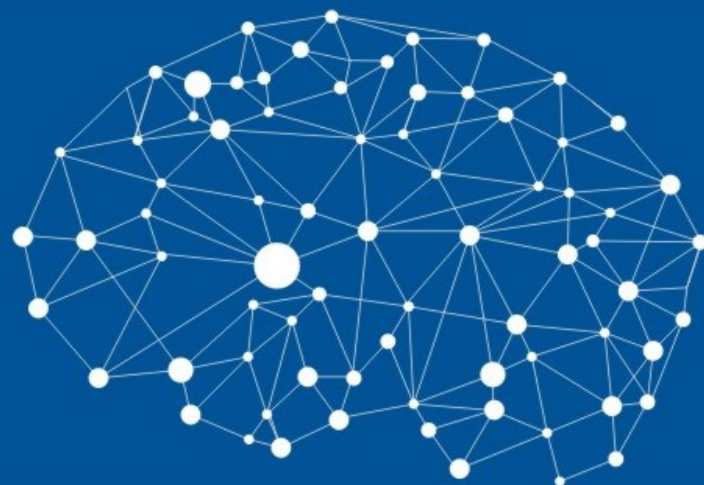




Caracterização dos Bolsistas de Produtividade em Pesquisa do CNPq a partir de Dados da Plataforma Lattes

Thiago Magela Rodrigues Dias¹; Leandro Santos²; Tales Moreira³; Gray Moita⁴

DIAS, T. M. R.; et al. Caracterização dos Bolsistas de Produtividade em Pesquisa do CNPq a partir de Dados da Plataforma Lattes In: ENCONTRO BRASILEIRO DE BIBLIOMETRIA E CIENTOMETRIA, 5., 2016, São Paulo. **Anais...** São Paulo: USP, 2016. p. A90

Caracterização dos Bolsistas de Produtividade em Pesquisa do CNPq a partir de Dados da Plataforma Lattes

Eixo temático: Produção e Produtividade Científica

Modalidade: Apresentação oral

1 INTRODUÇÃO

A extração de dados de produção científica e a modelagem e visualização efetiva de redes de coautoria são assuntos de pesquisa atuais [Dias et al. 2013]. Esse aumento da produção científica pode ser explicado pela maior disseminação de materiais de cunho científico na *Internet*. Serviços como repositórios bibliográficos têm sido cada vez mais utilizados por pesquisadores para armazenar sua trajetória acadêmica/científica [Dias e Moita, 2015]. Nesse âmbito, tem sido dado especial interesse à prospecção de dados em repositórios de produção científica, como a Plataforma Lattes [Mena-Chalco e Cesar-Junior, 2011].

Os currículos presentes na Plataforma Lattes são denominados currículos Lattes e são considerados um padrão nacional para análise do mérito e competência dos pesquisadores cadastrados, pois reúne toda a trajetória acadêmica do mesmo [Amorim, 2003]. Além disso, os currículos são monitorados pela própria comunidade científica, o que fortalece a qualidade e, conseqüentemente, a veracidade das informações contidas em cada currículo. Pode-se extrair diversos tipos de informação utilizando a Plataforma Lattes e ainda analisar grupos específicos de pesquisadores, como os bolsistas de produtividade em pesquisa do CNPq.

A bolsa de produtividade em pesquisa do CNPq é destinada aos pesquisadores que se destaquem entre seus pares na realização de pesquisas nas áreas científica e tecnológica. Dessa forma, o CNPq oferece um estímulo aos segmentos da pesquisa de excelência no país, valorizando a qualidade, o aprofundamento e a possível aplicação dos estudos.

Portanto, a relevância do presente trabalho se dá na comparação entre os pesquisadores bolsistas e não-bolsistas de um modo geral. A análise é realizada através de métodos estatísticos de sumarização dos dados. O perfil dos pesquisadores bolsistas de produtividade em pesquisa tem sido muito pesquisado em diversas áreas do conhecimento [Alves et al., 2014; Lima et al., 2015]. O presente trabalho fará primeiramente uma análise dos pesquisadores bolsistas de um modo geral, analisando seus perfis e produtividade.

2 MATERIAL E MÉTODOS

Para a análise do perfil dos Bolsistas de Produtividade do CNPq foi utilizado os dados da Plataforma Lattes como principal fonte de informação. A Plataforma Lattes, criada em 1997 pelo CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) com a participação do Grupo Stela do PPGE/UFSC, tornou-se conhecida da comunidade científica tanto no Brasil quanto em outros países americanos e europeus. O principal componente da plataforma é o Currículo Lattes, com aproximadamente 4 milhões de currículos.

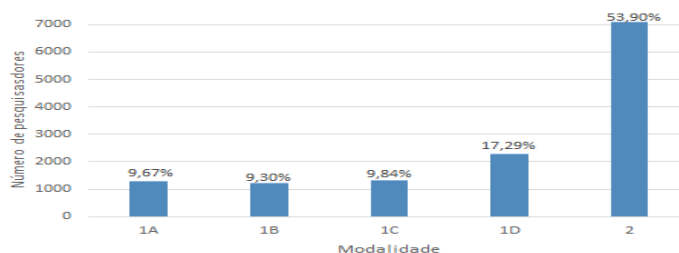
A Bolsa de Produtividade em Pesquisa é organizada em três categorias, 1, 2 e Sênior, sendo que a categoria 1 possui quatro níveis, em ordem crescente: 1D, 1C, 1B e 1A. Cada nível provê uma mensalidade crescente, havendo um salto maior entre a categoria 2 e 1D.

Para as análises aqui apresentadas, todo o processo de extração é realizado pelo *framework* proposto por [Dias et al., 2013]. Esse *framework* é responsável por realizar a extração de toda a base de Currículos Lattes ou de grupos específicos de currículos, como por exemplo dos bolsistas de produtividade, objeto de estudo deste trabalho.

No presente trabalho, é avaliado um conjunto bem definido de pesquisadores de acordo com diferentes aspectos. Para isso, foi necessário a criação de um conjunto de dados com diversas informações sobre cada pesquisador. As etapas realizadas ao construir o conjunto de dados que será utilizado para análise são as seguintes:

Etapa 1: *Levantamento de informações sobre as bolsas e os bolsistas de produtividade em pesquisa do CNPq.* Considerando todos os potenciais pesquisadores contidos na Plataforma Lattes, foi definido que o presente artigo trataria o subconjuntos dos pesquisadores que possuem bolsa de produtividade em pesquisa do CNPq. Tais bolsas, divididas em 5 níveis, são concedidas através de critérios avaliados por 46 Comitês de Assessoramento. Existem 14.154 pesquisadores de todas as áreas distribuídos entre os 5 níveis da bolsa, como é mostrado na Figura 1. Note de 46% dos pesquisadores estão nas modalidades 1A-1D e o restante, 54%, estão na modalidade 2. É importante salientar que a modalidade 2 é a porta de entrada para a bolsa, concentrando a maior parte dos bolsistas.

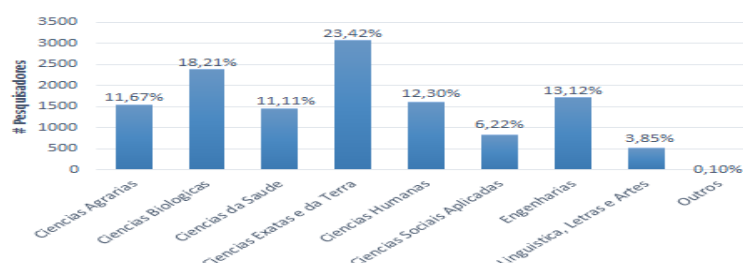
Figura 1: Distribuição dos bolsistas por modalidade



Fonte: Próprios autores

Etapa 2: Organização dos dados. Nessa etapa foi construído o perfil de cada um dos 14.154 pesquisadores extraindo dados da Plataforma Lattes. Sendo que o perfil traçado é composto pela formação acadêmica, produção científica e orientação de estudantes. A Figura 2 mostra a distribuição dos pesquisadores bolsistas por grande área do conhecimento.

Figura 2: Distribuição por grande área



Fonte: Próprios autores

Etapa 3: Tratamento do perfil dos pesquisadores. Os dados referentes à produção científica precisam ser tratados antes de serem analisados. As publicações de cada autor podem não ser agrupadas devido a alguma ambiguação de nomes: o mesmo autor pode aparecer com nomes diferentes, ou autores diferentes podem ter nomes similares, causando divisão ou mistura de citações. Para lidar com ambiguidade de nomes presentes no conjunto, foi utilizado um método estado-da-arte desenvolvido por Dias et al., 2015, denominado ISColl.

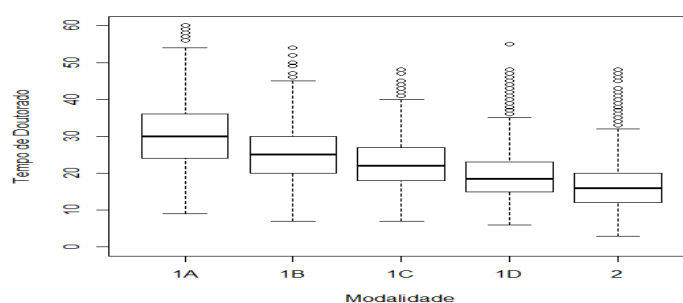
3 ANÁLISE DOS PESQUISADORES

A fim de caracterizar uma visão geral dos bolsistas PQ, são avaliados diversos aspectos relativos à carreira dos pesquisadores. Especificamente, duas variáveis são importantes e fáceis de definir: tempo de carreira e estudantes orientados. Primeiramente foi

analisado como o tempo de carreira impacta na produtividade do pesquisador. Consideremos que o tempo de carreira inicia-se com o ano em que o pesquisador terminou seu doutorado.

A Figura 3 mostra a distribuição de tempo desde que o pesquisador obteve o grau de doutor, agrupados pela modalidade da bolsa. Observa-se que o valor da mediana aumenta gradativamente da modalidade 2 em direção à modalidade 1A, formando um caminho de evolução. Consequentemente, pode-se dizer que os pesquisadores mais experientes pertencem à modalidades mais altas.

Figura 3: Tempo de Doutorado



Fonte: Próprios autores

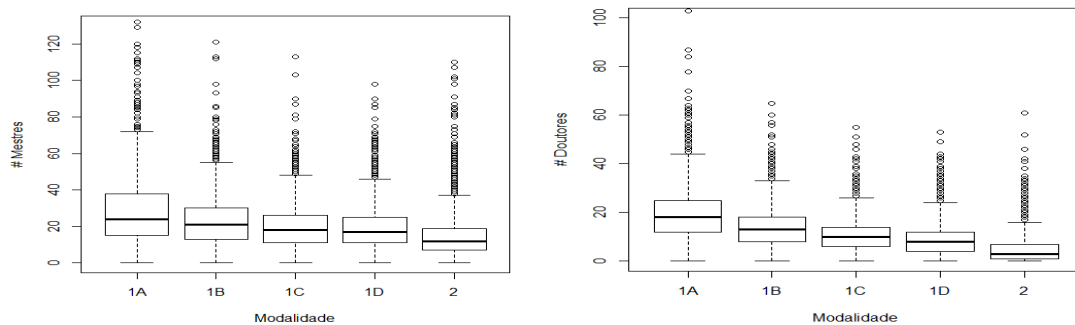
Observa-se que o intervalo entre as medianas das modalidades 2 até 1B são bem próximos. A diferença maior está entre a modalidade 1B e 1A. Para obter o nível 1A, o pesquisador deve ter contribuições nacionais e internacionais para a comunidade acadêmica, o que geralmente demora mais tempo e esforço.

A Figura 4 mostra o número de mestres e doutores orientados pelos pesquisadores, distribuídos entre as cinco modalidades da bolsa. Verifica-se que o valor da mediana diminui ao longo das modalidades, partindo da modalidade 1A até a 2, em ambos os gráficos. Na Figura 4a é possível observar que as medianas estão próximas em todas as modalidades. Já na Figura 4b, há uma diferença significativa entre as medianas das modalidades 1A e 1B.

Figura 4: Quantidade de Mestres e Doutores formados

(a) Número de Mestres formados

(b) Número de Doutores formados

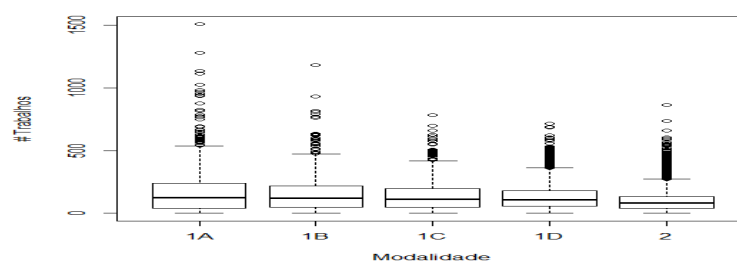


Fonte: Próprios autores

Outro aspecto importante para se caracterizar o perfil dos pesquisadores bolsistas é o volume da produção científica, que é composta por artigos em anais de congressos e em periódicos. A Figura 5 mostra a o volume da produção de artigos em anais de congressos, na qual observa-se que o valor das medianas estão bem próximas entre as modalidades 1A-1D, diminuindo consideravelmente na categoria 2.

A presença de *outliers*, embora poucos, em todas as modalidades da bolsa mostra que existem alguns pesquisadores com produção acima da média. Os Comitês de Assessoramento consideram outros fatores, além do volume de publicação, como critérios de avaliação.

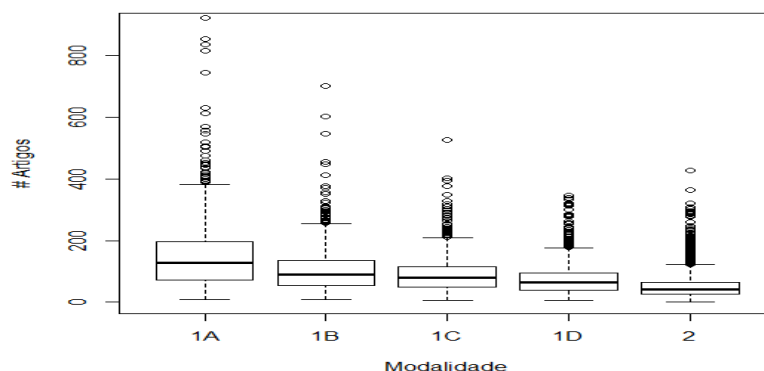
Figura 5: Artigos em anais de congressos



Fonte: Próprios autores

A diferença de produção da modalidade 1A para as demais fica ainda mais evidenciada ao observar a produção de artigos em periódicos, mostrada na Figura 6. Nota-se que o valor da mediana diminui gradativamente da modalidade 1A até a modalidade 2. Sendo que a diferença é mais expressiva entre as modalidades 1A e 1B. Também é importante salientar a presença de *outliers* em todas as modalidades.

Figura 6: Artigos em periódicos



Fonte: Próprios autores

Quanto mais alto o nível da bolsa, maiores são os critérios para se manter naquele nível. Dessa forma, a tendência natural é que os pesquisadores com níveis mais altos tenham uma média de publicações maior que os demais. Além disso, pesquisadores de nível 1A, por exemplo, tendem a orientar mais alunos de mestrado e doutorado, o que eleva a produção científica do pesquisador.

4 CONSIDERAÇÕES FINAIS

Este artigo apresentou uma visão geral do perfil dos bolsistas de produtividade em pesquisa do CNPq e o volume da produção científica em cada nível da bolsa. Foi observado que a maioria dos pesquisadores são bolsistas na modalidade 2. Tal modalidade é a porta de entrada dos pesquisadores no programa de bolsas fomentado pelo CNPq. Dos pesquisadores do nível 1A, grande parte possui de 20 a 40 anos de doutorado, sendo que 36,31% possuem entre 30 e 40 anos. Já os níveis 1B e 1C possuem um número maior de pesquisadores entre 15 e 20 anos de doutorado, entretanto se concentram na faixa de 20 a 25 anos. O nível 1B possui maior homogeneidade entre as faixas. O nível 1D possui pesquisadores entre 10 e 25 anos de doutorado. Por fim, a categoria 2 engloba pesquisadores entre 5 de 25 anos de doutorado, sendo que 33,62% possuem de 10 a 15 anos.

Em relação à distribuição da produção científica entre as modalidades, foi observado uma distribuição esperada na produção de artigos em periódicos, em que o valor da mediana diminui na direção dos níveis mais baixos, sendo que o nível 1A se destaca. Entretanto, ao considerar a produção de artigos em anais de congressos, os valores das medianas ficaram próximos nos níveis 1A-1D.

REFERÊNCIAS

- ALVES, A. D.; YANASSE, H. H.; SOMA, N. Y. Profile of CNPq Productivity Research Fellows in the Chemistry area based on the Lattes Plataform. **Química Nova**, SciELO Brasil, v. 37, p. 377-383, 2014.
- AMORIN, C. V. Curriculum Vitae Organization: The Lattes Software Platform. **Pesquisa Odontológica Brasileira**, SciELO Brasil, v. 17, p. 18-22, 2003.
- DIAS, T. M. R.; MOITA, G. F. A Method of the Identification of Collaboration in Large Scientific Databases. **Em Questão**, UFGRS, 2015.
- DIAS, T. M. R.; MOITA, G. F.; DIAS, P. M.; MOREIRA, T. H.; SANTOS, L. R.F S. Modelagem e Caracterização de Redes Científicas: Um Estudo sobre a Plataforma Lattes. In: **Brazilian Workshop on Social Network Analysis and Mining**. 2013.
- LIMA, H.; SILVA, T. H.; MORO, M. M.; SANTO, R. L.; JR, W.; LAENDER, A. H. Assessing the Profile of Top Brazilian Computer Science Researchers. **Scientometrics**, Springer, v. 103, n. 3, p. 879-896, 2015.
- MENA-CHALCO, J. P.; CESAR-JUNIOR, R. M. Prospecção de Dados Acadêmicos de Currículos Lattes Através de ScriptLattes. **Capítulo do livro Bibliometria e Cientometria: Reflexões Teóricas e Interfaces (in press)**. São Carlos, p. 1-20, 2011.