

Estudo do êxodo científico brasileiro a partir de dados em acesso aberto da Plataforma Lattes

Higor Alexandre Duarte Mascarenhas

Mestrando em Modelagem Matemática e Computacional pelo Centro Federal de Educação Tecnológica de Minas Gerais (CEFET/MG) - Belo Horizonte, MG - Brasil.

<http://lattes.cnpq.br/8723936697065308>

E-mail: higoralexandre1996@gmail.com

Thiago Magela Rodrigues Dias

Doutor em Modelagem Matemática e Computacional pelo Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG) - Belo Horizonte, MG - Brasil. Professor do Centro Federal de Educação Tecnológica de Minas Gerais (CEFET) - Divinópolis, MG - Brasil.

<http://lattes.cnpq.br/4687858846001290>

E-mail: thiagomagela@gmail.com

Patrícia Mascarenhas Dias

Doutoranda em Modelagem Matemática e Computacional pelo Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG) - Belo Horizonte, MG - Brasil. Professora da Universidade do Estado de Minas Gerais (UEMG) - Divinópolis, MG - Brasil.

<http://lattes.cnpq.br/6871965805554986>

E-mail: patricia.dias@uemg.br

Data de submissão: 13/09/2019. Data de aprovação no ConfOA: 12/06/2019. Data de publicação:

RESUMO

O êxodo de indivíduos por diversos motivos vem ao longo dos últimos anos crescendo significativamente no cenário brasileiro e mundial. Um dos motivos identificados no processo migratório é a formação acadêmica desses indivíduos que buscam se capacitar em instituições de ensino de melhor qualidade. Logo, o objetivo deste trabalho será analisar como ocorre o êxodo científico do Brasil, em que indivíduos de diversas localidades do país migram em busca de uma melhor formação acadêmica e científica. Logo, será caracterizada e analisada a ocorrência do êxodo científico brasileiro nas diversas áreas do conhecimento. Para isso foi necessária a extração dos dados acadêmicos, disponíveis em acesso aberto na Plataforma Lattes, selecionando todos os indivíduos com doutorado concluído, e totalizando 284.050 currículos. A escolha do grupo se justifica por se tratar do conjunto de mais alto nível de formação acadêmica, e que possui dados recentemente atualizados em seus currículos. Para tanto, filtraram-se os dados, selecionando os atributos relevantes à pesquisa e, finalmente, efetuou-se o tratamento dos dados com o intuito de encontrar a localização geográfica das instituições em que se capacitaram. Como resultado utilizado do conjunto de dados abertos extraídos da Plataforma Lattes, foi possível realizar um retrato sobre como ocorre o êxodo científico brasileiro com todo o conjunto de indivíduos analisados.

Palavras-chave: Plataforma Lattes. Fluxo migratório. Análise de dados em acesso aberto.

Study of the Brazilian scientific exodus based on data in access to the Lattes Platform

ABSTRACT

The spectator of man standing has been growing in recent years growing significantly in the Brazilian and world scene. One of the reasons identified in the migratory process is the academic formation of those who seek training in higher quality educational institutions. Therefore, the objective of this paper will be to analyze how the scientific successes of Brazil are taking place, in the case of some of the various localities of the country in search of a better logical and scientific form. Therefore, it will be characterized and analyzed from the first year in various areas of knowledge. This requires an extraction of academic data, available in open access on the Lattes Platform, with all individuals with a doctorate, and totaling 284,050 curricula. The double of this group is justified by having a high level concept of academic formation, and that meets the following criteria: To do so, filter the data, obtain the necessary information, and finally perform the data processing in order to find a geographical location of the institutions in which they trained. The resulting assisted data set code can be executed in conjunction with the data set occurring in Brazil.

Keywords: *Lattes platform. Migratory flow. Open access data analysis.*

Estudio del éxodo científico brasileño a partir de datos de acceso abierto de la plataforma Lattes

RESUMEN

El espectador del hombre de pie ha estado creciendo en los últimos años creciendo significativamente en la escena brasileña y mundial. Una de las razones identificadas en el proceso migratorio es la formación académica de quienes buscan capacitación en instituciones educativas de mayor calidad. Por lo tanto, el objetivo de este documento será analizar cómo se están produciendo los éxitos científicos de Brasil, en el caso de algunas de las diversas localidades del país en busca de una mejor forma lógica y científica. Por lo tanto, se caracterizará y analizará desde el primer año en diversas áreas del conocimiento. Esto requiere una extracción de datos académicos, disponibles en acceso abierto en la Plataforma Lattes, con todos los individuos con un doctorado, y un total de 284,050 planes de estudio. El doble de este grupo se justifica por tener un concepto de formación académica de alto nivel, y cumple con los siguientes criterios: para ello, filtre los datos, obtenga la información necesaria y, finalmente, realice el procesamiento de los datos para encontrar un Ubicación geográfica de las instituciones en las que se formaron. El código del conjunto de datos asistido resultante se puede ejecutar junto con el conjunto de datos que se produce en Brasil.

Palabras clave: *Plataforma Lattes. Flujo migratorio. Análisis de datos de acceso abierto.*

INTRODUÇÃO

A emigração de brasileiros para outros países e a migração para outros estados têm aumentado de forma significativa, de modo que, no Brasil, estudos demonstram que algumas cidades possuem taxa de 10 a 30% de migrantes que não vivem no seu estado de origem (ALMEIDA, 2017).

Um dos motivos para migração é a necessidade de capacitação em alto nível de formação. Das principais causas por optar pela mobilidade de indivíduos no território brasileiro, destacam-se a qualidade de ensino superior em outros estados, a busca de novas oportunidades e mais experiências em suas áreas (LOMBAS, 2017). Outro refúgio para esses estudantes é a ida para outros países, buscando geralmente intercâmbio cultural e melhor investimento em bolsas de pesquisa. A saída do aluno para outros países não é interessante somente ao discente, mas também às instituições de origem, pois ele retorna na maioria das vezes mais produtivo, com maior vivência e com uma rede de contato mais extensa.

Diante desse cenário, este trabalho apresentará um estudo sobre o êxodo de estudantes brasileiros que partiram do seu estado de nascimento para outros estados e/ou aqueles que foram para outros países em busca de capacitação. Para extração dos dados de estudantes brasileiros será utilizada o framework LattesDataXplorer (DIAS, 2016), ferramenta responsável por extrair e tratar dados curriculares disponíveis em acesso aberto na Plataforma Lattes. Atualmente, o repositório de currículos em acesso aberto da Plataforma Lattes é direcionado à área acadêmica/científica e profissional, possuindo cerca de 6.100.000 de currículos cadastrados.

Chaves *et al.* (2016) realizaram um estudo com dados da Plataforma Lattes de pesquisadores de pós-graduação, coletando informações da trajetória do pesquisador desde o nascimento até a sua última titulação. Tendo o foco principal em doutores, chegou-se à conclusão de que 95% são de origem dos estados Sul, Sudeste e Nordeste. Mencionou-se que 40% das primeiras formações dos doutores foram realizadas em suas cidades de origem, e que 87% daqueles que se deslocam para outras cidades não ultrapassam o limite de 1.000 km. Foi destacado também que a cidade com maior formação de doutores é São Paulo.

Já Pierro (2016) analisa a mobilidade de pesquisadores e estudantes brasileiros ao longo de sua formação acadêmica. Ressalta-se que 20% dos pesquisadores trabalham, entretanto, a mais de 500 km de distância na instituição de onde ingressou na trajetória acadêmica. Em contrapartida, a maioria trabalha a cerca de 100 km de distância. Essa mobilidade fez com que pesquisadores entrevistados se envolvessem em várias linhas de pesquisa, tornando assim seu trabalho mais conhecidos nos locais da sua trajetória.

O estudo indica que os estados da região Sudeste, principalmente São Paulo, são aqueles em que a maioria dos pesquisadores nasceram na própria região; já os outros estados brasileiros possuem um padrão migratório temporário. A mobilidade de pesquisadores pode contribuir para a produção científica brasileira, mesclando conhecimento entre os pesquisadores dos diversos estados da federação.

Logo, é notório que uma larga escala de indivíduos que optam por obter uma capacitação em alto nível de formação parte da sua cidade de origem para outra e, com menor escala, partem do seu país de origem procurando obter capacitação. Vale ressaltar também que, por se tratar de um repositório de muita importância para o estudo da produção científica brasileira, muitos dos trabalhos relacionados a este projeto preferem extrair dados dos currículos da Plataforma Lattes.

No entanto, ao contrário deste trabalho, todos os outros estudos tiveram ênfase em grupos específicos de indivíduos, analisando apenas pequenos conjuntos de currículos. Este, por sua vez, estuda todos os doutores com currículos cadastrados na Plataforma Lattes.

METODOLOGIA

No presente trabalho, a principal fonte de dados foi o repositório curricular disponível em acesso aberto da Plataforma Lattes, por se tratar de uma importante fonte de dados de alta qualidade com o intuito de medir e avaliar o desempenho acadêmico nacional.

Conforme descrito anteriormente, apesar da reconhecida relevância dos currículos cadastrados na Plataforma Lattes para análise e entendimento sobre a evolução da ciência brasileira, o acesso ao repositório de dados passa a ser um fator limitante para análises que considerem todos os indivíduos independentemente de suas áreas de atuação. Não obstante a viabilidade de acesso individual a cada um dos currículos ser possível por meio da interface de consulta no portal da Plataforma Lattes, a análise de grandes grupos de indivíduos passa a ser um fator limitante para análises abrangentes. Logo, no contexto deste trabalho, para extração de todo o conjunto de currículos a serem analisados, será utilizado o LattesDataXplorer (DIAS, 2016) para extração e tratamento dos dados.

A extração dos dados foi realizada em maio de 2018, totalizando aproximadamente 284.050 currículos de indivíduos com doutorado concluído. O framework utilizado é responsável por conter uma coleção de componentes que visam realizar a coleta e tratamento dos dados.

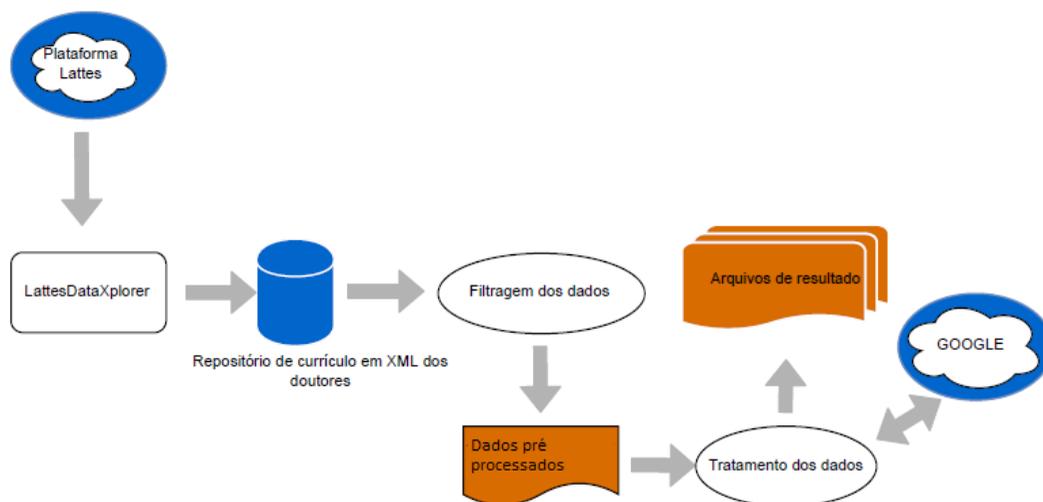
Utilizando o framework é possível fazer um refinamento da busca de currículos específicos, baseado em parâmetros como nome, titulação, idioma, nacionalidade, grande área e área de atuação, entre outros. Assim, são geradas listas de currículos que atendem aos parâmetros informados.

Todo o procedimento de extração e tratamento de dados realizado pelo LattesDataXplorer se inicia a partir da aquisição dos códigos de currículos da Plataforma Lattes, baseado no refinamento executado na busca, com o propósito de no futuro esses códigos serem armazenados localmente. A lista resultada da consulta refinada possui todos os códigos de identificação de todos os currículos cadastrados, possibilitando acesso individual em cada um dos existentes na Plataforma Lattes. A partir disso, será possível analisar somente os currículos selecionados para um determinado tipo de grupo.

Diante do exposto, foram coletados somente currículos de doutores, tanto por se tratar do grupo com o maior nível de formação, como por se tratar de currículos que estão sempre atualizados e grande parte dos parâmetros necessários para o trabalho estar registrada em seus currículos.

A fim de mapear o êxodo de indivíduos brasileiros cadastrados na Plataforma Lattes, foi efetuada a mineração de informações para filtrar os dados relevantes para esta pesquisa. Logo após eles serem filtrados, ocorrerá um tratamento com o intuito de enriquecer os dados para as análises a serem realizadas. A figura 1 apresenta uma visão geral do conjunto de componentes que foram desenvolvidos no intuito de obter as análises desejadas.

Figura 1 – Aspecto geral do conjunto de componentes utilizados



Fonte: Autores.

Após a extração dos dados, ocorre a filtragem. Essa fase é responsável por executar a mineração nos arquivos XML para extrair informações relevantes da pesquisa, armazenando-as em um extrato de dados formatados (dados pré-processados). Assim será definido o conjunto de dados centrais a serem processados. As informações dos currículos possuem: código do indivíduo, estado de nascimento, cidade de nascimento, grande área, área, código, nome e CEP do vínculo atual que o indivíduo trabalha. Apresentam também o código de identificação, nome da instituição, início da formação e fim da formação das primeiras formações concluídas, iniciando na graduação até o doutorado.

O Módulo “Tratamento de dados” tem o intuito de processar os dados coletados dos doutores, tratá-los e caracterizá-los, resultando em outros arquivos, a fim de facilitar as análises dos dados. Esse processo realizará, basicamente, três etapas: busca pela localização geográfica; limpeza e agrupamento de dados; e normalização dos dados.

- A etapa de “busca pela localização geográfica” é uma tarefa a ser realizada com a finalidade de geolocalizar uma instituição. Em linhas gerais, a etapa será efetuada da seguinte maneira: ao acessar a API (Application Programming Interface) de geolocalização do Google, será enviado o endereço da instituição, para posteriormente ter como retorno a localização geográfica da instituição.
- Na etapa “limpeza e agrupamento de dados” ocorre a realização de exclusão de possíveis termos irrelevantes para a pesquisa, com o intuito de diminuir o volume de dados a serem processados e analisados. Como exemplo: A remoção de stopWords. A normalização para extrair palavras acentuadas e substituí-las pelo seu equivalente sem acentuação.
- Já a etapa de “normalização dos dados” tem o intuito de reduzir a redundância de informações, descartando atributos de currículos com a ausência de dados.

Os “Arquivos de resultado” possuem todos os dados específicos para a realização de algumas análises após as ações realizadas.

Logo após toda a execução das etapas descritas anteriormente e com os arquivos formatados de modo a facilitar as análises, diversas métricas são aplicadas para compreensão sobre como tem ocorrido o êxodo científico brasileiro.

ANÁLISE E DISCUSSÃO DOS RESULTADOS

Como resultado, foi possível obter um mapeamento das distâncias percorridas pelos indivíduos ao longo de sua capacitação a cada nível de formação. Na Tabela 1 é demonstrado o quantitativo de registros obtidos, para ser medida a distância média percorrida pelos indivíduos de um nível de formação acadêmico até outro. Nela são apresentadas informações que possibilitam medir a distância de um nível até outro.

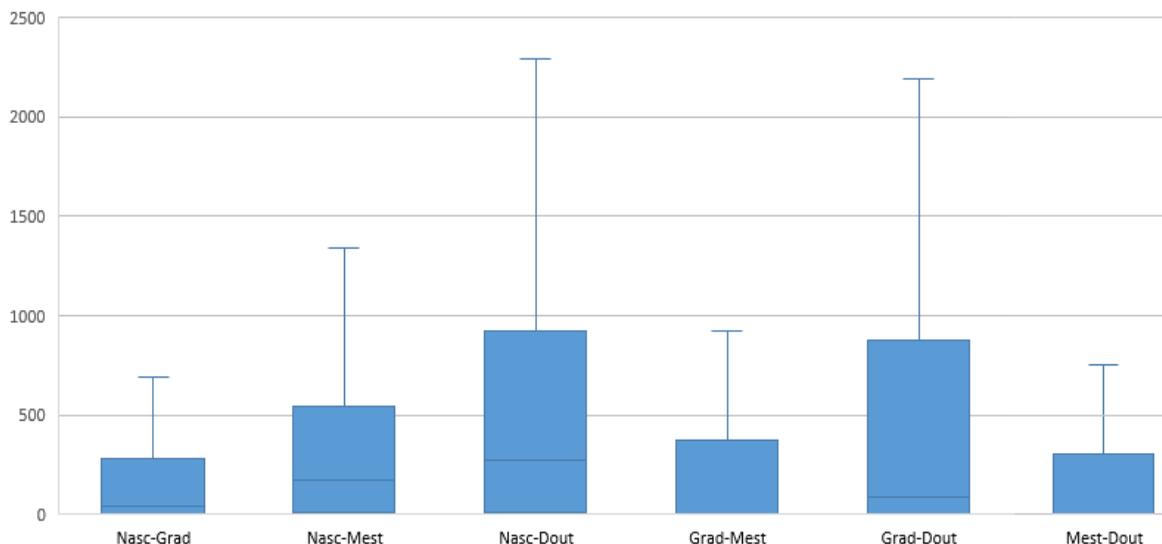
Tabela 1 – Quantidade de registros obtidos

Quantidade	Graduação	Mestrado	Doutorado
Nascimento	212.441 – (75%)	204.209 – (72%)	221.563 – (78%)
Graduação	-	188.752 – (66%)	202.715 – (71%)
Mestrado	-	-	203.295 – (71%)
Total	281.051		

Fonte: Elaborado pelo autores

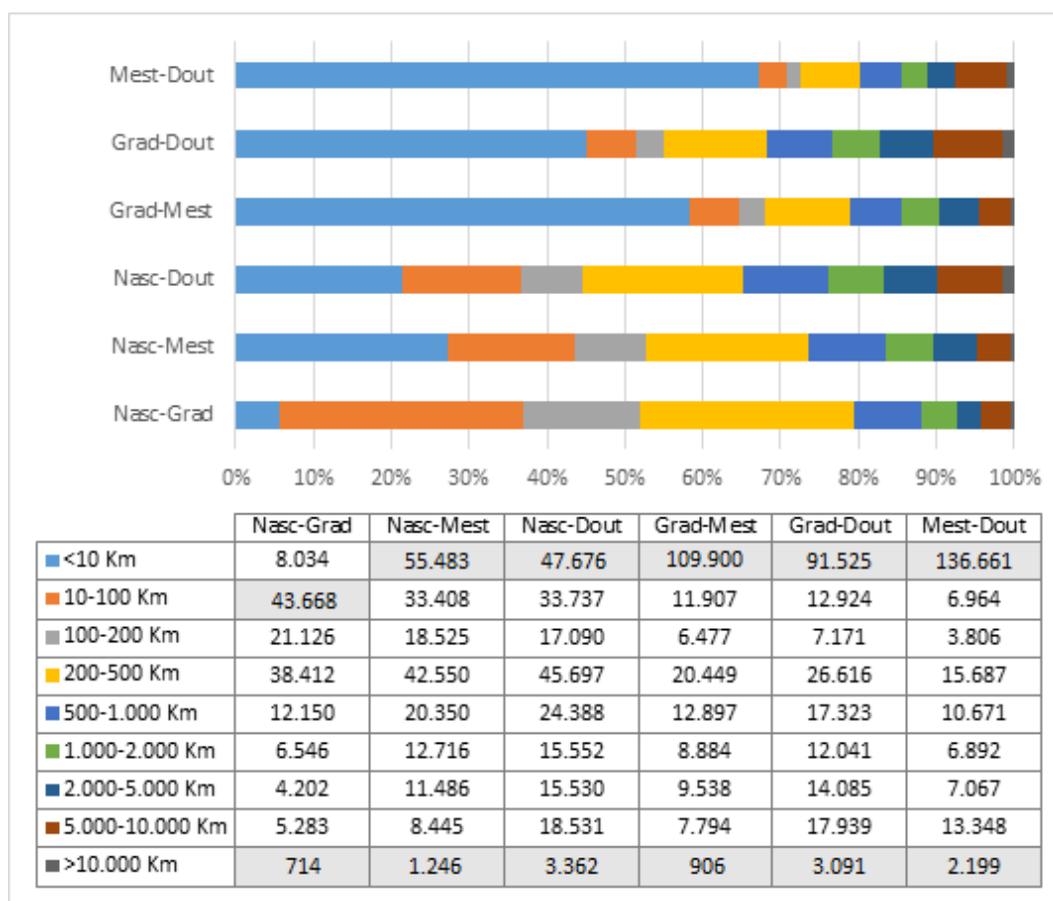
Quando se trata de quantidade de registros, deve-se destacar a perda quantitativa daqueles que não possibilitam a medição de um determinado nível de capacitação para seu nível posterior. Os registros de graduação-mestrado (188.752) têm uma perda considerada, justificando-se pela inserção incorreta de CEP das instituições que foram obtidas as formações acadêmicas. Em compensação, a menor perda sofrida são dos registros nascimento-doutorado. Isso se explica pelo fato de o nascimento ser um registro obrigatório na Plataforma Lattes e também porque o doutorado foi o grupo selecionado e extraído da Plataforma Lattes para a análise na pesquisa.

Figura 2 – Distância média entre os níveis de formação



Fonte: Elaborada pelos autores.

Figura 3 – Distância por faixas entre níveis de formação



Fonte: Elaborada pelos autores.

A maioria das distâncias entre um nível de formação e seu subsequente possuem uma dimensão menor, diferente do intervalo entre mestrado-doutorado, em que grande parte das distâncias é igual a zero, conforme pode ser observado na figura 2.

A figura 2 apresenta as medianas das distâncias percorridas pelos indivíduos brasileiros ao longo de sua formação acadêmica, sendo separado entre nascimento-graduação, nascimento-mestrado, nascimento-doutorado, graduação-mestrado, graduação-doutorado e mestrado-doutorado.

A maior mediana observada é a de nascimento-doutorado, por se tratar do local de nascimento do indivíduo até o mais alto nível de formação, resultando em uma representação de toda a sua vida acadêmica. Vale ressaltar também a distância mínima das medianas entre graduação-mestrado e mestrado-doutorado. Isso se justifica pelo fato de os estudantes decidirem seguir, na maior parte dos casos, na mesma instituição durante essas transições em seus processos formativos. Foi efetuada também uma análise que tornou possível a identificação de distâncias médias em quilômetro por faixas de distâncias (figura 3).

Como pode ser observado, grande parte dos indivíduos (aproximadamente 52%) na transição do nascimento para a graduação não ultrapassam a distância média de 200 quilômetros. Um dos motivos é que a maioria das cidades brasileiras oferecem cursos de graduação, minimizando o deslocamento do indivíduo.

Outra questão que vale ressaltar é a quantidade de indivíduos que se deslocam com uma distância inferior à de 10 quilômetros, principalmente nas etapas de graduação-mestrado e mestrado-doutorado, totalizando uma porcentagem de 57% e 66%, respectivamente, pois no geral os doutores, ao longo das etapas de formação, optam por dar continuidade na mesma instituição onde já estavam se capacitando.

Deve ser apontado também a menor parte dos indivíduos que optam por realizar um deslocamento superior a 10.000 quilômetros, podendo ser justificado pela pouca quantidade de indivíduos que vão se capacitar no exterior.

CONCLUSÕES

A partir dos resultados obtidos foi possível verificar a viabilidade em adotar os currículos cadastrados na Plataforma Lattes como fonte de dados para análises sobre como ocorre o Êxodo Científico Brasileiro.

A escolha do grupo de doutores se caracteriza como uma parcela significativa de todo o conjunto de dados cadastrados na Plataforma Lattes, tendo em vista que são os indivíduos com maior nível de formação acadêmica concluída. Percebeu-se também que em geral seus currículos são recentemente atualizados e a maioria possui endereço profissional cadastrado.

Como resultado deste estudo, foi possível identificar — pelos dados em acesso aberto extraídos dos currículos da Plataforma Lattes — a mediana das distâncias entre os níveis de formação, sendo observado que as distâncias percorridas foram pequenas, e que muitas das vezes os indivíduos optaram em se capacitar na mesma instituição onde estavam. Traçaram-se faixas de distâncias percorridas pelos doutores nos seus respectivos graus acadêmicos, podendo ser observado que a maior parte do comprimento de um grau de formação a outro chega a ser menor que 10 quilômetros, e que uma pequena parcela de indivíduos percorre uma distância acima de 10.000 quilômetros.

Consequentemente, diversos outros estudos poderão ser realizados, como por exemplo um estudo baseado em áreas de atuação, possibilitando assim observar quais estados possuem maior fluxo de migração de determinada área, ou de quais estados os indivíduos se deslocam para cursar determinado curso, o que permite uma possível abertura de cursos para minimizar os impactos com deslocamentos.

REFERÊNCIAS

ALMEIDA, G. Z. R. Fluxos migratórios: a distribuição da população de cada estado pelo país. *Nexo Jornal*, São Paulo, 1 dez. 2017. Disponível em: <https://www.nexojornal.com.br/grafico/2017/12/01/Fluxos-migrat%C3%B3rios-a-distribui%C3%A7%C3%A3o-da-popula%C3%A7%C3%A3o-de-cada-estado-pelo-pa%C3%ADs>.

CHAVES, L. C. R. *et al. Analisando a mobilidade de pesquisadores através de registros curriculares na plataforma lattes*. Paraíba: Universidade Federal da Paraíba, 2016.

DIAS, T. M. R. *Um Estudo Sobre a Produção Científica Brasileira a partir de dados da Plataforma Lattes*. 2016. 181p. (Doutorado) - Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte, 2016.

LOMBAS, M. L. D. S. A mobilidade internacional acadêmica: características dos percursos de pesquisadores brasileiros. *SciELO*, [S.l.], v. 19, n. 44, p. 308–333, jan 2017.

PIERRO, B. de. Circulação limitada. *Pesquisa Fapesp*, São Paulo, p. 36–39, jan 2016.