship, international co-authorship and citation impact of national research systems," *Journal of Informetrics*, vol. 12, no. 4, pp. 1251–1262, 2018.
- [8] S. Kumar, "Co-authorship networks: a review of the literature," *Aslib Journal of Information Management*, vol. 67, no. 1, pp. 55–73, 2015.
- [9] L. Lyu, W. Wu, H. Hu, and R. Huang, "An evolving regional innovation network: collaboration among industry, university, and research institution in china's first technology hub," *The Journal of Technology Transfer*, vol. 44, no. 3, pp. 659–680, 2019.
- [10] A. Cardillo, S. Scellato, and V. Latora, "A topological analysis of scientific coauthorship networks," *PHYSICA A-STATISTICAL MECHANICS AND ITS APPLICATIONS*, vol. 372, no. 2, pp. 333–339, DEC 15 2006, Internation Symposium on Nonlinearity, Nonequilibrium and Complexity - Questions and Perspectives in Statistical Physics, Tepotzlan, MEXICO, NOV 27-DEC 02, 2005.
- [11] H. B. Ghafouri, H. Mohammadhassanzadeh, F. Shokraneh, M. Vakilian, and S. Farahmand, "Social network analysis of Iranian researchers on emergency medicine: a sociogram analysis," *EMERGENCY MEDICINE JOURNAL*, vol. 31, no. 8, pp. 619–U117, AUG 2014.
- [12] A. Abbasi, L. Hossain, and L. Leydesdorff, "Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks," *JOURNAL OF INFORMETRICS*, vol. 6, no. 3, pp. 403–412, JUL 2012.
- [13] M. Perc, "The Matthew effect in empirical data," *JOURNAL OF THE ROYAL SOCIETY INTERFACE*, vol. 11, no. 98, SEP 6 2014.
- [14] Z. Neal, "Is the urban world small? the evidence for small world structure in urban networks," *Networks and Spatial Economics*, vol. 18, no. 3, pp. 615–631, 2018.
- [15] Y.-X. Liu, B. Lu, and Q. Zhang, "Empirical analysis of the coauthorship network based on DBLP," in *Machine Learning and Cybernetics (ICMLC), 2013 International Conference on*, vol. 3. IEEE, 2013, pp. 1070–1076.
- [16] J. Liu, Y. Li, Z. Ruan, G. Fu, X. Chen, R. Sadiq, and Y. Deng, "A new method to construct co-author networks," *Physica A: Statistical Mechanics and its Applications*, vol. 419, pp. 29–39, 2015.
- [17] X. Zhou, W. Liang, I. Kevin, K. Wang, R. Huang, and Q. Jin, "Academic influence aware and multidimensional network analysis for research collaboration navigation based on scholarly big data," *IEEE Transactions on Emerging Topics in Computing*, 2018.
- [18] D. Wang, K.-K. Yan, J. Rozowsky, E. Pan, and M. Gerstein, "Temporal Dynamics of Collaborative Networks in Large Scientific Consortia," *TRENDS IN GENETICS*, vol. 32, no. 5, pp. 251–253, MAY 2016.
- [19] C. Bird, P. Devanbu, E. Barr, V. Filkov, A. Nash, and Z. Su, *Structure and Dynamics of Research Collaboration in Computer Science*, pp. 826–837. [Online]. Available: <https://epubs.siam.org/doi/abs/10.1137/1.9781611972795.71>
- [20] M. Franceschet, "Collaboration in Computer Science: A Network Science Approach," *JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY*, vol. 62, no. 10, pp. 1992–2012, OCT 2011.
- [21] Y. Wu, S. Venkatramanan, and D. M. Chiu, "Research collaboration and topic trends in computer science based on top active authors," *PeerJ Computer Science*, vol. 2, p. e41, Jan. 2016. [Online]. Available: <https://doi.org/10.7717/peerj-cs.41>
- [22] J. Chen and J. A. Konstan, "Conference paper selectivity and impact," *Communications of the ACM*, vol. 53, no. 6, pp. 79–83, 2010.
- [23] V. Batagelj and M. Cerinšek, "On bibliographic networks," *Scientometrics*, vol. 96, no. 3, pp. 845–864, 2013.
- [24] S. Drożdż, A. Kulig, J. Kwapien, A. Niewiarowski, and M. Staszczek, "Hierarchical organization of h. eugene stanley scientific collaboration community in weighted network representation," *Journal of Informetrics*, vol. 11, no. 4, pp. 1114–1127, 2017.
- [25] P. Erdős and A. Rényi, "On random graphs," *Publicationes Mathematicae Debrecen*, vol. 6, pp. 290–297, 1959.
- [26] J.-C. Valderrama-Zurián, R. Aguilar-Moya, A. Cepeda-Benito, D. Melero-Fuentes, M.-Á. Navarro-Moreno, A. Gandía-Balaguer, and R. Alexandre-Benavent, "Productivity trends and collaboration patterns: A diachronic study in the eating disorders field," *PloS one*, vol. 12, no. 8, 2017.
- [27] J. C. Brunson, S. Fassino, A. McInnes, M. Narayan, B. Richardson, C. Franck, P. Ion, and R. Laubenbacher, "Evolutionary events in a mathematical sciences research collaboration network," *Scientometrics*, vol. 99, no. 3, pp. 973–998, 2014.
- [28] K. Li, J. Rollins, and E. Yan, "Web of science use in published research and review papers 1997–2017: a selective, dynamic, cross-domain, content-based analysis," *Scientometrics*, vol. 115, no. 1, pp. 1–20, 2018.
- [29] J. Scott, "Social network analysis: developments, advances, and prospects," *Social network analysis and mining*, vol. 1, no. 1, pp. 21–26, 2011.
- [30] E. Yan, Y. Ding, and Q. Zhu, "Mapping library and information science in china: A coauthorship network analysis," *Scientometrics*, vol. 83, no. 1, pp. 115–131, 2010.
- [31] G. T. Cantwell and M. Newman, "Mixing patterns and individual differences in networks," *Physical Review E*, vol. 99, no. 4, p. 042306, 2019.
- [32] V. Batagelj and M. Zaveršnik, "Fast algorithms for determining (generalized) core groups in social networks," *Advances in Data Analysis and Classification*, vol. 5, no. 2, pp. 129–145, 2011.
- [33] L. C. Freeman, "Centrality in social networks conceptual clarification," *Social Networks*, vol. 1, no. 3, pp. 215–239, 1978.

- [34] C. Zhang, Y. Bu, Y. Ding, and J. Xu, "Understanding scientific collaboration: Homophily, transitivity, and preferential attachment," *Journal of the Association for Information Science and Technology*, vol. 69, no. 1, pp. 72–86, 2018.
- [35] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.
- [36] B. Elango and P. Rajendran, "Authorship trends and collaboration pattern in the marine sciences literature: A scientometric study," *International Journal of Information Dissemination and Technology*, vol. 2, no. 3, 2012.
- [37] N. Amsaveni, M. Manikandan, and M. Manjula, "Authorship pattern and collaborative research in bioinformatics," *International Journals of Computer Science and Mobile Computing*, vol. 2, no. 11, pp. 230–238, 2013.
- [38] X. Kong, M. Mao, W. Wang, J. Liu, and B. Xu, "VOPRec: Vector representation learning of papers with text information and structural identity for recommendation," *IEEE Transactions on Emerging Topics in Computing*, 2018.