

ÁRVORES GENEALÓGICAS ACADÊMICAS COMO ESTRATÉGIA PARA ANÁLISES DO PROCESSO DE ORIENTAÇÃO

Email:
tales.info@gmail.com
thiagomagela@gmail.com
gray@dppg.cefetmg.br
patriciamdias@gmail.com

Tales Henrique José Moreira¹, Thiago Magela Rodrigues Dias¹, Gray Farias Moita¹, Patrícia Mascarenhas Dias¹

RESUMO

A relação orientador-orientado, que pode se caracterizar como uma forma de propagação do conhecimento, como resultado pode proporcionar um aumento na produção científica dos orientadores e orientados. Portanto, caracterizar como ocorre o processo de orientação e a produção científica resultante dessa relação é mais um importante meio de análise da colaboração científica nas diversas áreas do conhecimento. Neste trabalho, são utilizados os dados de orientações extraídos dos currículos cadastrados na Plataforma Lattes com o intuito de obter-se, além de uma visão geral do processo de orientação no Brasil, uma análise advinda desse processo de orientação, identificando dessa forma como o conhecimento científico Brasileiro tem se propagado. Buscando um maior entendimento sobre como tem se desenvolvido o processo de orientação acadêmica brasileira nas diversas áreas do conhecimento, esta pesquisa tem como objetivo geral caracterizar as orientações em programas de pós-graduação, a partir de análises bibliométricas e baseadas em análises de redes sociais realizadas sobre dados curriculares disponíveis na Plataforma Lattes.

Palavras-Chave: *Genealogia Acadêmica; Plataforma Lattes; Orientação.*

ABSTRACT

The guiding-oriented relationship, which can be characterized as a way of propagating knowledge, as a result can provide an increase in the scientific output of guiding and oriented. Therefore, characterizing how the orientation process occurs and the scientific production resulting from this relationship is another important means of analyzing scientific collaboration in the various areas of knowledge. In this work, the data of guidelines extracted from the curricula registered in the Lattes Platform are used in order to obtain, besides an overview of the orientation process in Brazil, an analysis derived from this orientation process, thus identifying the knowledge Brazilian science has spread. Seeking a greater understanding of how the process of Brazilian academic orientation has developed in the different areas of knowledge, this research has as general objective to characterize the orientations in postgraduate programs, based on bibliometric analyzes and based on analyzes of social networks carried out curricular data available on the Lattes Platform.

Keywords: *Academic Genealogy; Lattes Platform; Orientation.*

¹ Programa de Pós-Graduação em Modelagem Matemática e Computacional.
Pesquisa em andamento apresentada no XIX. GT 8 – Informação e Tecnologia
Acesso ao artigo: <http://enancib.marilia.unesp.br/index.php/XIXENANCIB/xixenancib/paper/view/1451>

1 INTRODUÇÃO

Nas últimas décadas do século XX, em razão da construção, manutenção e informatização de repositórios de dados científicos, tornou-se realidade a produção de indicadores bibliométricos de maior representatividade (MUGNAINI; JANNUZZI; QUONIAM, 2004). Neste mesmo período, diversos outros estudos têm procurado compreender a evolução da ciência e, principalmente, como ocorre a colaboração científica entre indivíduos. Assim sendo, recentemente tem aumentado o surgimento de técnicas, como, por exemplo, métricas baseadas em análise de redes sociais, análises bibliométricas e cientométricas com o objetivo de auxiliar na análise destes dados (DIAS, 2016).

Percebe-se neste contexto que a evolução das pesquisas científicas tem forte influência no processo de formação, em que pesquisadores orientadores inserem novos pesquisadores que contribuem para que novos estudos sejam realizados em diversas áreas do conhecimento. Destaca-se que grande parte dos trabalhos realizados com orientação no Brasil são decorrentes de Programas de Pós-Graduação (PPGs), impulsionados pela necessidade de capacitação e titulação de pesquisadores.

Para Sugimoto (2014), genealogia acadêmica é um estudo quantitativo da herança intelectual através da relação orientador-orientado. Para Ferreira, Furtado e Silveira (2009), o binômio (ou díade) orientador-orientado é indubitavelmente a base dos PPGs, o que determina o crescimento e a expansão dos cursos de Pós-Graduação (PG) e a demanda de orientação. Além disso, os autores ressaltam que o aluno de PG é um pesquisador em potencial em estágio avançado de desenvolvimento, ou seja, a caminho da autonomia científica, mas ainda dependente de um professor, o que justifica as atividades de orientação como efetivamente necessárias. Adicionalmente aos dados básicos que caracterizam os vínculos sobre o processo de orientação, informações como, por exemplo, publicações e área de atuação podem ser analisadas com o intuito de se compreender o perfil de orientação de um orientador em particular ou de uma área do conhecimento.

Em Leite Filho e Martins (2006), são verificadas as influências da relação orientador-orientado no processo de produção de teses e dissertações dos PPGs em Contabilidade da cidade de São Paulo. Os autores citam como justificativas para o estudo a importância de se analisar aspectos que teriam ligação com a construção do conhecimento, especificamente em se tratando da área de Contabilidade, e a tentativa de sinalizar a importância da temática.

Dores e Laender (2016) optaram por utilizar em seu trabalho a base de dados *NDLTD* (*Networked Digital Library of Theses and Dissertations*) para fornecer uma série de análises com base em estruturas de árvores genealógicas geradas. Apesar de ser uma base relevante, a mesma possui algumas desvantagens em relação aos currículos da Plataforma Lattes, sendo elas: menor quantidade de registros (já que esta é uma base de teses e dissertações, descartando os demais níveis de capacitação, além de depender das instituições submeterem seus dados, diferentemente da Plataforma Lattes, em que cada indivíduo é responsável por atualizar seus dados) e problema de desambiguação (sendo que o mesmo também ocorre com os currículos cadastrados na Plataforma Lattes, porém, em menor escala, já que na Plataforma Lattes existe a possibilidade de criação de vínculos entre indivíduos).

Tendo em vista as possibilidades de visualização e entendimento do histórico de orientação e, conseqüentemente, a difusão do conhecimento, realizar a modelagem e a caracterização de

árvores genealógicas acadêmicas surge como uma alternativa interessante para a análise de como a ciência brasileira tem se propagado utilizando-se para tanto, dados de orientações. Para isso, as árvores genealógicas acadêmicas podem também ser caracterizadas, facilitando o entendimento do processo de formação com a análise destas árvores.

Diante disso, este trabalho apresenta um estudo sobre os dados de orientação acadêmica registrados nos currículos cadastrados na Plataforma Lattes, propondo uma estratégia para identificar orientações que não estão implícitas nos currículos, permitindo caracterizar de forma inédita grandes árvores genealógicas, possibilitando, dessa maneira apresentar um estudo sobre o histórico de todos os indivíduos com orientações concluídas em cursos de pós-graduação. A presente abordagem é, até então, inédita, tendo em vista a abrangência de indivíduos e de dados analisados.

2 DESENVOLVIMENTO

Este estudo tem como principal fonte de dados os currículos cadastrados na Plataforma Lattes. Os currículos se tornaram um padrão nacional no registro da vida pregressa e atual dos estudantes e pesquisadores do país e é hoje adotado pela maioria das instituições de fomento, universidades e institutos de pesquisa do país. Por sua riqueza de informações e sua crescente confiabilidade e abrangência, se tornou elemento indispensável e compulsório à análise de mérito e competência dos pleitos de financiamentos na área de ciência e tecnologia (CNPQ, 2017).

De posse dos currículos, é possível aplicar a eles diversas transformações e processá-los, a fim de se obter resultados importantes. Para a coleta dos currículos que compõem a Plataforma Lattes utilizados neste trabalho, um arcabouço denominado LattesDataXplorer (DIAS, 2016) foi utilizado. Posteriormente, foi realizada uma expansão do mesmo para atender as necessidades do presente estudo (Figura 1).

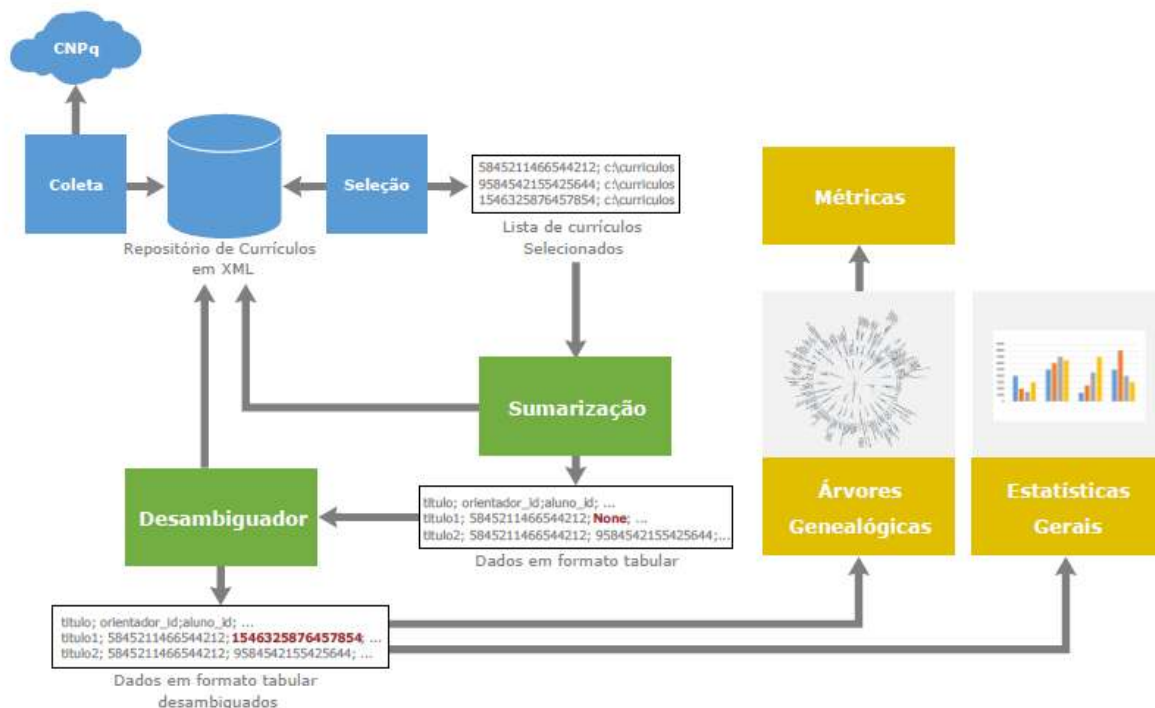


Figura 1: Visão geral do LattesDataXplorer expandido

O arcabouço inicialmente proposto, que engloba toda a parte de extração e seleção dos dados, foi expandido de tal maneira que passou a incorporar módulos para sumarização dos dados, desambiguação dos registros de orientação e produção de árvores genealógicas, além do cálculo de estatísticas gerais, que são fundamentais para a obtenção dos resultados aqui apresentados. Tais módulos são fundamentais para desambiguar os dados de orientação acadêmica e, conseqüentemente, produzir árvores genealógicas acadêmicas que incluem indivíduos antes desconhecidos. Estes mesmos dados desambiguados também são utilizados na obtenção de análises estatísticas gerais.

Os dados representados na Tabela 1 correspondem a todo o conjunto de currículos da Plataforma Lattes no momento da extração utilizada, se revelando uma grande fonte de dados que pode ser utilizada na área da descoberta de conhecimento. Um destes é a genealogia acadêmica. Esta pode ser identificada através dos dados de orientações, possibilitando a análise da vida pregressa do indivíduo. Além disso, tais dados podem ser associados aos dados de produção bibliográfica, por exemplo.

Tabela 1: Resultado da sumarização de todos os currículos analisados.

	Descrição	Quantidade Total
	Dados de informações pessoais	4.591.941
	Dados de formação acadêmica	10.488.174
	Dados de produção bibliográfica (periódicos e anais de congresso)	19.335.882
	Dados de orientações concluídas	8.138.267

A identificação de relacionamentos em orientações não é uma tarefa trivial. Atualmente, os registros de orientações dos currículos possuem uma opção de se realizar a vinculação manual do nome do orientado ou dos coautores a seus identificadores únicos na Plataforma Lattes. No entanto, tal vínculo não é automático e, em geral, relacionamentos antigos permaneceram sem seus vínculos com os identificadores, exibindo apenas o nome no registro de orientação ou coautoria. Diante disso, uma estratégia de identificação de relacionamento se faz necessária para que se possa caracterizar redes com a maior quantidade possível de indivíduos.

Para tanto, uma das principais contribuições deste trabalho é a proposta de um método eficiente para caracterização de vínculos de orientações acadêmicas em grandes repositórios de dados ambíguos, em que a primeira etapa da estratégia de identificação aqui colocada é a obtenção dos dados pessoais de cada indivíduo. Com base nestes dados, é possível obter informações como nome completo e nome em citações bibliográficas, utilizadas para produção do dicionário, conforme a Figura 2. Além de nomes de citações presentes nos currículos, também são produzidos, automaticamente, possíveis nomes de citação, a partir do nome completo de cada indivíduo.

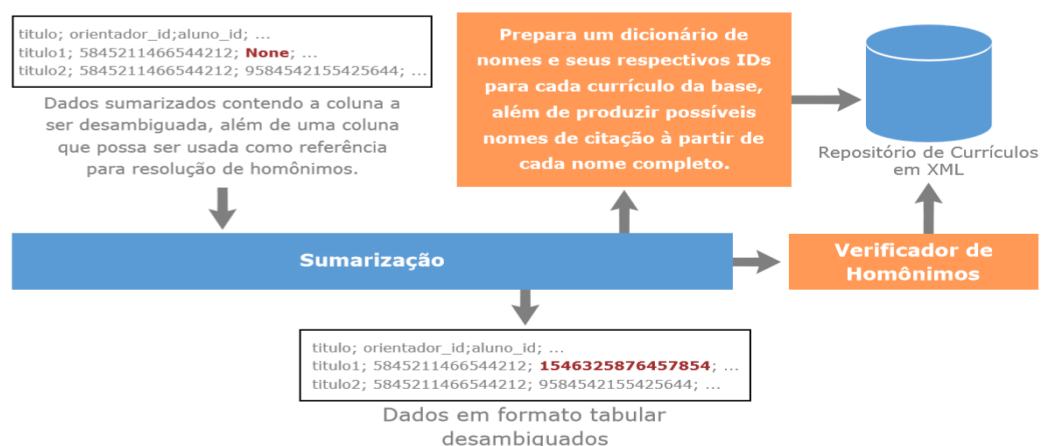


Figura 2: Processo de desambiguação de nomes da Plataforma Lattes

A Plataforma Lattes conta atualmente com mais de 5.200.000 currículos, totalizando cerca de 8.449.497 orientações, independentemente de sua natureza. Antes da execução do método de identificação proposto neste trabalho, apenas 621.384 possuíam relacionamentos vinculados implicitamente, conforme a Tabela 2. Ou seja, apenas estas orientações foram devidamente vinculadas com os orientados pelo orientador.

Tabela 2: Resultado do desambiguador

Descrição	Quantidade	%
Não identificados	4.397.692	52,04
Não identificados entre homônimos	247.742	2,93
Subtotal:	4.645.434	54,97
Previamente vinculados	621.384	7,35
Únicos identificados	2.889.545	34,19
Identificados com	293.134	3,46

Homônimos		
Subtotal:	3.804.063	45,03
Total:	8.449.497	100

Após a execução do método proposto, 3.804.063 relacionamentos novos foram identificados, um total de 45,03%, valor bem superior aos 7,35% encontrados antes do processo de desambiguação.

3 RESULTADOS

Como resultados das etapas de tratamentos dos dados, são produzidos arquivos de dados pré-processados em formato tabular, podendo ou não estar desambiguados, contendo todas as informações necessárias para o cálculo de diversas métricas bibliométricas e baseadas em análise de redes sociais, como pode ser observado na arquitetura proposta. Os cálculos realizados a partir destes arquivos são facilitados, já que nestes constam todos os dados sem a necessidade de acesso e busca dos dados em cada um dos currículos.

Os dados aqui utilizados foram coletados em setembro de 2017 e correspondem a 5.152.148 currículos cadastrados na Plataforma Lattes, distribuídos em diversas áreas do conhecimento e nos mais diversos níveis de capacitação. Em todo o conjunto em análise, 481.624 indivíduos orientaram em algum nível de formação, ou seja, apenas 9,348% do total de todos os indivíduos com currículos cadastrados na Plataforma Lattes. As orientações consideradas para esta caracterização cobrem o período de 1900 (orientação mais antiga cadastrada) até dezembro de 2016, período este considerado, tendo em vista a coleta realizada em 2017.

As orientações também podem ser distribuídas de acordo com o seu nível, conforme apresentado na Figura 3, sendo que grande parte destas são orientações de Graduação (44%), seguidas pelas orientações de Especialização (14%), as de Iniciação Científica (14%) e as de Outra Natureza (13%).

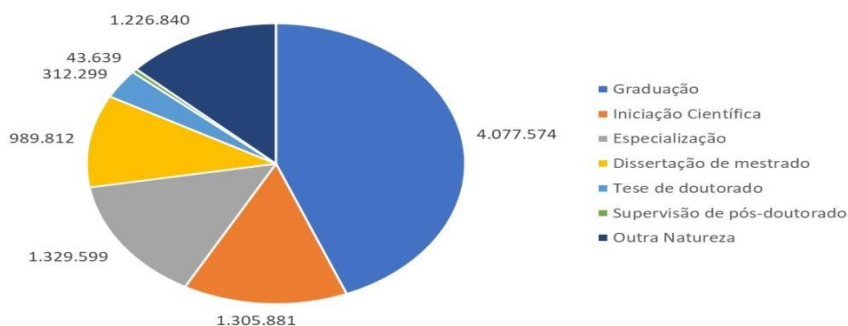


Figura 3: Distribuição das orientações por natureza de todos os indivíduos

Conforme descrito, apesar da baixa quantidade de orientações selecionadas para as análises aqui apresentadas (1.345.750 registros de orientações em programas de Pós-graduação) se comparado ao total geral de orientações, este conjunto possui informações sobre orientações com registros mais consistentes, tornando um interessante objeto de estudo.

Os dados analisados também foram distribuídos por grandes áreas do conhecimento, conforme apresenta a Figura 4. Do total de orientações, mais de 50% não possuem grande área informada. Entre as demais grandes áreas, a grande área com maior quantidade de orientações é a de Ciências Sociais Aplicadas.

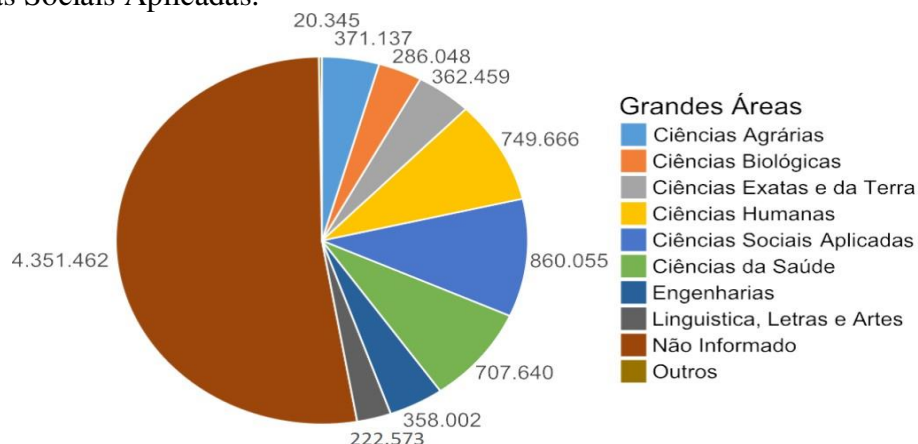


Figura 4: Gráfico de orientações por grandes áreas

Também foram produzidas árvores genealógicas dos orientadores que mais orientam de acordo com os registros dos currículos. Devido à quantidade de dados, são utilizados apenas dados de orientações dos cursos de pós-graduação, facilitando não só a análise, como também a visualização das árvores. Assim, são utilizados 1.196.916 registros de orientações selecionados a partir de todo o conjunto identificado.

A Tabela 3 apresenta as maiores redes identificadas, ranqueadas pela maior rede encontrada e pelo maior número de gerações de descendentes, ambas com uma quantidade relativamente baixa de orientações diretas.

Tabela 3: Indivíduos com maiores redes do conjunto de dados.

Tipo	Grande Área	Gerações	Tamanho da Rede
Maior Quantidade de Gerações	Ciências Biológicas	15	5716
	Ciências Biológicas	15	5442
	Ciências Biológicas	15	5426
	Ciências Humanas	6	8318
Tamanho máximo de Rede	Ciências Humanas	6	8318

A árvore da Figura 5 possui a maior quantidade de descendentes, apesar de possuir apenas 6 gerações (muito abaixo das 15 encontradas na árvore da Figura 23) e cerca de 30% da quantidade de orientados diretos em relação ao indivíduo da primeira árvore apresentada. O nó raiz (orientador principal) possui como titulação máxima doutorado em Filosofia da Educação e atua na Universidade Estadual de Campinas (UNICAMP), tendo como grande área de atuação Ciências Humanas. Suas orientações começaram em 1974 e são distribuídas em Doutorado

(47%), Mestrado (28%) e Pós-doutorado (18%), além de algumas poucas orientações em outros níveis.

Diferentemente das árvores anteriores, esta última é a única com supervisões de pós-doutorado, o que pode influenciar no tamanho da rede, visto que orientados de pós-doutorado já possuem doutorado concluído, e, conseqüentemente, já podem estar atuando em programas de pós-graduação a mais tempo que os demais.

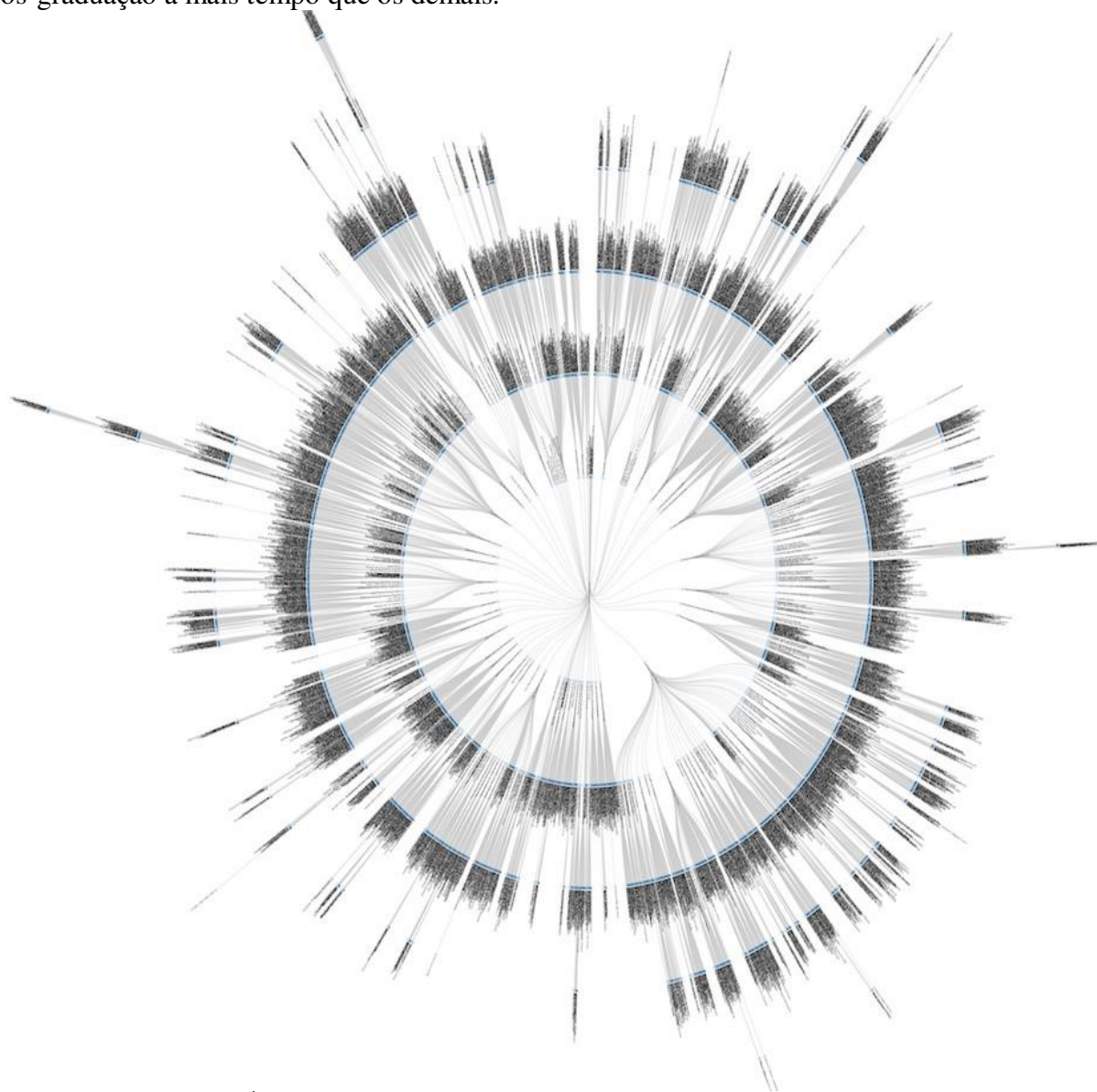


Figura 5: Árvore genealógica do orientador com a maior quantidade de descendentes

Até o momento foram exibidas árvores genealógicas dos orientadores com relação à quantidade de orientados diretos, tamanhos das árvores (com base nos descendentes) e quantidade de gerações. Além disso, diversas outras métricas baseadas em redes podem ser aplicadas para melhor caracterização das mesmas.

5 CONSIDERAÇÕES FINAIS

Buscando um maior entendimento sobre como tem se desenvolvido o processo de orientação acadêmica brasileira, este trabalho busca caracterizar as orientações acadêmicas, a partir de análises bibliométricas e baseadas em análises de redes realizadas sobre dados de registros de orientações acadêmicas dos currículos cadastrados na Plataforma Lattes.

Os resultados obtidos foram distribuídos por grandes áreas do conhecimento, para assim caracterizar cada uma delas. Neste contexto, destaca-se a área de Ciências Sociais Aplicadas, seguida por Ciências Humanas e Ciências da Saúde. Porém, apesar da maior quantidade de orientações da grande área Ciências Sociais Aplicadas, a grande maioria pertence a cursos de Graduação, tendo outras áreas maior destaque em Pós-graduação, como, por exemplo, Ciências Humanas e Ciências da Saúde.

REFERÊNCIAS

CNPQ. Sobre a plataforma Lattes. 2017. Disponível em: <<http://lattes.cnpq.br/>>.

DIAS, T. M. R. Um Estudo Sobre a Produção Científica Brasileira a Partir de Dados da Plataforma Lattes. 181 p. Tese (Doutorado) — Centro Federal de Educação Tecnológica de Minas Gerais, 2016.

DORES, W.; LAENDER, A. H. F. Extracting Academic Genealogy Trees from the Networked Digital Library of Theses and Dissertations. In: Proceedings of the 16th ACM/IEEE-CS on Joint Conference on Digital Libraries - JCDL '16. New York, New York, USA: ACM Press, 2016. p. 163–166. ISBN 9781450342292. Disponível em: <<http://dl.acm.org/citation.cfm?doid=2910896.2910916>>.

FERREIRA, L. M.; FURTADO, F.; SILVEIRA, T. S. Relação Orientador-Orientando. O Conhecimento Multiplicador. Acta Cirúrgica Brasileira, v. 24, n. 3, p. 170–172, 2009.

LEITE FILHO, G. A.; MARTINS, G. D. A. Relação orientador-orientando e suas influências na elaboração de teses e dissertações. Revista de Administração de Empresas, SciELO Brasil, v. 46, n. spe, p. 99–109, 2006.

MUGNAINI, R.; JANNUZZI, P. d. M.; QUONIAM, L. Indicadores bibliométricos da produção científica brasileira: uma análise a partir da base Pascal. Ciência da Informação, SciELO Brasil, v. 33, n. 2, p. 123–131, 2004.

NICHOLAS, D.; RITCHIE, M. Literature and bibliometrics. [S.l.]: C. Bingley, 1978. ISBN 0208015418.

SUGIMOTO, C. R. Academic genealogy. In: Beyond bibliometrics. Harnessing multidimensional indicators of scholarly impact. MIT Press, 2014. p. 365–382.