Journal of Informetrics 11 (2017) 564–582

Contents lists available at ScienceDirect

Journal of Informetrics

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

Genealogical index: A metric to analyze advisor–advisee relationships

Luciano Rossi, Igor L. Freire, Jesús P. Mena-Chalco*

Center for Mathematics, Computation and Cognition, Federal University of ABC, Santo André, Brazil

ARTICLE INFO

Article history:
Received 14 October 2016
Received in revised form 1 April 2017
Accepted 2 April 2017
Available online 22 April 2017

Keywords:
Genealogical index
Academic genealogy
Advisor–advisee relationship
Genealogical graph

ABSTRACT

Academic genealogy can be defined as the study of intellectual heritage that is undertaken through the relationship between a professor (advisor/mentor) and student (advisee) and on the basis of these ties, it establishes a social framework that is generally represented by an academic genealogy graph. Obtaining relevant knowledge of academic genealogy graphs makes it possible to analyse the academic training of scientific communities, and discover ancestors or forbears who had special skills and talents. The use of metrics for characterizing this kind of graph is an active form of knowledge extraction. In this paper, we set out a formal definition of a metric called 'genealogical index', which can be used to assess how far researchers have affected advisor–advisee relationships. This metric is based on the bibliometrics h-index and its definition can be broadened to measure the effect of researchers on several generations of scientists. A case study is employed that includes an academic genealogy graph consisting of more than 190,000 Ph.D.s registered in the Mathematics Genealogy Project. Additionally, we compare the genealogical indices obtained from both the Fields Medal and Wolf Prize winners, and found that the latter has had a greater impact than the former.

1. Introduction

Scientific evolution traces economic and social development that is carried out through scientific research, technological innovation and patenting. The University is the natural environment for scientific evolution since it involves academic mentoring that seeks to develop high quality human resources. The existence of this phenomenon is the driving-force behind the search for variables/metrics that can measure it. In this paper, we introduce a metric that is designed to assess the effects of academic mentoring on the achievements of the scientific community by means of academic genealogy graphs. The reason for applying this metric is to answer a fundamental research question: does academic supervision affect the performance of scientists?

Academic genealogy (AG) can be defined as a quantitative study of intellectual inheritance that has been perpetuated by generations of researchers through academic advising by mentors to their students (Sugimoto, 2014). The AG allows an analysis to be conducted about the dissemination of scientific knowledge and the progress made by scientific communities.

* Corresponding author.
E-mail addresses: luciano.rossi@ufabc.edu.br (L. Rossi), igor.freire@ufabc.edu.br (I.L. Freire), jesus.mena@ufabc.edu.br (J.P. Mena-Chalco).

http://dx.doi.org/10.1016/j.joi.2017.04.001
1751-1577/

The varied work of academic advisors influences the training programs of researchers and encourages future generations of researchers to continue their activities (Malmgren, Ottino, & Amaral, 2010). The AG provides a means of assessing and quantitatively analyzing the way these training schemes are conducted (Sugimoto, Ni, Russell, & Bychowski, 2011).

The analysis of academic and scientific communities has attracted a good deal of attention among researchers. Particular importance has been attached to the classification and identification of researchers involved in several knowledge areas. These include the adoption and improvement of quantitative metrics that support this analysis, as well as the study of the acquisition of scientific knowledge. The studies in this area mostly entail carrying out an in-depth analysis of academic publications, but there are a few studies devoted to evaluating research from the perspective of AG. In most cases, these studies are used to identify the forebears and descendants of an individual researcher (i.e., to compile an egocentric genealogy) simply to honor them (Kobayashi, 2015; van der Kruit, 2015).

A number of studies have been carried out to characterize the AG with the aim of analyzing specific knowledge areas, such as Neuroscience (David & Hayden, 2012), Chemistry (Andraos, 2005), Mathematics (Chang, 2011; Gargiulo, Caen, Lambiotte, & Carletti, 2016; Malmgren et al., 2010), Physiology (Bennett & Lowe, 2005; Jackson, 2011), Meteorology (Hart & Cossuth, 2013), Primatology (Kelley & Sussman, 2007), Bibliometry and Information Science (Russell & Sugimoto, 2009), and Protozoology (Elias, Floeter-Winter, & Mena-Chalco, 2016), and many others.

It should be noted that these studies converge insofar as they share the following common objectives: (i) prospecting, structuring and storing data about academic genealogy (i.e., historical records) (Andraos, 2005; Bennett & Lowe, 2005; Chang, 2011; Hart & Cossuth, 2013), (ii) characterizing knowledge areas and/or disciplines (egocentric and honorific genealogy), by analyzing genealogical frameworks using basic descriptive statistics (David & Hayden, 2012; Elias et al., 2016; Gargiulo et al., 2016; Malmgren et al., 2010; Russell & Sugimoto, 2009), and (iii) making the information available (i.e., publishing it) to members of the community and interested parties (David & Hayden, 2012; Hart & Cossuth, 2013). However, the majority of these studies does not focus their analysis in topological structures neither use metrics to find out key academic groups or individuals. Among the few studies that use metrics to characterize the topology of a network, it is worth highlighting the work of David and Hayden (2012) who employed fecundity metrics to characterize the neuroscientific community, Rossi and Mena-Chalco (2015) whose work examined the basis of the genealogical index and its main applications, and Lü, Zhou, Zhang, and Stanley (2016) where the h-index was used for the characterization of scientific networks.

There have also been initiatives that rely on web applications to document and share the academic genealogy of researchers across several fields, such as the following: The Mathematics Genealogy Project,[1] The Neurotree Project,[2] The Academic Family Tree,[3] and the Academic Genealogy Wiki.[4] These projects, which register and document the names of individuals, make it possible to study the influence of generations of researchers on the academic world, through academic mentoring.

According to Sugimoto (2014), the academic genealogy is mainly used by researchers interested in discovering and describing their own origins. These studies have tended to be neglected by those who are studying a branch of science from a historical, philosophical, sociological or scientific perspective. The real importance of academic genealogy is that it provides quantitative and qualitative inputs that can assist in measuring interactions at different levels. The academic genealogy allows science to be analyzed from the standpoint of a transmission of scientific knowledge through generations of researchers.

The academic genealogy can be analyzed by topological metrics that represent different features and provide useful information on the training of academic communities as well as by revealing the names of key researchers who have made a significant contribution to areas of knowledge. In this paper, a topological metric is used, called genealogical index, which can be applied to measure the academic influence of researchers by means of academic genealogy graphs. We use the term 'rank' to define the limited coverage of the generations that must be included for the calculation of the genealogical index.

This approach makes a formal adjustment to the bibliometric h-index in the academic genealogy project. This is a system to measure the influence of a researcher quantitatively from the perspective of human resources training, rather than simply concentrating on publications, citations, coauthorships, or research projects.

Finally, it should be emphasized that this work is aligned with the epistemology of big data analysis (Big Data), in the form of data-driven science and seeks to discover knowledge about universally-accepted scientific theories, as described by Frické (2014).

2. Method

2.1. Academic genealogy graphs

The advisor–advisee relationships can be represented in the form of a graph that can be useful for the study of academic genealogy. The structure used is called a genealogical tree. In fact, the structures built from academic genealogy data cannot

1 http://genealogy.math.ndsu.nodak.edu, last accessed on January 15, 2017.
2 http://neurotree.org/neurotree, last accessed on January 18, 2017.
3 http://academictree.org/physics, last accessed on January 20, 2017.
4 http://phdtree.org, last accessed on February 02, 2017.

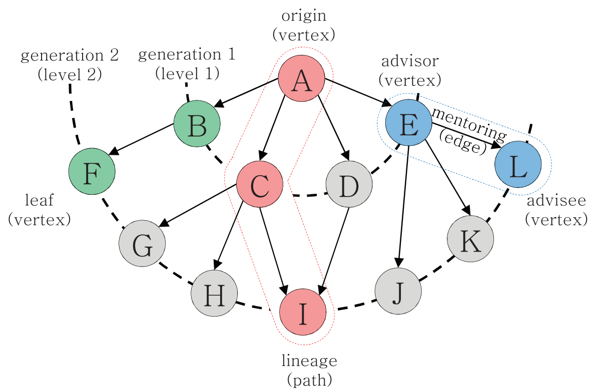


Fig. 1. Example of a genealogical graph with the definition of its main features.

be categorized as 'trees', since they may have more than one route between two vertices in the graph. Thus, in our work, we used the term 'academic genealogy graph' (a directed acyclic graph).

A directed graph $\vec{G}$ is a pair of (V, E), where V is a finite set of vertices and E, edges, is a binary relation in V. Vertices (V) represent individuals (researchers) and the directed edges (E) represent the mentoring relationships.

The academic lineage is a path in the genealogy graph that connects with forbears or descendants. Formally, a path with length k from a vertex source (u) to a vertex destination (u') in a directed graph ($\vec{G}$), is a sequence of vertices $(v_0, v_1, v_2, \ldots, v_k)$ such that $u = v_0$, $u' = v_k$ and $(v_{i-1}, v_i) \in E$ for $i = 1, 2, 3, \ldots, k$. Fig. 1 shows an example of an academic genealogy graph.

2.2. Genealogical index

In the area of Bibliometrics/Scientometrics, the h-index is a measure of impact, as proposed by Hirsch (2005). It can be defined as follows: a researcher has an h-index if the researcher has h papers with at least h citations each. Although there are different questions about the efficiency of the h-index (Waltman & Eck, 2012; Yong, 2014), this measure is widely used to evaluate research because it combines the quantity (number of publications) and relative quality (number of citations) of the scientific production (indexed works).

Several studies were based on the Hirsch metric. For instance, Petersen and Succi (2013) make a generalization of the h-index metric to reduce the potential impact caused by the original measure which may have been overlooked. Additionally, Tol (2013) mentions some problems caused by the measure of the Hirsch index. These works suggest that academics with long careers have benefited from the h-index to the detriment of younger academics and refer to the kernel quantile regression for identifying exceptional researchers.

The bibliometric h-index is a widely used impact indicator and, because of this, has raised a number of questions. Perry et al. (2013) points to distortions in the ranking by the bibliometric h-index and provide a measure based on Euclidean distance. The arguments raised by the authors are valid within the context of bibliometrics, where the structure used represents different elements (author, paper, and citations). In this case, both papers and citations refer to the author and any addition can be attributed to the standard of the author's performance. In academic genealogy, the structure consists of a single feature that is academic. The inclusion of a new academic can only be credited to another academic if there is a direct link between them.

The adapting of the bibliometric h-index to the graphs to characterize the academic genealogy was initially carried out by Rossi and Mena-Chalco (2015), although its definition was not formalized. This *genealogical index* makes it possible to study academic advisors with regard to their performance in human resources management training by examining several researchers in different generations.

We shall now introduce some new concepts which have proved to be quite useful and important in our work. Thus as well as exploring them in a rigorous way, in view of their importance, we shall also illustrate and discuss them in pedagogical terms by referring to Figs. 1 and 2. We believe that a clear understanding of them will assist the reader to interpret our results and clarify the follow-up discussions.

In what follows, $\mathbb{N}$, $|A|$, $\vec{G}(V, E)$ and $v \in V$ denote the set of positive integers, the number of elements of the set A, a genealogy graph and a vertex of interest. We are now able to introduce key concepts to our paper.

- The direct descendants of the vertex v in $\vec{G}$ is defined by:

$$D(v) = \{u \in V : (v, u) \in E\}. \quad (1)$$

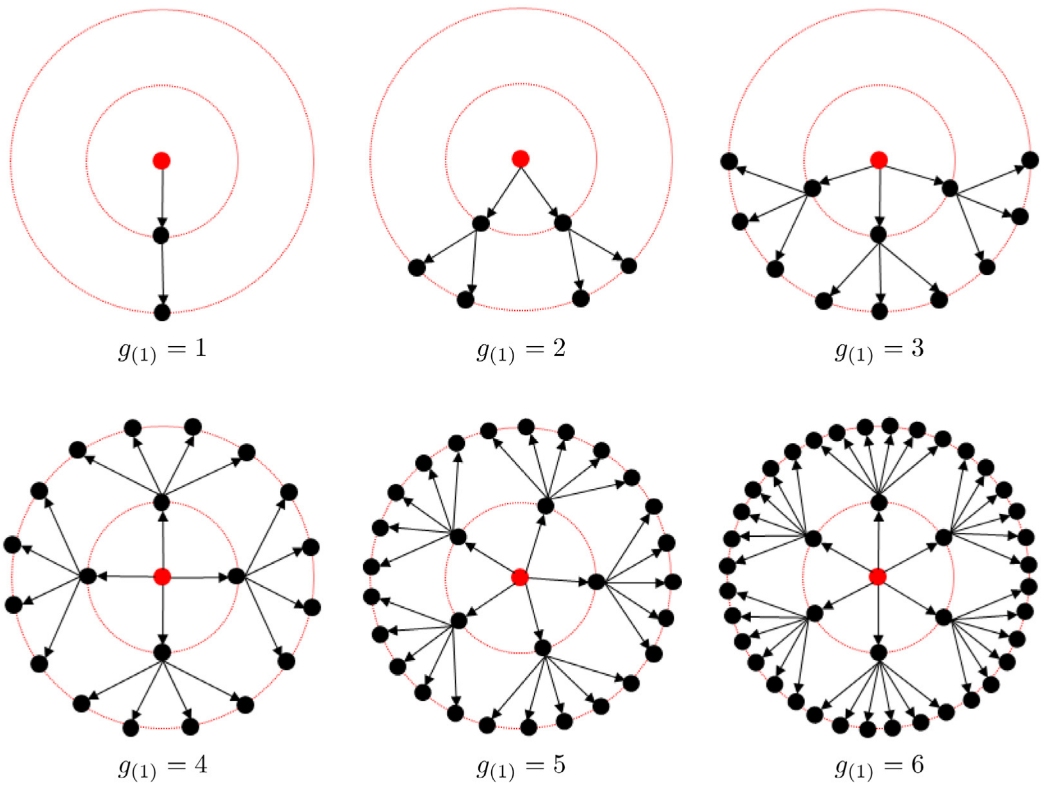


Fig. 2. Examples of academic genealogy graphs with genealogical index rank 1. The vertex of interest is shown in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

- The number of adjacent vertices to v is given by

$$l(v) = |D(v)|. \tag{2}$$

The quantity $l(v)$ measures the number of advisees directly guided by the advisor represented by v. We can interpret this as the *direct fertility* of a researcher, that is, those researchers which were advised by the person represented by v.

If we take Fig. 1 into account and consider $v = A$, then $D(v) = \{B, C, D, E\}$ and $l(v) = 4$. On the other hand, if we turn our attention to vertex C, we would then conclude $D(v) = \{G, H, I\}$ and $l(v) = 3$.

- The vertex v is called k-fertile if $l(v) = k$, but $l(v) < m$, for each $m > k$, that is, the vertex has k, and only k vertices in its adjacency.
- Let $m \in \mathbb{N}$. The *m-fertile descendants* of v are

$$D^{(m)}(v) = \{u \in D(v) : l(u) \geq m\}. \tag{3}$$

Given $m \in \mathbb{N}$, we can define the m-fertility of v as

$$l^{(m)}(v) = |D^{(m)}(v)|. \tag{4}$$

The set $D^{(m)}(v)$ is therefore those descendants of the advisor represented by v who advised at least m other advisees, given by the expression (4). This is a type of measure for the researchers directly advised by the person represented by v.

In Fig. 1, if we assume $v = A$, then $D^{(1)}(v) = \{B, C, D, E\}$, $D^{(2)}(v) = \{C, E\}$ and $D^{(3)}(v) = \{C, E\}$. We then have $l^{(1)}(u) = 4$, $l^{(2)}(u) = l^{(3)}(u) = 2$ and $l^{(k)}(u) = 0$ for $k > 3$.

- The genealogical index of the vertex v is defined by:

$$g_{(1)}(v) = \max\{k \in \mathbb{N} : l(v) \geq k \;\text{ and }\; l^{(k)}(v) \geq k\}. \tag{5}$$

The meaning of $g_{(1)}(v) = g$ is as follows: first, v represents an advisor who had at least g descendants. This derives from the fact that $l(v) \geq g$. Secondly, at least g of these descendants also advised other g advisees. This can be measured by the fact that $l^{(k)}(v) \geq g$.

Intuitively, the genealogical index defines a geometric progression and the total number of vertices belonging to the sub-graph n-ary complete is $\sum_{i=0}^{d+1} (g_{(d)})^i$, where d is the rank that indicates the number of levels considered $(d+1)$. Note that, the genealogical index is $d = 1$ (rank 1).

Assuming that the value of g to a vertex v is equal to one ($g_{(1)} = 1$) then there is at least a complete unary sub-graph in the offspring of the vertex v. Hence, $g_{(1)} = 2$ indicates a complete binary sub-graph and $g_{(1)} = n$ a complete n-ary sub-graph, all with two levels of coverage. The scope of the genealogical index is indicated by its rank, which when it is equal to one indicates two levels of coverage. Fig. 2 shows different genealogy graphs with a genealogical index value ranging from one to six. Higher values of $g_{(1)}$ indicate a higher density graph in its first two consecutive levels.

The value obtained for $g_{(1)}(v)$ is a lower bound, as there is, at least one sub-graph n-ary complete and there is not a sub-graph $(n+1)$-ary complete for coverage up to the second level of the graph. It is worth noting that it is fairly common to find advisees under the scientific guidance of more than one advisor. In these cases the advisee will be assigned to each mentor individually. The genealogical index is an important measure for identifying the influence of the mentor on the academic community, and only includes academic mentoring relationships. However, the metric only includes two genealogical graph levels, which limits its capacity for characterization. In view of this, the procedure outlined above can be applied in a recursive manner to expand the territory of the analyzed vertex.

- For the sake of consistency, we define $g_{(0)}(v) = l(v)$ and $D^{(0)}(v) = D(v)$. For each $d \in \mathbb{N}$ and $m \in \mathbb{N} \cup \{0\}$, let

$$A^{(m)}_{(d)}(v) = \{u \in D(v) : g_{(d-1)}(u) \geq m\} \tag{6}$$

and

$$g_{(d)}(v) = \max\{k \in \mathbb{N} : l(v) \geq k \ \text{ and } \ |A^{(k)}_{(d)}(v)| \geq k\}. \tag{7}$$

We observe that $A^{(m)}_{(1)}(v) = D^{(m)}(v)$, for each $m \in \mathbb{N} \cup \{0\}$.

The set $A^{(m)}_{(d)}(v)$ corresponds to at least m descendants of v whose descendants also have m descendants successively up to $d-1$ generation. The meaning of $g_{(d)}(v) = g$ is the following: it means v has at least g descendants, with each one having at least other g descendants successively until the end of its graph.

The genealogical index can be used to characterize academic genealogy graphs where the scope is only limited by the topology. Thus, all the levels of the graph are taken into account. Fig. 3 illustrates this definition representing a genealogy graph with a single root vertex and a depth equal to four in three different contexts.

In Fig. 3(a) the vertices are labeled with the respective genealogical index of rank one. The root of the graph (vertex of interest) has $g_{(1)} = 4$. The territory of this vertex (central vertex in the figure) contains at least one complete quaternary sub-graph with two levels in its territory. A complete quaternary graph is one in which its vertices have an out-degree equal to four, except the vertices belonging to the last level considered.

Fig. 3(b) shows a graph with its vertices labeled with the genealogical index of rank two, and in the case of the root, its value is $g_{(2)} = 3$. This means that in the territory from the root there is, at least, one complete ternary sub-graph with three levels of coverage. It should be noted that there is another example of a valid full ternary sub-graph in the graph under consideration, but there is no other quaternary sub-graph for this case.

In Fig. 3(c), the genealogical indices of the vertices refer to rank three. The root vertex has $g_{(3)} = 2$, which shows at least a complete binary sub-graph with four scale levels contained in the graph considered. The definition can be applied until the last level of the graph ($d+1$).

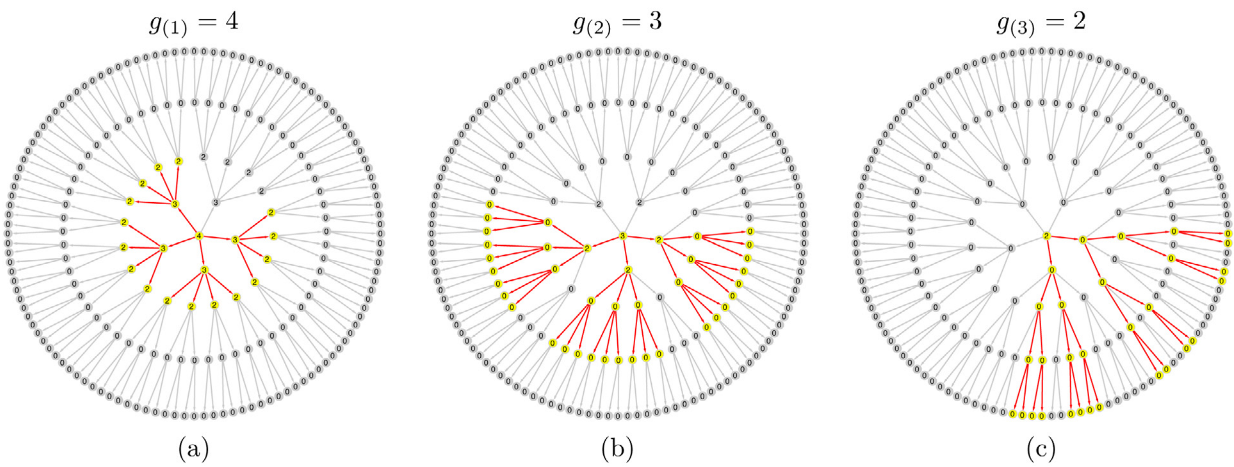


Fig. 3. Examples of genealogy graphs with the representation of $g_{(d)}$ with ranks between one and three. The labels of the vertex have a value of $g_{(d)}$ for their ranks.

2.3. Algorithm to calculate the genealogical index

The genealogical index of rank d (Eq. (7)) can be implemented through a recursive approach. The pseudo-code shown below was designed to calculate the genealogical index. The procedure includes as input three parameters: the genealogy graph $\vec{G}(V, E)$, a vertex of interest (v) and rank (d).

```
    GENEALOGICAL-INDEX(G⃗, v, d)
1   if d = 0
2     return v.g(d) ← |G⃗.adj[v]|
3   else
4     for each u ∈ G⃗.adj[v]
5       u. g(d−1) = GENEALOGICAL-INDEX(G⃗, u, d − 1)
6     breadth ← |G⃗.adj[v]| + 1
7     do
8       adjacency ←0
9       breadth ← breadth −1
10      for each u ∈ G⃗.adj[v]
11        if breadth ⩽u. g(d−1)
12          adjacency ← adjacency +1
13  while breadth > adjacency
14  return v.g(d) ← breadth
```

First, the algorithm checks if the rank is 0. If this is the case, the vertex breadth value is returned as $g_{(0)}(v)$ (lines 1–2). Otherwise, the algorithm is recursively called for all adjacent vertices to vertex v. In this case, we are interested in the previous rank, i.e., $d-1$ (lines 4–5). Then the vertex breadth, i.e., $l(v)$, is compared with the value of $g_{(d-1)}$ of its adjacent ones, counting the occurrence of $l_{(v)} \leqslant g_{(d-1)}$ (lines 10–12). This procedure is repeated while the breadth value is less than or equal to the count of the adjacent ones (line 13), and reduces the breadth value in each iteration. Finally, the procedure returns the adjusted breadth value (line 14).

3. Results

The applicability of the genealogical index was tested by using it to measure all the Ph.D.s. in mathematics and its mentoring academic relationships. These data were gathered from the genealogy project on mathematicians (*Mathematic Genealogy Project – MGP*).

3.1. Dataset

The MGP was designed by Harry Coonce, who was a mathematics professor at the *North Dakota State University*, in the early 1990s (Jackson, 2007) and its aims is to compile information about all the mathematicians in the world. It uses a web platform to carry out the historical record of individuals who obtained a Ph.D. in mathematics (or equivalent) and their students/doctors whose training has been completed. The data available in the MGP on this select group consists of the following: (i) the full *name* of the mathematician; (ii) the *institution* and *country* where the degree was obtained; (iii) the *year* in which the degree was obtained; (iv) the *title* of thesis or work; (v) the *classification* from the area of operation (*Mathematics Subject Classification*[5]); (vi) the *advisors* and direct *advisees*; and (vii) the total number of *descendants*.

The records of the MGP are obtained by an identity number (ID) for each mathematician. The data, which are the object of this paper, were obtained by recursive queries to the MGP (*web crawling*). This is recognized as one of the most complete and detailed source of valuable information about Ph.D.'s (Gargiulo et al., 2016) and their subject-areas, although it may be subject to distortions, especially when restricted to data in the distant past (Gargiulo et al., 2016). This is the price for using a database where the earliest information relies on information from the 14th Century or earlier. However, since that time, the data has become more complete. Additionally, MGP is also a source of inspiration for similar projects in different areas of knowledge (Tenn, 2016).

In August 2015, we obtained 191,276 mathematical records and revealed 202,147 academic mentoring relationships. These were spread over 185 countries, sometimes in combinations (this is due to the fact that two countries can share the same system for awarding degrees) and 2671 institutions (or combinations). The most representative country in the MGP database is the United States, which accounts for 47.40% of total registrations. This is followed by Germany (13.03%), the UK (5.74%), France (4.08%), Canada (3.25%), the Netherlands (3.03%) and Russia (2.60%). For the other countries, the rate of representation is less than 2.00%. In the database, 56 mathematicians received the Fields Medal and 58 winners of the Wolf Prize[6] are listed. These select groups are also the object of study in this manuscript.

[5] Alphanumeric classifier designed by the *American Mathematical Society* and used to categorize subjects of mathematics, available at: http://www.ams.org/msc/msc2010.html, last accessed on August 29, 2016.

[6] The Wolf Prize winner Kiyosi Itô (1987) was not considered because the spelling does not match in MGP.

The representation of mathematicians as vertices and their academic mentoring relationships as edges resulted in an academic genealogy graph. The obtained graph was made up of 10,994 connected components. The largest connected component contains approximately 88.72% of the total number of vertices (158,548), whereas, the second connected component has only 0.08% (141). This genealogical graph has 7542 isolated vertices. Altogether on average every vertex in the graph has 1.05 neighbors.

Genealogical data are usually obtained by spontaneous statement, such as MGP database. This way of filling in information in the database may result in an incomplete set of data. In the case of the MGP database, about 6% of the records do not have the attributes of locality, year or country degree. In addition, the declared attributes may include inconsistencies because, e.g., changing the global geopolitical structure can affect the attributes declared. Works that are based on the description of a group in terms of their attributes (Gargiulo et al., 2016) can use additional databases to enhance their data, which is not the purpose of our work.

It is possible that mentoring relationships, are not fully represented in the database. The record of relationships depends on the interest of the individual who possesses the information. On the other hand, it is common for people to seek to identify and record possible connections with scientists that have the highest reputation. This behavior is termed egocentric genealogy. Finally, we could not find additional databases that could add information of value on academic mentoring relationships for the mathematical community.

3.2. *Quantification of the genealogical index*

The genealogical index was applied to the dataset extracted from MGP. The characterization of the dataset was undertaken on the basis of two dimensions of $g_{(d)}$. The first dimension is the result of the metric that shows $g_{(d)} = n$ to $n = 0, 1, 2, 3, \ldots$. This dimension represents the amplitude of the complete graph n-ary, i.e., the number of direct descendants for each vertex of the graph, except the vertices of the last level. The second dimension includes the rank d which specifies the $d+1$ levels or generations from the analyzed vertex. The calculation of the genealogical index was made to rank 10 (11 levels); this limit was chosen because from level 6 ($d = 6$) only complete unary graphs (paths) were observed, or the maximum result for $d > 6$ is $g_{(d)} = 1$. It is noteworthy that the size of the largest existing path in this dataset is equal to 39.

An important feature that differentiates a genealogical index from other bibliometric measures is its evolutionary patterns. In short, the genealogical index is more heavily influenced by the academic life cycle of the scientist. Fig. 4 shows the evolutionary pattern of the metric. For values before 1900, it is not possible to identify a consistent pattern, mainly because of the difficulty of obtaining complete genealogical data for this period. The number of records for this period is less than 2500 mathematicians (1.3% of the considered database). After 1900, there is a period of stability followed by a decrease in the average values of the genealogical index according to the year of the degree of the scientist. The decreasing might be a consequence of the people killed during the World War I and II, which affected the number of people defending their thesis during the 1930s and 1940s (as consequence of the first War) and during the 1950s and 1960s (as consequence of the second War).

Table 1 shows the size of the groups according to the genealogical index that was applied in the mathematical database. The columns represent the significant ranks of the genealogical index. The rows represent the different possible values for each grade. The intersections show the number of individuals belonging to each group. For example, $g_{(5)} = 2$ indicates a complete binary sub-graph that is 6 levels deep; this kind of sub-graph contains 127 vertices and, in the dataset, there are 9 academics with these features, while $g_{(1)} = 12$ means a complete 12-ary sub-graph that is 2 levels deep, which was only achieved by a single academic. See Table 3 in the Appendix for the size of all the other groups.

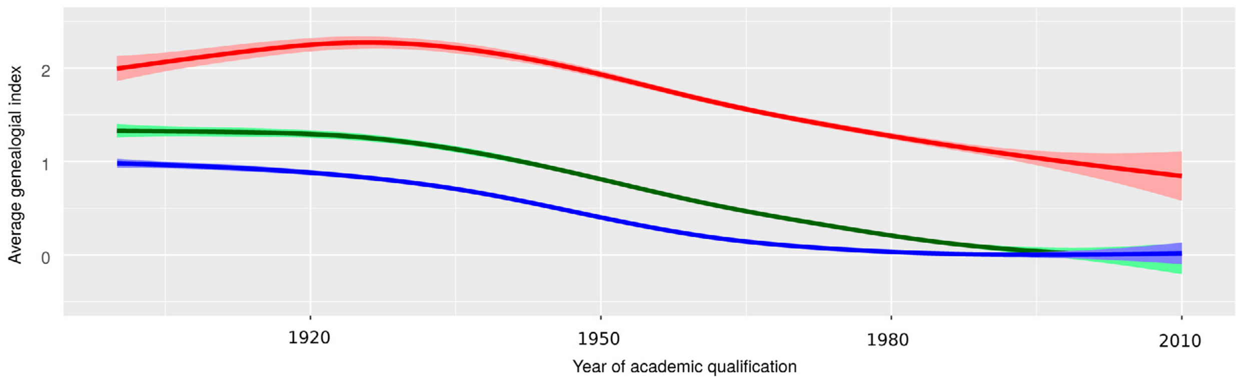


Fig. 4. Evolution of the genealogical index. The horizontal axis refers to the year of the degree of the scientist. The vertical axis refers to the average value of the genealogical index for groups of authors with the same degree year. Each line covers a different rank (1 is red, 2 is green and 3 is blue). The shadow around each line represents the confidence interval. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Summary of the results found by applying the genealogical index in the MGP database. The table shows the number of mathematicians in each class. The classes refer to $g_{(d)}$ = result, for rank d = 1, . . ., 10 and results = 0, . . ., 12. There are no significant observations for the parameters above the limits used. The groups marked in bold are those that appear in the Appendix.

Result	Genealogical index rank									
	1	2	3	4	5	6	7	8	9	10
0	173,827	182,932	186,696	188,500	189,434	189,943	190,237	190,425	190,560	190,652
1	12,338	7287	4363	2736	1833	1333	1039	851	716	624
2	2990	850	201	**40**	**9**	0	0	0	0	0
3	1105	170	**16**	0	0	0	0	0	0	0
4	523	**28**	0	0	0	0	0	0	0	0
5	252	**8**	0	0	0	0	0	0	0	0
6	113	**1**	0	0	0	0	0	0	0	0
7	**51**	0	0	0	0	0	0	0	0	0
8	**30**	0	0	0	0	0	0	0	0	0
9	**24**	0	0	0	0	0	0	0	0	0
10	**15**	0	0	0	0	0	0	0	0	0
11	**7**	0	0	0	0	0	0	0	0	0
12	**1**	0	0	0	0	0	0	0	0	0

A search for larger rank d and the results should be conducted simultaneously to find the most representative academics (on the basis of their ability to disseminate knowledge). In the case of this dataset, a representative genealogy sub-graph can be found from the vertex that represents the German mathematician Heinz Hopf (these values are featured in bold in Table 1). Hopf has $g_{(2)} = 6$ and is the only person that has a complete hex-anal sub-graph. This sort of sub-graph has 259 vertices in its offspring.

Fig. 5 shows the genealogy sub-graph originated from Heinz Hopf, who has the highest genealogical index (rank 2), Hopf also has 6 direct descendants who also have 6 descendants each, in according to academic mentoring.

The importance of Heinz Hopf in academic mentoring can be confirmed by the indices for other ranks $g_{(d)} = [12, 6, 3, 2, 1, 1, 1]$ to d ranging from 1 to 7. It is worth noting that the comparison between different individuals is only effective when we use the same rank d for comparative purposes. Although Hopf is the only mathematician with $g_{(1)} = 12$ and $g_{(2)} = 6$, there are 16 individuals with $g_{(3)} = 3$. For $g_{(4)} = 2$ there are 40 individuals in total. Other mathematicians who have been prominent for propagating knowledge, according to the genealogical index, can be seen in Appendix.

3.3. *Notable mathematicians from the standpoint of the genealogical index*

An important question about studies based on academic genealogy arises from the existence of a correlation between excellence in scientific research and the development of human resources. To illustrate this point, we use two significant subsets comprising outstanding figures in the field of mathematics Fields medalists and Wolf Prize winners.

The Fields Medal is a quadrennial prize awarded to no more than four mathematicians below the age of 40, who have made outstanding contributions to Mathematics. It is regarded as the highest honor that a mathematician can receive. In contrast the Wolf Prize is a set of awards established by the Wolf Foundation in Israel. It has been awarded nearly every year since 1978 and is recognized as one of the most important distinctions that can be awarded to a mathematician. Fig. 9, in the Appendix, shows the members of these selected groups who are also listed in the MGP.

We compared the results in the genealogical index for the following groups: MGP, Fields medalists and Wolf Prize winners. The last of these did not include mathematicians who received both awards. We treated these double winners as a new group (intersection). The Wolf group contain 43 mathematicians, Fields contains 42 and there are 14 mathematicians in the Wolf ∩ Fields group. The comparison of the groups was made by checking the percentage of mathematicians for each result of the

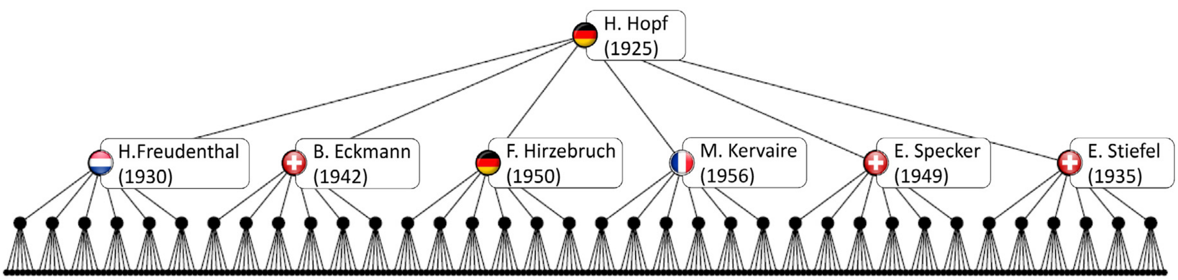


Fig. 5. Sub-graph genealogy of Heinz Hopf, identified by genealogical index with 3 generations of descendants (rank 2) and $g_{(2)} = 6$. For each mathematician mentioned, there appears his name, the graduation year and the country of origin (flag).

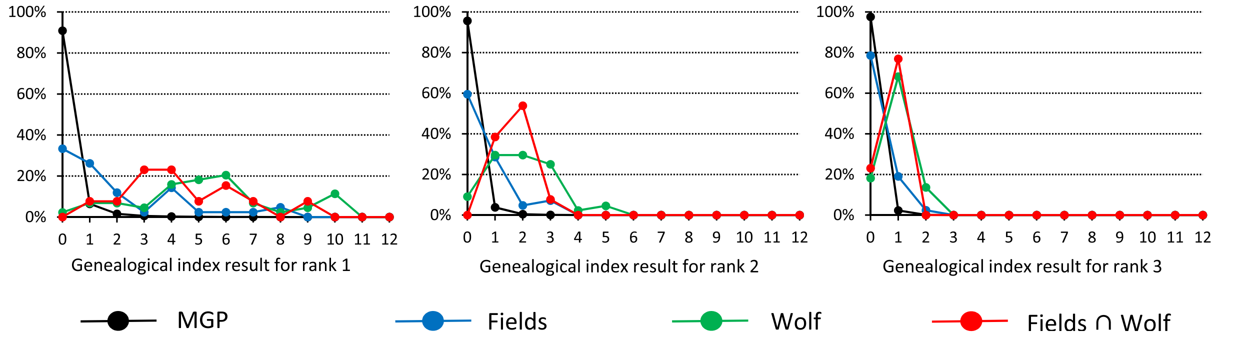


Fig. 6. Percentage distribution of mathematicians in function of $g_{(d)}$ for the groups: MGP, Fields Medal winners, Wolf Prize winners and intersections of the last two.

genealogical index. For example, in the case of the set of the MGP and the genealogical index with rank equal to 1, we found that 90.9% of mathematicians have $g_{(1)} = 0$, for $g_{(1)} = 1$ and $g_{(1)} = 2$ the percentages are 6.5% and 1.6% respectively. In the case of the other classes, the percentages are not representative (note that there are intersections between the sets).

Fig. 6 shows the percentage results for each of the 4 groups, and includes a genealogical index with rank 1, 2 and 3. In all the representative cases, the groups Wolf ∩ Fields and Wolf Prize winners outperform the others. The Wolf group achieves better results than the panel of Fields medalists and this result may be attributed to the fact that there are no age limits for being nominated for a Wolf Prize, which is not the case of the Fields Medal. The group of Fields Medal winners also achieved results that were better than the total group of mathematicians. For the other ranks see Table 18 in the Appendix.

When the median of the genealogical index values of the four groups is examined, it can again be noted that the Wolf and Wolf ∩ Fields groups have a better performance than the others (see Fig. 7).

Comparative results between the four groups included (i.e., Fields medalists, Wolf winners, Wolf ∩ Fields group and MGP) may suggest that there is a correlation between academic mentoring and scientific performance. However, there are no other bases that can support this hypothesis. In this case, we are faced with an ambiguous question: does academic mentoring reflect scientific performance or is the opposite true? We shall deal with this intriguing question in a forthcoming work.

We believe that the genealogical index, which relies on records of academic performance, can be useful to investigate those researchers who are involved in increasing scientific output, and who are recognized by the international academic community.

3.4. The genealogical index from other perspectives

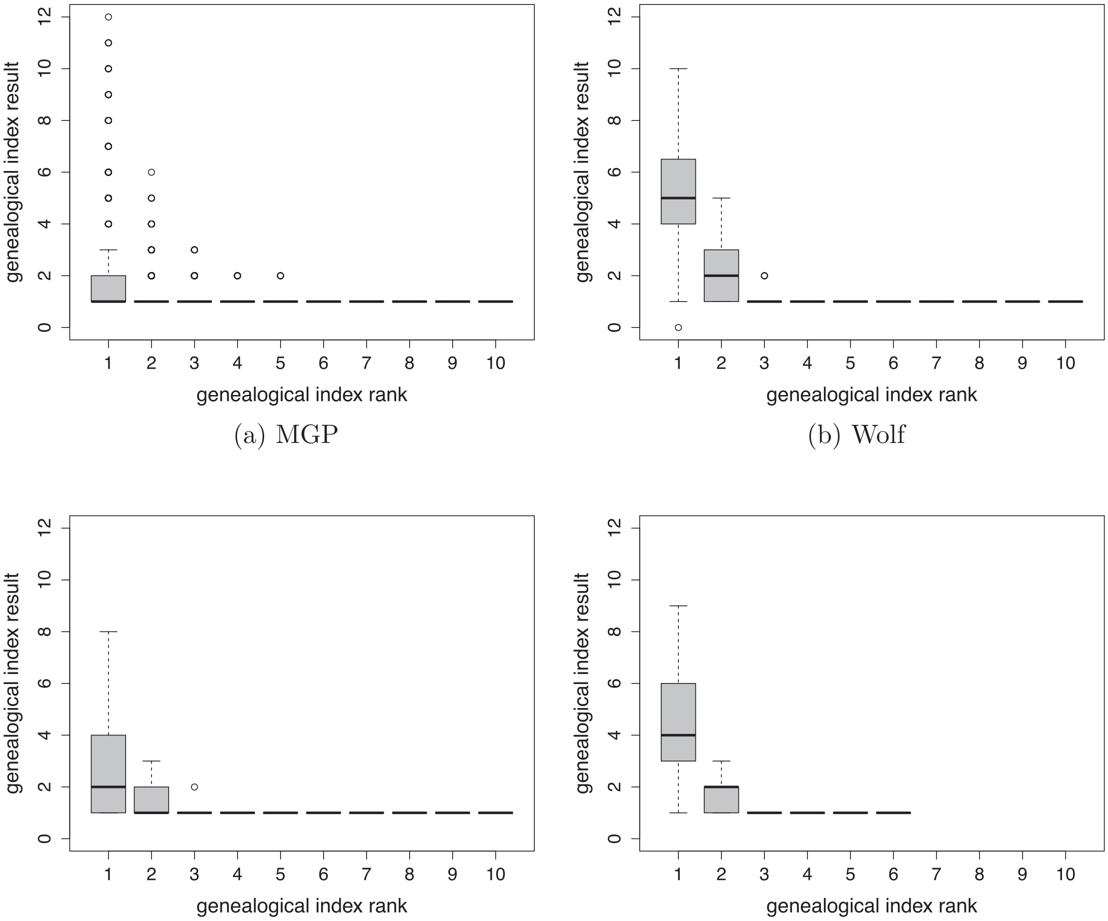
The whole group of mathematicians analyzed (MGP) has other remarkable features and their graph helps to explain how they behave. Fig. 8 shows relationships between the genealogical index of rank 1 and other measures of the group examined in this work. For instance, Fig. 8(a) shows that most of the mathematicians were guided by a single advisor, whereas Fig. 8(b) shows the genealogical index of the advisor and its relation with the number of direct descendants. It is worth noticing that by the phrase “the number of direct descendants of a certain advisor” we mean the total number of mathematicians who were awarded their Ph.D.s under the scientific guidance of that advisor.

The “fecundity” of an individual can be defined as the number of descendants that a person has, and includes mathematicians in the levels of the genealogy graph. The rise in fertility is due to the genealogical index of rank 1, but it is worth noting that in this group the fecundity is not necessarily high. Academics who have many descendants may not have higher genealogical index values (Fig. 8(c)).

Fig. 8(d) shows the relationship between the genealogical index values and the year of the degree of the mathematicians under consideration. On average, this group has degrees in the middle of the 20th Century, and those with higher genealogical index values are potentially older. Although in the complete degree system of the database, there are records going back to 1363, no representative values in the genealogical index for mathematicians of this period have been found (see Section 3.1).

With regard to the complete genealogical graph, the depth of a given vertex in the graph represents the size of the largest path between it and a leaf vertex.[7] Additionally, this measure represents the number of generations descending from a vertex of interest. On the other hand, the height of a given vertex in a genealogical graph is the length of the longest path that exists between it and an origin vertex[8] thus representing the number of generations of an upward vertex. Fig. 8(e) and (f) show the relationship between the genealogical index and measures the depth and height, respectively. Note that

[7] A leaf vertex is defined as a vertex that has no other adjacent vertices in its descendants.
[8] An origin vertex is defined as a vertex that has no other adjacent vertices in its ascendancy.



(c) Fields (d) Wolf ∩ Fields

Fig. 7. Boxplots for the four groups considered. The horizontal axis represents the rank of the genealogical index, while the vertical axis represents its result.

there is a relationship between the values of the genealogical index and the depth of the graph. However, given that the longest path in the complete graph is 39, mathematicians that have a considerable academic lineage do not necessarily have representative values for the genealogical index. Mathematically this group has a height equal to 30, or on average each has 30 previous generations.

Finally, we compared the main bibliometric measures and the genealogical indices with a rank of 1–10, by means of a correlation analysis. The bibliometric performance measures were as follows: (i) citations number, (ii) bibliometric genealogical index, (iii) number of coauthors, and (iv) number of articles. These data were obtained from the Scopus database that is an online subscription-based database maintained by Elsevier and contains bibliographic and citation data.[9] The correlation can be defined as being a quantity that indicates the strength and direction of a linear relationship between two random variables (Todeschini & Baccini, 2016). The Spearman rank correlation coefficient (ρ) was used to measure the correlation between these groups. Table 2 shows the correlation coefficients between the measures for the two groups in question – Fields medalists and Wolf Prize winners. The results indicate that there is no significant correlation between the bibliometric measures and genealogical index. Thus, in our results we found evidence that, academic performance as measured in terms of training human resource only has a limited relationship with the number of scientific publications.

The independence of the genealogical index is an important feature as it adds a new dimension to the analysis of scientific impact and may also be applied to fields other than mathematics. The use of genealogical measures complements the traditional bibliometric analyses, by improving the accuracy of the results, although in this case it is a measure to evaluate long careers and not young scientists.

[9] https://www.scopus.com/freelookup/form/author.uri, last accessed on August 29, 2016.

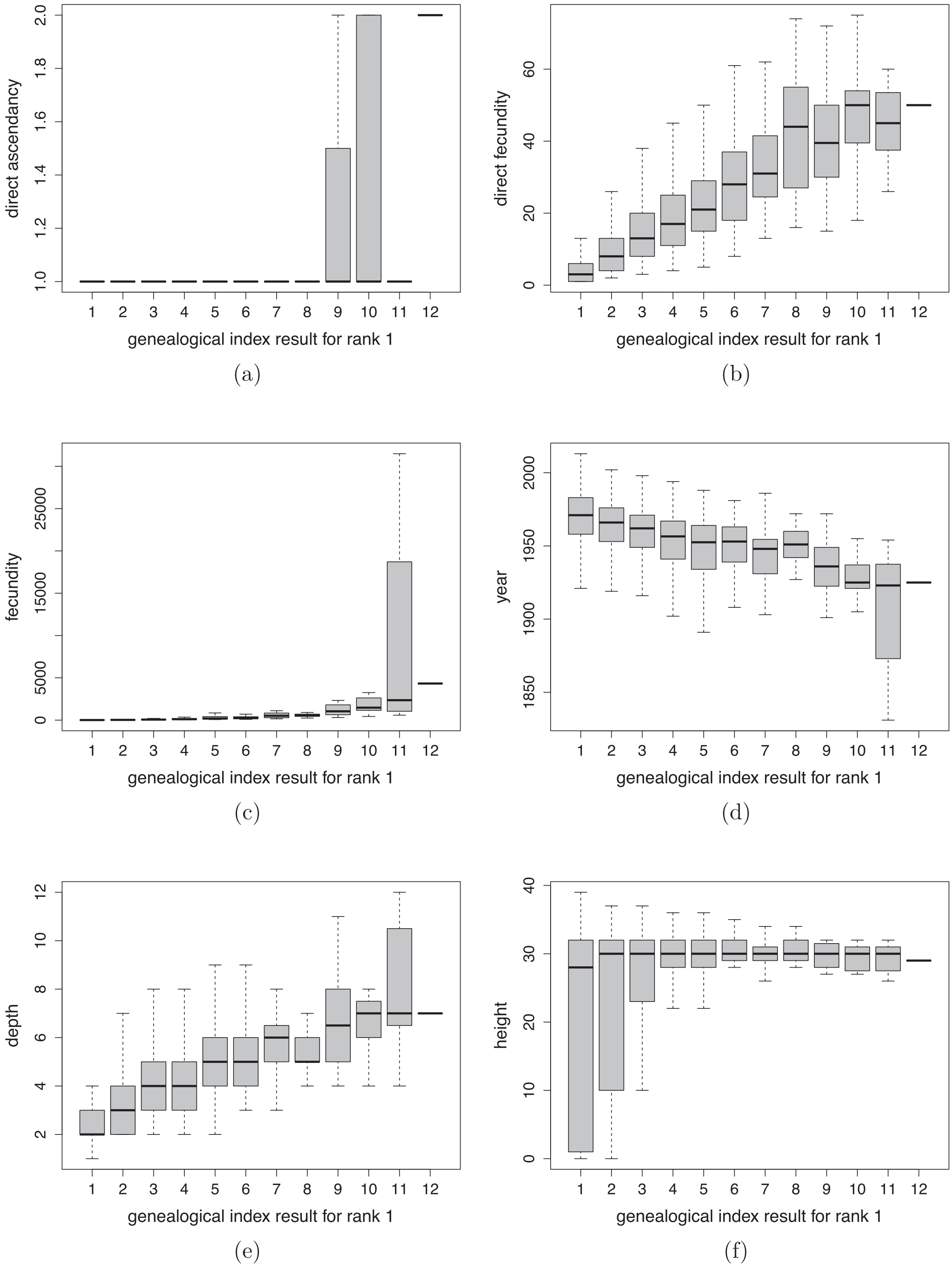


Fig. 8. Distribution of genealogical index from the perspective of: (a)direct ascendancy, (b) direct descendants, (c) fecundity, (d) year of degree, (e) number of later generations (depth), and (f) number of previous generations (height).

Table 2
Correlation coefficients (Spearman – ρ) between the bibliometric measures and the genealogical index (ranks 1–10).

Group	Measure	Genealogical index rank									
		1	2	3	4	5	6	7	8	9	10
Fields	Citations	0.29	0.20	−0.05	−0.26	−0.28	−0.12	−0.16	−0.16	−0.16	−0.16
	Bibliometric h-index	0.24	0.14	−0.06	−0.31	−0.32	−0.20	−0.22	−0.22	−0.22	−0.22
	Coauthors	−0.03	−0.13	−0.26	−0.46	−0.32	−0.26	−0.18	−0.18	−0.18	−0.18
	Articles	0.09	0.02	−0.14	−0.29	−0.20	−0.23	−0.22	−0.22	−0.22	−0.22
Wolf	Citations	−0.15	−0.14	−0.11	−0.05	−0.22	−0.05	0.00	−0.03	−0.03	−0.03
	Bibliometric h-index	−0.07	−0.16	−0.23	−0.10	−0.36	−0.27	−0.13	−0.05	−0.05	−0.05
	Coauthors	−0.02	−0.19	−0.27	−0.24	−0.47	−0.30	−0.17	−0.06	−0.06	−0.06
	Articles	−0.04	−0.17	−0.18	−0.11	−0.35	−0.23	−0.05	−0.01	−0.01	−0.01

4. Conclusion

The academic genealogy is a valuable source for the analysis of publications and citations, since it is largely responsible for everything that is known about the origin and development of disciplines, the dissemination of high-impact knowledge (Didegah & Thelwall, 2013) and the evolution of science.

The broadening of genealogical studies helps to supplement the analysis of a wide range of scientific publications. The value of this addition to research studies is that academic genealogy provides an historical background for assessing the value of researchers in the dissemination of scientific knowledge, as well as being useful in predicting the possible evolution of scientific knowledge areas (Russell & Sugimoto, 2009).

Among several metrics used to characterize lineage structures, the genealogical index is a measure with a significant semantic intuition that provides information about the scope of the mentoring relationships and complements the characterization provided by bibliometrical measures such as number of citations, bibliometric h-index, and number of coauthors and articles. The application of the genealogical index metric is an accurate mean of assessing the contribution of academics to a type of human resources training that is more innovative than the other measures that are used.

In this paper, possible connection is established between scientific production and academic orientation, through the genealogical index which is not correlated with the bibliometric measures that are commonly used.

The genealogical index allows mentoring to be carried out at several levels or generations to quantify the propagation of scientific knowledge and assess its impact on the academic community. The development of topological metrics, such as the genealogical index, and its application in the academic genealogy graphs, can be regarded as an effective way of measuring and analyzing the intellectual influence of academic advisors on their respective communities as well as the way this influence affects different generations.

There is thus evidence that genealogical index is a good measure to assess the knowledge transmission capacity of scientists and, thus, might be useful to evaluate researchers with long-time careers. On the other hand, it is not a good measure for assessing scientists in the early stages of their professional careers, for an obvious reason: the index has an extremely low number of scientists with less than 10 years experience career.

Similarly, the study of academic genealogy provides the academic background necessary for the discovery of the way science has progressed, by tracing its origins and the evolution of its disciplines, which has resulted in the present state of scientific knowledge.

Acknowledgements

The authors would like to express their thanks to CNPq and CAPES for their financial support for this work.

Author contributions

Conceived and designed the analysis: Luciano Rossi; Igor L. Freire; Jesús P. Mena-Chalco.
Collected the data: Luciano Rossi.
Contributed data or analysis tools: Jesús P. Mena-Chalco.
Performed the analysis: Luciano Rossi.
Wrote the paper: Luciano Rossi; Igor L. Freire; Jesús P. Mena-Chalco.
Other contribution: Igor L. Freire.

Appendix

In this section, we provide some additional information to supplement the article and further the discussion of the subject.

Table 3
Class size for the possible combinations between rank d and $g_{(d)}$ result.

Result	Genealogical index rank									
	1	2	3	4	5	6	7	8	9	10
0	1	1	1	1	1	1	1	1	1	1
1	3	4	5	6	7	8	9	10	11	12
2	7	15	31	63	127	255	511	1023	2047	4095
3	13	40	121	364	1093	3280	9841	29,524	88,573	265,720
4	21	85	341	1365	5461	21,845	87,381	349,525	1.4×10^6	5.6×10^6
5	31	156	781	3906	19,531	97,656	488,281	2.4×10^6	1.2×10^7	6.1×10^7
6	43	259	1555	9331	55,987	335,923	2.0×10^6	1.2×10^7	7.3×10^7	4.4×10^8
7	57	400	2801	19,608	137,257	960,800	6.7×10^6	4.7×10^7	3.3×10^8	2.3×10^9
8	73	585	4681	37,449	299,593	2.4×10^6	1.9×10^7	1.5×10^8	1.2×10^9	9.8×10^9
9	91	820	7381	66,430	597,871	5.4×10^6	4.8×10^7	4.4×10^8	3.9×10^9	3.5×10^{10}
10	111	1111	11,111	111,111	1.1×10^6	1.1×10^7	1.1×10^8	1.1×10^9	1.1×10^{10}	1.1×10^{11}
11	133	1464	16,105	177,156	1.9×10^6	2.1×10^7	2.4×10^8	2.6×10^9	2.9×10^{10}	3.1×10^{11}
12	157	1885	22,621	271,453	3.2×10^6	3.9×10^7	4.7×10^8	5.6×10^9	6.8×10^{10}	8.1×10^{11}

Table 3 shows the size of the expected graph for the classes obtained by combining the genealogical index and rank. For example, in the case of vertices that have $g(3) = 5$, it is known that there is at least one subgraph with 781 vertices from the vertex of interest. The number of graphs found in MGP is available in Table 1 in the main text.

Tables 4–17 provides details of the most representative mathematicians according to the genealogical index, and includes the rank and the result of the metric. In each table, there is also the name, institution, year and country in which the mathematician was awarded his academic degree.

Table 4
Sub-set of mathematicians registered in the Mathematics Genealogy Project with $g_{(5)} = 2$.

Name	Institution	Degree	Country
E. H. Moore	Yale University	1885	United States
David Hilbert	University of Königsberg	1885	Germany
Erhard Schmidt	Georg-August-Universität Göttingen	1905	Germany
C. Felix Klein	Rheinische Friedrich-Wilhelms-Universitat Bonn	1868	Germany
C. L. Ferdinand Lindemann	Friedrich-Alexander-Universität Erlangen-Nürnberg	1873	Germany
Karl Weierstrass	Westfälische Wilhelms-Universität Münster	1841	Germany
Pafnuty Chebyshev	St. Petersburg State University	1849	Russia
Ernst Kummer	Martin-Luther-Universitat Halle-Wittenberg	1831	Germany
Gaston Darboux	École normale supérieure - Paris	1866	France

Table 5
Sub-set of mathematicians registered in the Mathematics Genealogy Project with $g_{(4)} = 2$.

Name	Institution	Degree	Country
Oswald Veblen	The University of Chicago	1903	United States
E. H. Moore	Yale University	1885	United States
Ferdinand Frobenius	Universitat Berlin	1870	Germany
George Birkhoff	The University of Chicago	1907	United States
David Hilbert	University of Königsberg	1885	Germany
Erhard Schmidt	Georg-August-Universität Göttingen	1905	Germany
Richard Courant	Georg-August-Universität Göttingen	1910	Germany
C. Felix Klein	Rheinische Friedrich-Wilhelms-Universitat Bonn	1868	Germany
C. L. Ferdinand Lindemann	Friedrich-Alexander-Universität Erlangen-Nürnberg	1873	Germany
Philipp Furtwangler	Georg-August-Universität Göttingen	1896	Germany
Solomon Lefschetz	Clark University	1911	United States
Karl Weierstrass	Westfälische Wilhelms-Universität Münster	1841	Germany
Hermann Schwarz	Universitat Berlin	1864	Germany
Leopold Fejer	Eötvös Loránd University	1902	Hungary
Marcel Riesz	Eötvös Loránd University	1912	Hungary
Edmund Landau	Universitat Berlin	1899	Germany
Salomon Bochner	Universitat Berlin	1921	Germany
Friedrich Schottky	Universitat Berlin	1875	Germany
Nikolai Luzin	Moscow State University	1915	Russia
Lazarus Fuchs	Universitat Berlin	1858	Germany

Table 6
Sub-set of mathematicians registered in the Mathematics Genealogy Project with $g_{(4)} = 2$ (complement).

Name	Institution	Degree	Country
Ernest Barnes	University of Cambridge	1907	United Kingdom
Pafnuty Chebyshev	St. Petersburg State University	1849	Russia
Andrei Markov	St. Petersburg State University	1884	Russia
Georgy Voronoy	St. Petersburg State University	1896	Russia
Waclaw Sierpinski	Uniwersytet Jagielloński	1906	Poland
Stefan Mazurkiewicz	University of Lwów	1913	Poland
Heinz Hopf	Universitat Berlin	1925	Germany
Ernst Kummer	Martin-Luther-Universitat Halle-Wittenberg	1831	Germany
Edmund Whittaker	University of Cambridge	1895	United Kingdom
Jacques Hadamard	École normale supérieure - Paris	1892	France
Karl Pearson	University of Cambridge	1879	United Kingdom
Arnold Sommerfeld	University of Königsberg	1891	Germany
Charles Hermite	École Polytechnique	1841	France
Gaston Darboux	École normale supérieure - Paris	1866	France
C. Emile Picard	École normale supérieure - Paris	1877	France
Emile Borel	École normale supérieure - Paris	1893	France
Alexander Korkin	St. Petersburg State University	1860	Russia
Gregor Wentzel	Ludwig-Maximilians-Universität München	1921	Germany
Dmitry Grave	St. Petersburg State University	1896	Russia
Georges Valiron	Université de Paris	1914	France

Table 7
Sub-set of mathematicians registered in the Mathematics Genealogy Project with $g_{(3)} = 3$.

Name	Institution	Degree	Country
Oswald Veblen	The University of Chicago	1903	United States
E. H. Moore	Yale University	1885	United States
Ferdinand Frobenius	Universitat Berlin	1870	Germany
David Hilbert	University of Königsberg	1885	Germany
Erhard Schmidt	Georg-August-Universität Göttingen	1905	Germany
Richard Courant	Georg-August-Universität Göttingen	1910	Germany
C. Felix Klein	Rheinische Friedrich-Wilhelms-Universitat Bonn	1868	Germany
Solomon Lefschetz	Clark University	1911	United States
Karl Weierstrass	Westfälische Wilhelms-Universität Münster	1841	Germany
Leopold Fejer	Eötvös Loránd University	1902	Hungary
Marcel Riesz	Eötvös Loránd University	1912	Hungary
Edmund Landau	Universitat Berlin	1899	Germany
Heinz Hopf	Universitat Berlin	1925	Germany
G. H. Hardy	University of Cambridge	1903	United Kingdom
Jacques Hadamard	École normale supérieure - Paris	1892	France
C. Emile Picard	École normale supérieure - Paris	1877	France

Table 8
Sub-set of mathematicians registered in the Mathematics Genealogy Project with $g_{(2)} = 4$.

Name	Institution	Degree	Country
R. L. Moore	The University of Chicago	1905	United States
Oswald Veblen	The University of Chicago	1903	United States
Saunders Mac Lane	Georg-August-Universität Göttingen	1934	Germany
Marshall Stone	Harvard University	1926	United States
Irving Segal	Yale University	1940	United States
Hellmuth Kneser	Georg-August-Universität Göttingen	1921	Germany
C. Felix Klein	Rheinische Friedrich-Wilhelms-Universitat Bonn	1868	Germany
Karl Weierstrass	Westfälische Wilhelms-Universität Münster	1841	Germany
Leopold Fejer	Eötvös Loránd University	1902	Hungary
Marcel Riesz	Eötvös Loránd University	1912	Hungary
Alonzo Church	Princeton University	1927	United States
Stefan Mazurkiewicz	University of Lwów	1913	Poland
Heinrich Behnke	Universitat Hamburg	1923	Germany
John Tukey	Princeton University	1939	United States
Ernst Kummer	Martin-Luther-Universitat Halle-Wittenberg	1831	Germany
Leon Lichtenstein	Technische Universitat Berlin	1908	Germany
Beno Eckmann	ETH Zürich	1942	Switzerland
Ernst Holder	Universitat Leipzig	1926	Germany
Lothar Collatz	Universitat Berlin	1935	Germany
Konrad Knopp	Universitat Berlin	1907	Germany
Karl-Heinrich Weise	Friedrich-Schiller-Universitat Jena	1934	Germany
Edmund Hlawka	Universitat Wien	1938	Austria
Arnold Sommerfeld	Universität Königsberg	1891	Germany
Helmut Hasse	Philipps-Universitat Marburg	1922	Germany
Jacques-Louis Lions	Université Henri Poincaré Nancy 1	1954	France
Haim Brezis	Université de Paris	1972	France
Franz Rellich	Georg-August-Universität Göttingen	1929	Germany
Carl Siegel	Georg-August-Universität Göttingen	1920	Germany

Table 9
Sub-set of mathematicians registered in the Mathematics Genealogy Project with $g_{(2)} = 5$.

Name	Institution	Degree	Country
George Birkhoff	The University of Chicago	1907	United States
David Hilbert	Universität Königsberg	1885	Germany
Erhard Schmidt	Georg-August-Universität Göttingen	1905	Germany
Richard Courant	Georg-August-Universität Göttingen	1910	Germany
Solomon Lefschetz	Clark University	1911	United States
Salomon Bochner	Universitat Berlin	1921	Germany
Andrei Kolmogorov	Moscow State University	1925	Russia
Oscar Zariski	Universitá di Roma La Sapienza	1925	Italy

Table 10
Sub-set of mathematicians registered in the Mathematics Genealogy Project with $g_{(2)} = 6$.

Name	Institution	Degree	Country
Heinz Hopf	Universitat Berlin	1925	Germany

Table 11
Sub-set of mathematicians registered in the Mathematics Genealogy Project with $g_{(1)} = 7$.

Name	Institution	Degree	Country
R H Bing	University of Texas at Austin	1945	United States
Oswald Veblen	The University of Chicago	1903	United States
E. H. Moore	Yale University	1885	United States
Lars Ahlfors	Helsingin yliopisto	1932	Finland
Marvin Minsky	Princeton University	1954	United States
C. L. Ferdinand Lindemann	Friedrich-Alexander-Universität Erlangen-Nürnberg	1873	Germany
Helmut Schaefer	Universitat Leipzig	1951	Germany
Richard Bruck	University of Toronto	1940	Canada
Robert Constable	University of Wisconsin-Madison	1968	United States
Nikolai Luzin	Moscow State University	1915	Russia
William Thurston	University of California, Berkeley	1972	United States
Arthur Wightman	Princeton University	1949	United States
Wilhelm Magnus	Johann Wolfgang Goethe-Universitat Frankfurt	1931	Germany
Erich Lehmann	University of California, Berkeley	1946	United States
Alfred Tarski	Uniwersytet Warszawski	1924	Poland
Joseph Keller	New York University	1948	United States
Stephen Davis	Rensselaer Polytechnic Institute	1964	United States
Israil Gelfand	Moscow State University	1935	Russia
Frank Smithies	University of Cambridge	1937	United Kingdom
Ernst Holder	Universitat Leipzig	1926	Germany
Joseph Harris	Harvard University	1978	United States
Johann Cigler	Universitat Wien	1960	Austria
Friedrich Hirzebruch	Westfalische Wilhelms-Universitat Münster	1950	Germany
Heinz Bauer	Friedrich-Alexander-Universität Erlangen-Nürnberg	1953	Germany
Wilhelm Klingenberg	Christian-Albrechts-Universitat zu Kiel	1950	Germany

Table 12
Sub-set of mathematicians registered in the Mathematics Genealogy Project with $g_{(1)} = 7$ (complement).

Name	Institution	Degree	Country
Max Deuring	Georg-August-Universität Göttingen	1931	Germany
Erhard Heinz	Georg-August-Universität Göttingen	1951	Germany
Ernst Peschl	Ludwig-Maximilians-Universität München	1931	Germany
Emanuel Sperner	Universitat Hamburg	1928	Germany
Karl-Heinrich Weise	Friedrich-Schiller-Universitat Jena	1934	Germany
Georgiy Shilov	Moscow State University	1954	Russia
Kurt Friedrichs	Georg-August-Universität Göttingen	1925	Germany
Ernst Specker	ETH Zürich	1949	Switzerland
Gerald Sacks	Cornell University	1961	United States
Igor Shafarevich	Steklov Institute of Mathematics	1946	Russia
Karl Stein	Westfalische Wilhelms-Universitat Münster	1937	Germany
Peter Neumann	University of Oxford	1966	United Kingdom
Albert Pfluger	ETH Zürich	1935	Switzerland
Walter Saxer	ETH Zürich	1923	Switzerland
Arnold Sommerfeld	University of Königsberg	1891	Germany
William Feller	Georg-August-Universität Göttingen	1926	Germany
Peter Bickel	University of California, Berkeley	1963	United States
Gene Golub	University of Illinois at Urbana-Champaign	1959	United States
Shokichi Iyanaga	Tokyo Imperial University	1931	Japan
Karl Zeller	Eberhard Karls Universität Tübingen	1950	Germany
Jacques Neveu	Faculté des Sciences, Paris	1955	France
Willem van Zwet	Universiteit van Amsterdam	1964	Netherlands
Martin Vetterli	École Polytechnique Fédérale de Lausanne	1986	Switzerland
David Cox	University of Leeds	1949	United Kingdom
Dennis Sciama	University of Cambridge	1953	United Kingdom
Joos (Joseph) Vandewalle	Katholieke Universiteit Leuven	1976	Belgium

Table 13
Sub-set of mathematicians registered in the Mathematics Genealogy Project with $g_{(1)} = 8$.

Name	Institution	Degree	Country
Saunders Mac Lane	Georg-August-Universität Göttingen	1934	Germany
Paul Halmos	University of Illinois at Urbana-Champaign	1938	United States
John Moore	Brown University	1952	United States
Robion Kirby	The University of Chicago	1965	United States
Jim Douglas, Jr.	Rice University	1952	United States
Hubert Wall	University of Wisconsin-Madison	1927	United States
William Browder	Princeton University	1958	United States
Roger Temam	Université de Paris	1967	France
Manuel Blum	Massachusetts Institute of Technology	1964	United States
Louis Nirenberg	New York University	1949	United States
Peter Lax	New York University	1949	United States
Eugene B. Dynkin	Moscow State University	1948	Russia
Friedrich Kasch	Westfälische Wilhelms-Universität Münster	1950	Germany
Konrad Jacobs	Ludwig-Maximilians-Universität München	1954	Germany
Roland Bulirsch	Technische Universitat München	1961	Germany
Eduard Stiefel	ETH Zürich	1935	Switzerland
Graham Higman	University of Oxford	1941	United Kingdom
George Mackey	Harvard University	1942	United States
E. Christopher Zeeman	University of Cambridge	1955	United Kingdom
Albert Meyer	Harvard University	1972	United States
Michael Atiyah	University of Cambridge	1955	United Kingdom
John Wheeler	The Johns Hopkins University	1933	United States
George Dantzig	University of California, Berkeley	1946	United States
Henry McKean, Jr.	Princeton University	1955	United States
Laurent Schwartz	Université Louis Pasteur - Strasbourg I	1943	France
Azriel Rosenfeld	Columbia University	1960	United States
C. R. Rao	University of Cambridge	1948	United Kingdom
Thomas Kailath	Massachusetts Institute of Technology	1961	United States
Jeffrey Ullman	Princeton University	1966	United States
Hans Liepmann	Universität Zürich	1938	Switzerland

Table 14
Sub-set of mathematicians registered in the Mathematics Genealogy Project with $g_{(1)} = 9$.

Name	Institution	Degree	Country
Irving Kaplansky	Harvard University	1941	United States
Nathan Jacobson	Princeton University	1934	United States
Stephen Smale	University of Michigan	1957	United States
George Birkhoff	The University of Chicago	1907	United States
Irving Segal	Yale University	1940	United States
Antoni Zygmund	Uniwersytet Warszawski	1923	Poland
Richard Courant	Georg-August-Universität Göttingen	1910	Germany
Solomon Lefschetz	Clark University	1911	United States
Richard Brauer	Universitat Berlin	1926	Germany
Norman Steenrod	Princeton University	1936	United States
Lipman Bers	Charles University	1938	Czech Republic
Alonzo Church	Princeton University	1927	United States
Samuel Karlin	Princeton University	1947	United States
Issai Schur	Universitat Berlin	1901	Germany
Garrett Birkhoff	University of Cambridge	unknown	United Kingdom
Philip Hall	University of Cambridge	1926	United Kingdom
Beno Eckmann	ETH Zürich	1942	Switzerland
Friedrich Bauer	Ludwig-Maximilians-Universität München	1951	Germany
Gunter Hotz	Georg-August-Universität Göttingen	1958	Germany
Phillip Griffiths	Princeton University	1962	United States
Helmut Hasse	Philipps-Universitat Marburg	1922	Germany
Haim Brezis	Université de Paris	1972	France
Carl Siegel	Georg-August-Universität Göttingen	1920	Germany
Alan Oppenheim	Massachusetts Institute of Technology	1964	United States

Table 15
Sub-set of mathematicians registered in the Mathematics Genealogy Project with $g_{(1)} = 10$.

Name	Institution	Degree	Country
R. L. Moore	The University of Chicago	1905	United States
Shiing-Shen Chern	Universitat Hamburg	1936	Germany
Elias Stein	The University of Chicago	1955	United States
David Hilbert	Universität Königsberg	1885	Germany
Hellmuth Kneser	Georg-August-Universität Göttingen	1921	Germany
C. Felix Klein	Rheinische Friedrich-Wilhelms-Universitat Bonn	1868	Germany
John Tate, Jr.	Princeton University	1950	United States
Emil Artin	Universitat Leipzig	1921	Germany
Salomon Bochner	Universitat Berlin	1921	Germany
Andrei Kolmogorov	Moscow State University	1925	Russia
John Tukey	Princeton University	1939	United States
Oscar Zariski	Università di Roma La Sapienza	1925	Italy
Lothar Collatz	Universitat Berlin	1935	Germany
Pavel Aleksandrov	Moscow State University	1927	Russia
Edmund Hlawka	Universitat Wien	1938	Austria

Table 16
Sub-set of mathematicians registered in the Mathematics Genealogy Project with $g_{(1)} = 11$.

Name	Institution	Degree	Country
Reinhold Baer	Georg-August-Universität Göttingen	1927	Germany
Erhard Schmidt	Georg-August-Universität Göttingen	1905	Germany
Karl Weierstrass	Westfälische Wilhelms-Universität Münster	1841	Germany
Heinrich Behnke	Universitat Hamburg	1923	Germany
Ernst Kummer	Martin-Luther-Universitat Halle-Wittenberg	1831	Germany
Jacques-Louis Lions	Université Henri Poincaré Nancy 1	1954	France
Mark Krasnoselskii	Institute of Mathematics, Kiev	1948	Ukraine

Table 17
Sub-set of mathematicians registered in the Mathematics Genealogy Project with $g_{(1)} = 12$.

Name	Institution	Degree	Country
Heinz Hopf	Universitat Berlin	1925	Germany

Table 18
Percentage representation of the distribution of mathematicians in according to $g_{(d)}$ for the 4 groups included.

Result	Group	Genealogical index rank									
		1	2	3	4	5	6	7	8	9	10
	MGP	90.9%	95.6%	97.6%	98.5%	99.0%	99.3%	99.5%	99.6%	99.6%	99.7%
0	Fields	33.3%	59.5%	78.6%	85.7%	95.2%	97.6%	97.6%	97.6%	97.6%	97.6%
	Wolf	2.3%	9.1%	18.2%	50.0%	75.0%	86.4%	95.5%	97.7%	97.7%	97.7%
	Wolf ∩ Fields	0.0%	0.0%	23.1%	69.2%	84.6%	92.3%	100.0%	100.0%	100.0%	100.0%
	MGP	6.5%	3.8%	2.3%	1.4%	1.0%	0.7%	0.5%	0.4%	0.4%	0.3%
1	Fields	26.2%	28.6%	19.0%	14.3%	4.8%	2.4%	2.4%	2.4%	2.4%	2.4%
	Wolf	6.8%	29.5%	68.2%	50.0%	25.0%	13.6%	4.5%	2.3%	2.3%	2.3%
	Wolf ∩ Fields	7.7%	38.5%	76.9%	30.8%	15.4%	7.7%	0.0%	0.0%	0.0%	0.0%
	MGP	1.6%	0.4%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
2	Fields	11.9%	4.8%	2.4%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	Wolf	6.8%	29.5%	13.6%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	Wolf ∩ Fields	7.7%	53.8%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	MGP	0.6%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
3	Fields	2.4%	7.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	Wolf	4.5%	25.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	Wolf ∩ Fields	23.1%	7.7%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	MGP	0.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
4	Fields	14.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	Wolf	15.9%	2.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	Wolf ∩ Fields	23.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	MGP	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
5	Fields	2.4%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	Wolf	18.2%	4.5%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	Wolf ∩ Fields	7.7%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	MGP	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
6	Fields	2.4%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	Wolf	20.5%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	Wolf ∩ Fields	15.4%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	MGP	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
7	Fields	2.4%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	Wolf	6.8%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	Wolf ∩ Fields	7.7%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	MGP	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
8	Fields	4.8%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	Wolf	2.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	Wolf ∩ Fields	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	MGP	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
9	Fields	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	Wolf	4.5%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	Wolf ∩ Fields	7.7%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	MGP	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
10	Fields	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	Wolf	11.4%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	Wolf ∩ Fields	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	MGP	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
11	Fields	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	Wolf	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	Wolf ∩ Fields	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	MGP	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
12	Fields	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	Wolf	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	Wolf ∩ Fields	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

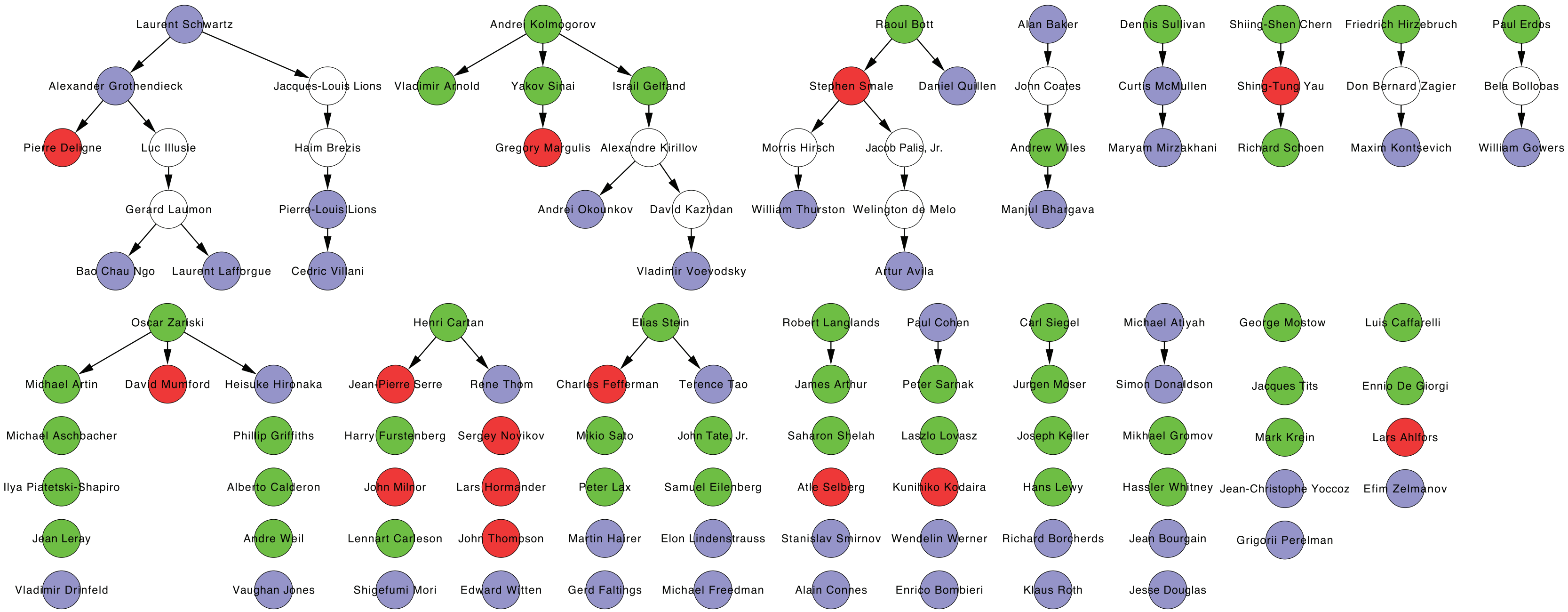


Fig. 9. Graph representing the relationships between notable mathematicians. The vertices in green indicate Wolf Prize winners. Blue represents Fields medalists. Red represents mathematicians who hold both prizes (Wolf ∩ Fields). The colorless vertices indicate transitional vertices. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In the main text, we provide charts with the percentage distributions of each group that are in accordance with the rank and the results of the genealogical index (Fig. 6). Since the chart shows distributions up to Rank 3, Table 18 supplements the information by expanding the parameters.

Finally, we provide the graph of the mathematicians that were included in the analyses (Fig. 9). The following groups are highlighted (MGP, Wolf, Fields and Wolf ∩ Fields), together with their components and the academic mentoring relationship between them.

References

Andraos, J. (2005). Scientific genealogies of physical and mechanistic organic chemists. *Canadian Journal of Chemistry*, *83*, 1400–1414.

Bennett, A. F., & Lowe, C. (2005). The academic genealogy of George A. Bartholomew. *Integrative and Comparative Biology*, *45*, 231–233.

Chang, S. (2011). *Academic genealogy of mathematicians*. World Scientific.

David, S. V., & Hayden, B. Y. (2012). Neurotree: A collaborative, graphical database of the academic genealogy of neuroscience. *PLoS One*, *7*, e46608.

Didegah, F., & Thelwall, M. (2013). Which factors help authors produce the highest impact research? Collaboration, journal and document properties. *Journal of Informetrics*, *7*, 861–873.

Elias, M., Floeter-Winter, L. M., & Mena-Chalco, J. P. (2016). The dynamics of Brazilian protozoology over the past century. *Memórias do Instituto Oswaldo Cruz*, *111*, 67–74.

Frické, M. (2014). Big data and its epistemology. *Journal of the Association for Information Science and Technology*,

Gargiulo, F., Caen, A., Lambiotte, R., & Carletti, T. (2016). The classical origin of modern mathematics. *EPJ Data Science*.

Hart, R. E., & Cossuth, J. H. (2013). A family tree of tropical meteorology's academic community and its proposed expansion. *Bulletin of the American Meteorological Society*, *94*, 1837–1848.

Hirsch, J. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 16569–16572.

Jackson, A. (2007). A labor of love: The mathematics genealogy project. *Notices of the AMS*, *54*, 1002–1003.

Jackson, D. C. (2011). Academic genealogy and direct calorimetry: A personal account. *Advances in Physiology Education*, *35*, 120–127.

Kelley, E. A., & Sussman, R. W. (2007). An academic genealogy on the history of American field primatologists. *American Journal of Physical Anthropology*, *132*, 406–425.

Kobayashi, H. (2015). Academic genealogy of Shoshichi Kobayashi and individuals who influenced him. In *Geometry and analysis on manifolds*. pp. 17–38. Springer.

van der Kruit, P. C. (2015). Studies in Utrecht. In *Jacobus Cornelius Kapteyn*. pp. 27–52. Springer.

Lü, L., Zhou, T., Zhang, Q.-M., & Stanley, H. E. (2016). The h-index of a network node and its relation to degree and coreness. *Nature Communications*, 7.

Malmgren, R., Ottino, J., & Amaral, L. (2010). The role of mentorship in protégé performance. *Nature*, *465*, 622–626.

Perry, M., Reny, P. J., et al. (2013). How to count citations if you must. *American Economic Review*.

Petersen, A. M., & Succi, S. (2013). The z-index: A geometric representation of productivity and impact which accounts for information in the entire rank-citation profile. *Journal of Informetrics*, *7*, 823–832.

Rossi, L., & Mena-Chalco, J. P. (2015). Indice-h genealogico expandido: Uma medida de impacto em grafos de orientação acadêmica. In *Brazilian Workshop on Social Network Analysis and Mining (BraSNAM)* (pp. 1–12).

Russell, T. G., & Sugimoto, C. R. (2009). MPACT family trees: Quantifying academic genealogy in library and information science. *Journal of Education for Library and Information Science*, 248–262.

Sugimoto, C. R. (2014). Academic genealogy. In B. Cronin, & C. R. Sugimoto (Eds.), *Beyond bibliometrics: Harnessing multidimensional indicators of scholarly impact* (1st ed., pp. 365–382). MIT Press.

Sugimoto, C. R., Ni, C., Russell, T. G., & Bychowski, B. (2011). Academic genealogy as an indicator of interdisciplinarity: An examination of dissertation networks in library and information science. *Journal of the American Society for Information Science and Technology*, *62*, 1808–1828.

Tenn, J. S. (2016). *Introducing astrogen: The astronomy genealogy project*. , arXiv preprint: arXiv:1612.08908.

Todeschini, R., & Baccini, A. (2016). *Handbook of bibliometric indicators volume 1*. Wiley-VCH Verlag GmbH & Co. KGaA.

Tol, R. S. J. (2013). Identifying excellent researchers: A new approach. *Journal of Informetrics*, *7*, 803–810.

Waltman, L., & Eck, N. J. V. (2012). The inconsistency of the h-index. *Journal of the American Society for Information Science and Technology*, *63*, 406–415.

Yong, A. (2014). Critique of Hirsch's citation index: A combinatorial Fermi problem. *Notices of the American Mathematical Society*, *61*, 1040–1050.