Author(s): Semenov, Alexander; Veremyev, Alexander; Nikolaev, Alexander; Pasiliao, Eduardo L.; Boginski, Vladimir

Title: Network-based indices of individual and collective advising impacts in mathematics

Year: 2020

Version: Published version

Please cite the original version:

Semenov, A., Veremyev, A., Nikolaev, A., Pasiliao, E. L., & Boginski, V. (2020). Network-based
indices of individual and collective advising impacts in mathematics. Computational Social
Networks, 7(1). https://doi.org/10.1186/s40649-019-0075-0

RESEARCH

Network-based indices of individual and collective advising impacts in mathematics

Alexander Semenov[1], Alexander Veremyev[2], Alexander Nikolaev[3], Eduardo L. Pasiliao[4] and Vladimir Boginski[2*]

*Correspondence:
vladimir.boginski@ucf.edu
[2] Department
of Industrial Engineering
and Management Systems,
University of Central Florida,
12800 Pegasus Drive,
Orlando, FL, USA
Full list of author information
is available at the end of the
article

Abstract

Advising and mentoring Ph.D. students is an increasingly important aspect of the academic profession. We define and interpret a family of metrics (collectively referred to as "*a*-indices") that can potentially be applied to "ranking academic advisors" using the academic genealogical records of scientists, with the emphasis on taking into account not only the number of students advised by an individual, but also subsequent academic advising records of those students. We also define and calculate the extensions of the proposed indices that account for student co-advising (referred to as "adjusted *a*-indices"). In addition, we extend some of the proposed metrics to ranking universities and countries with respect to their "collective" advising impacts, as well as track the evolution of these metrics over the past several decades. To illustrate the proposed metrics, we consider the social network of over 200,000 mathematicians (as of July 2018) constructed using the Mathematics Genealogy Project data.

Keywords: Social networks, Big data, Scientific advising impact, *a*-Indices, Mathematics Genealogy Project

Introduction

In recent years, universities and other research institutions have put a lot of emphasis on assessing and enhancing the productivity of their faculty. One aspect that has been traditionally deemed important in these efforts is the number and quality of a researcher's publications. The popular metrics of publication productivity include various quantities based on an individual's citation record (e.g., total number of citations, weighted citations, $i10$-index, h-index, etc.), typically accounting for the "prestige" measures of publication outlets (e.g., journal impact factors, 5-year impact factors, SNIP, CiteScore, etc.). However, besides publication output, another—possibly equally important—aspect of the academic profession success is associated with advising and mentoring Ph.D. students. One can argue that a successful academician is not only the one who publishes many highly cited articles, but also the one who successfully advises students, and further, whose students in turn become successful academic advisors, thus ensuring the continuity and prosperity of an academic discipline. Indeed, in the modern era, many universities emphasize the importance of effective mentorship and post-graduation academic productivity of their Ph.D. students.

This paper makes contributions towards a systematic network-based analysis of large-scale Ph.D. student advising data. We define and interpret a family of new network-based metrics (collectively referred to as "*a*-indices") that can be used for "ranking academic advisors" using the academic genealogical records of scientists. We rely on the well-known web-based Mathematics Genealogy project resource that has collected a vast amount of data on Ph.D. student advising records in mathematics-related fields.

Due to its popularity and public availability, MathGenealogy dataset has been used as a testbed in several previous studies. The basic characteristics of the MathGenealogy network snapshot from 2011, as well as those of the underlying network of countries, were presented in [1]. In [2], the authors analyzed the performance of students of those individuals who were near the beginning versus near the end of their academic careers and revealed interesting insights. Another study [3] used the data of Ph.D. degrees granted after 1973 and used it to compose a network of universities, where some of the universities were then labeled as strong sources ("authorities") of Ph.D. production, while the others were labeled as strong destinations ("hubs"). The authors of [4] presented a comprehensive analysis of the MathGenealogy network with respect to the classification of mathematics-related subjects, as well as most influential countries in terms of the Ph.D. graduates output. Further, they revealed the major "families" of mathematicians that originated in certain root nodes ("fathers" of mathematics' genealogical families), in the different "eras", covered by the project data. A new concept of eigenvector-based centrality was defined and tested on the MathGenealogy network in [5]. In [6], the authors proposed the so-called "genealogical index" for measuring individuals' advising records. As it will be seen below, one of the indices proposed in this paper can be viewed as a special case of the "genealogical index" proposed in [6].

This paper takes a further step towards studying and ranking academic advising impact using MathGenealogy social network. The emphasis of this study is on taking into account not only the number of students advised by an individual but also subsequent academic advising records of those students, while providing the respective metrics that are easy to calculate, understand, and interpret. It should also be noted that this study does not aim to explicitly compare the proposed indices with other metrics/results available in the aforementioned related literature. However, we believe that the presented approaches and results provide a new perspective on this interesting subject and further demonstrate the utility of social network analysis tools in the considered context.

The paper is organized as follows. In the next section, we briefly describe the MathGenealogy dataset and provide its basic characteristics along with definitions and notations that will be used in the paper. Next, we define and interpret the family of "*a*-indices" that we propose for ranking academic advisors. We then extend these definitions to take into account co-advising. Finally, we present the results obtained on the most recent snapshot of the MathGenealogy dataset, as well as investigate the evolution of individual and collective *a*-indices over the past several decades.

Semenov *et al. Comput Soc Netw*   (2020) 7:1

Page 3 of 18

Data description, notations, and basic characteristics of MathGenealogy network

To facilitate further discussion, we first describe the MathGenealogy dataset and provide its basic characteristics, as well as define graph-theoretic concepts that will be used in the paper.

Data description

The data were collected from the Mathematics Genealogy Project website[1] using a web-crawler software. The dataset contains the records about nearly 231,000 mathematicians (as of July 2018). The information for each mathematician in the database includes name, graduation year, university, country, Ph.D. thesis topic and its subject classification, as well as the list of students advised by this individual. This available data allowed us to construct the directed network of advisor–advisee relationships.

Related graph-theoretic concepts

Due to the fact that the considered dataset is a directed network, it is represented by a directed acyclic graph $G = (N, \mathcal{A})$, with a set of n nodes, $N = \{1, \ldots, n\}$, and a set of m arcs (links) $\mathcal{A}$, where the mathematicians are represented by the nodes of the graph, and the relation "i is an advisor of j" is represented by an arc from i to j. The in-degree ($\deg^{in}(i)$) and out-degree ($\deg^{out}(i)$) of node i are the numbers of the arcs coming into and going out of node i, respectively. Clearly, the in-degree of node i is the number of this individual's Ph.D. dissertation advisors (equal to one for many nodes in the network, although a substantial fraction of nodes do have higher in-degrees), whereas the out-degree of node i is the number of Ph.D. students that this individual has successfully graduated. Node j is said to be reachable from node i if there exists a directed path from i to j. The number of links in the shortest path from i to j is referred to as the distance between these nodes and denoted by $d(i, j)$ ($d(i,j) = +\infty$ if there is no such path). A group of nodes is said to form a weakly connected component if any two nodes in this group are connected via a path and no other nodes are connected to the group nodes, where the directions of arcs in a path are ignored.

The *harmonic centrality* of node i is defined as $C_h(i) = \sum_{j \in N} \frac{1}{d(i,j)}$ [7, 8]. The *decay centrality* of node i is defined as $C_d(i) = \sum_{j \in N} \delta^{d(i,j)}$ [9, 10], where the parameter $\delta \in (0,1)$ is user-defined, although it is often set at $\delta = 1/2$, which is the value used in this study (it is assumed that $1/d(i,j) = \delta^{d(i,j)} = 0$ if $d(i,j) = +\infty$).

Basic characteristics of MathGenealogy network

The retrieved network had 12,263 weakly connected components, with the giant weakly connected component having 208,526 nodes and 238,212 arcs (thus containing about 90% of all the nodes in the network). All the computational results presented below were obtained for this giant component. Further in the text, we will use the term "network" implying this giant weakly connected component.

---

[1] http://www.genealogy.ams.org//

Fig. 1 Out-degree distribution (log–log scale) of the MathGenealogy network
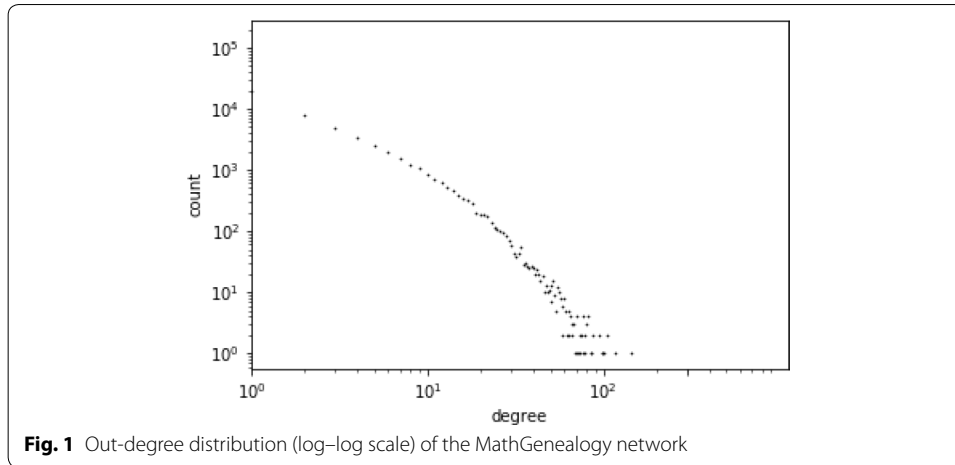
The analysis of many basic characteristics of an earlier snapshot of this network was conducted in [1]. Since such analysis is not the main focus of this study, we report only some of these basic characteristics for the most recent snapshot that are relevant to the material presented in this paper. The distribution of out-degrees in this network is presented in Fig. 1. As one can observe, it does resemble a power law, although it is not a "pure" power law, which is consistent with observations for many other real-world networks [11].

The out-degree correlation for all "tail-head" (or, "advisor–student") pairs of nodes corresponding to all arcs (directed links) in the considered directed network was calculated as follows. Consider an ordered list of all directed links $l \in \{1, \ldots, |\mathcal{A}|\}$ in the network, let i and j be the head and tail nodes of link l, and let $\deg_l^{out}(i)$ and $\deg_l^{out}(j)$ be their out-degrees, respectively. Thus, we have an array of size $|\mathcal{A}|$ of head nodes (denote the average out-degree of all nodes in this array by $\overline{\deg^{out}(i)}$) and an array of size $|\mathcal{A}|$ of tail nodes (denote the average out-degree of all nodes in this array by $\overline{\deg^{out}(i)}$). Then, the out-degree correlation (also sometimes referred to as the out-assortativity) can be calculated as:

$$r_{out} = \frac{\sum_{l=1}^{|\mathcal{A}|} (\deg_l^{out}(i) - \overline{\deg^{out}(i)})(\deg_l^{out}(j) - \overline{\deg^{out}(j)})}{\sqrt{\sum_{l=1}^{|\mathcal{A}|} (\deg_l^{out}(i) - \overline{\deg^{out}(i)})^2} \sqrt{\sum_{l=1}^{|\mathcal{A}|} (\deg_l^{out}(j) - \overline{\deg^{out}(j)})^2}}$$

The value of the out-degree correlation for this network was found to be approximately 0.055. This implies that on average there is a very minor correlation between the mentorship productivity of an advisor and a student. Therefore, we believe that in the proposed metrics and rankings of academic advisors it makes sense to "reward" those prolific advisors whose students are also successful academic mentors.

As for the in-degree distribution, it is not surprising that the majority of the nodes have in-degree equal to one. However, the network contains over 30,000 nodes with in-degree greater than one, which means that a substantial fraction (about 15%) of the mathematicians in the dataset had more than one Ph.D. advisor. Therefore, it is important to take into account the effects of co-advising, which is why we define "adjusted" versions of the proposed metrics (indices).

Advising impact metrics

In this section, we define four metrics ("*a*-indices") that we believe are appropriate for quantifying an individual's advising impact, with a focus on taking into account the mentoring success of an individual's students (going beyond just the number of the Ph.D. students that an individual has graduated). One way to address this is to consider the numbers of students and students-of-students, whereas another approach is to take into account all the academic descendants of an individual. These considerations are reflected in the following definitions.

Definition 1 (*a-index*) The *a*-index[2] of an individual i is the largest integer number n such that the individual i has advised n students (Ph.D. graduates) each of whom has advised at least n of their own students (Ph.D. graduates). Equivalently, this is the largest number n of out-neighbors of node i in the directed network such that each of these neighbors has out-degree of at least n.

Definition 2 (*a_∞-index*) The a_∞-index of an individual i is the total number of their academic descendants, computed as the largest number of distinct nodes that are reachable from node i through a directed path.

Definition 3 (*a_1-index*) The a_1-index of an individual i is the harmonic centrality of the corresponding node i in the directed network: $a_1(i) = C_h(i) = \sum_{j \in N} \frac{1}{d(i,j)}$.

Definition 4 (*a_2-index*) The a_2-index of an individual i is the decay centrality (with $\delta = \frac{1}{2}$) of the corresponding node i in the directed network: $a_2(i) = C_d(i) = \sum_{j \in N} \frac{1}{2^{d(i,j)}}$.

It can be seen from Definitions 1–4 that the *a*-index is a measure of the most "immediate" advising impact of an individual, which takes into account their advising success simultaneously with the advising success of their students.[3] Note that the *a*-index is similar to the *h*-index well-accepted for citations record evaluation; however, it turns out that it is rather hard to achieve a double-digit value of the *a*-index over one's academic career due to the fact that graduating a Ph.D. student is generally a less frequent event than publishing a paper. As it can be seen in Table 1, the highest *a*-index value in the considered dataset is 12 (achieved by only four mathematicians). Note that a relevant study [6] reported only one mathematician with the value of *a*-index ($g_{(1)}$ measure in their terminology) equal to 12. Overall, the *a*-index may be applicable as a metric of the advising impact for middle- to late-career academic scientists.

Note that the *a*-index can be extended in a straightforward fashion to reflect a more "long-term" advising impact of an individual by considering third, fourth, etc., generations of an individual's students as it was proposed in the definition of the "genealogical index" in [6]. However, the main issue with this approach is that close to 100% of the

---

[2] To be more consistent with the notation for the rest of the "*a*-indices" defined here, one may denote this index as a_0-index; however, for simplicity, throughout the paper we will call this metric the "*a*-index" (which may be viewed as an analogy to the *h*-index widely used as a citation metric).

[3] Of course, the out-degree of a node, that is, the number of advised students, is the simplest measure that assesses immediate advising impact; however, it is not in the scope of this study as it does not reflect the advising impact of descendants.

Table 1 Top individuals by *a*-index, with the *a*-index of at least 10 and their corresponding adjusted *a*-index

a-index	Name	Grad. year	Country of Ph.D.	Adjusted a-index
12	Heinz Hopf	1925	Germany	11
12	Jacques-Louis Lions	1954	France	11
12	Mark Aleksandrovich Krasnoselskii	1948	Ukraine	12
12	Erhard Schmidt	1905	Germany	10
11	Andrei Nikolayevich Kolmogorov	1925	Russia	10
11	C. Felix (Christian) Klein	1868	Germany	10
11	Heinrich Adolph Behnke	1923	Germany	9
11	Karl Theodor Wilhelm Weierstrass	1841	Germany	9
11	John Torrence Tate, Jr.	1950	United States	11
11	Ernst Eduard Kummer	1831	Germany	10
11	Reinhold Baer	1927	Germany	8
11	Salomon Bochner	1921	Germany	11
11	David Hilbert	1885	Germany	10
10	Lothar Collatz	1935	Germany	9
10	Günter Hotz	1958	Germany	10
10	Pavel Sergeevich Aleksandrov	1927	Russia	10
10	Edmund Hlawka	1938	Austria	9
10	Phillip Augustus Griffiths	1962	United States	9
10	Michael Francis Atiyah	1955	United Kingdom	9
10	Haim Brezis	1972	France	10
10	Thomas Kailath	1961	United States	10
10	R. L. (Robert Lee) Moore	1905	United States	10
10	Alan Victor Oppenheim	1964	United States	10
10	Shiing-Shen Chern	1936	Germany	10
10	Elias M. Stein	1955	United States	10
10	Richard Courant	1910	Germany	9
10	Hellmuth Kneser	1921	Germany	9
10	Emil Artin	1921	Germany	10
10	Lipman Bers	1938	Czech Republic	9
10	Issai Schur	1901	Germany	8
10	Roger Meyer Temam	1967	France	9
10	John Wilder Tukey	1939	United States	9
10	Philip Hall	1926	United Kingdom	10
10	Beno Eckmann	1942	Switzerland	9
10	Oscar Ascher Zariski	1925	Italy	10

mathematicians in the considered dataset would have zero values of such index, which would not allow one to effectively rank advisors' long-term impacts using this metric.

Therefore, in order to provide more practically usable quantifications of "long-term" advising impacts of individuals, especially for those scientists who are in the late stages of their careers and for those who have lived and worked centuries ago, we propose the a_1, a_2, and a_∞ indices. The a_∞-index essentially assigns equal weights to all the academic descendants of an individual, whereas the a_1 and a_2 indices prioritize (with different weights) the immediate (directly connected) students and students-of-students while still giving an individual some credit for more distant descendants. Possible practical interpretations of these indices are as follows.

The a_∞-index is appropriate for ranking the "root nodes" of the mathematics genealogy network, that is, nodes with zero in-degrees, which essentially correspond to "fathers" of mathematics' "genealogical families", such as those described in [4]. It is not practically significant to calculate this index for nodes with non-zero in-degree values, since their predecessors in the network would obviously have higher values of this index. Thus, the a_∞ index is interesting primarily from the perspective of history of mathematics, although it can certainly be calculated very easily for any contemporary mathematician.

On the other hand, the a_1-index and a_2-index do not necessarily possess the aforementioned property of the a_∞-index: the values of these indices may be higher for contemporary mathematicians than for the "fathers" of genealogical families due to the fact that an individual's immediate students and any other early-generation students attain higher index values than do any distant descendants. These indices are based on the well-known concepts of harmonic and decay centralities, which makes them easy to calculate and interpret, and hence, attractive from a practical perspective. These indices can be applied to an academic advisor from any era, thus providing a universal tool of assessing the academic advising impact. However, it is still likely that the advisors in the late stages of their careers would have higher values of these indices (especially the a_1-index that gives higher weights to distant descendants) than those in early-to-mid-stages of their careers. This is not surprising, since these indices are designed to assess the long-term advising impact beyond the number of immediate students.

Further, note that there are several natural extensions of these definitions. First, all of these indices can be adjusted by taking into account the effects of co-advising, that is, giving a special treatment to the cases when multiple individuals have advised the same student j (that is, with node j having multiple incoming links). These particular extensions are addressed in greater detail in the next section. Second, the a-index can also be defined for a specific country or university (similarly to the h-index of a journal among citations metrics), that is, considering the respective country or university as a "super-node", with the outgoing links directed to all the Ph.D. graduates ever produced (or produced during a specific time frame) by this country or university, respectively. The resulting *collective* advising impact values for universities and countries, based on MathGenealogy dataset, will also be presented below.

Advising impact metrics adjusted for co-advising

In this section, we define the extensions of our basic indices (Definitions 1–4) to handle the cases of co-advising, that is, the situations where one Ph.D. student was co-advised by more than one individual. It makes practical sense to introduce these definitions due to the fact that a substantial fraction of the individuals in the considered dataset were advised by more than one advisor. The basic assumption that we make in the definitions below is that the credit for advising such a student is split equally between each of the co-advisors (i.e., if there are n listed co-advisors for a student, then each of the co-advisors receives $1/n$ credit for graduating the student).

Adjusted a_∞, a_1, a_2 indices

The definitions of a_∞, a_1, a_2 indices can be modified to take into account co-advising as follows.

Definition 5 (*adjusted a_∞-index*) The adjusted a_∞-index of an individual i is the total number of their academic descendants weighted by the reciprocals of their in-degrees, that is, $a_{\infty,adj}(i) = \sum_{j \in N} \frac{1}{deg^{in}(j)} \mathbb{1}_{\{d(i,j) < +\infty\}}$, where $\mathbb{1}_{\{d(i,j) < +\infty\}}$ is the indicator function corresponding to the condition that node j is reachable from node i through a directed path.

Definition 6 (*adjusted a_1-index*) The adjusted a_1-index of an individual i is defined as $a_{1,adj}(i) = \sum_{j \in N} \frac{1}{deg^{in}(j)} \frac{1}{d(i,j)}$.

Definition 7 (*adjusted a_2-index*) The adjusted a_2-index of an individual i is defined as $a_{2,adj}(i) = \sum_{j \in N} \frac{1}{deg^{in}(j)} \frac{1}{2^{d(i,j)}}$.

As one can clearly see from these definitions, the values of these adjusted indices are always less than or equal to the respective values of their "regular" counterparts, as common sense would suggest.
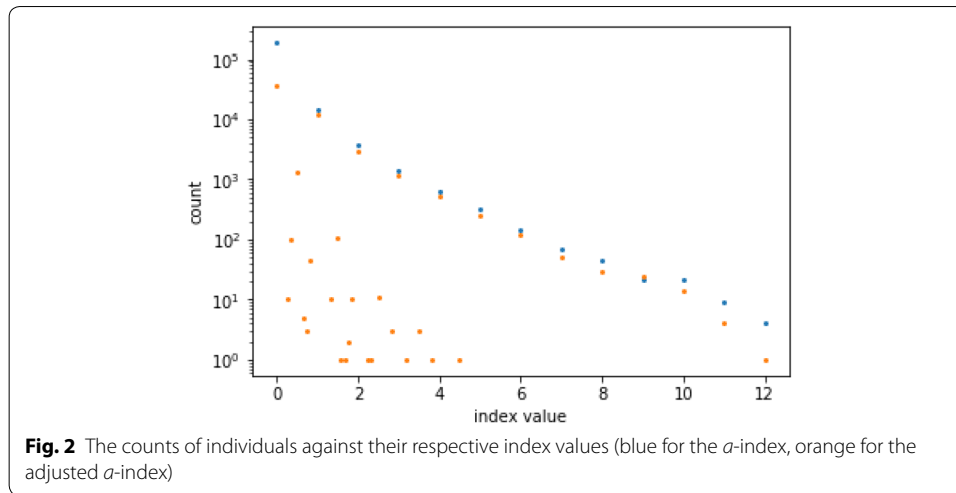
Adjusted *a*-index

The above definition of *a*-index can also be modified to take into account co-advising, although this extension is not as straightforward as those in the previous subsection. The "adjusted *a*-index" of node i can be calculated as follows:

1. Calculate the "adjusted" out-degree of node i: $deg^{out}_{adj}(i) = \sum_{j:(i,j) \in \mathcal{A}} \frac{1}{deg^{in}(j)}$. Clearly, this value can be fractional and is reduced to simply the out-degree of node i if none of the students of the corresponding individual i were co-advised.
2. Compute and sort the adjusted out-degrees (defined as indicated above) of all nodes $\{j : (i,j) \in \mathcal{A}\}$ in the non-increasing order. Denote this sorted array as $D_1, D_2, \ldots$ and let D_k be the kth element of this array such that k is the largest integer satisfying $\lceil D_k \rceil \geq k$. Calculate $\min\{D_k, k\}$.
3. Calculate the *adjusted a-index* of node i, $a_{adj}(i)$, as the minimum over the values obtained in the steps 1 and 2 above.

This computational procedure ensures that the adjusted *a*-index of any node i is always less than or equal to its "regular" *a*-index, whereas the possibility of fractional values of the adjusted *a*-index provides a more diverse set of its possible values. This would potentially allow one to create a more "diversified" ranking of academic advisors based on their own productivity and productivity of their students, while taking into account co-advising.

Results for MathGenealogy dataset

In this section, we present the results obtained on the MathGenealogy network using the metrics proposed above. Figure 2 shows the distribution of the values of the *a*-index and the adjusted *a*-index over the entire network. One can observe that while the "regular" *a*-index is always integer by definition, the adjusted *a*-index does often take fractional values, especially for lower spectrum values of the index, thus providing a more diverse set of possible values in a ranking. Further, Table 1 provides

Fig. 2 The counts of individuals against their respective index values (blue for the *a*-index, orange for the adjusted *a*-index)

a ranking of top academic advisors with an *a*-index of at least 10, many of whom are prominent mathematicians from the nineteenth and twentieth centuries (note that none of the mathematicians who worked before the nineteenth century made it into this ranking). Their respective adjusted *a*-index values are also given in the same table for comparison. One can observe that this ranking would change if it was done using the adjusted *a*-index, thus showing that co-advising is indeed a significant factor to consider in this context.
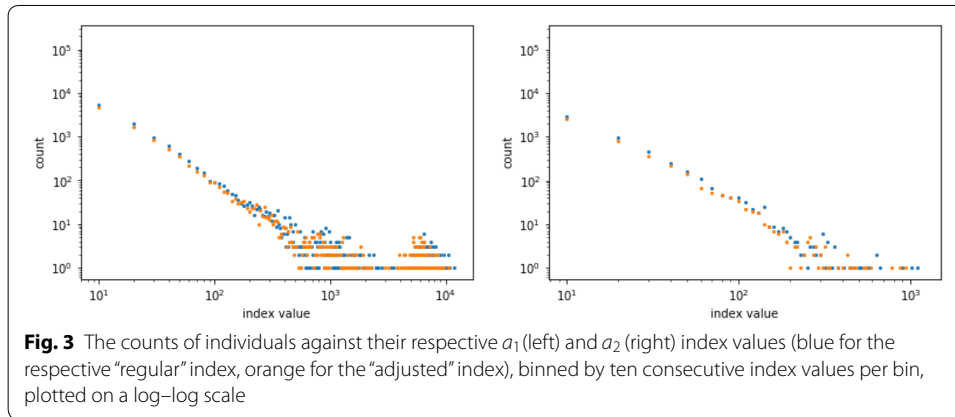
Table 2 presents the collective advising impact rankings of universities and countries based on their respective values of *a*-index. It can be observed that universities and countries with prominent reputation in mathematics-related research fields lead these rankings, which shows that (i) not surprisingly, there is correlation between collective university-scale and country-scale research and advising impacts, and (ii) the *a*-index appears to be a realistic and appropriate metric for collective advising impact of a university or a country. Note that we do not consider adjusted *a*-index in this case (although it would be possible), since it is rare in the dataset that an individual's co-advisors come from different universities or countries.

Figure 3 shows the distribution of regular and adjusted a_1 and a_2 indices in the network. It appears that both of these distributions are close to power-law, whereas the range of values of the a_1-index is larger than that of the a_2-index, which follows from the respective definitions. Tables 3 and 4 present the rankings of the top 25 advisors by regular versus adjusted a_1 and a_2 indices. For each index, mostly the same group of advisors appears in the regular versus adjusted index rankings, although their order slightly changes in both tables. Moreover, one can observe that the a_1-index-based ranking favors earlier generations of mathematicians (those from sixteenth, seventeenth, and eighteenth centuries), whereas the a_2-index-based ranking features mathematicians from the nineteenth and the twentieth centuries. This is a direct consequence of the impact of the different weights given by these indices to distant academic descendants of an individual.

The ranking of individuals with in-degree zero in the network (that is, "fathers" of genealogical families) by their a_∞ and adjusted a_∞-index values is given in Table 5. The

Table 2 Top universities and countries by *a*-index

University name	*a*-index	Country	*a*-index
Harvard University	31	United States	54
Princeton University	30	Germany	45
University of California, Berkeley	29	United Kingdom	33
Massachusetts Institute of Technology	28	Russia	31
Stanford University	28	Netherlands	29
The University of Chicago	25	France	26
Lomonosov Moscow State University	25	Switzerland	25
University of Cambridge	24	Austria	22
Columbia University	24	Canada	21
ETH Zürich	24	Belgium	19
Georg-August-Universität Göttingen	22	India	19
University of Wisconsin-Madison	22	Sweden	18
California Institute of Technology	22	Ukraine	17
University of Michigan	21	Australia	17
University of Oxford	21	Romania	17
Universiteit van Amsterdam	21	Poland	17
Yale University	20	Spain	17
University of Illinois at Urbana-Champaign	20	Israel	17
Universität Berlin	20	Japan	16
Ludwig-Maximilians-Universität München	20	Italy	15
Carnegie Mellon University	20	Finland	15

Fig. 3 The counts of individuals against their respective a_1 (left) and a_2 (right) index values (blue for the respective "regular" index, orange for the "adjusted" index), binned by ten consecutive index values per bin, plotted on a log–log scale

top-ranked scientist with respect to both of these indices is Sharaf al-Din al-Tusi, who lived in the twelfth century and currently has 149,942 academic descendants.

Evolution of individual and collective *a*-indices in MathGenealogy dataset

As a natural further step in the analysis of individual and collective advising impacts using *a*-indices, we consider the dynamics of year-by-year evolution of the aforementioned indices over the past several decades. Specifically, we consider the time period starting from 1900 till 2017 (which was the last full year for which MathGenealogy data was collected in this study). The main reasons for considering only the data starting from 1900 are that (i) the growth of mathematics as a major research field

Table 3　Top 25 individuals ranked by the a_T-index (left) and adjusted a_T-index (right)

Name	Year	a_T-index	Name	Year	Adj. a_T-index
Simeon Denis Poisson	1800	11800.58	Simeon Denis Poisson	1800	10486.00
Abraham Gotthelf Kästner	1739	10719.19	Abraham Gotthelf Kästner	1739	9509.77
Joseph Louis Lagrange		10557.30	Joseph Louis Lagrange		9380.81
Pierre-Simon Laplace		10555.30	Pierre-Simon Laplace		9379.31
Jakob Thomasius	1643	10254.40	Jakob Thomasius	1643	9175.63
Leonhard Euler	1726	9969.036	Emmanuel Stupanus	1613	8852.32
Emmanuel Stupanus	1613	9907.44	Leonhard Euler	1726	8836.15
Christian August Hausen	1713	9712.28	Christian August Hausen	1713	8621.04
Johann Friedrich Pfaff	1786	9601.81	Friedrich Leibniz	1622	8565.26
Friedrich Leibniz	1622	9569.92	Giovanni Beccaria		8491.40
Giovanni Beccaria		9556.55	Jean Le Rond d'Alembert		8491.15
Jean Le Rond d'Alembert		9555.55	Johann Friedrich Pfaff	1786	8479.49
Carl Friedrich Gauss	1799	9395.80	Carl Friedrich Gauss	1799	8264.47
C. Felix (Christian) Klein	1868	9316.05	Petrus Ryff	1584	8262.28
Petrus Ryff	1584	9245.31	C. Felix (Christian) Klein	1868	8111.17
Johann Bernoulli	1690	9126.35	Johann Bernoulli	1690	8090.71
Johann Andreas Planer	1686	8885.37	Johann Andreas Planer	1686	7889.29
J. C. Wichmannshausen	1685	8882.86	J. C. Wichmannshausen	1685	7887.46
Johann Elert Bode		8707.06	Felix Plater	1557	7748.45
Felix Plater	1557	8669.28	Johann Elert Bode		7693.68
Jacob Bernoulli	1676	8400.77	Jacob Bernoulli	1676	7447.01
Nikolaus Eglinger	1660	8398.77	Nikolaus Eglinger	1660	7445.01
Julius Plücker	1823	8210.41	Johannes W. von Andernach	1527	7322.90
Johann Pasch	1683	8189.74	Guillaume Rondelet		7296.20
Rudolf Jakob Camerarius	1684	8189.74	Otto Mencke	1665	7274.00

occurred during the twentieth century with many new Ph.D. degrees awarded during that time frame, and (ii) the collected data itself is more reliable and complete for this most recent time frame, which makes the results on a-indices evolution corresponding to this time interval more interesting. We should also note that some of the plots presented below reflect the data starting from 1950, which is done for visual clarity purposes.

Figure 4 shows the year-by-year evolution of a-index values of top 10 mathematicians (according to their a-index value at present, as indicated in Table 1) starting from 1900 and until 2017. An interesting observation is that for most of these individuals, it took around 20–30 years to grow their a-index from 0 to 1 (that is the time period from the year an individual received his/her own Ph.D. degree to the year when his/her first student successfully graduates a student of their own). Further, it took another ~30 years to grow their respective a-index value from 1 to around 10. The overall time period of 50–60 years to grow the a-index from zero to a high value of 10 or more is on the same order of magnitude as the length of a lifetime academic career (i.e., from the receipt of a Ph.D. degree till retirement). This shows that most of these "high-impact" advisors followed a similar temporal pattern of their careers. This observation is also consistent with the intent for this index to reflect an individual's career-long rather than short-term advising impact.

**Table 4 Top 25 individuals ranked by the a_2-index (left) and adjusted a_2-index (right)**

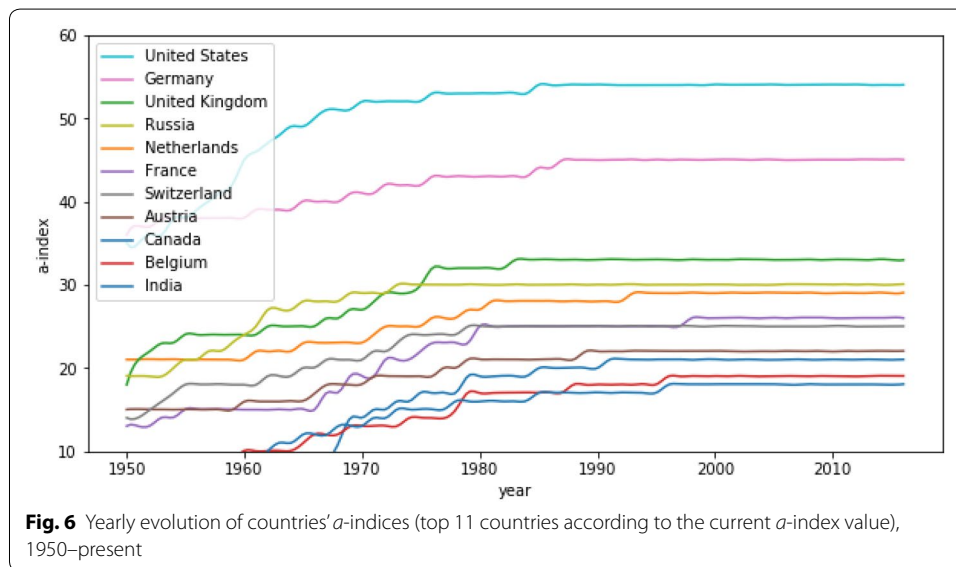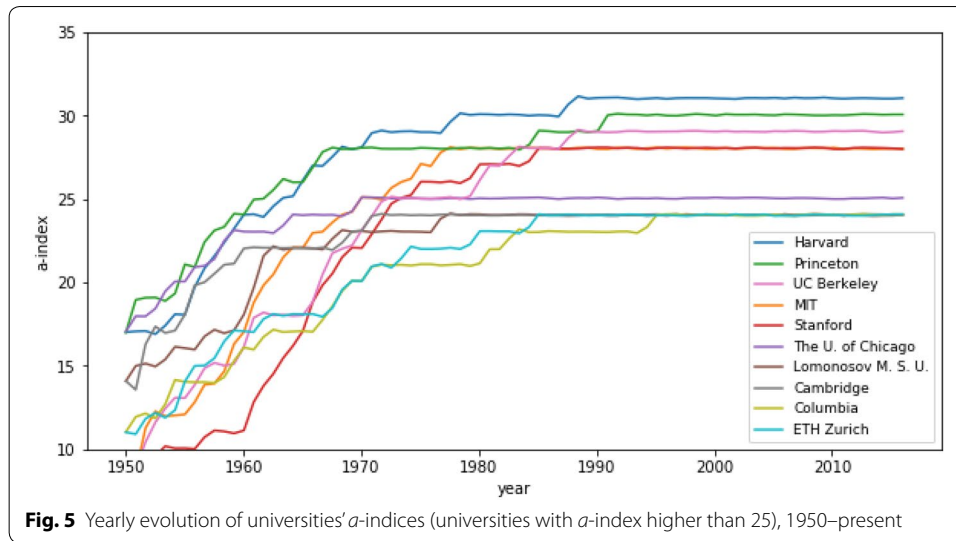| Name	Year	a_2-index	Name	Year	Adj. a_2-index
David Hilbert	1885	1099.72	David Hilbert	1885	949.74
C. Felix Klein	1868	1016.04	C. Felix Klein	1868	873.30
C. L. Ferdinand Lindemann	1873	907.25	C. L. Ferdinand Lindemann	1873	780.80
Erhard Schmidt	1905	667.56	E. H. Moore	1885	597.00
E. H. Moore	1885	639.77	Erhard Schmidt	1905	550.32
Ernst Eduard Kummer	1831	636.78	Ernst Eduard Kummer	1831	535.59
K.T.W. Weierstrass	1841	575.10	K.T.W. Weierstrass	1841	484.32
Julius Plucker	1823	522.36	Solomon Lefschetz	1911	466.55
Solomon Lefschetz	1911	510.06	Julius Plucker	1823	449.19
R. O. S. Lipschitz	1853	508.52	R. O. S. Lipschitz	1853	436.90
Oswald Veblen	1903	474.34	Oswald Veblen	1903	431.94
Richard Courant	1910	458.12	Richard Courant	1910	400.25
Heinz Hopf	1925	446.18	George David Birkhoff	1907	388.12
George David Birkhoff	1907	415.53	Heinz Hopf	1925	349.33
Jacques-Louis Lions	1954	385.92	Nikolai Nikolayevich Luzin	1915	335.95
Nikolai Nikolayevich Luzin	1915	366.49	Jacques-Louis Lions	1954	329.44
Simeon Denis Poisson	1800	362.25	A. N. Kolmogorov	1925	326.89
A. N. Kolmogorov	1925	361.89	Simeon Denis Poisson	1800	320.97
Ferdinand Georg Frobenius	1870	354.50	Gaston Darboux	1866	311.96
Gaston Darboux	1866	346.64	Michel Chasles	1814	309.62
Michel Chasles	1814	337.36	H. A. Newton	1850	301.44
G. P. L. Dirichlet	1827	335.53	Ferdinand Georg Frobenius	1870	294.73
Ludwig Bieberbach	1910	334.29	C. Emile Picard	1877	287.67
Edmund Landau	1899	330.46	G. P. L. Dirichlet	1827	283.56
H. A. Newton	1850	323.17	Edmund Landau	1899	280.94

It should also be noted that the "outliers" in this plot are E.E. Kummer and K.T.W. Weierstrass, who received their own Ph.D. degrees substantially earlier than the other individuals in this list, and their a-indices were already equal to 6 in 1900. Interestingly, both of their a-index values have "saturated" at 11 around 1920 and have not changed since then. This is most likely due to the fact that all of their direct descendants (students) finished their own academic careers by that time; therefore, they did not produce any more students after that, which means that the respective a-index cannot increase anymore. Thus, the a-index is a good measure of an individual's lifetime advising impact; however, it does not reflect any further advising impact that an individual might achieve after the end of his/her own and his/her students' academic careers. On the other hand, an individual's a_1, a_2, and a_∞ indices clearly can grow indefinitely, even decades or centuries after the end of one's career (as it will be illustrated below). Therefore, a long-term advising impact may need to be evaluated by considering a combination of metrics (such as the indices defined in this paper) rather than by taking into account only one metric.

It is also worth mentioning that a *collective a-index* of a university or a country does not exhibit the "saturation" behavior that was mentioned above for an *individual a-index*. Indeed, a university or a country would typically keep producing Ph.D. graduates indefinitely (unless a university/country ceases to exist). Figures 5 and 6

Table 5 Top 25 individuals with zero in-degrees ranked by the a_∞-index (left) and adjusted a_∞-index (right)
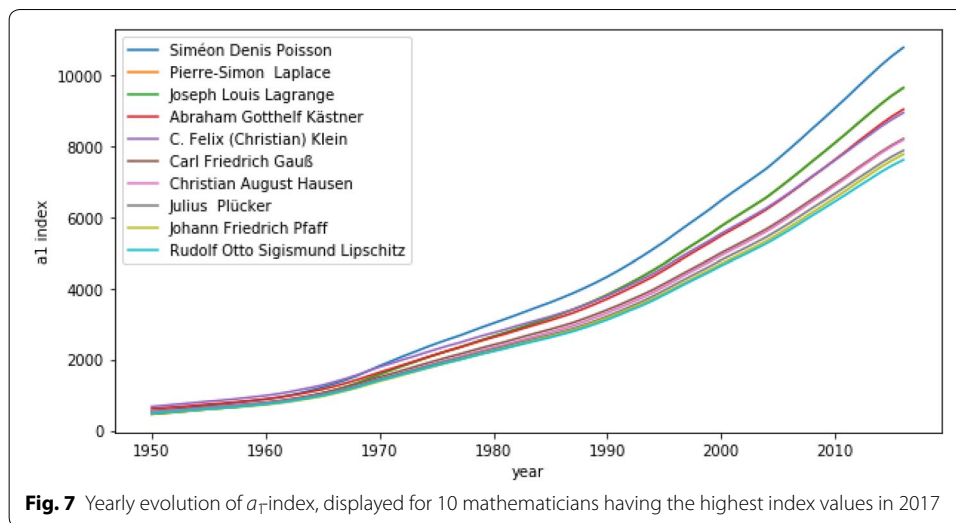
Name	Year	a_∞-index	Name	Year	Adj. a_∞-index
Sharaf al-Din al-Tusi		149942	Sharaf al-Din al-Tusi		134493.04
Elissaeus Judaeus		149909	Elissaeus Judaeus		134464.20
Jan Standonck	1474	149852	Jan Standonck	1474	134420.70
Cristoforo Landino		149821	Cristoforo Landino		134398.29
G. G. M. Groote		149820	Moses Perez		134395.79
F. F. R. Radewyns		149820	G. G. M. Groote		134393.87
Moses Perez		149818	F. F. R. Radewyns		134393.87
Ulrich Zasius	1501	149817	Ulrich Zasius	1501	134392.37
G. Hermonymus		149807	G. Hermonymus		134385.84
Jean Tagault		149750	Francois Dubois	1516	134341.45
J. ben Jehiel Loans		149750	Jean Tagault		134341.45
Francois Dubois	1516	149750	J. ben Jehiel Loans		134340.62
Johannes Stoffler	1476	149719	Johannes Stoffler	1476	134316.79
Nicole Oresme		147113	Nicole Oresme		131926.62
Luca Pacioli		147109	Luca Pacioli		131923.12
Bonifazius Erasmi	1509	147108	Bonifazius Erasmi	1509	131923.12
L. von Dobschutz	1489	147108	L. von Dobschutz	1489	131922.62
Johann Hoffmann		147073	Johann Hoffmann		131897.42
Friedrich Leibniz	1622	146244	Friedrich Leibniz	1622	131125.67
Thomas Cranmer	1515	142171	Thomas Cranmer	1515	127225.92
Ludolph van Ceulen		137372	Ludolph van Ceulen		122765.09
Marin Mersenne	1611	127076	Marin Mersenne	1611	113204.87
Paolo da Venezia		125964	Paolo da Venezia		112062.20
Sigismondo Polcastro		125961	Sigismondo Polcastro		112059.20
Matthaeus Adrianus		125959	Matthaeus Adrianus		112058.70
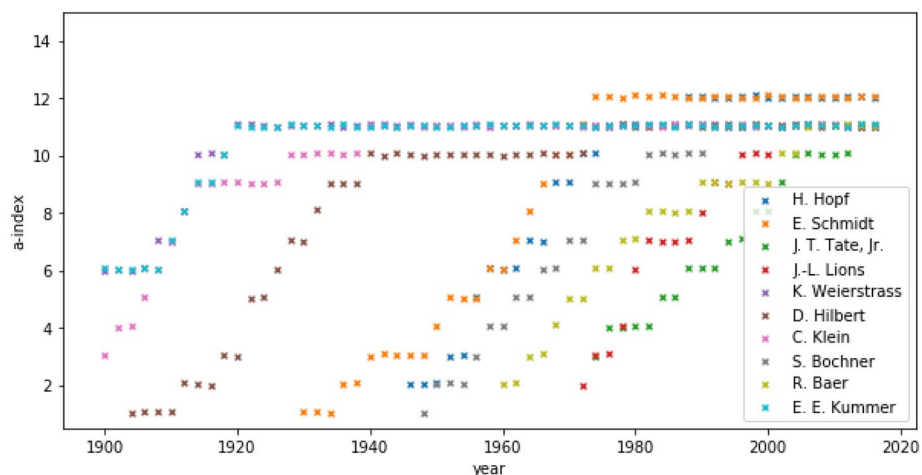
Fig. 4 Yearly evolution of individual mathematicians' *a*-indices: 10 out of top 13 individuals according to the current *a*-index value, as indicated in Table 1 (the plots for M.A. Krasnoselskii, A.N. Kolmogorov, and H.A. Behnke are not shown due to missing data on graduation years of some of their descendants), 1900–present

Fig. 5 Yearly evolution of universities' *a*-indices (universities with *a*-index higher than 25), 1950–present

Fig. 6 Yearly evolution of countries' *a*-indices (top 11 countries according to the current *a*-index value), 1950–present

illustrate the evolution of collective *a*-index values corresponding to top universities and countries (according to their current *a*-index values as shown in Table 2). For visual clarity, these plots are shown starting from 1950 rather than 1900.

For universities' collective *a*-index values, there were several lead changes during 1900–1950 (not pictured), with Princeton being top-ranked for most of the 1950s and 1960s (briefly overtaken by the University of Chicago in mid-1950s), whereas in 1968 Harvard took the top-ranked position, which it has held till now. It should be also noted that Stanford has made a big jump from number 10 to number 4 in the *a*-index ranking during the past half-century. As for countries' *a*-indices evolution, the United States passed Germany as number one in the collective *a*-index ranking in 1956 and has held this top position since then.
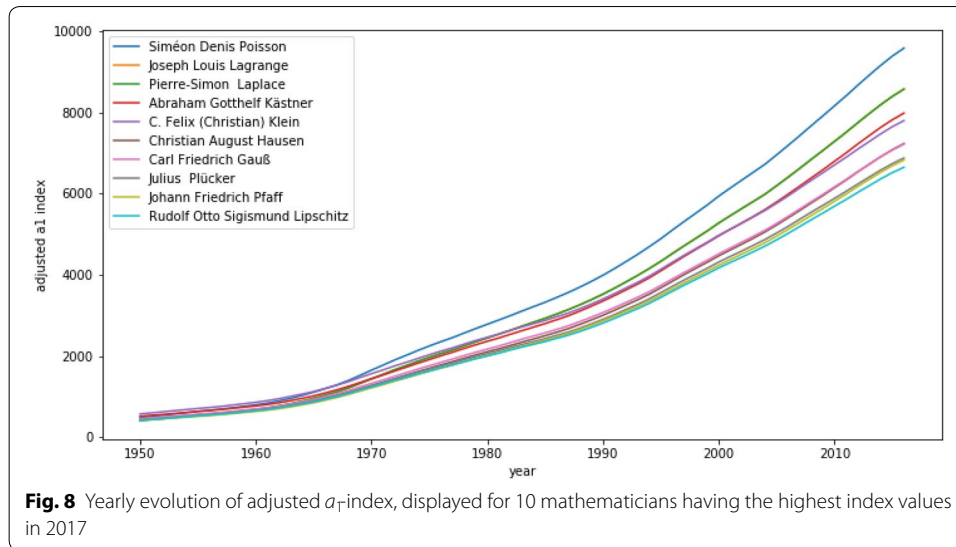
Fig. 7 Yearly evolution of a_1-index, displayed for 10 mathematicians having the highest index values in 2017

Note that the collective *a*-indices of most universities and countries that are depicted in Figs. 5 and 6 have not changed since 1990s. This may be explained by the fact that it becomes harder and harder to increase the *a*-index when it has already reached high values (similarly to what happens to the *h*-index in research citations). Another factor may be that not all Ph.D. graduates from the past 10–20 years have been added to the MathGenealogy dataset yet. However, as mentioned above, this "temporary" saturation behavior of collective *a*-indices is not the same as the one we observed for individual *a*-indices, since the production of Ph.D. graduates by a city or a country is not limited by the lengths of academic careers of individual advisors.

Further, we consider the evolution of a_1 and a_2 indices (along with their adjusted versions) of the top advisors listed in Tables 3 and 4. The respective plots are shown in Figs. 7, 8, 9, and 10. Note that the evolution of a_∞-index is not depicted here, since the plots corresponding to all top advisors according to this index exhibit a highly similar pattern and thus would look indistinguishable in a figure.

Interestingly, from Figs. 7 and 8 one can observe that both the a_1 and the adjusted a_1 index values for all of the top advisors were very close to each other up to around 1970, which is a very recent date compared to the dates of their respective careers. However, in the past 3–4 decades, these indices have increased substantially and started to spread over a broader range of values, approximately between 6000 and 10,000. The ranking of advisors according to both the a_1 and the adjusted a_1 index has been stable over the past decades, with S.D. Poisson holding the top spot.
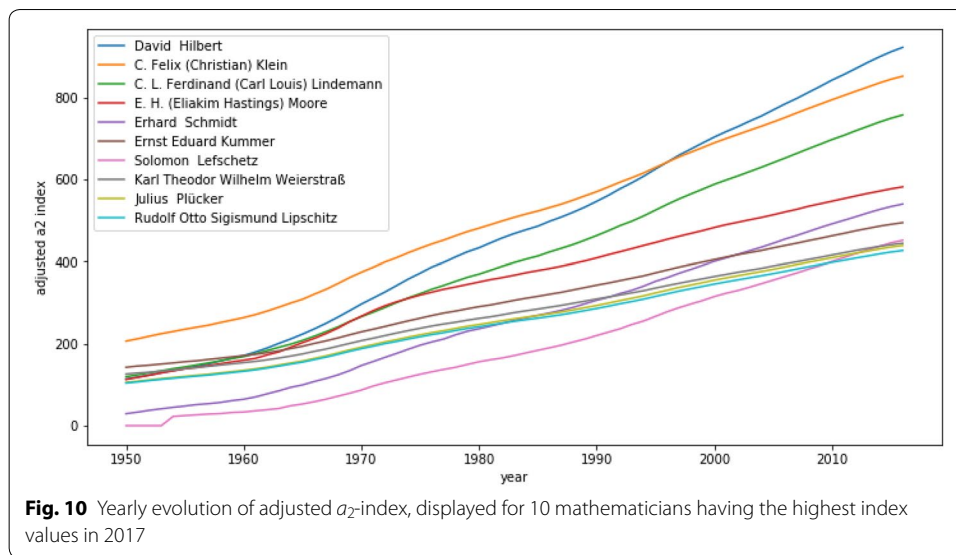
As for the evolution of a_2 and adjusted a_2 indices, there has been much more diversity and changes in the ranking of top advisors over the past decades (compared to a_1-index). As noted above, the a_2-index gives lower weights to distant descendants of an individual, which results in a lower order of magnitude of this index compared to the a_1-index. Nevertheless, despite a narrower range of values for this index, there have been several changes in the ranks of top advisors according to this index in the twentieth century (although many of these mathematicians worked in the nineteenth century). Notably, D. Hilbert has assumed the top spot in the a_2-index-based ranking only in the 1990s despite

Fig. 8 Yearly evolution of adjusted a_1-index, displayed for 10 mathematicians having the highest index values in 2017

Fig. 9 Yearly evolution of a_2-index, displayed for 10 mathematicians having the highest index values in 2017
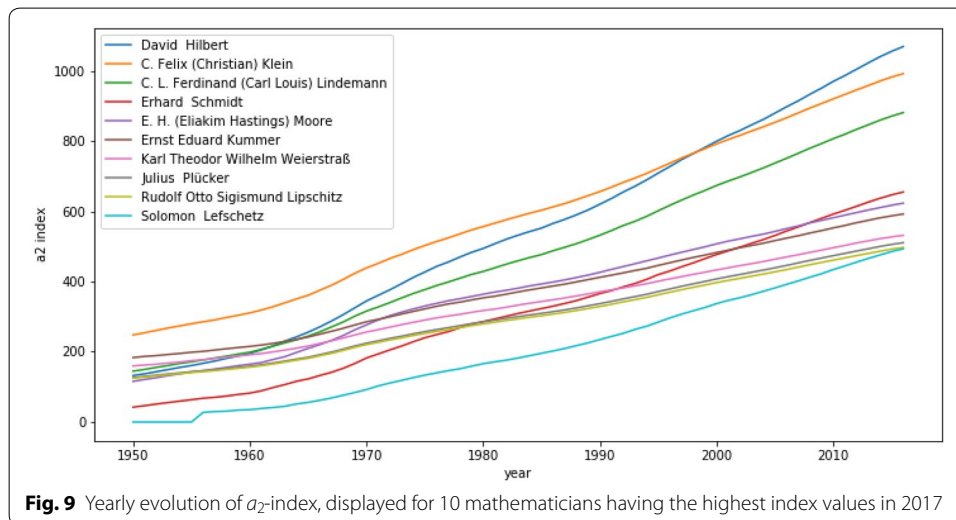
the fact that he received his own Ph.D. degree more than 100 years prior. Thus, the a_2-index can be viewed as a meaningful metric of an individual's advising impact that lasts well beyond the end of one's career and still can increase considerably in subsequent decades or even centuries.

As a concluding remark of this section, we should note that all of the aforementioned results should be viewed in the light of the fact that MathGenealogy dataset is not necessarily complete for the considered time interval, and some Ph.D. graduates (as well as some Ph.D. graduation year information, as mentioned above) may not have been added to the database yet. This may lead to discrepancies in the results presented here with those that may be obtained in future studies when more entries are added to the database. Nevertheless, the considered dataset is still rather large and comprehensive, and the presented results reveal interesting temporal patterns of the proposed individual and collective advising indices.

Fig. 10 Yearly evolution of adjusted a_2-index, displayed for 10 mathematicians having the highest index values in 2017

Concluding remarks

We proposed a family of network-based advising impact metrics (a-indices) that are easy to calculate and interpret, as well as provided a flexible framework for quantifying advising impacts of individuals from different "eras" and stages of their academic careers, as well as collective advising impacts of countries and universities. Although we illustrated our approaches on MathGenealogy dataset only, these approaches are certainly applicable to other scientific domains where comprehensive advisor–student datasets may become available.

Due to the fact that we focus on the advising impact beyond the number of immediate students of an individual, this approach is not intended for measuring advising impacts of early-career scientists (simply calculating an out-degree for "young" advisors would still be a viable option). However, one may argue that a true impact of an academic advisor is evident towards later stages of career when one's students achieve their own advising success. Therefore, we believe that these indices can be used in practical settings, for instance, by universities in order to quantify and promote individual and collective advising successes of their faculty members. This study shows the applicability of network-based techniques for these purposes. As one of the possible directions of future research, it could be of interest to look at "groups of influential advisors", for instance, using optimization-based techniques that identify "central" groups of nodes in a network [12].

It should also be noted that this study is not intended to build direct comparisons or preferences between different metrics of advising impact, including those proposed here or those proposed in other related studies. Instead, we believe that long-term individual or collective advising impact should be considered in the context of an "ensemble" of various quantitative metrics, including the proposed a-indices. Similarly to debates regarding citation indices (e.g., whether the h-index or some other quantitative metrics of citations are the most appropriate to measure citation impact), there is no definitive answer to the question about the "best" metric for

advising impact. We hope that this study will stimulate further research in this interesting research direction.

Author details
[1] Faculty of Information Technology, University of Jyväskylä, Jyväskylä, Finland. [2] Department of Industrial Engineering and Management Systems, University of Central Florida, 12800 Pegasus Drive, Orlando, FL, USA. [3] Department of Industrial and Systems Engineering, University at Buffalo, 312 Bell Hall, Buffalo, NY, USA. [4] Air Force Research Laboratory, Eglin AFB, FL, USA.

References
1. Arslan E, Gunes MH, Yuksel M. Analysis of academic ties: a case study of mathematics genealogy. In: GLOBECOM Workshops (GC Wkshps), 2011. IEEE, 2011. p. 125–9.
2. Malmgren RD, Ottino JM, Amaral LAN. The role of mentorship in protégé performance. Nature. 2010;465(7298):622.
3. Myers SA, Mucha PJ, Porter MA. Mathematical genealogy and department prestige. Chaos. 2011;21(4):041104.
4. Gargiulo F, Caen A, Lambiotte R, Carletti T. The classical origin of modern mathematics. EPJ Data Sci. 2016;5(1):26.
5. Taylor D, Myers SA, Clauset A, Porter MA, Mucha PJ. Eigenvector-based centrality measures for temporal networks. Multiscale Model Simul. 2017;15(1):537–74.
6. Rossi L, Freire IL, Mena-Chalco JP. Genealogical index: a metric to analyze advisor–advisee relationships. J Inform. 2017;11(2):564–82.
7. Boldi P, Vigna S. Axioms for centrality. Internet Math. 2014;10(3–4):222–62.
8. Marchiori M, Latora V. Harmony in the small-world. Physica A. 2000;285(3–4):539–46.
9. Jackson MO. Social and economic networks. Princeton: Princeton University Press; 2010.
10. Tsakas N. On decay centrality. BE J Theor Econ. 2016;19:1–18.
11. Broido AD, Clauset A. Scale-free networks are rare. Nat Commun. 2019;10(1):1017.
12. Veremyev A, Prokopyev OA, Pasiliao EL. Finding groups with maximum betweenness centrality. Optim Methods Softw. 2017;32(2):369–99.

Publisher's Note