
UM RETRATO DA PRODUÇÃO CIENTÍFICA BRASILEIRA BASEADO EM DADOS DA PLATAFORMA LATTES

A snapshot of the Brazilian scientific production based on the Lattes platform

Thiago Magela Rodrigues Dias (1) Gray Farias Moita (2)

(1) CEFET-MG, thiagomagela@gmail.com (2) gray@dppg.cefetmg.br

Resumo

Os estudos sobre dados de produções científicas têm recebido atenção de pesquisadores, de diversas áreas, que visam obter conhecimento sobre a evolução das pesquisas em geral. Tais estudos possibilitam a análise da produção científica para diversos propósitos e um dos desafios neste tipo de análise está na diversidade de repositórios contendo dados em formatos e estruturas distintas. Além disso, problemas como redundância dos dados, ambiguidade e dados incorretos tornam as análises ainda mais complexas. Neste artigo, é realizada uma análise dos dados curriculares de pesquisadores brasileiros cadastrados na Plataforma Lattes. Para isso, foram coletados os currículos utilizando um framework de coleta e tratamento dos dados. A quantidade de dados analisados possibilita obter uma visão de todo o conjunto que é composto por milhões de currículos, e que contém dados sobre publicações científicas, nível de formação, áreas de atuação, dentre outros. Nos resultados apresentados, é possível obter uma visão sobre a produção científica brasileira e qual o perfil da produtividade científica dos doutores que tem atuado no Brasil.

Palavras-chave: Bibliometria; Extração de Dados; Bibliotecas Digitais; Publicações Científicas.

1 Introdução

A produção e a publicação de trabalhos científicos têm aumentado significativamente nos últimos anos, sendo a World Wide Web (Web) o principal facilitador para acesso e distribuição das publicações científicas. Serviços como bibliotecas digitais, redes sociais acadêmicas e sítios para registro individual de produção científica são alguns exemplos de ferramentas tecnológicas que têm contribuído para aumentar a quantidade de trabalhos publicados, permitindo que usuários não apenas acessem o conteúdo disponível, mas também possam registrar a sua própria produção técnica e científica interagindo diretamente com essas ferramentas. Com isso, trabalhos publicados e disponibilizados eletronicamente podem ser acessados de forma instantânea, contribuindo de forma significativa para a disseminação do conhecimento científico.

Exemplos desses repositórios são a DBLP¹ (Digital Bibliography & Library Project), ArnetMiner², Google Scholar³ e Microsoft Academic Search⁴ que contém

Abstract

Studies on scientific production data have received attention from researchers in various fields to gain insight into the evolution of research in general. Such studies allow the analysis of scientific production for various purposes and one of the challenges in this type of analysis lies in the diversity of repositories containing data in different formats and structures. In addition, problems such as data redundancy, ambiguity, and incorrect data make the analysis even more complex. In this article, an analysis of the curricular data of Brazilian researchers enrolled in the Lattes Platform is carried out. The amount of data analyzed makes it possible to obtain a view of the whole set consisting of millions of curricula, and contains data on scientific publications, training level, areas of activity, among others. In the results presented, it is possible to obtain an overview of the Brazilian scientific production and the profile of the productivity of the doctors who have worked in Brazil.

Keywords: Bibliometrics; Data Extraction; Digital Libraries; Scientific Publications.

dados de publicações científicas, além da PubMed⁵ que contém dados de citações para literatura biomédica, e Scopus⁶, repositório de resumos e citações de artigos, bem como as redes sociais acadêmicas, como ResearchGate⁷ e academia.edu⁸ que possibilitam o compartilhamento de dados e a interação entre os seus membros.

No Brasil, um importante repositório de dados científicos e acadêmicos é a Plataforma Lattes⁹, mantido pelo CNPq¹⁰ e que se caracteriza como fonte de informação usado durante o processo de avaliação de projetos e solicitações de auxílio a pesquisa. Nele cada indivíduo registra em seus currículos Lattes, de forma individual, sua produção científica, bem como dados sobre sua formação acadêmica, áreas de atuação e orientações. Atualmente a Plataforma Lattes contém aproximadamente 5.500.000 currículos cadastrados.

Segundo Lane, em artigo publicado na revista Nature (Lane, 2010), medir e avaliar o desempenho acadêmico passa a ser um fator crucial para a vida científica. Vá-

rios fatores para esta avaliação e medição estão vinculados aos cálculos de métricas. Porém, os sistemas atuais de medição são insuficientes para determinar respostas confiáveis. No trabalho supracitado, são descritos os vários problemas na adoção das métricas normalmente utilizadas para, por exemplo, classificar grupos ou instituições de pesquisa e os desafios nas propostas de métricas eficazes. A autora apresenta esforços empregados para a construção de repositórios confiáveis de dados científicos que poderiam permitir análises com o objetivo de explorar e compreender como a ciência tem evoluído. Embora todos os esforços sejam úteis, alguns apresentam problemas que comprometem o sucesso de tais iniciativas. Exemplo disso são informações dependentes de softwares específicos, repositórios dependentes de financiamentos transitórios (e esforços para poder combiná-los) e bancos de dados sem estruturas coesas e confiáveis. Nesse cenário, a Plataforma Lattes é citada como um poderoso exemplo de boas práticas para o fornecimento de dados de alta qualidade, e de como uma plataforma de coleta de dados tem sido incentivada por órgãos federais, instituições e órgãos financiadores. Por fim, é destacado que a Plataforma Lattes é uma das fontes de dados de pesquisadores mais confiáveis existente.

Digiampietri et al. (2014) apresentam uma estratégia para caracterizar os programas de pós-graduação em Ciência da Computação brasileiros e a relação entre eles. No trabalho, são apresentadas análises dos programas sob diversas perspectivas. São descritas a classificação sob cada uma destas perspectivas, a correlação entre as métricas de avaliação utilizadas, além de como os programas se relacionam, e quais os aspectos que aumentam a produtividade de pesquisa dos programas. Foram analisados 37 Programas de Pós-graduação acadêmicos em Ciência da Computação avaliados pela CAPES nos triênios 2004-2006 e 2007-2009. Foram identificados 732 professores nos programas avaliados, obtidos pelos relatórios trienais da CAPES, e de posse dos nomes dos professores foi possível obter o identificador para o currículo de todos os professores, para, posteriormente, serem extraídos da Plataforma Lattes. Após o processo de extração, os currículos passaram por um pré-processamento e seus dados foram inseridos em um banco de dados relacional. Após a inserção dos dados curriculares, estes foram enriquecidos com informações disponíveis em diversos repositórios contendo dados sobre conferências e periódicos de publicações científicas. Essas informações descrevem o nível de cada um dos periódicos ou conferências, a quantidade de citações de um determinado trabalho publicado, dentre outros. Para a identificação de colaborações, foi feita a comparação dos títulos das publicações, dos currículos em análise com todos os demais, visando encontrar títulos iguais ou similares e, desta forma, caracterizar uma colaboração.

Já Mena-Chalco et al. (2014) utilizam dados dos currículos da Plataforma Lattes para identificar e caracterizar a rede de colaboração de pesquisadores brasileiros. O trabalho objetivou extrair os dados de currículos cadastrados na Plataforma Lattes, identificar automaticamente a colaboração baseada em informações bibliométricas, produzindo uma rede de colaboração, e aplicar métricas baseadas em análise topológica para compreender como ocorre a interação entre os pesquisadores.

Lima et al. (2015) fazem uma avaliação do desempenho dos principais pesquisadores que atuam em Ciência da Computação. São utilizados dados extraídos dos currículos da Plataforma Lattes de 406 pesquisadores bolsistas de produtividade em pesquisa do CNPq na área de Ciência da Computação nas cinco modalidades da bolsa (1A, 1B, 1C, 1D e 2). A avaliação considerou três dimensões centrais, sendo: tempo de carreira do pesquisador, em que foi considerada a quantidade de anos após a conclusão do doutorado; quantidade de alunos orientados; e produtividade científica, referente ao volume de publicações e citações. Com relação ao tempo de carreira dos pesquisadores, observa-se que aqueles que possuem os níveis mais elevados de bolsa também são detentores do maior tempo de carreira. Já com relação às orientações, existe uma distinção entre a quantidade de orientações de mestrado e doutorado, em que há, aproximadamente, uma orientação de doutorado para cada ano de carreira entre todos os bolsistas. Ao se considerar as orientações de mestrado, este valor cai para menos de 0,5 orientação para cada ano após a conclusão do doutorado. Ao se avaliar a produção científica dos pesquisadores, observa-se que o volume de publicações aumenta com o nível das bolsas, exceto no caso do nível 1A, que se assemelha ao 1C, por incluir pesquisadores mais antigos, estes não estavam vinculados à programas com certo nível de maturidade. Estas avaliações demonstram coerência entre os itens avaliados e as modalidades das bolsas em que os pesquisadores se enquadram na área de Ciência da Computação.

Sidone, Haddad e Mena-Chalco (2016) apresentam o papel da geografia na evolução da produção e colaboração científica no Brasil entre 1992 e 2009. Para os estudos, foram utilizados dados dos currículos da Plataforma Lattes de um milhão de pesquisadores. Os autores destacam o processo de desaceleração da produção científica brasileira a partir dos últimos triênios analisados. No trabalho, foi possível identificar que o padrão de crescimento das grandes áreas é semelhante nos triênios iniciais da análise, quando praticamente todas apresentaram produção em taxas duplicadas entre 1992-1994 e 1995-1997. No entanto, percebe-se que o crescimento desacelerado ao longo do período é comum a todas as grandes áreas, porém em intensidades distintas. Como exemplo, verifica-se que a desaceleração no ritmo da produção científica ocorreu de maneira

mais forte nos últimos triênios, primordialmente nas grandes áreas de Engenharias, Ciências Exatas e da Terra, Ciências da Saúde e Ciências Biológicas.

Diversos outros trabalhos têm explorado a Plataforma Lattes como principal fonte de dados para análise bibliográfica e de redes de colaboração, dentre os quais os de Arruda et al. (2009), Alves, Yanasse e Soma (2011), Nascimento et al. (2011), Costa et al. (2013), Colugnati et al. (2014), Roos et al. (2014), Dias e Moita (2015), Furtado et al. (2015), e Brito, Quoniam e Mena-Chalco (2016).

Logo, diante do exposto, este trabalho tem como objetivo realizar uma análise dos dados curriculares cadastrados na Plataforma Lattes, para que dessa forma seja possível apresentar uma visão do perfil dos indivíduos com currículos cadastrados. O restante do artigo está organizado da seguinte forma. Em seguida, é realizada uma caracterização geral dos dados e analisada a produção científica dos indivíduos com doutorado concluído. Finalmente, conclui-se o estudo apresentando as principais considerações.

2 Caracterização Geral

Uma grande parte dos editais de financiamento de projetos de pesquisa realizados por agências de amparo à pesquisa, por instituições de ensino, bem como, pelo próprio CNPq, utiliza dados cadastrados nos currículos da Plataforma Lattes dos proponentes como uma das formas de avaliação das propostas. Isto passou a ser um significativo incentivo para que os pesquisadores mantenham seus currículos com informações atualizadas, tornando a Plataforma Lattes uma fonte extremamente rica para análise da produção científica brasileira. Além disso, jovens estudantes que pretendem realizar parte de seus estudos em instituições estrangeiras, financiados por agências governamentais brasileiras, também devem possuir currículos cadastrados na Plataforma Lattes, contribuindo de forma significativa para o aumento na quantidade de currículos atualmente dispostos na Plataforma Lattes. A Figura 1 apresenta o crescimento da quantidade de currículos cadastrados durante os meses de maio de 2014 a junho de 2016. Nesse período, houve um aumento de aproximadamente 28% na quantidade de currículos armazenados.

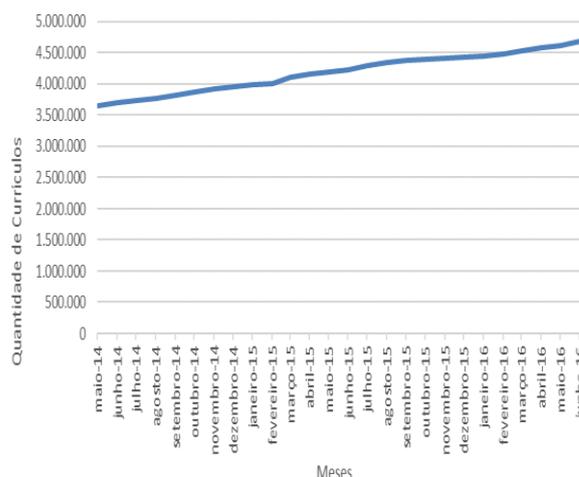


Figura 1. Crescimento da quantidade de currículos cadastrados na plataforma lattes

É possível identificar que a Plataforma Lattes vem agregando, a cada dia, novos indivíduos que registram dados sobre suas atividades profissionais, acadêmicas e de pesquisa, se apresentando como uma fonte extremamente rica e consistente de informações que podem revelar como têm sido realizadas as pesquisas científicas brasileiras nas diversas áreas do conhecimento. Logo, explorar seus dados com o objetivo de se obter uma visão detalhada de todo o repositório, pode proporcionar uma visão inédita sobre a produção científica brasileira e como tem ocorrido a colaboração entre os pesquisadores brasileiros, fornecendo subsídios para análises e comparações sobre as diversas áreas de atuação dos indivíduos que as compõem.

Uma caracterização geral de todo o repositório de dados curriculares na proporção atual e com significativa abrangência de estudos não foi tratada anteriormente, já que os trabalhos correlatos utilizaram apenas grupos específicos de currículos, tendo em vista que a manipulação de todos os currículos da Plataforma Lattes não é uma tarefa trivial. A dificuldade envolve problemas de recuperação de informação e algoritmos eficientes para manipulação de grande volume de dados. Para a obtenção dos dados utilizados neste trabalho, o framework de extração e tratamentos dos dados proposto por Dias (2016), e denominado LattesDataXplorer (Figura 2 no apêndice) foi utilizado.

Os dados utilizados nesta caracterização foram coletados em abril de 2015 e correspondem a 4.156.635 currículos, distribuídos nas diversas áreas do conhecimento. Esses dados foram coletados utilizando o LattesDataXplorer, que possibilita realizar uma caracterização geral dos dados contidos em todos os currículos da Plataforma Lattes. Os indivíduos com endereço profissional informado fora do país foram excluídos das análises aqui apresentadas. As publicações consideradas para esta caracterização cobrem o período de 1961 (trabalho mais antigo cadastrado) até dezembro de

2014. Logo, apesar de a coleta ter sido realizada no início de abril de 2015, somente publicações até o fim do ano de 2014 foram consideradas. Dessa forma, espera-se que os diversos trabalhos finalizados no final de 2014 já estejam registrados nos currículos de seus respectivos autores, tendo em vista que tais trabalhos são incorporados geralmente após a sua conclusão, como no caso de dissertações e teses, ou publicação, como artigos em anais de congresso ou periódico.

Estatísticas gerais, como distribuição geográfica e titulação dos pesquisadores, podem ser obtidas pela interface de consulta da Plataforma Lattes, no entanto, nenhuma informação sobre a produção científica dos pesquisadores, crescimento do repositório de dados e atualização dos currículos que compõem a Plataforma Lattes pode ser obtido pela sua interface de consulta. Dessa forma, estatísticas gerais sobre a produção científica e dados sobre orientações apresentadas neste trabalho são inéditas, já que não é possível extrair tais informações pelas interfaces de consultas disponibilizadas, e nenhum outro trabalho analisou todo o repositório de currículos Lattes em detalhes como neste trabalho.

É importante destacar a diversidade dos dados que estão registrados neste conjunto de currículos, como artigos publicados em anais de congresso e periódico, apresentação de trabalhos, participação em eventos, nível de formação, orientações concluídas, dentre outros. Ressalta-se que um determinado trabalho pode estar inserido em currículos distintos, já que este pode ter sido realizado em colaboração. Logo, no repositório da Plataforma Lattes, um trabalho pode aparecer várias vezes, tendo em vista que ele pode ser registrado por cada um de seus autores. A Tabela I apresenta o quantitativo geral de todos os trabalhos registrados na Plataforma Lattes contabilizando os trabalhos de todos os currículos, bem como os totais únicos, em que são desconsiderados títulos de trabalhos iguais. Na coluna “Único”, são considerados apenas trabalhos com títulos distintos, para que, dessa forma, um mesmo trabalho registrado em vários currículos seja contabilizado uma única vez.

<i>Tipo de Publicação</i>	<i>Total</i>	<i>Únicos</i>
Artigos em Anais de Congressos	11.591.142	7.437.979
Artigos em Periódicos	4.560.921	2.806.910
Capítulos de Livros	1.055.388	820.732
Livros	455.447	360.902

Tabela I. *Quantitativo de publicações*

Os dados apresentados correspondem aos registros de publicações de todos os currículos, independentemente de sua área de atuação ou de sua data de atualização. A quantidade de dados registrada corrobora a importância da Plataforma Lattes, confirmando ser um dos maiores repositórios de dados científicos mundiais da atualidade (Lane, 2010), caracterizando-se como uma fonte

extremamente rica para análise da produção científica brasileira. É possível observar a tendência de publicação de artigos em anais de congresso, seguida em menor número pela publicação de artigos em periódico. Além dos artigos, estão registrados outros tipos de publicação, tais como capítulos de livro, livros, textos em jornais e revistas, trabalhos técnicos, outras produções bibliográficas que não se enquadram nas categorias anteriores e demais trabalhos que um indivíduo julgar serem relevantes e tenha interesse em cadastrar, bem como, dados de orientações nos mais diversos níveis de formação.

Ressalta-se que cada indivíduo possui um único currículo cadastrado, não existindo a possibilidade de dados duplicados sobre a produção de um mesmo indivíduo, exceto em caso de coautorias. Além disso, em cada currículo, é possível identificar a data de sua última atualização. Tendo em vista que a atualização dos currículos é realizada pelos próprios indivíduos, frequentemente são encontrados currículos que estão há algum tempo sem atualização. Foi possível verificar que, apesar de existirem currículos com data da última atualização em 1997, a quantidade é baixa (201 currículos) se comparada com o total de currículos atualizados nos anos posteriores, principalmente a partir de 2004 (39.248). A não atualização desses currículos pode ter motivos diversos e de difícil identificação. Porém, a maioria dos currículos possui data de atualização recente. Cerca de 33% dos currículos (1.379.057) possuem data de última atualização em 2014 e 51,34% (2.134.154) foram atualizados nos últimos dois anos. Tendo em vista que, como a coleta dos dados foi realizada em abril de 2015, 368.177 currículos têm data de atualização já em 2015. Ou seja, além da quantidade e diversidade de dados presentes em cada um dos currículos, estes, em geral, têm data de atualização recente, justificando todo o seu potencial de análise.

Uma informação relevante para compreender o perfil dos indivíduos que possuem currículos cadastrados diz respeito ao seu nível de formação. Nos currículos da Plataforma Lattes, é possível informar o histórico acadêmico que se inicia no ensino básico e vai até o pós-doutorado, ou seja, um indivíduo pode registrar em seu currículo todos os seus níveis de formação em andamento ou concluído. A Figura 3 apresenta o nível mais alto de formação concluído dos indivíduos.

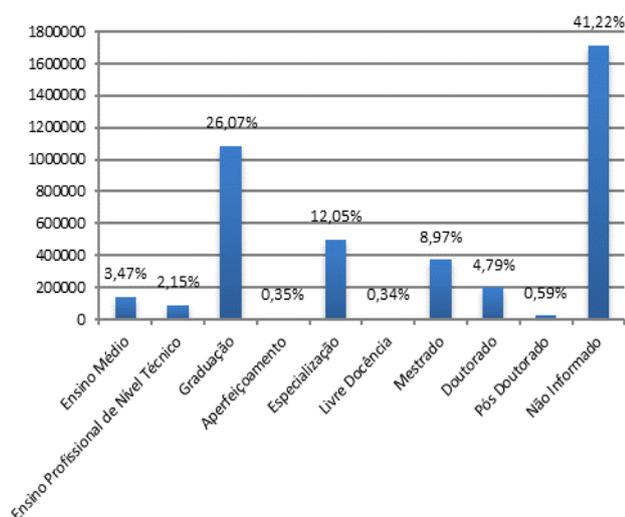


Figura 3. Nível mais alto de formação concluída

Com relação às informações correspondentes aos mais altos níveis de formação concluída informadas nos currículos da Plataforma Lattes, aproximadamente 26% dos indivíduos têm como maior titulação a graduação. Este significativo número de indivíduos com graduação concluída tem relação com as exigências de ingresso em grande parte dos cursos de pós-graduação e recentes programas de incentivo à capacitação no exterior por órgãos governamentais, que exigem que os candidatos tenham seus currículos cadastrados na Plataforma Lattes. Existe uma queda natural entre os níveis de Mestrado e Doutorado, no entanto, é importante destacar que, como significativa parte dos pesquisadores com esses níveis de formação estão vinculados a instituições de ensino ou pesquisa, por exigência, eles também devem ter seus currículos cadastrados na Plataforma Lattes, o que significa que representam quase a totalidade dos indivíduos que detêm esse nível de formação. Diferentemente, os níveis inferiores de capacitação como Ensino Médio e Ensino Profissional de Nível Técnico possuem pequena quantidade de indivíduos cadastrados, pois, em geral, esses indivíduos ainda estão em formação e não cadastram seus currículos. Ressalta-se que 41,22% dos currículos não tem nenhum registro sobre algum nível de formação concluído, logo, foram categorizados como Não Informado.

Uma informação presente nos currículos diz respeito à grande área, área, subárea e especialidade de atuação do indivíduo. Este dado opcional visa possibilitar que um determinado indivíduo informe suas áreas de atuação. Com a indicação da grande área, área, subárea e especialidade de atuação de cada indivíduo, é possível realizar agrupamentos e verificar como os currículos estão distribuídos nas diversas áreas do conhecimento. Como é possível inserir mais de uma grande área e, conseqüentemente, áreas e subáreas, para a classificação adotada neste trabalho, o primeiro registro contendo esta informação foi utilizado, tendo em vista que,

neste primeiro momento, os currículos foram classificados em apenas uma grande área. A distribuição dos indivíduos por grande área de atuação pode ser observada na Figura 4.

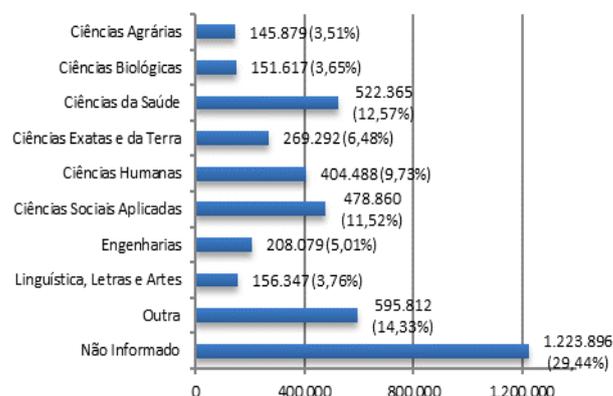


Figura 4. Distribuição por grande área de atuação

Aproximadamente, 29% dos currículos não indicam uma grande área e, conseqüentemente, a área específica de atuação não é informada. A exemplo dos currículos sem níveis de formação informados, em sua maioria, os currículos sem grande área informada não possuem data de atualização recente e, em geral, possuem apenas dados básicos como dados de identificação e endereço eletrônico para acesso ao currículo. Ou seja, tais currículos não apresentam dados relevantes além da identificação do indivíduo, já que várias outras seções não foram preenchidas. Aproximadamente, 71% dos currículos sem grande área informada não são atualizados há mais de dois anos. A grande área de Ciências da Saúde e a opção Outra se caracterizam como aquelas com maior quantidade de currículos cadastrados na Plataforma Lattes, totalizando 522.365 (12,57%) e 595.812 (14,33%), respectivamente. Já a grande área de Engenharias possui a maior quantidade de áreas (quatorze), contendo, no entanto, apenas 208.079 (5,01%) currículos cadastrados, o que mostra que a quantidade de áreas não está relacionada com a quantidade de indivíduos nas grandes áreas.

Cada uma das grandes áreas possui indivíduos com perfis distintos com relação ao maior nível de formação informado nos currículos, como pode ser observado na Tabela II (no apêndice). Currículos com a opção Outra, são aqueles com maior percentual sem área informada (93,18%), e também com o maior percentual sem formação informada (79,87%), além de corresponderem aos indivíduos com o maior nível de formação em cursos técnico-profissionalizantes. Ressaltam-se, ainda, as grandes áreas de Ciências Exatas e da Terra e Engenharias, com significativa quantidade de currículos neste nível de formação, sendo uma hipótese para tal fenômeno, a característica das áreas que as compõem na formação de pessoal de nível técnico. A grande área de Ciências da Saúde possui a maior quantidade de currí-

culos com o maior nível de formação em graduação e, apesar de ser a segunda grande área com o maior número de currículos cadastrados, não possui a mesma representatividade nos níveis mais altos de formação. Neste aspecto, destaca-se a grande área de Ciências Humanas, que possui a maior quantidade de indivíduos com mestrado e doutorado concluídos, mesmo sendo apenas a quarta grande área com maior quantidade de currículos.

Independentemente da área de atuação, a identificação dos pesquisadores mais produtivos em cada conjunto de currículos analisados pode indicar autoridades e quais destes conjuntos são mais produtivos, no contexto deste trabalho considerando a quantidade de publicações. É possível identificar que apesar da significativa quantidade de currículos cadastrados na Plataforma Lattes, somente uma parcela desses possuem dados suficientes que possibilitam realizar análises específicas como as análises bibliométricas, análises baseadas em redes de colaboração científica e/ou sobre a qualidade e impacto de suas publicações. A identificação de uma parcela que represente os pesquisadores mais produtivos não é uma tarefa trivial, tendo em vista a dificuldade da definição dos critérios para a seleção. Logo, neste trabalho uma análise considerando o maior nível de formação dos indivíduos foi realizada (Tabela III, no apêndice).

Como pode ser observado, os indivíduos com maior nível de formação concluída são, conseqüentemente, os mais produtivos. Existe um crescimento gradual em praticamente todos os tipos de publicação entre os níveis de formação mais baixos para os mais elevados, em que o conjunto que possui titulação de doutor se apresenta bem mais produtivo que todos os outros. Apesar de ser o grupo com o segundo menor número de currículos, os doutores possuem o maior percentual de publicações dentre todos os grupos considerados. Destaca-se a quantidade de publicações de artigos em periódico dos doutores, que detêm 74,51% de todos os artigos deste tipo de publicação cadastrados nos currículos da Plataforma Lattes. Esta quantidade de artigos faz com que a média de artigos em periódico para cada doutor (16,55) seja bem superior à dos outros grupos. Logo, diante do exposto, o restante deste trabalho irá se concentrar no conjunto de indivíduos com doutorado concluído.

3 Produção científica dos doutores

Para análises detalhadas da produção científica brasileira, considerou-se o conjunto de doutores com currículos cadastrados na Plataforma Lattes. O motivo da seleção desse conjunto de indivíduos para análises específicas tem como motivação os resultados apresentados na Tabela III, no qual foi possível observar que, apesar desse conjunto corresponder a somente 5,38% do total de indivíduos, ele é responsável por 74,51% dos artigos publicados em periódico e 64,67% dos artigos publica-

dos em anais de congresso. Além disso, em geral esse conjunto de currículos possui data de atualização recente e praticamente em sua totalidade possui algum tipo de publicação registrada. Este grupo de indivíduos que, em sua maioria, tem atuado em pesquisas, seja em instituições de ensino seja em institutos de ciência e tecnologia, ainda é responsável pela formação dos alunos nos principais programas de pós-graduação stricto sensu no Brasil, e vários deles são reconhecidos por sua elevada produção científica.

Aproximadamente 70% dos currículos dos doutores foram atualizados há menos de um ano, destes, 39% já nos primeiros meses de 2015. Os currículos não atualizados recentemente, em geral, são de indivíduos que informaram apenas informações básicas e possuem poucos registros sobre publicações ou dados sobre atuações profissionais. A vantagem em se analisar um conjunto de currículos recentemente atualizado como o dos doutores é que trabalhos recém-publicados provavelmente já estarão registrados nesses currículos, proporcionando uma visão atual da produção científica brasileira.

As grandes áreas com o maior número de doutores são Ciências da Saúde e Ciências Humanas, respectivamente, com quantidades muito próximas. Especificamente no caso da grande área de Ciências Sociais Aplicadas, que é a segunda grande área com maior número de indivíduos com currículos cadastrados na Plataforma Lattes, esta surge como a quinta grande área com o maior número de doutores, mostrando não haver qualquer correlação entre a quantidade de indivíduos e a quantidade de doutores dentro de uma mesma grande área. Além disso, apenas cerca de 6% dos doutores não informaram sua grande área em seus currículos, mostrando que a grande maioria dos doutores teve o cuidado de especificar suas áreas de atuação, possibilitando análises mais confiáveis. Tendo em vista a pouca quantidade de currículos com a opção Outra, optou-se por desconsiderá-la para as análises específicas aqui apresentadas.

A partir da Figura 5, é possível verificar que a quantidade de doutores tem aumentado consideravelmente a cada ano. O doutorado concluído mais antigo identificado no conjunto de indivíduos analisado é de 1932, seguido por outro concluído em 1939 e, posteriormente, um terceiro em 1944. Após o número de conclusões aumenta de forma muito modesta até meados de 1970 com algumas dezenas de doutores formados por ano. De 1975 até 1996, ocorreu um crescimento constante no número de doutores formados e, já em 1997, houve um aumento significativo nesse número que não se manteve no ano seguinte, mas que voltou a se manifestar de forma constante a partir de 1998. Entre 2011 e 2013, a taxa de crescimento aumenta novamente de forma significativa, sofrendo, no entanto, uma queda em 2014. Esse período coincide exatamente com a expansão dos cursos de doutorado ocorrido nos últimos

anos no país, conforme destacado nas avaliações trienais da CAPES de 2007 e 2010.

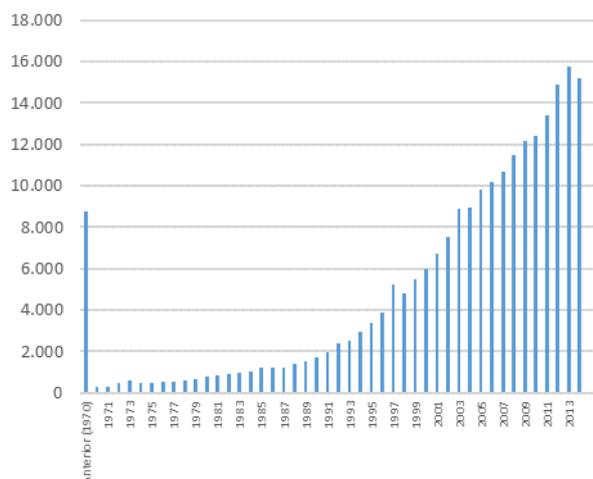


Figura 5. *Quantidade de doutores formados por ano*

Tendo em vista a expansão dos cursos de doutorado ocorrida a partir do final da década de 1990, é apresentada uma análise da produção científica dos doutores cadastrados na Plataforma Lattes, com ênfase nos artigos publicados em anais de congresso e em periódico. A exemplo do que ocorre com a produção científica de todo o conjunto de indivíduos com currículos cadastrados na Plataforma Lattes (Dias, 2016), a produção científica dos doutores também tem sofrido queda em todos os tipos de publicação analisados, sendo mais significativa no caso dos artigos em anais de congresso (Figura 6 no apêndice).

Como os doutores são responsáveis pela maioria dos trabalhos publicados no país, era esperado que esse comportamento fosse mesmo semelhante ao do conjunto de todos os indivíduos com currículos cadastrados na Plataforma Lattes. Assim, em praticamente todos os tipos de publicação, há uma queda acentuada a partir de 2012, ano em que se encerra um período de crescimento constante, iniciado na década de 1990. Importante destacar que, para todos os tipos de publicação, só foram considerados trabalhos únicos, ou seja, trabalhos em colaboração que estão cadastrados em mais de um currículo são contabilizados uma única vez.

Os artigos em periódico, cuja produção teve crescimento considerável a partir de 1992, atingem seu ápice em 2012, com 145.148 publicações. No entanto, a partir de 2013 inicia-se uma queda, sendo que em 2014 foram registradas apenas 126.480 publicações, queda de 11% em relação a 2013, ano com maior número de doutorados concluídos. Já os artigos em anais de congresso possuem características distintas dos demais tipos de publicação, tendo seu crescimento acentuado a partir de 1990, mantendo-se até 2005, ano em que há uma pequena queda e, conseqüentemente, um comportamento irregular em suas taxas de crescimento, com variações

positivas e negativas bem próximas de 0. No entanto, a partir de 2011, percebe-se uma queda acentuada, reduzindo o número de publicações em, aproximadamente, 40% em um período de apenas três anos. Para melhor compreensão e, ainda, para tentar encontrar justificativas que expliquem as razões da recente queda em todos os tipos de publicação, foi calculada a produção per capita do conjunto de doutores (Figura 7 no apêndice).

Para o cálculo da produção per capita foi considerado o número de publicações em cada um dos tipos em cada ano, tendo como universo de pesquisadores aqueles que tiveram alguma publicação daquele tipo no ano em questão. A divisão desses valores aqui denominada de produção per capita representa a taxa de publicação, desconsiderando todos os pesquisadores que não publicaram em um determinado ano. Como pode ser observado, há um crescimento nos últimos anos em todos os tipos de publicação, exceto no caso dos artigos em anais de congresso no último ano (2014). Baseado na quantidade de publicações e considerando o conjunto dos doutores analisados, era de se esperar que a tendência dos valores da produção per capita reduzisse ao longo dos últimos anos, devido à queda geral na produção científica. Porém, não é o que ocorre, ou seja, apesar de recentemente ter havido uma queda geral na produção científica, a produção per capita aumentou nesse mesmo período, exceto em 2014 no caso de artigos em anais de congresso, resultado de uma queda na quantidade de doutores que têm publicado a cada ano, superior à de trabalhos publicados.

Com o intuito de verificar a hipótese sobre a queda do número de autores, foi realizada uma comparação entre o número de publicações e a quantidade de autores que publicaram no ano em questão. Foi observado que a quantidade de autores que publicaram nos últimos anos também sofre uma redução considerável a exemplo do que acontece com a quantidade de publicações. Neste cálculo, um autor somente é contabilizado se teve, pelo menos, uma publicação naquele referido ano. Percebe-se que a quantidade de autores que publicaram em anais de congresso atinge seu ápice em 2007 (115.809) e, a partir deste ano, sofre uma queda anual constante. Já com relação aos autores que publicaram em periódicos, a queda se inicia após 2012, ano com a maior quantidade de autores (96.079), sofrendo queda considerável nos dois próximos anos. Assim, com o passar do tempo, o número de doutores com produção científica tem diminuído.

Diante disso, as produções científicas de dois subconjuntos dos doutores brasileiros foram analisadas com o propósito de melhor compreender considerável redução no número de publicações produzidas nos últimos anos. Para a seleção dos dois subconjuntos, foi utilizado o LattesDataXplorer, que analisou as informações de duas listagens, contendo a relação de bolsistas de Produtividade em Pesquisa do CNPq, e dos atuais docentes em Programas de Pós-graduação. O primeiro

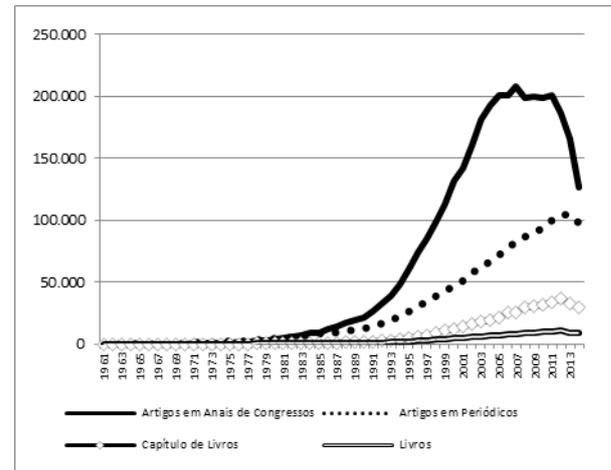
subconjunto analisado corresponde a 63.300 docentes que atuam em programas de pós-graduação no Brasil. Além disso, um segundo subconjunto, contendo 14.394 bolsistas de Produtividade em Pesquisa do CNPq, também foi considerado, sendo que, do total de bolsistas, 98,5% também são docentes em programas de pós-graduação, e conseqüentemente, também estão presentes no subconjunto de docentes. A escolha desses grupos tem como motivação, possibilitar obter uma visão sobre o perfil dos docentes, que, atualmente, são responsáveis pela formação dos mestres e doutores no Brasil, bem como, comparar a produção científica desses docentes com a dos principais pesquisadores brasileiros que recebem auxílio pela excelência de suas pesquisas em suas áreas de atuação.

O primeiro conjunto foi selecionado a partir da relação dos docentes de programas de pós-graduação, obtida a partir dos dados do Coleta CAPES 2013¹¹. Apesar de tal listagem ser referente aos docentes que estavam atuando nos programas no final de 2013, todos os dados a serem analisados foram extraídos dos currículos coletados em maio de 2015.

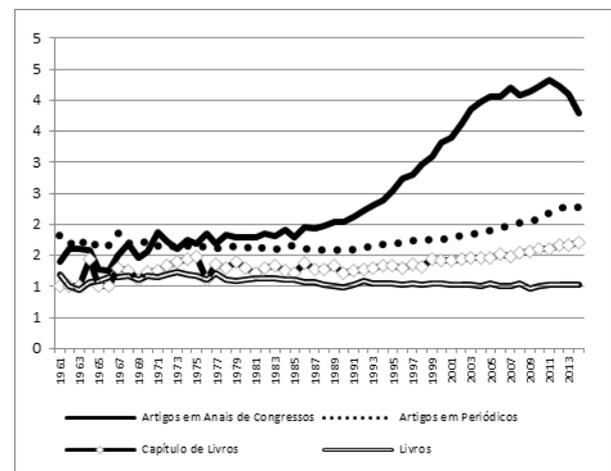
Já o segundo subconjunto analisado considera todos os atuais pesquisadores que recebem bolsas de produtividade em pesquisa do CNPq. Eles foram selecionados a partir da informação inserida automaticamente nos currículos pelo próprio CNPq. A bolsa de produtividade em pesquisa é destinada aos pesquisadores que se destacam entre seus pares, valorizando sua produção científica segundo critérios normativos definidos pelo CNPq e pelos seus Comitês de Assessoramentos (CAs). Dentre os principais critérios para a concessão da bolsa, o solicitante deve ter o título de doutor ou perfil científico equivalente e dedicar-se a atividades de pesquisa. As bolsas são divididas em modalidades, sendo a modalidade 1 subdividida nos níveis, A, B, C e D. Já na modalidade 2 não há subdivisão de nível. O acesso às bolsas por um determinado pesquisador se inicia pela modalidade 2 e, ao longo do tempo, pode ser promovido por progressões até chegar à modalidade 1A, mais elevado nível, dedicado a pesquisadores que tenham demonstrado excelência continuada na sua produção científica e na formação de recursos humanos, e que liderem grupos de pesquisa consolidados. Logo, o grupo de pesquisadores bolsistas de produtividade em pesquisa se caracteriza por ser um segmento significativo da excelência científica brasileira.

Ao se analisar a produção científica dos dois subconjuntos selecionados, destaca-se que, com relação aos docentes de pós-graduação, após o início das orientações, significativa parte de seus trabalhos passa a ser realizada, cada vez mais, de forma colaborativa com alunos ou com seus pares que compõem seus grupos de pesquisa. Logo, uma análise da produção científica desses dois subconjuntos de pesquisadores se faz importante. A Figura 8 apresenta a produção geral (8a) e

per capita (8b) dos docentes de pós-graduação ao longo dos anos.



(a)



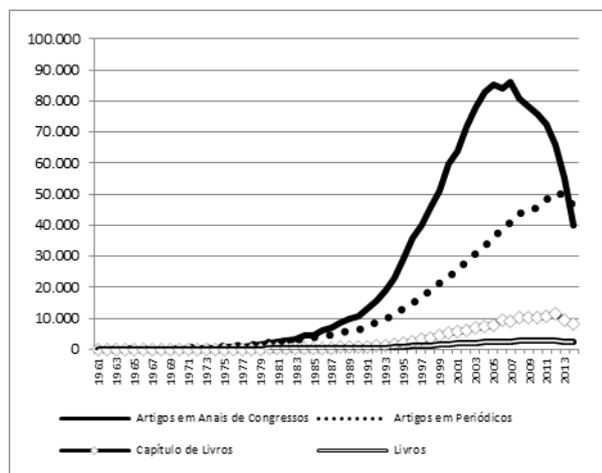
(b)

Figura 8. Produção científica dos docentes de pós-graduação

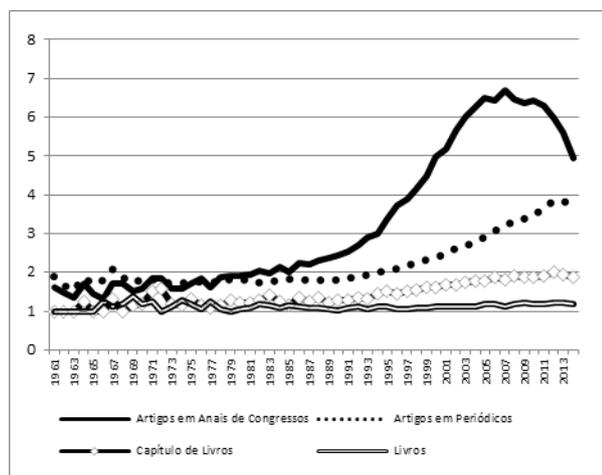
A exemplo do que ocorre com a análise da produção científica dos doutores, e tendo em vista que os docentes de pós-graduação são um subconjunto relevante dos doutores, observa-se também uma queda na produção científica desse subconjunto em todos os tipos de publicação nos últimos anos. No entanto, a queda mais acentuada manifesta-se nos artigos em anais de congresso. Esta categoria de publicação tem seu ápice em 2007. Em seguida, sofre uma queda acentuada a partir de 2011 e, após três anos, apresenta uma redução de aproximadamente 37%. Fato este que pode ser melhor observado com relação a produção per capita dos artigos em anais de congresso, que se caracteriza como a única com queda acentuada nos últimos anos, resultado de que a cada ano, menos docentes deixam de publicar este tipo de produção científica. Dentre as hipóteses para explicar tal fenômeno, a que melhor se aplica para os docentes dos programas de pós-graduação é a que diz respeito às avaliações da CAPES, que deixam de considerar os artigos em anais de congresso, dando

ênfase aos artigos em periódico. Enfatiza-se, aqui, que a avaliação dos programas tem dependência direta com a produção científica dos docentes. Com relação aos artigos em periódico, o ano de 2012 apresenta a maior produção, sofrendo uma sensível queda em 2013, que se mostra ainda mais acentuada em 2014. Porém, é preciso destacar que o ano de 2014 se caracteriza como aquele em que a diferença entre a quantidade de artigos em anais de congresso e artigos em periódico é a menor dos últimos anos (29.742), caracterizando uma tendência dos docentes de pós-graduação em concentrar a publicação de seus artigos em periódico.

Ao analisar a produção científica dos bolsistas de produtividade em pesquisa, que, conforme já mencionado, são os pesquisadores de excelência em suas áreas de atuação e se caracterizam por publicações de maior impacto, nota-se comportamento semelhante à produção científica dos docentes, mas com queda na publicação de artigos em anais de congresso, ainda mais intensa (ver Figura 9a produção geral e 9b produção per capita).



(a)



(b)

Figura 9. Produção científica dos bolsistas de produtividade

Como o conjunto de bolsistas é praticamente um subconjunto dos docentes de pós-graduação, era de se esperar um comportamento similar ao observado na análise de todo o conjunto de docentes e, consequentemente, dos doutores. Porém, vale destacar a queda acentuada na quantidade de artigos em anais de congresso, tornando os artigos em periódico o principal tipo de publicação dos bolsistas de produtividade em pesquisa a partir de 2013. Este fato, que inicialmente ocorre com os bolsistas que passam a publicar mais artigos em periódico do que artigos em anais de congresso, pode representar uma tendência em relação aos outros conjuntos de pesquisadores aqui analisados. Tal fato poderá influenciar diretamente no perfil da produção científica brasileira, tendo em vista que os bolsistas de produtividade em pesquisa são os pesquisadores de excelência responsáveis pela formação de grande parte dos novos pesquisadores e, por isso, também influenciam o perfil de publicações desses novos pesquisadores.

4 Considerações Finais

Neste trabalho, inicialmente foi apresentada uma análise de todo o conjunto de dados curriculares da Plataforma Lattes. Os currículos da Plataforma Lattes têm sido objeto de estudo de diversos trabalhos que objetivam verificar como tem evoluído a produção científica brasileira. Foi possível analisar a produção científica, bem como área de atuação e formação acadêmica de todo o conjunto de currículos cadastrados na Plataforma Lattes, apresentando uma visão da comunidade científica brasileira de forma ampla. Após analisar a produção científica dos doutores, que são responsáveis pela maioria das publicações cadastradas nos currículos da Plataforma Lattes, foi possível apresentar uma caracterização do perfil de publicação e área de atuação desses pesquisadores.

Na análise da produção científica dos doutores brasileiros, foi identificado que, apesar de corresponderem à apenas 5,38% do conjunto total de indivíduos cadastrados na Plataforma Lattes, eles são detentores de aproximadamente, 74% dos artigos publicados em periódicos e 65% dos artigos em anais de congressos de todo o repositório extraído. Logo, em um estudo que vise analisar todo o conjunto de dados da Plataforma Lattes, este conjunto se torna bem representativo, apesar de ser uma pequena fração de todos os indivíduos. Em análise a suas publicações nos últimos anos, a quantidade de trabalhos publicados tem sofrido uma considerável queda dentre os vários tipos de publicação, em especial, artigos em anais de congresso e artigos em periódico.

A exemplo do que ocorre com a produção científica do conjunto de todos os doutores, há uma queda, também, na produção do conjunto de docentes de pós-graduação e bolsistas de produtividade em pesquisa do CNPq, queda esta que se mostra mais acentuada nos artigos em anais de congresso. Neste aspecto, destacam-se os

bolsistas que, a partir de 2013, com a queda acentuada na publicação de artigos em anais de congresso, passam a ter como principal tipo de publicação artigos em periódico. Tal resultado parece ter relação direta com os critérios de avaliação da CAPES, conforme já discutido, que influenciam diretamente os grupos analisados, tendo em vista que estão vinculados a programas de pós-graduação e a produção científica de tais programas depende, diretamente, da produção de seus docentes.

Conclui-se que apesar de um crescimento significativo ao longo dos anos na produção de artigos científicos, percebe-se que nos dois últimos, uma queda acentuada é identificada. Tal queda pode estar relacionada a atualização dos trabalhos, o que se torna uma tarefa complexa de ser comprovada, tendo em vista que alguns autores não mantêm seus currículos atualizados com frequência.

Notas

- (1) DBLP: <http://dblp.uni-trier.de/db/>
- (2) ArnetMiner: <http://arnetminer.org/>
- (3) Google Scholar: <http://scholar.google.com>
- (4) Microsoft Academic:
<http://academic.research.microsoft.com/>
- (5) PubMed: <http://www.ncbi.nlm.nih.gov/pubmed>
- (6) Scopus: <http://www.elsevier.com/online-tools/scopus>
- (7) ResearchGate: <http://www.researchgate.net/>
- (8) Academia.edu: <http://www.academia.edu>
- (9) Lattes Platform: <http://lattes.cnpq.br>
- (10) CNPq: <http://www.cnpq.br>
- (11) <http://www.capes.gov.br/avaliacao/plataforma-sucupira/modulo-coleta-de-dados>

Referencias

- Alves A D, Yanasse H H, Soma N Y. (2011). LattesMiner: a multilingual DSL for information extraction from lattex platform. // Proceedings of the compilation of the co-located workshops on DSM'11, TMC'11, AGERE'11, AOOPEs'11, NEAT'11, VMIL'11. Portland, Oregon, USA: ACM: p. 85-92.
- Arruda, D. et al. (2009). Brazilian computer science research: Gender and regional distributions. // *Scientometrics* 79:3 (2009) 651-665.
- Brito, A. G. C., Quoniam, L., Mena-Chalco, J. P. (2016). Exploração da Plataforma Lattes por assunto: proposta de metodologia. // *TransInformação* 28:1 (2016) 77-86.
- Colugnati F. A. B, Firpo S., Castro P. F. D., Sepulveda J. E., Salles-Filho S. L. M. (2014). A propensity score approach in the impact evaluation on scientific production in Brazilian biodiversity research: the BIOTA Program. // *Scientometrics* 101: 85-107.
- Costa B. M. G, Pedro E. S. E Macedo G. R. (2013). Scientific collaboration in biotechnology: The case of the northeast region in Brazil. // *Scientometrics* 95: 571-592.
- Dias T. M. R. (2016). Um estudo da produção científica brasileira a partir de dados da plataforma lattex. 181p. // Programa de Pós-Graduação em Modelagem Matemática e Computacional, Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte (Doutorado).
- Dias T. M. R, Moita G. F. (2015). A method for the identification of collaboration in large scientific databases. // *Em Questão* 21 (2015) 140-161.
- Digiampietri L. A., Mena-Chalco J. P., Melo P. O. S. V., Malheiro A. P. R., Meira D. N. O., Franco L. F., Oliveira L. B. (2014). BraX-Ray: An X-Ray of the Brazilian Computer Science Graduate Programs. // *PLoS ONE* 9, 2014.
- Furtado, C. A. et al. (2015). A Spatiotemporal Analysis of Brazilian Science from the Perspective of Researchers' Career Trajectories. // *PLoS ONE* 10(10): e0141528 (2015). <https://doi.org/10.1371/journal.pone.0141528>
- Lane J. (2010). Let's make science metrics more scientific. // *Nature* 464 (2010) 488-489.
- Lima H., Silva T. H. P., Moro M. M., Santos R. L. T., Wagner J. M., Laender A. H. F. (2015). Assessing the profile of top Brazilian computer science researchers. // *Scientometrics* 103 (2015) 879-896.
- Mena-Chalco J. P., Digiampietri L. A., Lopes F. M., Cesar-Junior R. M. (2014). Brazilian bibliometric coauthorship networks. // *J Assoc Inform Sci Technol* 65 (2014) 1424-1445.
- Nascimento C., Laender A. H. F., Silva A. S, Gonçalves M. A. A. (2011). Source independent framework for research paper recommendation. // *Proceedings of the 11th Ann. Int. ACM/IEEE Joint Conf. Digital libraries, ACM*. 296-306 p.
- Roos, D. H. et al. (2014). Brazilian scientific production in areas of biological sciences: a comparative study on the modalities of full doctorate in Brazil or abroad. // *Scientometrics*. 98: 1 (2014) 415-427.
- Sidone, O. J. G.; Haddad, E. A.; Mena-Chalco, J. P. (2016). A Ciência nas Regiões Brasileiras: Evolução da Produção e das Redes de Colaboração Científica. *Transinformação*. 28:1 (2016) 15-31.

Copyright: © 2018. Dias e Moita. This is an open access article distributed under the terms of the Creative Commons CC Attribution-ShareAlike (CC BY-SA), which permits use, distribution, and reproduction in any medium, under the identical terms, and provided the original author and source are credited.

Received:2018-04-18 Accepted: 2018-12-02

Apêndice

Tabela II.

Grande Área / Maior Nível de Formação	Total de Currículos	Sem Formação Declarada	Curso Técnico / Profissionalizante	Graduação	Especialização	Mestrado	Doutorado
Ciências Agrárias	145.879	54.108	6.512	35.484	7.478	22.362	19.935
Ciências Biológicas	151.617	43.966	2.606	38.820	12.118	27.399	26.708
Ciências da Saúde	522.365	169.120	7.786	173.058	88.183	50.390	33.828
Ciências Exatas e da Terra	269.292	86.754	10.584	74.939	30.373	36.717	29.925
Ciências Humanas	404.488	105.751	3.677	120.314	78.520	62.240	33.986
Ciências Sociais Aplicadas	478.860	138.765	6.050	162.496	92.073	59.488	19.988
Engenharias	208.079	79.807	9.826	56.818	15.529	27.144	18.955
Linguística, Letras e Artes	156.347	46.493	1.812	48.436	25.580	22.432	11.594
Outra	595.812	475.904	12.092	87.858	13.502	4.688	1.768
Total	2.932.739	1.200.668	60.945	798.223	363.356	312.860	196.687

Tabela II. Distribuição dos currículos por grandes áreas de atuação e maior nível de formação

Tabela III.

Produção / Maior Nível de Formação	Sem Formação Declarada / Ensino Médio	Curso Técnico / Profissionalizante	Graduação	Especialização	Mestrado	Doutorado
Total de Currículos	1.857.980 44,70%	89.280 2,15%	1.083.439 26,07%	529.440 12,73%	373.019 8,97%	223.477 5,38%
Artigos em Anais de Congressos	283.367 2,44%	34.534 0,31%	802.993 6,93%	433.959 3,74%	2.540.123 21,91%	7.496.166 64,67%
Artigos em Periódicos	32.285 0,71%	3.692 0,08%	162.556 3,56%	130.813 2,87%	833.275 18,27%	3.398.300 74,51%
Capítulos de Livros	7.191 0,68%	587 0,06%	36.145 3,42%	37.406 3,54%	163.144 15,46%	810.915 76,84%
Livros	3.675 0,80%	315 0,07%	19.288 4,23%	26.357 5,79%	85.776 18,84%	320.036 70,27%
Textos em Jornais e Revistas	25.357 1,87%	2.297 0,17%	107.834 7,97%	132.890 9,82%	310.819 22,96%	774.448 57,21%
Média de Artigos em Periódicos	0,02	0,04	0,15	0,25	2,23	15,20

Tabela III. Distribuição científica por nível de formação

Figura 2.

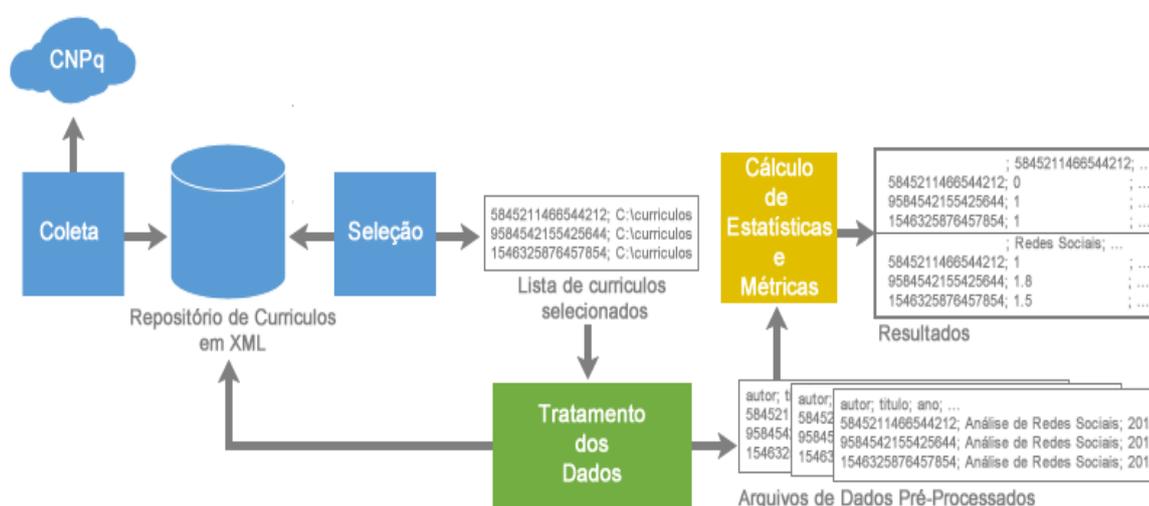


Figura 2. Visão Geral do LattesDataExplorer (Dias, 2016).

Figura 6.

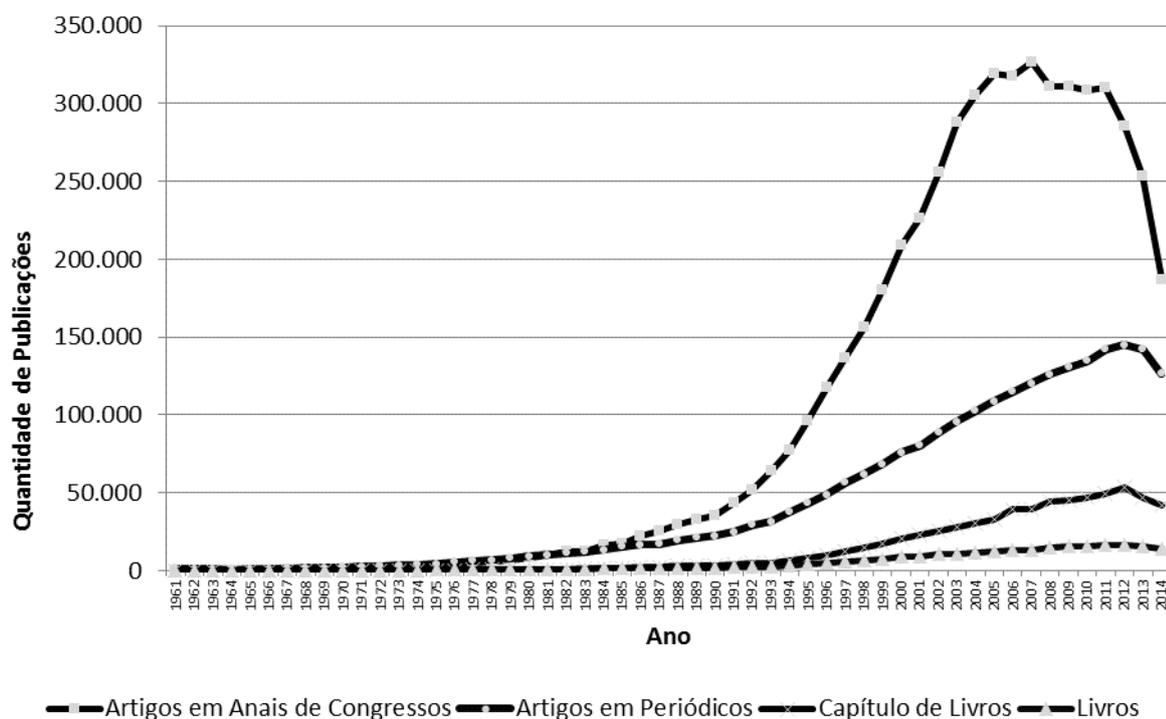


Figura 6. Quantidade de publicações ao longo dos anos

Figura 7.

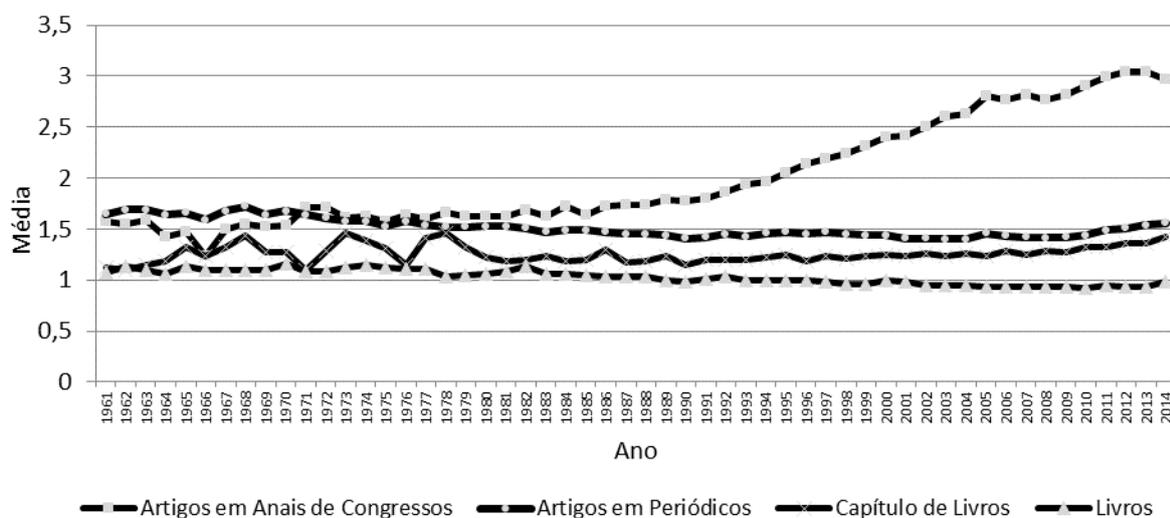


Figura 7. Produção per capita do conjunto analisado