Genealogia acadêmica: Um novo olhar sobre impacto acadêmico de pesquisadores

Modalidade: Artigo

Eixo: Metodologias de avaliação e mensuração de impactos de programas, políticas e ações de CT&I

Tema: Ferramentas ou procedimentos de análise de dados e informações na CT&I

Resumo

A genealogia acadêmica considera as relações de orientação acadêmica, em diferentes níveis, para a avaliação e classificação de grupos e indivíduos atuantes na academia, respectivamente. Os resultados obtidos por meio da aplicação da genealogia acadêmica podem ser considerados de forma a complementar às tradicionais análises sobre produção bibliográfica. A relevância da genealogia acadêmica recai na possibilidade de descoberta de novas informações que auxiliem às agências de fomento em CT&I para documentar, identificar, e analisar o impacto acadêmico de um pesquisador ou um grupo de pesquisadores. O conhecimento obtido a partir da genealogia acadêmica está relacionado com a produção de recursos humanos, com a estruturação de comunidades acadêmicas e com seus respectivos desdobramentos acadêmicos.

Palavras-chave - Genealogia. Indicadores de ciência e tecnologia. Cientometria.

Abstract

The academic genealogy considers the relations of academic mentoring, at different levels, for the evaluation and classification of groups and individuals acting in the academy, respectively. The results obtained through the application of the academic genealogy can be considered in a way that complements the traditional analysis of scholarly output. The relevance of academic genealogy lies in the possibility of discovering new information that will assist ST&I development agencies to document, identify, and analyze the academic impact of a researcher or a group of researchers. The knowledge obtained from academic genealogy is related with the propagation of human resources, with the structuring of academic communities and with their unfoldings.

Keywords - Genealogy. Science and technology Indicators. Scientometrics.

1. INTRODUÇÃO

A busca da humanidade por suas origens é tema recorrente e pode ser observado desde os registros no Livro de Gênesis até iniciativas contemporâneas no sentido de documentar a história de comunidades acadêmicas por meio de suas relações. A genealogia acadêmica (GA) busca, através do registro de acadêmicos em diversas áreas e seus relacionamentos de orientação, documentar de forma perene a história de comunidades acadêmico-científicas e, assim, oferecer um melhor entendimento sobre como o passado impactou na formação do presente e qual a tendência para o futuro destas comunidades (Malmgren *et al.*, 2010).

A genealogia é uma ciência auxiliar da história que estuda a origem, evolução e disseminação de grupos inter-relacionados por laços familiares ou outro tipo de relacionamento qualquer. O objeto de pesquisa da genealogia são os ascendentes e os descendentes de um indivíduo. O processo de pesquisa envolvido na genealogia abrange a identificação de parentesco formal ou informal entre indivíduos por meio de registros históricos como certidões de nascimento, casamento, óbito, registro de propriedades e outros documentos que possam comprovar uma ligação entre eles (Chao *et al.*, 1992).

Um objetivo importante em um estudo genealógico é obter conhecimento que permita um entendimento do futuro com base no passado, por vezes, desconhecido. Esse estudo possibilita traçar a memória de diferentes famílias (Gargiulo *et al.*, 2016).

O desenvolvimento da internet e de novas tecnologias e plataformas de informação mudou radicalmente a forma pela qual pesquisadores na área de genealogia conduzem suas pesquisas, o acesso e o compartilhamento de dados é uma das evoluções observadas no mundo digital. Pesquisadores na área de genealogia, seja em qual for o contexto, têm como ferramenta fundamental de trabalho o compartilhamento amplo e assertivo de dados referentes aos indivíduos pesquisados e seus relacionamentos.

A atividade de orientação/supervisão acadêmica é uma forma de relacionamento que promove a evolução do orientado, da instituição, da ciência e da sociedade de termos gerais. Atualmente pode-se observar diferentes iniciativas, por parte de comunidades acadêmicas específicas, no sentido de documentar, analisar e classificar estruturas de genealogia acadêmica. Analisar os relacionamentos de orientação acadêmica sob a forma de uma estrutura genealógica (e.g., grafo ou árvore de genealogia acadêmica), permite um maior entendimento sobre a comunidade científica,

a caracterização do acadêmico por meio de seus relacionamentos e a identificação do impacto gerado por estes acadêmicos na formação dessa comunidade.

Segundo Lane (2010), a Plataforma Lattes[1], projetada e desenvolvida pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), apresenta-se como o maior e mais relevante repositório curricular de acadêmicos relacionados com a Ciência Brasileira. Desse conjunto de dados, além de explorar a produção científica nacional de todo indivíduo/grupo acadêmico atuante no Brasil pode-se explorar as relações formais de orientação tanto nos níveis de mestrado quanto de doutorado. No contexto de genealogia, a propagação/projeção acadêmica de um indivíduo/grupo pode ser mensurada objetivamente usando os registros curriculares da Plataforma Lattes, um valioso conjunto de dados pouco explorado nessa perspectiva.

O objetivo deste trabalho é apresentar a definição do conceito de genealogia acadêmica, os bancos de dados existentes, e os principais benefícios para mensuração quantitativa e qualitativa da ciência através de aplicações. Adicionalmente, é apresentado um novo sistema de consulta sobre a genealogia acadêmica do Brasil denominado Plataforma Acácia[2]. Acreditamos que este novo olhar permitirá, de forma complementar a outros olhares cientométricos, evidenciar importantes atores em CT&I no Brasil.

2. GENEALOGIA ACADÊMICA

Segundo Cronin e Sugimoto (2014), a Genealogia Acadêmica (GA) é definida como um estudo quantitativo da herança intelectual perpetuada por meio de relacionamentos de orientação acadêmica entre professores (orientadores/supervisores) e seus alunos (orientados). Os relacionamentos de orientação acadêmica promovem a propagação de conhecimento científico por meio da interação entre orientador, com diferentes desempenhos em orientação, e seus orientados, que são influenciados pelas características de seus orientadores. Neste contexto, a GA fornece meios para mensurar e analisar essas interações formais de transmissão de conhecimento.

[1] http://lattes.cnpq.br. Último acesso em 1 de junho de 2018.
[2] http://plataforma-acacia.org. Último acesso em 10 de Junho de 2018.

2.1. TIPOLOGIA

A GA vem despertando um interesse crescente na comunidade científica, motivando o desenvolvimento de trabalhos com objetivos distintos. Os objetivos de sua aplicação podem ser utilizados como parâmetros para uma classificação geral, porém não definitiva, em cinco tipos de genealogia, a saber:

- *Honorífica* - estudo da descendência de um indivíduo de interesse (ou grupo de indivíduos) com o objetivo de homenageá-lo por meio de sua linhagem, evidenciando sua relevância devido à contribuição exercida na formação da comunidade em função de suas orientações acadêmicas;
- *Egocentrista* - tipo de genealogia que utiliza a ascendência de um indivíduo de interesse afim de identificar ancestrais ilustres e comprovar uma ligação, ainda que remota, entre ambos. Comumente a motivação para esse tipo de estudo está relacionada com curiosidade ou autopromoção;
- *Histórica* - utilizada como ferramenta complementar a estudos que visam a identificação de personagens históricos importantes para determinadas áreas do conhecimento. Permite contextualizar as relações entre os indivíduos identificados;
- *Paradigmática* - considerado como um dos tipos de GA mais inserido no contexto científico. Utiliza os relacionamentos orientador-orientado para estudar a extensão do conhecimento transmitido entre essas relações; e
- *Analítica* - devido ao crescimento do número de bases de dados genealógicos e o desenvolvimento de análises estatísticas, a genealogia analítica surge como ferramenta para a avaliação e, por vezes, predição de padrões entre os membros de comunidades acadêmicas que são objeto de estudo. Comumente os resultados desse tipo de estudo são possibilitados pela aplicação de métricas topológicas (Rossi *et al.*, 2017).

2.2. DADOS GENEALÓGICOS

Atualmente observa-se um crescente interesse pela GA que se reflete nos esforços empreendidos, por diferentes comunidades, em desenvolver projetos que busquem a captação de dados de genealogia (Elias *et al.*, 2016). As iniciativas identificadas geralmente disponibilizam um conjunto de dados genealógicos via *web* de forma semi-estruturada.

Exemplos de grandes conjuntos de dados, associados a projetos de genealogia acadêmica no nível de Doutorado, são:

- Mathematics Genealogy Project (*genealogy.math.ndsu.nodak.edu*);
- The Neurotree Project (*neurotree.org/neurotree*);
- The Academic Family Tree (*academictree.org*);
- Academic Genealogy Wiki (*phdtree.org*); e
- Software Engineering Academic Genealogy (*web.engr.illinois.edu/~taoxie/sefamily.htm*).

Além dos conjuntos de dados listados, que tratam de dados genealógicos internacionais, pode-se contar com o conjunto de currículos acadêmicos disponíveis na Plataforma Lattes e com o catálogo de teses e dissertações da CAPES[3].

Para todos os acadêmicos registrados na Plataforma Lattes são encontradas informações correspondentes à produção científica, técnica e artística, sendo que a informação sobre orientação/supervisão está disponível para acadêmicos com atuação na formação de alunos. Como apontado por Mena-Chalco e Cesar-Jr (2013), essa informação registrada não mantém um padrão e, em muitos casos, é incompleta e sujeita a erros (e.g., registros com nomes abreviados tanto do orientador quanto do orientado, e sem associação do número identificador - ID Lattes). Assim, a obtenção de uma estrutura hierárquica correta que represente o parentesco acadêmico torna-se um problema não-trivial.

Aqui é importante destacar que, a própria coleta de dados, sua estruturação e análise representam atuais oportunidades de pesquisa e aplicação que envolvem tanto as áreas de Ciência da Computação, Ciência da Informação, assim como a de Política Científica.

2.3. APLICAÇÕES

As pesquisas desenvolvidas no âmbito da GA podem ter seus objetivos orientados pela tipologia, a qual foi descrita anteriormente. Comumente, estas pesquisas baseiam-se em um grafo de genealogia que se caracteriza como principal objeto de estudos acadêmicos-genealógicos. Um ponto em comum na maioria dos trabalhos em GA é a possibilidade de se registrar, de forma perene, o contexto histórico de indivíduos e sua importância na formação da comunidade acadêmica na qual está inserido.

[3] http://catalogodeteses.capes.gov.br. Último acesso em 1 de Junho de 2018.

Na Figura 1 é apresentado um exemplo de grafo de GA onde são representados Matemáticos ilustres e seus relacionamentos de orientação acadêmica no nível de doutorado. Os vértices que ilustram os respectivos países de titulação dos Matemáticos representados se agrupam em famílias acadêmicas. Este tipo de estruturação é útil, por exemplo, para o estudo da formação destas famílias sob o ponto de vista geopolítico, por meio da observação da influência de estrangeiros na formação de famílias em um determinado país.

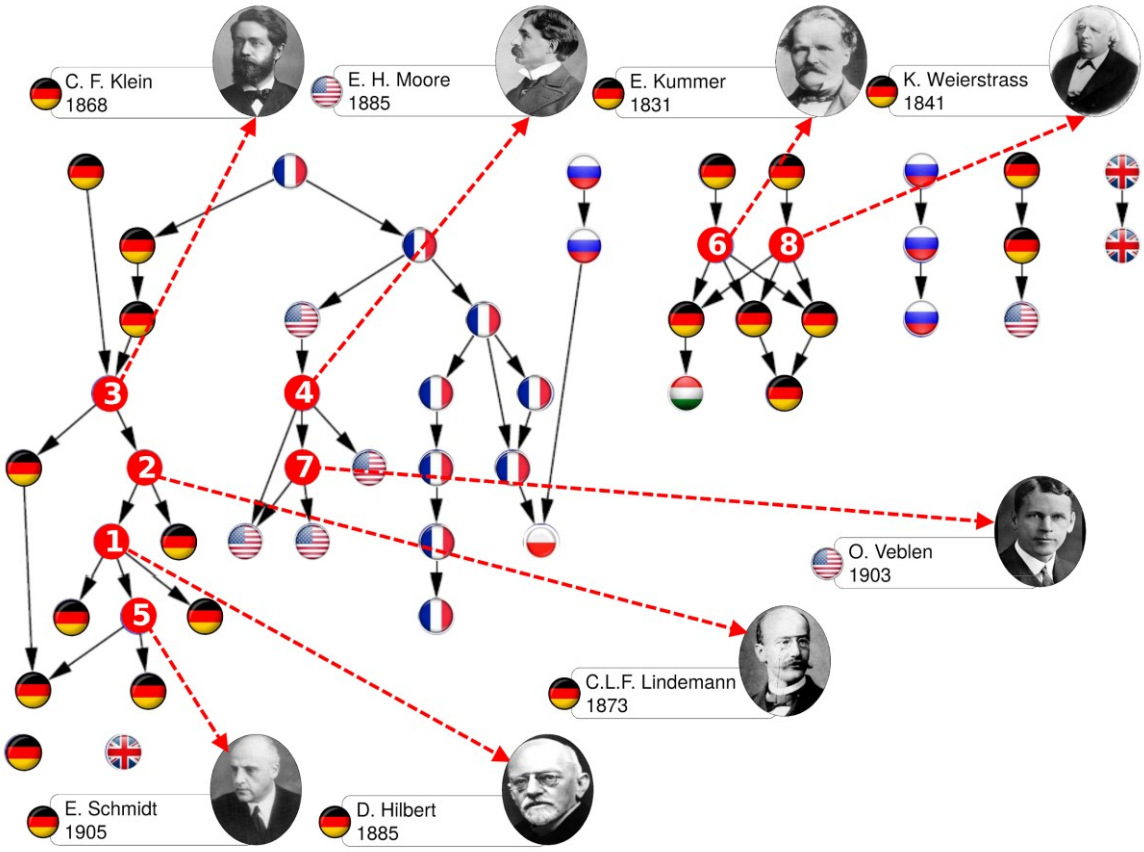

Figura 1: *Exemplo de grafo de GA que representa os relacionamentos de orientação, a nível de doutorado, entre Matemáticos ilustres. Cada vértice representa um pesquisador. Cada ligação a orientação formal no nível de doutorado. O conjunto de dados considerado corresponde ao Projeto* Mathematics Genealogy *coletado em 2017. Fonte: Elaborado pelos autores.*

Além do contexto histórico, é comum a utilização da GA para prestar homenagem a acadêmicos ilustres evidenciando sua importância para o desenvolvimento da área na qual está inserido. Na Figura 2 é apresentada a descendência acadêmica do Professor Cesare Mansueto Giulio Lattes, mais conhecido como César Lattes, renomado físico brasileiro que empresta seu nome ao maior

repositório acadêmico-curricular do Brasil, a Plataforma Lattes. No grafo, os vértices rotulados em verde representam os orientados diretos do Prof. Lattes, ou seja, seus filhos acadêmicos. Os vértices azuis claro completam a sua descendência que, no total, apresenta 425 acadêmicos (coleta de Nov. de 2017).

A identificação da descendência de um acadêmico pode ser útil para outras descobertas. Considerando atributos externos, como por exemplo a área de atuação dos acadêmicos pertencentes à descendência, é possível observar características como a divergência entre áreas de atuação do orientador e do orientado. Este tipo de padrão é especialmente interessante pois contraria o senso comum da perpetuação dos acadêmicos de uma mesma linha na mesma área de atuação (evidência de multidisciplinaridade).

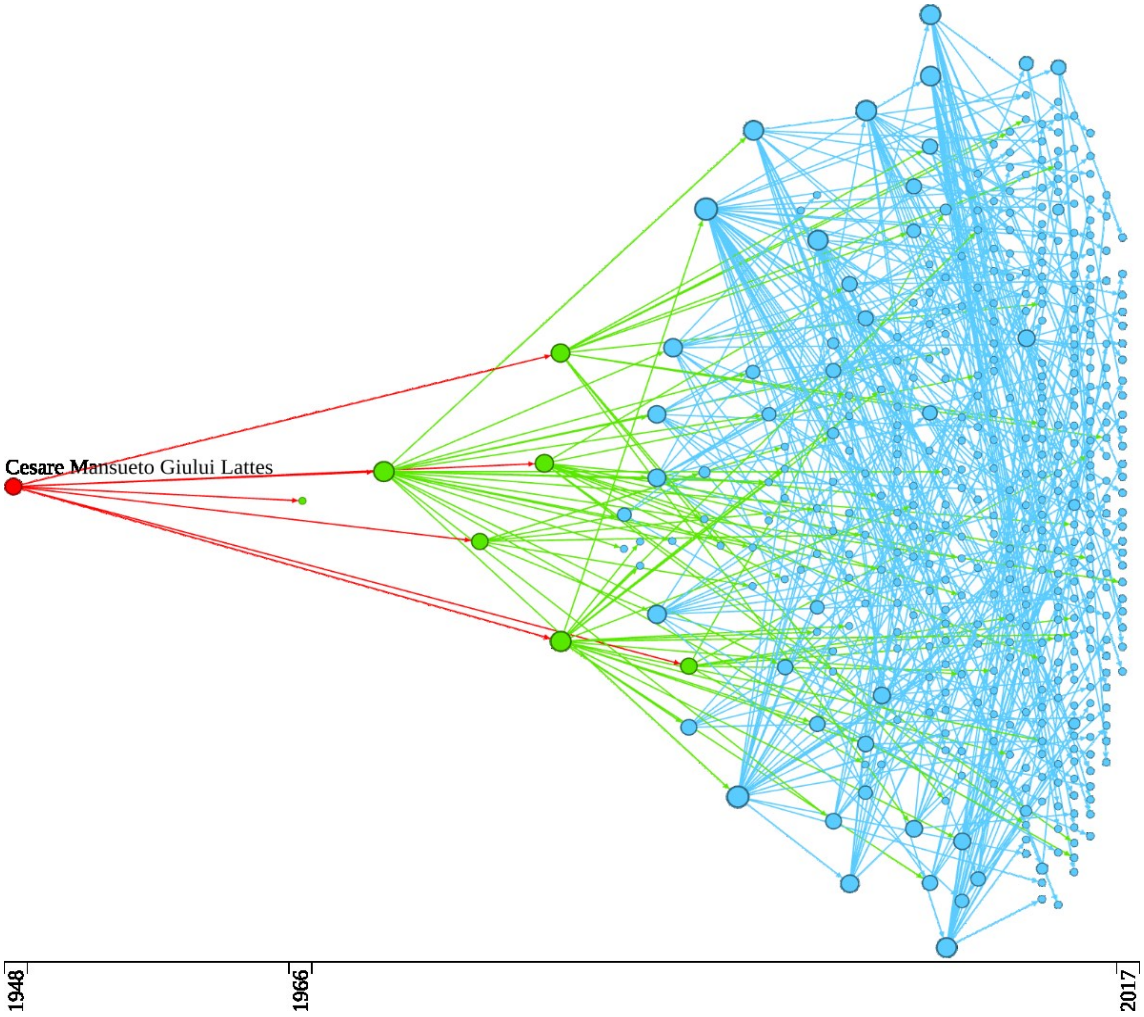

Figura 2: *Descendência acadêmica do físico brasileiro César Lattes (vértice vermelho). Os orientados diretos e indiretos são representados na cor verde e azul, respectivamente. O eixo horizontal representa o ano de titulação (Mestrado e/ou Doutorado) dos acadêmicos. Fonte: Elaborado pelos autores.*

As pesquisas alinhadas com a GA do tipo analítica são aquelas que buscam prover atributos quantitativos aos grafos de genealogia. Estes atributos podem descrever características sobre a configuração topológica do grafo, de modo que seja possível, por exemplo, classificar os acadêmicos com base em suas medidas topológicas (i.e., medidas que apenas observem as relações de orientação). Nesse contexto, na Figura 3 ilustramos algumas possíveis medidas topológicas que podem ser consideradas para a caracterização de acadêmicos. Essas medidas também podem ser tratadas como indicadores (métricas) quantitativos.

As medidas de Fecundidade e de Descendência representam, respectivamente, o número de orientados diretos e indiretos de um acadêmico de interesse (Figura 3a e 3b). Estas medidas são úteis para a verificação da produtividade/projeção dos acadêmicos, no contexto da formação de recursos humanos, e para evidenciar sua contribuição na formação da comunidade na qual está inserido. A medida Gerações indica o número de máximo de gerações que foram influenciadas por um determinado acadêmico (Figura 3c). Finalmente, a medida relacionada com o número de Folhas reflete a quantidade de acadêmicos não fecundos (aqueles acadêmicos que ainda não orientaram alunos) que existem na descendência de um acadêmico (Figura 3d).

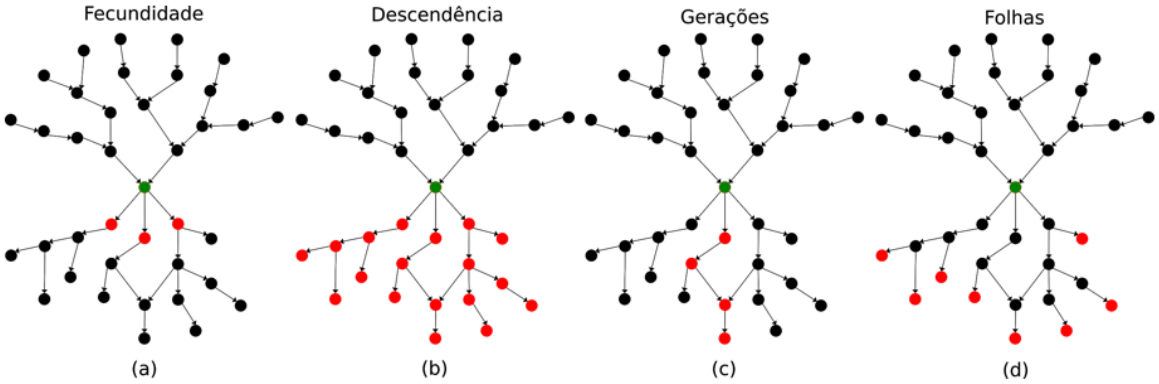

Figura 3: *Exemplos de vértices (em vermelho) associados às medidas topológicas sobre grafos de GA. As medidas refletem características quantitativas sobre a topologia do grafo e referem-se ao vértice rotulado em verde. Fonte: Elaborado pelos autores.*

As medidas topológicas são importantes para evidenciar características específicas dos grafos de GA. Estas características podem ser utilizadas para a identificação de padrões de desenvolvimento destas estruturas, relacionando as influências recebidas (medidas ascendentes) e exercidas (medidas descendentes), respectivamente. Neste contexto, acreditamos ser possível prever (e comparar) o

impacto que um determinado acadêmico exercerá em sua comunidade com base na observação de suas influências prévias. Isto é importante, por exemplo, para a definição de políticas de fomento à atividade acadêmico-científica.

3. A PLATAFORMA ACÁCIA

A história da Ciência no Brasil é muito recente, comparado a outros países com tradições seculares na formação de novos acadêmicos/pesquisadores que promovem a expansão das fronteiras do conhecimento pelo mundo. Por outro lado, a Plataforma Lattes é uma iniciativa ímpar em registrar as atividades de milhões de acadêmicos, a qual não encontra empreendimentos similares em outros países. Por meio de currículos acadêmicos, a Plataforma Lattes disponibiliza informações sobre formação, produção científica, participação em eventos, dentre outros registros acadêmicos relevantes.

Porém, no que se refere a informações de GA, os dados estão, em parte, incompletos, nos campos onde são descritos os ascendentes e descendentes. Isso impossibilita visualizar vínculos que existem na realidade mas não estão preenchidos na plataforma. Ainda, por manter uma organização curricular, a estrutura de dados é voltada para a exibição de currículos e não para a realização de análises genealógicas.

A Plataforma Acácia é uma iniciativa de um grupo de pesquisadores brasileiros e tem como objetivo fornecer informações acadêmico-genealógicas sobre a comunidade científica no Brasil. Os dados disponibilizados pela plataforma são resultado da prospecção de currículos oriundos da Plataforma Lattes, os quais são submetidos à processos computacionais que visam corrigir inconsistências e incompletudes observadas nos currículos, de forma a tornar as informações mais assertivas (Damaceno *et al.*, 2017).

Além de fornecer uma estrutura hierárquica, baseada nos relacionamentos de orientação acadêmica nos níveis de Mestrado e Doutorado, na forma de listas de orientadores e orientados, a Plataforma Acácia disponibiliza informações sobre: (i) Grande Área, (ii) Área, (iii) instituição de atuação do acadêmico, (iv) a maior titulação obtida por ele, e (v) o ano na qual a orientação ocorreu. Adicionalmente, a plataforma disponibiliza o resultado de três medidas sobre a GA: (i) Descendência, (ii) Fecundidade e (iii) Índice Genealógico[4].

[4] O Índice Genealógico de um acadêmico é o maior número *g* de descendentes que possuem, no mínimo, *g* descendentes cada um.

Para exemplificar, na Tabela 1 apresenta-se o número de mestres e doutores por geração dos acadêmicos identificados na Plataforma Acácia. Observa-se que existem ao todo 13 gerações, isto é, do acadêmico mais antigo identificado na plataforma ao mais recente, existem 12 níveis de separação. A maior parte dos acadêmicos estão entre as gerações 3 e 5, que concentram 652.576 mestres e doutores, representando 58,71% do total. Note que, esses dados mostram que a Ciência no Brasil ainda é jovem, dado que mais da metade dos cientistas estão em gerações relativamente recentes.

Tabela 1. Número de mestres e doutores por geração. Dados da Plataforma Acácia. *Fonte: Elaborado pelos autores.*

Geração	**Número de acadêmicos**		
	Mestres	**Doutores**	**Total**
1	70.012	39.779	109.791
2	41.947	36.600	78.547
3	144.735	78.658	223.393
4	197.980	100.197	298.177
5	172.551	73.929	246.480
6	79.756	28.163	107.919
7	24.318	7.179	31.497
8	6.569	1.975	8.544
9	2.373	1.022	3.395
10	1.601	900	2.501
11	936	226	1.162
12	125	9	134
13	4	0	4
Total	742.907	368.637	1.111.544

4. CONSIDERAÇÕES FINAIS

A orientação acadêmica é parte das atividades que integram um ambiente de pesquisa. A documentação histórica dessas atividades possibilita a criação de métodos para compreender as relações de interação da comunidade que constitui o ambiente: os

professores orientadores, os alunos orientados, as instituições onde se formaram os alunos, as áreas mais proeminentes/carentes na formação, dentre outras características.

Analisar as relações de orientações, através de genealogia acadêmica, pode revelar informações relevantes respeito da evolução da produção científica de um país. Conhecer o cenário atual da área de pesquisa acadêmico-científica e como chegou-se a este cenário podem contribuir para a tomada de decisões mais assertivas de futuro para a comunidade.

Acreditamos que o aprofundamento na pesquisa de modelos alternativos (i.e., novos olhares à ciência) contribuirá na criação de insumos importantes para responder questões, pouco exploradas atualmente, que permeiam a CT&I.

REFERÊNCIAS

G. T. Chao, P. Walz, e P. D. Gardner. Formal and informal mentorships: A comparison on mentoring functions and contrast with nonmentored counterparts. *Personnel psychology*, 45(3), 619-636, 1992.

B. Cronin e C. R. Sugimoto. Beyond bibliometrics: *Harnessing multidimensional indicators of scholarly impact*. MIT Press, 2014.

R. J. P. Damaceno, L. Rossi e J. P. Mena-Chalco. Identificação do grafo de genealogia acadêmica de pesquisadores: Uma abordagem baseada na Plataforma Lattes. In *Proceedings of the 32nd Brazilian Symposium on Databases,* 1:76-87, 2017.

M. C. Elias, L. M. Floeter-Winter, e J. P. Mena-Chalco. The dynamics of brazilian protozoology over the past century. *Memórias do Instituto Oswaldo Cruz*, 111:67–74, 01 2016.

F. Gargiulo, A. Caen, R. Lambiotte, e T. Carletti, T. The classical origin of modern mathematics. *EPJ Data Science*, 5(1), 26, 2016.

J. Lane. Let's make science metrics more scientific. Nature, 464(7288):488-489, 2010.

R. D. Malmgren, J. M. Ottino, e L. A. N. Amaral. The role of mentorship in protégé performance. *Nature*, 465(7298), 622, 2010.

J. P. Mena-Chalco e R. C. Junior. Prospecção de dados acadêmicos de currículos lattes através de scriptlattes. In *Bibliometria e Cientometria: reflexões teóricas e interfaces*, pages 109–128. 2013.

L. Rossi, I. L. Freire, e J. P. Mena-Chalco. Genealogical index: A metric to analyze advisor–advisee relationships. *Journal of Informetrics*, 11(2):564–582, 2017.