

Árvore da Ciência: Uma Plataforma para Exploração da Genealogia Acadêmica Brasileira

João M. M. C. Cota, Alberto H. F. Laender, Raquel O. Prates¹

¹Departamento de Ciência da Computação
Universidade Federal de Minas Gerais (UFMG)
31270-901 – Belo Horizonte – MG – Brasil
{joaocota,laender,rprates@dcc.ufmg.br}

Abstract. *Identifying and studying the formation of researchers over the years is a challenging task, as current repositories of theses and dissertations are cataloged in a decentralized manner in different digital libraries. In this paper, we take a step forward towards building a large repository to record the Brazilian academic genealogy. For this, we collected data from the Lattes platform, maintained by the Brazilian National Council for Scientific and Technological Development (CNPq), and developed a user-oriented platform to generate the academic genealogy trees of Brazilian researchers from them, also providing additional data on the main properties of such trees. Our effort has identified interesting aspects related to the academic career of the Brazilian researchers, which highlight the importance of cataloging their academic genealogy trees.*

Resumo. *Identificar e estudar a formação dos pesquisadores ao longo dos anos é uma tarefa desafiadora, pois os atuais repositórios de teses e dissertações são catalogados de forma descentralizada em diferentes bibliotecas digitais. Neste artigo, damos um passo à frente na construção de um grande repositório para registro da genealogia acadêmica brasileira. Para isso, coletamos dados da plataforma Lattes, mantida pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), e desenvolvemos uma plataforma voltada a usuários finais para gerar as árvores genealógicas acadêmicas dos pesquisadores brasileiros a partir dos dados coletados, fornecendo também dados adicionais sobre as principais propriedades dessas árvores. Nosso esforço identificou aspectos interessantes relacionados à trajetória acadêmica dos pesquisadores brasileiros, que destacam a importância de se catalogar as suas árvores genealógicas acadêmicas.*

1. Introdução

Uma maneira relevante de caracterizar o impacto de um pesquisador é por meio de sua genealogia acadêmica [Damaceno et al. 2019, Hirshman et al. 2016, Malmgren et al. 2010]. O termo genealogia acadêmica ou científica refere-se à organização de uma árvore genealógica de cientistas e estudiosos estruturada de acordo com as relações de orientação acadêmica (mestrado e doutorado) ao longo de suas respectivas carreiras [Jackson 2007]. Apesar dos inúmeros trabalhos que procuram analisar a dinâmica da produção científica em termos das redes de colaboração científica [Barabâsi et al. 2002, Digiampietri et al. 2019, Mena-Chalco et al. 2014, Newman 2004, Pessoa Junior et al. 2020], ainda são poucos aqueles que visam estudar a genealogia acadêmica e o processo de mentoria de modo geral.

Dentre os trabalhos pioneiros, Chang [2003] apresenta um estudo sobre a genealogia acadêmica dos físicos americanos, enquanto Jackson [2007] reporta os primeiros esforços

realizados para a construção da árvore genealógica da Matemática e que deu origem ao *Mathematics Genealogy Project*. Vale salientar ainda que a busca por uma representação ou até mesmo a investigação sobre as relações orientador-orientado não são recentes. Tais iniciativas têm sido realizadas por comunidades acadêmicas específicas no sentido de analisar, classificar e documentar essas relações de orientação com o objetivo de prover uma visão específica de como se dá a formação de novos pesquisadores [Hart and Cossuth 2013, Hirshman et al. 2016, Malmgren et al. 2010, Tuesta et al. 2015].

A construção de árvores genealógicas acadêmicas apresenta desafios variados tais como a coleta dos dados que muitas vezes não estão disponíveis em um formato padronizado, o tratamento dos dados coletados para que possam ser aplicados para a geração das árvores e, particularmente, a desambiguação dos nomes dos pesquisadores [Ferreira et al. 2012]. Assim, neste artigo, apresentamos a plataforma *Árvore da Ciência* que gera as árvores genealógicas acadêmicas dos pesquisadores brasileiros a partir de dados extraídos de seus currículos cadastrados na plataforma Lattes, permitindo não só a visualização e exploração dessas árvores, mas também a respectiva análise a partir de métricas que expressam indicadores quantitativos sobre a atuação desses pesquisadores no processo de formação de mestres e doutores. Vale ainda ressaltar que a plataforma *Árvore da Ciência* possui uma interface trilingue, o que permite a seus usuários acessá-la em português, inglês ou espanhol¹.

O restante deste artigo está organizado da seguinte forma. A Seção 2 descreve algumas iniciativas similares, enquanto a Seção 3 apresenta uma visão geral da plataforma desenvolvida. A seguir, a Seção 4 apresenta um breve exemplo de sua utilização e a Seção 5 reporta resultados obtidos na avaliação realizada com usuários típicos da plataforma. Finalmente, a Seção 6 apresenta algumas conclusões e sugestões para trabalhos futuros.

2. Iniciativas Similares

Além de iniciativas pessoais ou relacionadas a departamentos e sociedades acadêmicas, algumas plataformas têm sido criadas especificamente com o intuito de possibilitar a exploração da genealogia acadêmica de forma mais ampla. Nesta seção, apresentamos algumas dessas plataformas que contam com grande número de pesquisadores cadastrados e têm se mostrado de grande relevância para a área ou comunidade científica que representam.

Conforme já mencionado, o *Mathematics Genealogy Project*² [Jackson 2007] é um dos mais antigos esforços no sentido de prover uma plataforma para visualização das árvores genealógicas acadêmicas de uma comunidade científica específica. Como descrito em sua página, essa plataforma tem como objetivo compilar dados sobre os matemáticos de todo o mundo, além de permitir que instituições que realizam pesquisas na área de matemática cadastrem seus pesquisadores para que possam ter os seus dados divulgados amplamente. Atualmente a plataforma conta com dados de mais de 253 mil matemáticos das mais diversas instituições ao redor do mundo.

A plataforma *The Academic Family Tree*³ originou-se de outra iniciativa similar, denominada *Neurotree*⁴. A versão original, iniciada em janeiro de 2005, tinha como objetivo documentar as relações de orientação na área de Neurociência [David and Hayden 2012],

¹<https://sciencetree.net>, acesso em Abril de 2021.

²<https://www.genealogy.math.ndsu.nodak.edu>, acesso em Abril de 2020.

³<https://academicctree.org>, acesso em Abril de 2020.

⁴<http://neurotree.org>, acesso em Abril de 2020.

exibindo-as no formato de uma árvore genealógica. Logo após a criação da *Neurotree*, os seus mantenedores perceberam que a Neurociência possuía relações com outras áreas da ciência e que essas relações poderiam ser tão importantes quanto as da própria Neurociência, fazendo com que novas áreas fossem incorporadas à plataforma que passou a ser denominada *The Academic Family Tree*.

A Universidade do Estado da Flórida também apresenta uma iniciativa interessante que é o sítio da árvore genealógica acadêmica dos meteorologistas americanos⁵, mantido pela Sociedade Meteorológica Americana. Nesse sítio é possível explorar as árvores de diferentes comunidades de meteorologistas americanos. Além disso, ele permite que um pesquisador solicite a sua inclusão na árvore por meio de uma função específica denominada *Add Yourself/Someone Else*. Entretanto, o sítio não oferece uma forma automática de geração ou visualização das árvores, de modo que elas são compiladas de tempos em tempos e disponibilizadas em formato de imagem para os seus usuários.

Uma iniciativa pioneira no contexto brasileiro é a plataforma Acácia⁶ que visa documentar as relações formais de orientação dos nossos programas de pós-graduação. A partir de dados coletados da plataforma Lattes do CNPq são construídos grafos de genealogia acadêmica, em que cada vértice representa um pesquisador e cada aresta uma relação de orientação concluída entre o orientador e seu orientado. A plataforma Acácia foi lançada em Abril de 2018 e é mantida pelo grupo de pesquisa liderado pelo Prof. Jesús Mena-Chalco da UFABC. Detalhes do seu algoritmo para geração dos grafos de genealogia acadêmica e uma análise quantitativa de seus dados foram apresentados por Damaceno et al. [2019].

Finalmente, a plataforma *The Gold Tree*⁷ tem por finalidade gerar uma visualização da genealogia acadêmica dos pesquisadores brasileiros que possuem seus currículos cadastrados na plataforma Lattes do CNPq [Madeira et al. 2019]. Segundo dados disponíveis no próprio sítio da plataforma, foram utilizados dados de mais de 570 mil dissertações e teses para gerar as respectivas árvores. Entretanto, ela não apresenta qualquer análise específica relativa às árvores geradas, possibilitando apenas a exploração visual delas. A plataforma foi desenvolvida pelo Grupo de Pesquisa em Gerenciamento de Informações da Universidade Federal do Rio Grande.

De modo a possibilitar uma comparação geral entre as plataformas retratadas, a Tabela 1 apresenta as principais características de cada uma delas, considerando a forma de representação da genealogia, a forma de visualização dos dados, o número de métricas disponíveis e o foco da plataforma considerando a sua cobertura. Como podemos observar, quatro dessas plataformas representam a genealogia acadêmica de forma tabular, sendo que a plataforma *The Academic Family Tree* também provê uma representação gráfica semelhante a uma árvore. A plataforma *The Gold Tree* também provê uma representação gráfica das árvores. Entretanto, a visualização dos dados nessa plataforma, em razão da representação adotada, é bastante limitada e não permite uma exploração mais ampla das relações de orientação. Em relação às métricas, a plataforma *The Academic Family Tree* é a que provê o maior número delas (15), sendo que duas das demais plataformas, a *AMS Academic Family Tree* e a *The Gold Tree*, não provêm nenhuma métrica. Finalmente, apenas duas das plataformas, a *The Academic Family Tree* e o *Mathematics Genealogy Project*, possuem um foco global, já que ambas se utilizam dos dados fornecidos pelos próprios pes-

⁵<http://moe.met.fsu.edu/familytree>, acesso em Abril de 2020.

⁶<http://plataforma-acacia.org>, acesso em Abril de 2020.

⁷<http://thegoldtree.c3.furg.br>, acesso em Abril de 2020.

quisadores, ou por suas respectivas instituições ao redor do mundo, para manter os seus repositórios atualizados, ao contrário das plataformas *Acácia* e *The Gold Tree*, por exemplo, que fazem uso dados de um repositório local, a Plataforma Lattes, para gerar as árvores dos pesquisadores brasileiros.

Plataforma	Representação da Genealogia (Tabular ou Gráfica)	Visualização dos Dados (Textual ou Gráfica)	Número de Métricas Disponíveis	Foco da Plataforma
Mathematics Genealogy Project	Tabular	Textual	2	Global
The Academic Family Tree	Ambas	Ambas	15	Global
AMS Academic Family Tree	Tabular	Textual	Nenhuma	Nacional
Plataforma Acácia	Tabular	Textual	7	Nacional
The Gold Tree	Gráfica	Gráfica	Nenhuma	Nacional

Tabela 1. Principais características das cinco iniciativas analisadas.

3. Descrição Funcional da Plataforma *Árvore da Ciência*

Nesta seção apresentamos uma visão geral do funcionamento da plataforma *Árvore da Ciência*, descrevendo o seu fluxo de dados e a sua arquitetura interna. Vale ressaltar que a versão atual da plataforma originou-se de um estudo pioneiro sobre a genealogia acadêmica brasileira realizado por Dores et al. [2017] e que disponibilizou uma ferramenta que permitia apenas visualizar as árvores individualmente.

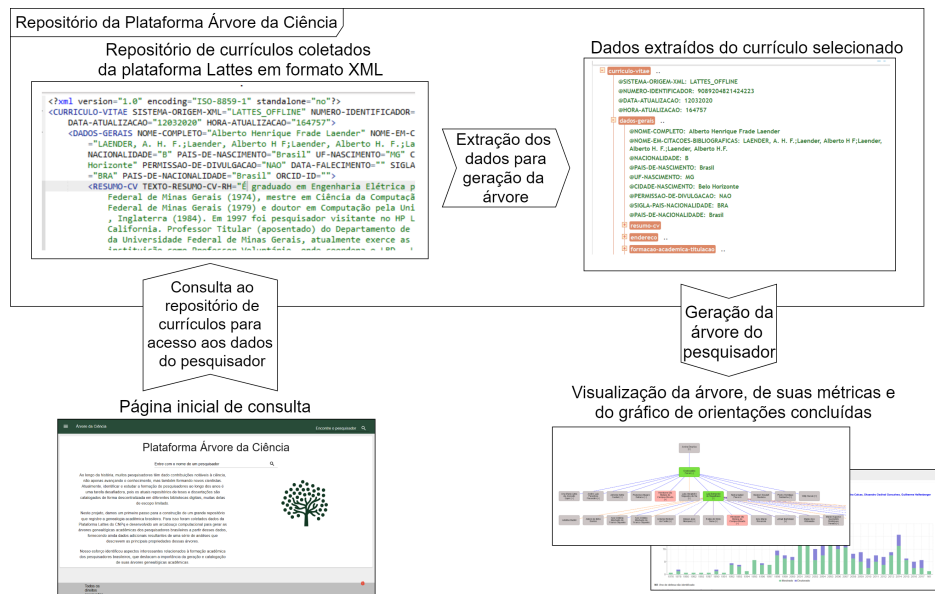


Figura 1. Fluxo de Dados da Plataforma *Árvore da Ciência*.

A Figura 1 descreve o fluxo de dados proposto para geração das árvores genealógicas acadêmicas na plataforma *Árvore da Ciência*. Como pode ser observado, a partir do nome de um pesquisador é feita uma consulta ao repositório de currículos coletados da plataforma Lattes para extração dos dados necessários para geração de sua respectiva árvore genealógica acadêmica. Extraídos os dados, é gerada a árvore do pesquisador com suas respectivas métricas. Finalmente, a árvore gerada pode ser visualizada juntamente com dados específicos sobre o pesquisador, como nome, formação acadêmica e instituição de vínculo, métricas que expressam diferentes aspectos de sua genealogia acadêmica e um gráfico temporal de suas orientações concluídas conforme registradas na plataforma Lattes.

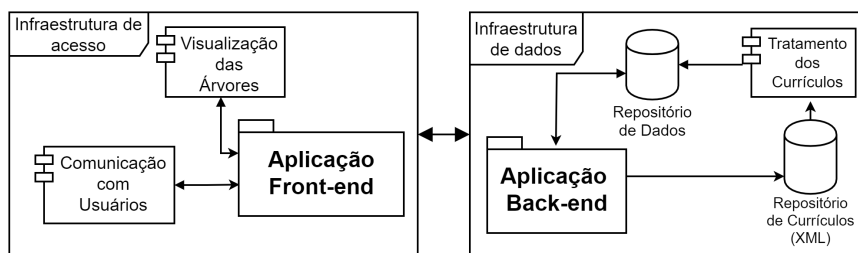


Figura 2. Arquitetura Interna da Plataforma Árvore da Ciência.

Conforme descrito na Figura 2, a plataforma é composta de duas aplicações principais: *Front-end* e *Back-end*. A aplicação *Front-end* estabelece a interface com os usuários, processando suas consultas e estruturando os dados que poderão ser acessados durante o uso da plataforma, por exemplo, ao navegar pelas árvores genealógicas acadêmicas visualizando os dados dos respectivos pesquisadores ou quando desejarem ter acesso aos valores de determinadas métricas que expressam a estrutura de uma árvore específica. Por outro lado, a aplicação *Back-end* intercepta as solicitações da aplicação *Front-end* e acessa o Repositório de Dados da plataforma para calcular, por exemplo, os valores das métricas genealógicas fornecidas para um pesquisador, e gerar o seu gráfico de supervisões concluídas. Vale ressaltar que esses dados são mantidos nos repositórios da própria plataforma e ficam disponíveis publicamente para acesso pelos usuários por meio da World Wide Web.

A interface de usuário da plataforma Árvore da Ciência foi concebida a partir de um *design* minimalista, com poucos elementos visuais e textos objetivos. Essas características auxiliam no entendimento dos dados apresentados aos usuários, além de diminuir o tempo de navegação. Além disso, como a plataforma disponibiliza várias métricas que permitem quantificar diferentes aspectos da genealogia acadêmica dos pesquisadores, é importante que a sua interface permita aos seus usuários obter uma visão mais analítica do desempenho de cada um deles em relação ao papel que exercem como orientadores. Assim, ela foi implementada empregando uma linguagem e bibliotecas de programação atuais como *.Net Core* (escrita em C#) que é utilizada para realizar as buscas dos pesquisadores e gerar as suas métricas, *ReactJs*⁸ utilizada na construção da interface com usuário, e *VisJs*⁹ utilizada para desenhar as árvores genealógicas acadêmicas e capturar as interações dos usuários com as árvores, ambas desenvolvidas em *JavaScript*. A forma de acesso à plataforma se dá por meio de um dispositivo eletrônico (computador, *smartphone* ou *tablet*) conectado à Internet e com acesso a um navegador.

Conforme já mencionado (ver Figura 1), as árvores genealógicas acadêmicas dos pesquisadores são geradas a partir do pré-processamento de seus currículos em formato XML. Para isso é utilizado o algoritmo proposto por Dores et al. [2017], escrito em *Python*. Acessado o repositório de currículos Lattes, é extraído, para cada pesquisador, os dados referentes às suas respectivas orientações (título das respectivas teses e dissertações, instituição onde ocorreram as orientações e o ano de concessão dos respectivos títulos). Com base nesses dados, é realizada a desambiguação dos nomes dos autores para a correta geração das respectivas relações de orientação. Após esse pré-processamento, as relações de orientação geradas são armazenadas no Repositório de Dados na forma de um grafo dirigido, para que sejam posteriormente visualizadas na forma de árvores genealógicas acadêmicas e para a geração dos valores de suas métricas correspondentes.

⁸<https://reactjs.org/>, acesso em Abril de 2021.

⁹<https://visjs.org/>, acesso em Abril de 2021.

4. Exemplo de Uso da Plataforma

Para ilustrar o funcionamento da plataforma Árvore da Ciência, nesta seção apresentamos um breve exemplo de sua utilização envolvendo a genealogia acadêmica do Prof. Crowdowaldo Pavan (1919-2009), um dos mais renomados geneticistas brasileiros. Graduado em 1941 em História Natural pela Faculdade de Filosofia, Ciências e Letras da Universidade de São Paulo (USP). O Prof. Pavan obteve o doutorado, também pela USP, em 1944 na área de Genética, investigando a adaptação ambiental e a evolução dos peixes cegos das cavernas de Iporanga, no interior de São Paulo. Ele foi também um dos responsáveis por introduzir na biologia o estudo da citogenética da mosca *Rhynchosciara americana*, conhecida por seus cromossomos gigantes, facilitando assim a determinação de locus gênicos. Esse trabalho contribuiu para demonstrar que a estrutura dos genes e dos cromossomos é alterada por infecções. Além de suas contribuições científicas como docente da USP, o Prof. Pavan ocupou vários cargos administrativos importantes ao longo de sua carreira, entre eles o de Presidente do CNPq entre 1996 e 2000.

O currículo do Prof. Pavan foi incluído na plataforma Lattes em 2008, pouco antes de seu falecimento em 2009, como uma homenagem pelo seu papel na consolidação da comunidade acadêmica brasileira, contendo apenas os dados básicos de sua formação acadêmica e uma única orientação de doutorado. Apesar disso, como será mostrado a seguir, a sua árvore na plataforma Árvore da Ciência inclui 6327 descendentes diretos e indiretos. Para a sua geração, foram utilizados não só os dados de seu próprio currículo, mas também de outros 10 currículos coletados da plataforma Lattes, cujos pesquisadores indicaram o Prof. Pavan como orientador, totalizando assim 11 orientações.

A Figura 3 a seguir mostra a página do Prof. Pavan na plataforma Árvore da Ciência que exibe os três primeiros níveis de sua árvore genealógica acadêmica com seus nodos expandidos na cor verde e com a aba de métricas aberta logo abaixo. Dentre as métricas que expressam diretamente o esforço de formação de um pesquisador, ressaltamos a largura da árvore (quantidade de descendentes diretos), a fertilidade (quantidade de descendentes diretos com pelo menos uma orientação) e o número médio de orientações concluídas por ano que, no caso do Prof. Pavan, são iguais a 11, 9 e 0,3, respectivamente. Entretanto, embora tais métricas possam parecer pouco significativas, a importância do Prof. Pavan na formação de novos pesquisadores em sua área de atuação é claramente indicada pela profundidade de sua árvore que é igual a 8 e pelo seu expressivo número de descendentes diretos e indiretos, acima de 6300 pesquisadores. Por fim, o gráfico de orientações concluídas por ano é uma forma complementar de expressar o esforço de orientação de um pesquisador por meio de um gráfico de barras que indica o total de orientações de doutorado (cor azul) e mestrado (cor verde) por ele concluídas a cada ano.

5. Avaliação com Usuários

O processo de desenvolvimento de software envolve geralmente uma série de etapas distintas, tais como análise de requisitos, projeto, implementação, teste e produção [Pressman and Maxim 2019]. Entre os diferentes tipos de teste de software, existem aqueles que visam analisar a perspectiva dos usuários em relação a um sistema ou aplicativo desenvolvido, bem como aqueles que avaliam a sua funcionalidade e o projeto geral. No caso da plataforma Árvore da Ciência, duas avaliações distintas com potenciais usuários foram realizadas para coletar dados relacionados à percepção da utilidade da plataforma proposta e à facilidade de interagir com ela, a fim de identificar dificuldades existentes e fornecer sugestões para melhorias futuras conforme proposto por Cota [2020].



Figura 3. Árvore do Prof. Crodowaldo Pavan expandida até o seu terceiro nível para mostrar um caminho específico de sua linhagem acadêmica (nós verdes). As linhas azuis significam orientações de doutorado e as linhas laranjas orientações de mestrado. A aba abaixo mostra as métricas relacionadas à sua árvore e o seu gráfico de supervisões concluídas por ano.

A primeira avaliação foi realizada por meio de um questionário e teve como objetivo efetuar uma coleta de dados mais ampla, focada principalmente na obtenção de indicadores quantitativos de aspectos relacionados à usabilidade da plataforma. A segunda avaliação envolveu a observação dos participantes interagindo com a plataforma e uma entrevista sobre suas respectivas opiniões e experiências ao utilizá-la. Essas duas técnicas complementares foram usadas para permitir uma análise ampla da perspectiva dos usuários em relação à plataforma *Árvore da Ciência*, permitindo identificar tanto indicadores de como ela seria recebida pela comunidade científica, como também aspectos relacionados à sua interface e à sua forma de interação que poderiam ser melhorados [Lazar et al. 2017].

Para a avaliação por questionário, foi enviado um convite de participação para uma lista de 5834 e-mails de pesquisadores que atuaram no programa dos Institutos Nacionais de Ciência e Tecnologia (INCTs) em diversas instituições do Brasil. Ao final, 286 pesquisadores responderam ao questionário. Destes, 179 (62,5%) se autodeclararam como do gênero

masculino, 103 (36,0%) do gênero feminino e 4 (1,5%) optaram por não declarar o gênero. Em relação às idades dos respondentes, a maioria deles, 129 (45,1%), ficou concentrada no grupo de idades entre 50 e 59 anos.

O questionário era composto de 13 questões em escala Likert de cinco pontos sobre a plataforma, organizadas em quatro grupos: Grupo 1 (Qualidade da apresentação dos dados e das métricas - 3 questões): *A visualização das métricas referentes ao pesquisador é clara, As métricas fornecidas pela plataforma Árvore da Ciência sobre a atuação do pesquisador como orientador são úteis, Quando um pesquisador é selecionado os dados sobre ele são apresentados de forma clara*; Grupo 2 (Aspectos visuais - 4 questões): *A plataforma é esteticamente agradável, A visualização das árvores é clara, As cores utilizadas nas páginas e nas legendas não causam confusão, As opções de filtros disponíveis são claras*; Grupo 3 (Facilidade de uso - 3 questões): *A plataforma é de fácil utilização, Foi fácil aprender a utilizar a plataforma, A plataforma proporciona uma navegação intuitiva*; Grupo 4 (Utilidade da plataforma - 3 questões): *A plataforma é útil no seu dia-a-dia como pesquisador, A visualização da relação orientador-orientado por meio de uma árvore genealógica é útil, A exploração das árvores genealógicas acadêmicas é útil para se ter uma ideia da atuação acadêmica dos pesquisadores*.

Por uma questão de espaço, apresentamos aqui os resultados obtidos na nossa avaliação de forma agregada. Dentre as 13 afirmativas apresentadas, para 11 delas os respondentes, na sua maioria, indicaram uma atitude positiva (i.e., marcaram as opções superiores na escala Likert). Assim, para o Grupo 1 63% dos respondentes concordaram com as afirmativas, para o Grupo 3 67%, para três das quatro afirmativas do Grupo 2 56% e para duas das três afirmativas do Grupo 4 75%. Apenas as afirmativas sobre a clareza dos filtros (Grupo 2) e sobre a utilidade da plataforma no dia-a-dia do pesquisador (Grupo 4) não obtiveram concordância da maioria dos respondentes. Em relação às opções de filtros, o resultado indicou um problema no elemento da interface utilizado para representá-las, seja com a sua localização na tela ou com a representação gráfica utilizada para essa opção, o que foi prontamente corrigido.

Para a segunda etapa de avaliação, foram convidados participantes que possuíssem algum envolvimento com pesquisa e orientação na pós-graduação. Assim, foram enviados convites para 27 pesquisadores de diversas áreas do conhecimento e diferentes níveis da carreira acadêmica (recém contratados, seniores e titulares), dos quais sete responderam aceitando o convite. As sessões de avaliação eram individuais e foram conduzidas presencialmente entre fevereiro e março de 2020. Cada sessão envolvia a utilização da plataforma tanto para a realização de algumas tarefas solicitadas, quanto para exploração livre da plataforma pelos participantes. Ao final das sessões, os participantes eram entrevistados sobre a sua visão e experiência com a plataforma.

Em relação aos pontos positivos levantados pelos participantes da avaliação, destacam-se as métricas, como a fertilidade, e o gráfico de orientações por ano. Os dados complementares extraídos do currículo Lattes e a facilidade de uso da plataforma foram outros pontos relevantes mencionados pelos participantes. Em relação a pontos de melhoria da plataforma, foi mencionada a necessidade de prover instruções de uso da plataforma, bem como melhorar a visualização de alguns componentes, como, legendas e filtros. A inclusão de uma explicação sobre as métricas e a melhoria, ou mesmo mudança, no formato de representação das árvores geradas (particularmente no caso daquelas com muitos nodos) foram outras sugestões apresentadas.

A conclusão obtida a partir dessas duas avaliações é que, embora possua pontos a serem melhorados, a plataforma *Árvore da Ciência* é de fácil utilização e se apresenta como uma ferramenta relevante para analisar a formação de pesquisadores no Brasil, conforme resultados experimentais reportados por Cota [2020]. É importante também salientar que algumas limitações levantadas pelos usuários já foram resolvidas, como a inclusão de uma página de explicação das métricas e o aumento do tamanho da fonte dos textos de ajuda. Entretanto, ainda há alguns aspectos específicos da plataforma que precisam ser melhorados como, por exemplo, a visualização de árvores com muitos níveis e grande número de nodos (orientados), e o posicionamento de determinados elementos na tela, como os filtros utilizados em alguns tipos de consulta.

6. Conclusões e Trabalhos Futuros

Nos últimos anos, tem havido um grande esforço visando o desenvolvimento de plataformas e aplicações específicas voltadas para a construção e manutenção de árvores genealógicas acadêmicas de pesquisadores das mais diversas áreas do conhecimento [Damaceno et al. 2019, David and Hayden 2012, Dores et al. 2017, Hart and Cossuth 2013, Jackson 2007, Madeira et al. 2019]. Em sua maioria, essas iniciativas adotam uma estratégia de colaboração coletiva, em que tais plataformas crescem com dados inseridos por diversos colaboradores, o que nem sempre provê uma clara visão de como se dá a formação de uma determinada comunidade em razão de um eventual desbalanceamento na cobertura dos dados utilizados. No caso da plataforma *Árvore da Ciência*, são utilizados dados de orientações acadêmicas extraídos dos currículos dos pesquisadores com título de doutor, coletados da plataforma Lattes no formato XML, o que garante uma cobertura mais abrangente dessas orientações. Além disso, diferentemente de outras iniciativas, como o *Mathematics Genealogy Project* e a *AMS Academic Family Tree*, que se limitam apenas a uma área do conhecimento, a plataforma *Árvore da Ciência* permite construir árvores genealógicas acadêmicas de pesquisadores de diversas áreas do conhecimento.

A plataforma *Árvore da Ciência*, entretanto, não se limita apenas a exibir as árvores geradas para os pesquisadores, mas também possibilita caminhar por elas à medida que são expandidas com os nós de seus descendentes acadêmicos. Essa é uma diferença fundamental quando comparada diretamente com as plataformas *Acácia* e *The Gold Tree*, que também utilizam dados extraídos da plataforma Lattes, uma vez que a primeira não provê uma visualização gráfica das respectivas árvores, enquanto a segunda apenas possibilita a visualização individual da árvore de cada pesquisador. A plataforma *Árvore da Ciência* provê ainda métricas específicas relacionadas às árvores geradas e às orientações correspondentes, além de exibir dados de interesse extraídos dos currículos Lattes dos respectivos pesquisadores. Assim, por meio dela é possível visualizar o papel dos nossos pesquisadores na formação acadêmica brasileira, bem como dos seus orientados a partir de suas próprias orientações. Muitos desses pesquisadores, como o Prof. Crodowaldo Pavan cuja árvore genealógica acadêmica é apresentada na Seção 4, são pioneiros em suas respectivas áreas de pesquisa, de modo que tiveram um papel preponderante na formação de importantes gerações de pesquisadores no Brasil.

Assim sendo, a plataforma *Árvore da Ciência*, disponibilizada por meio de um portal *online*, permite a visualização, navegação e exploração das árvores genealógicas acadêmicas dos pesquisadores brasileiros, além de prover métricas que quantificam o impacto desses pesquisadores como orientadores de pós-graduação, apresentando ainda características es-

pecíficas que tornam a sua utilização mais marcante quando comparada a outras iniciativas similares de acordo com os quesitos da Tabela 1. Além disso, vale ressaltar a sua avaliação realizada com dois grupos de usuários, o primeiro deles composto por 286 pesquisadores que responderam a um questionário, e o segundo, composto por sete pesquisadores com experiência em pesquisa e orientação de pós-graduação, que aceitaram participar de uma avaliação presencial por meio de uma entrevista, durante a qual eles executaram uma série de tarefas pré-estabelecidas utilizando a plataforma. Os resultados dessas duas avaliações possibilitaram não só validar as principais facilidades oferecidas pela plataforma, como também identificar novos recursos que poderiam ser adicionados a ela.

Em relação a trabalhos futuros, faz-se necessário prover a atualização automática do repositório de dados a partir de coletas periódicas realizadas na plataforma Lattes, de modo a tornar a plataforma *Árvore da Ciência* ainda mais atrativa para a comunidade científica brasileira. Outro ponto a ser mencionado é a melhoria do algoritmo de geração das árvores desenvolvido por Dores et al. [2017] a fim de tornar mais efetiva a desambiguação de pesquisadores com nomes semelhantes ou que tenham os seus nomes registrados na plataforma Lattes com diferentes grafias, o que pode ser realizado usando alguns dos métodos previamente desenvolvidos pelo nosso grupo de pesquisa [Ferreira et al. 2020]. Finalmente, vale ressaltar a importância de se ampliar a divulgação da plataforma a fim de aumentar a sua utilização e a demanda por novas funcionalidades.

Agradecimentos

Este trabalho conta com apoio da FAPEMIG (processo APQ-02302-17) e do CNPq (processo 308528/2019-0). Os autores agradecem a todos os pesquisadores que participaram das avaliações com usuários reportadas neste artigo.

Referências

- Barabási, A.-L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., and Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A: Statistical mechanics and its applications*, 311(3-4):590–614.
- Chang, S. (2003). Academic genealogy of american physicists. *AAPPS Bull.*, 13(6):6–41.
- Cota, J. M. M. C. (2020). *Uma Plataforma para Exploração da Genealogia Acadêmica Brasileira*. Dissertação de Mestrado, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil.
- Damaceno, R. J. P., Rossi, L., Mugnaini, R., and Mena-Chalco, J. P. (2019). The Brazilian academic genealogy: evidence of advisor-advisee relationships through quantitative analysis. *Scientometrics*, 119(1):303–333.
- David, S. V. and Hayden, B. Y. (2012). Neurotree: A collaborative, graphical database of the academic genealogy of neuroscience. *PLoS One*, 7(10).
- Digiampietri, L., Mugnaini, R., Trucolo, C., Delgado, K., Mena-Chalco, J., and Köhler, A. (2019). Geographic and disciplinary distribution of the Brazilian's PhD community. *Brazilian Journal of Information Science: Research Trends*, 13(4):113–131.
- Dores, W., Soares, E., Benevenuto, F., and Laender, A. H. F. (2017). Building the Brazilian Academic Genealogy Tree. In *Research and Advanced Technology for Digital Libraries - 21st International Conference on Theory and Practice of Digital Libraries, TPDL 2017, Thessaloniki, Greece, Sept. 18-21, 2017, Proceedings*, pages 537–543.

- Ferreira, A. A., Gonçalves, M. A., and Laender, A. H. F. (2012). A Brief Survey of Automatic Methods for Author Name Disambiguation. *SIGMOD Record*, 41(2):15–26.
- Ferreira, A. A., Gonçalves, M. A., and Laender, A. H. F. (2020). *Automatic Disambiguation of Author Names in Bibliographic Repositories*. Synthesis Lectures on Information Concepts, Retrieval, and Services. Morgan & Claypool Publishers, San Rafael, CA.
- Hart, R. E. and Cossuth, J. H. (2013). A family tree of tropical meteorology's academic community and its proposed expansion. *Bulletin of the American Meteorological Society*, 94(12):1837–1848.
- Hirshman, B. R., Tang, J. A., Jones, L. A., Proudfoot, J. A., Carley, K. M., Marshall, L., Carter, B. S., and Chen, C. C. (2016). Impact of medical academic genealogy on publication patterns: an analysis of the literature for surgical resection in brain tumor patients. *Annals of Neurology*, 79(2):169–177.
- Jackson, A. (2007). A Labor of Love: The Mathematics Genealogy Project. *Notices of the AMS*, 54(8):1002–1003.
- Lazar, J., Feng, J. H., and Hochheiser, H. (2017). *Research methods in human-computer interaction*. Morgan Kaufmann, Burlington, Massachusetts, USA.
- Madeira, G., Borges, E. N., Lucca, G., Santos, H., and Dimuro, G. (2019). An Information System for Analyzing Academic Genealogy. In *Proceedings of the 21st International Conference on Enterprise Information Systems, ICEIS 2019, Heraklion, Crete, Greece, May 3-5, 2019, Volume 1*, pages 114–120. Springer.
- Malmgren, R. D., Ottino, J. M., and Amaral, L. A. N. (2010). The role of mentorship in protégé performance. *Nature*, 465(7298):622.
- Mena-Chalco, J. P., Digiampietri, L. A., Lopes, F. M., and Cesar, R. M. (2014). Brazilian bibliometric coauthorship networks. *Journal of the Association for Information Science and Technology*, 65(7):1424–1445.
- Newman, M. E. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences*, 101(suppl 1):5200–5205.
- Pessoa Junior, G. J., Dias, T. M. R., Silva, T. H. P., and Laender, A. H. F. (2020). On interdisciplinary collaborations in scientific coauthorship networks: the case of the Brazilian community. *Scientometrics*, 124(3):2341–2360.
- Pressman, R. S. and Maxim, B. R. (2019). *Software Engineering: A Practitioner's Approach*. McGraw Hill, New York, USA, 9th edition.
- Tuesta, E. F., Delgado, K. V., Mugnaini, R., Digiampietri, L. A., Mena-Chalco, J. P., and Pérez-Alcázar, J. J. (2015). Analysis of an advisor-advisee relationship: An exploratory study of the area of exact and earth sciences in Brazil. *PloS One*, 10(5):e0129065.