

UMA ANÁLISE DA MOBILIDADE ACADÊMICA BASEADO EM DADOS DE DOUTORES DA ÁREA DE CIÊNCIA DA INFORMAÇÃO

AN ANALYSIS OF ACADEMIC MOBILITY BASED ON DATA FROM DOCTORS IN THE INFORMATION SCIENCE AREA

Higor Alexandre Duarte Mascarenhas^a

Thiago Magela Rodrigues Dias^b

RESUMO

Objetivo: Analisar como ocorre o êxodo científico no Brasil baseado em doutores da área da Ciência da Informação, em que indivíduos de diversas localidades do país migram em busca de uma melhor formação acadêmica. **Metodologia:** utilizou-se como fonte de dados a Plataforma Lattes, extraindo todos os currículos com a ferramenta *LattesDataExplorer* e a partir de então foi efetuada todo o tratamento de dados para análises posteriores. **Resultados:** realizou-se cálculos de distâncias médias percorridas pelos doutores da Ciência da Informação, bem como foi possível destacar os fluxos dos estados ao longo da formação acadêmica, bem como as principais cidades habitadas por estes indivíduos. **Conclusões:** observou-se que os indivíduos tendem a percorrer baixas distâncias ao longo do processo de formação, a localização geográfica possui forte influência sobre o processo migratório para capacitação e que a cidade de São Paulo é preferida pelos indivíduos no processo de formação.

Descritores: Mobilidade acadêmica. Ciência da Informação. Análise de dados.

1 INTRODUÇÃO

A emigração de brasileiros para outros países e para outros estados tem aumentado de forma significativa, de modo que, no Brasil, estudos demonstram que em algumas cidades possuem taxa de 10 a 30% de migrantes que não vivem

^a Doutorando em Modelagem Matemática e Computacional pelo Centro Federal de Educação Tecnológica de Minas Gerais (CEFET), Belo Horizonte, Brasil. E-mail: higoralexandre1996@gmail.com

^b Doutor em Modelagem Matemática e Computacional pelo Centro Federal de Educação Tecnológica de Minas Gerais (CEFET). Docente do Programa de Pós-Graduação em Modelagem Matemática e Computacional pelo Centro Federal de Educação Tecnológica de Minas Gerais (CEFET), Belo Horizonte, Brasil. E-mail: thiagomagela@gmail.com

no seu estado de origem (ALMEIDA; ZANLORENSSI, 2017). Em muitos casos, brasileiros saem em busca de emprego, ou estudos, visando sempre qualidade de vida.

Dentre os principais motivos para migração está a necessidade de capacitação em alto nível de formação. Das principais causas por optar pela mobilidade de indivíduos no território brasileiro, refere-se à qualidade de ensino superior em outros estados, a busca de novas oportunidades e mais experiências em suas áreas (LOMBAS, 2017). Outro refúgio para tais estudantes condiz na ida para outros países, buscando assim, intercâmbio cultural e melhor investimento em bolsas de pesquisa. A saída do estudante para outros países não é interessante somente ao discente, mas também às instituições de origem, pois, o mesmo retorna na maioria das vezes mais produtivo, com rede de contato mais extensa, maior vivência, e podendo futuramente compartilhar suas experiências com outros estudantes da instituição de origem.

De acordo com (DEMARTINI, 2017) a cada dia tem se tornado mais difícil produzir pesquisa científica no Brasil, devido a cortes de investimentos destinados a bolsas. Um dos principais motivos para a emigração de pesquisadores brasileiros para outros países, pode ser apontado pela falta de apoio do governo. Logo, com esse cenário, pesquisadores brasileiros saem do país, dificultando assim o retorno pela falta de oportunidades. Grande parte dos cientistas brasileiros que voltam para o Brasil não conseguem emprego na sua área de formação, fazendo assim que não progridam nas suas carreiras.

Um programa que facilitou e auxiliou bastante o ingresso de estudante às instituições com sedes em outros países foi o Ciência Sem Fronteiras, por se referir a um programa que amparou estudantes, oferecendo bolsas de estudos. Em 2015, o governo pretendia alcançar 101.000 bolsas de estudos para pesquisadores, graduandos, doutorandos, alunos ingressados no pós-doutorado, incentivando os discentes a se capacitarem em instituições de reconhecida relevância (AVEIRO, 2014). Recentemente, o Programa perdeu bastante influência no ingresso de estudantes para outros países, por motivos de cortes de investimento.

Logo, surge o interesse do aprofundamento em estudos direcionados em

mobilidade acadêmica, visando a oportunidade de compreender o cenário atual, e através dessa compreensão levar a adoção de medidas para promover possíveis aberturas de novos cursos de graduação ou pós-graduação em regiões em que se identifica um déficit de cursos em determinadas áreas do conhecimento.

Neste contexto, este trabalho tem como objetivo efetuar um estudo focando na mobilidade acadêmica de doutores brasileiros com área de atuação em Ciência da Informação (CI), que partiram do seu estado de nascimento para outros estados, em busca de melhor formação.

2 TRABALHOS RELACIONADOS

Ramos e Velho (2011) efetuaram um estudo apresentando uma análise de como vem se desenvolvendo o “*brain drain*”, desde os anos 1960 até na data da pesquisa. Os autores destacam que anteriormente a mobilidade internacional era tratada como uma perda de capital humano, principalmente se tratando de indivíduos altamente qualificados, porém, com o passar dos anos, a mobilidade internacional passou a ser vista com bons olhos, uma vez que os indivíduos retornariam com uma ampla rede de contatos, ampliando seus conhecimentos. Em um estudo se tratando de doutores brasileiros, os autores utilizaram dados divulgados pela CAPES e CNPq, destacando assim que do ano de 1996 até o ano de 2004 a quantidade de estudantes brasileiros que migraram para outros países para se capacitar utilizavam em maior quantidade a modalidade de “Doutorado Pleno no Exterior” comparado a modalidade de “Doutorado sanduiche no exterior”, porém no ano de 2008 ocorre o inverso, isso ocorre, pois as instituições de fomento optam por diminuir o tempo de bolsa no exterior visando que os indivíduos não criem laços fortes com as instituições no exterior, fazendo assim que os estudantes brasileiros retornem ao país de origem para auxiliar as posteriores pesquisas.

O trabalho de Cañibano, Otamendi e Solís (2011) tem como objetivo analisar a mobilidade de pesquisadores de um sistema público de pesquisa da Espanha com informações cadastradas no Sistema de Informação Científica da Andaluzia (SICA), bem como as características das mobilidades. Estes

currículos possuem informações acadêmicas dos indivíduos, e dados referentes a mobilidade dos pesquisadores, sendo assim os autores efetuaram um tratamento dos dados coletados, tendo currículos de doutores de uma determinada faixa temporal para mensurar a imigração dos indivíduos, dos anos 1950 a 2009, também estipularam que o tempo que o indivíduo exercia a mobilidade deveria ser superior a uma semana. Inicialmente os autores subdividiram os doutores em dois grupos, como móveis (representando indivíduos que ficaram no exterior de uma semana a dois anos) e não móveis (aqueles que nunca viajaram para o exterior, ou permaneceram por um tempo superior a dois anos). A taxa de mobilidade geral dentro da população estudada é de 37,8%, o que significa que mais de 60% dos doutorados declararam sem visitas internacionais de pesquisa. Calcularam que a idade média de conclusão do doutorado é de 31,4 anos. Se tratando do “*brain drain*” observou-se que 86,6% das visitas são menores que 6 meses e 68,5% menores que 3 meses, apesar de que indivíduos de determinadas áreas variam o tempo de estadia em certas visitas, como por exemplo indivíduos das áreas de humanidades e ciências sociais, com duração média de 65 e 64 dias respectivamente. Já se tratando de destino das visitas por destino a pesquisa aponta que 60% são para países da Europa Ocidental, seguida pela América Latina (18,5%), América do Norte (14,8%) e Europa Oriental (3,5%). Já África e Ásia-Pacífico somam cerca de 3% das visitas, tendo como os cinco primeiros países destino: Reino Unido, EUA, França, Itália e Alemanha, que respondem por 60% de todas as visitas.

Dubois, Rochet e Schlenker (2014) efetuaram um estudo com o intuito de analisar a produtividade e mobilidade acadêmica de um grupo de matemáticos espalhados por todo o mundo, totalizando 32.574 matemáticos ativos do período de 1984 a 2006. Inicialmente, para análise de mobilidade, os autores levaram em conta as filiações dos matemáticos para entender o processo de mobilidade acadêmica destes indivíduos, ou aqueles que não possuem filiação, foi recuperado o país de localização da primeira publicação. Já sobre a produção científica, eles extraíram os autores que publicaram em 98 dos periódicos mais relevantes da área. Os autores destacaram que a maior porcentagem dos matemáticos tende a publicar nos Estados Unidos. Com relação a mobilidade,

os autores observaram que os principais países que atraem os matemáticos nas primeiras publicações são: Canadá, Israel e Reino Unido. Uma vez analisando a média que os matemáticos ocuparam locais diferentes ao longo da análise, isto é, baseado em dados de departamentos ocupados pelos matemáticos, os autores destacaram que a média é equivalente a 1,87. A partir de análises de departamentos em universidades, comparando assim a questão econômica, os autores apontaram que as universidades mais ricas tendem a atrair melhores pesquisadores. Como conclusão, destacaram que incentivar a mobilidade parece ser uma forma de melhorar tanto a qualidade de um departamento onde um indivíduo está alocado quanto a produção científica de seus membros. Por outro lado, incentivar os membros de um departamento a colaborar mais não parece ser eficiente, exceto se a colaboração for com colegas de áreas diferentes. Sugerem, então que grupos de leitura ou seminários reunindo matemáticos de diferentes especialidades podem ser uma forma de ampliar seus interesses e melhorar sua produção.

Madeira e Marengo (2016) efetuaram um trabalho com o objetivo de mapear dimensões de internacionalização das carreiras dos docentes vinculados aos programas de pós-graduação em ciência política do Brasil. Os autores se basearam em formação acadêmica e a publicação em periódicos internacionais dos docentes, visando mensurar o impacto que a formação em instituições estrangeiras afetaria nas futuras publicações internacionais. Para que as publicações fossem contabilizadas, elas teriam que estar sediadas fora do Brasil. Os autores utilizaram o repositório de dados da Plataforma Lattes, extraindo currículos desde 1960 a 2015. Logo, os autores mensuraram a quantidade de doutores que formaram no exterior, justificando que a escolha de alguns indivíduos por optarem em se capacitar no exterior em meados dos anos de 1960 a 1990 é devido à falta de oferta de cursos na área analisada, além de destacarem que a principal opção destes indivíduos por se capacitar, ou daqueles que se tornaram um professor visitante foi nos Estados Unidos. Os doutores publicaram com maior frequência em periódicos sediados nos Estados Unidos. Os autores também destacaram que os indivíduos que tenham experiência internacional, tendem a publicar cerca de 1,4 vezes mais que os

demais indivíduos. Concluindo assim que metade dos doutores analisados publicaram no exterior, e a presença dominante dos Estados Unidos no âmbito da circulação internacional de textos acadêmicos.

Com um estudo baseado em economistas espalhados pelo mundo todo, Albarrán, Carrasco e Ruiz-Castillo (2017) em seu trabalho tinham um objetivo de analisar a mobilidade geográfica e produtividade de pesquisa de economistas a partir de uma seleção dos principais departamentos de economia mundial, logo selecionaram 81 principais departamentos de economia. Sendo assim, obtiveram dados da primeira graduação, o doutorado dos indivíduos e atuação profissional no ano de 2007. Após isto os autores subdividiram o conjunto de indivíduos em grupo 1, 2 e 3, de acordo com a quantidade e qualidade de publicações efetuadas pelos indivíduos, respectivamente. Para as análises foi utilizado grupo 2 (elite) e a amostra total. Os autores destacaram a maioria dos indivíduos tanto da elite como da amostra total optaram por se localizar nos Estados Unidos, seguido da União Europeia para efetuar a graduação, o doutorado e o atuar profissionalmente. O principal país em que os indivíduos optam por habitar é o Estados Unidos nestas três etapas citadas anteriormente. Já se tratando da elite, a quantidade de indivíduos que habitam nos estados unidos em todas as três etapas é superior ao somatório daqueles indivíduos que habitam em outros países. A respeito da mobilidade geográfica, percebe-se que 47,7% dos economistas saem do seu país de origem. No entanto, a divisão daqueles que saem de seu país de origem e os que permanecem é muito diferente em cada área geográfica, por exemplo: os EUA conseguem reter a maioria de seus graduados para vínculos empregatícios, bem como a maioria de seus doutores atuam em instituições dos EUA. Assim, 92% dos economistas nascidos nos EUA são permanentes. Já em outros países do mundo, a situação é exatamente oposta: os que permanecem no país representam apenas 26,3%, enquanto a fuga de cérebros representa 61,2%. Ao ser analisado a elite, a primeira mudança a ser observada é o aumento dos permanecem versus os que se movem. No entanto, devemos enfatizar que, enquanto os residentes nos EUA aumentam em aproximadamente 18%, os residentes em outros países reduzem 9%.

Sachini *et al.* (2020) realizaram um estudo com o intuito de analisar a mobilidade de pesquisadores gregos utilizando análises bibliométricas (número de publicações, afiliação de uma unidade de pesquisa). Foi utilizado a base de dados do *Scopus* para obtenção dos dados para a pesquisa, como forma de visualização de dados, utilizou-se a análise de redes para representar como estão distribuídos os pesquisadores. Como resultados gerais, os autores destacaram que a maior parte dos pesquisadores está estática no país de origem, Grécia (74,6%). Os autores confirmaram que os pesquisadores móveis constituem uma parcela menor (21,7%) da população total, enquanto os viajantes (60%) - autores que ganharam afiliações adicionais enquanto mantêm afiliação com seu país de origem, representam a maioria dos pesquisadores móveis. Por outro lado, os migrantes - autores que deixaram seu país de origem - representam 40% dos pesquisadores móveis. A maior parte dos pesquisadores (86,2%) e principalmente a elite dos pesquisadores (90,5%) mantinham vínculos sólidos científicos com a Grécia.

No trabalho de Verginer e Riccaboni (2020) os autores tinham a finalidade de investigar a mobilidade dos pesquisadores espalhados pelo mundo, porém não se tratando de migração, pois a análise destes autores foi baseada em localidades onde os pesquisadores publicaram seus trabalhos. Os resultados possibilitaram aos autores destacar quais foram as principais cidades/países em que os pesquisadores optam por publicar seus trabalhos, bem como uma análise temporária de opção por escolha dos demais países/cidades no ato de publicação. Os autores também descreveram algumas interações entre origem/destino entre países, e destacaram a interação entre países como por exemplo, Brasil e Portugal, que têm maiores quantidades de vínculos, diferente de Portugal e Espanha que estão no mesmo continente. Os autores também efetuaram métricas para avaliar o quão é relevante um país, se tratando do âmbito científico, que dependente da classificação de um periódico este seria mais relevante do que os demais, neste caso, a China se destaca.

Koszytán *et al.* (2021) apresentaram um trabalho com o intuito de avaliar o desempenho de universidades a partir da integração em redes de mobilidade e colaboração, utilizando dados disponíveis em diversas plataformas. Utilizaram

assim, dados relativos aos beneficiários dos programas *Erasmus* e *FP7* financiados pela União Europeia, ambos baseados na colaboração com outros atores acadêmicos e não acadêmicos. Ao analisarem a correlação entre mobilidade e redes de colaboração em pesquisa, os autores descobriram que a mobilidade Erasmus não tem um efeito substancial nas colaborações científicas e de projetos do *FP7*, apesar de os objetivos do *FP7* e do programa *Erasmus* estejam intimamente ligados, uma vez que ambos visam apoiar a União Europeia a se tornar uma sociedade baseada no conhecimento líder mundial e promover a cooperação transnacional. Os autores também citaram que encontraram complementaridade, mas não sinergia, entre os programas *Erasmus* e *FP7*, apesar de esperara que a mobilidade de professores, levasse à colaboração em pesquisa. A comparação da integração da rede de mobilidade e colaboração com o impacto científico demonstra aos formuladores de políticas que o uso desses programas deve ter um efeito de reforço mútuo, levando a uma maior sinergia entre esses programas.

Percebe-se que uma grande quantidade de indivíduos que decidem obter capacitação em alto nível de formação opta por partir da sua cidade de origem para outra, além daqueles indivíduos que com menor representatividade, partem do seu país de origem em busca de capacitação no exterior.

3 METODOLOGIA

Como principal fonte de dados foi utilizado o repositório curricular da Plataforma Lattes. A justificativa da escolha da Plataforma se dá por: (1) registrar a trajetória e a contribuição de cada estudante, técnico e pesquisador brasileiros cadastrados (MARQUES, 2015); (2) representa a experiência do CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) na integração de base de dados de currículos e de instituições da área de Ciência e tecnologia (SILVA; SMIT, 2009); (3) por se tratar de uma importante fonte de dados de alta qualidade com o intuito de medir e avaliar o desempenho acadêmico nacional (LANE, 2010).

Apesar da reconhecida relevância dos currículos cadastrados na Plataforma Lattes para análise e entendimento sobre a evolução da ciência

brasileira conforme descrito anteriormente, 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, ou nível de formação acadêmica. Apesar da viabilidade de acesso individual a cada um dos currículos ser possível através de interface de consulta dos currículos 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, foi utilizado o *LattesDataXplorer* (DIAS, 2016) para extração e tratamento dos dados.

A extração dos dados foi realizada em maio de 2019 totalizando 308.317 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, dentre outros. Sendo gerado assim 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 estes códigos sejam armazenados localmente. A lista resultante da consulta refinada, possui todos os códigos de identificação de todos os currículos cadastrados, possibilitando ter o acesso individual em cada um destes existentes na Plataforma Lattes.

Todos os currículos possuem uma URL (*Uniform Resource Locator*) a fim de permitir o acesso individual a cada um destes. A URL possui o endereço de busca da plataforma concatenada com o código do currículo do indivíduo. O componente de extração de Ids e Data tem a responsabilidade de acessar cada um dos currículos cujos códigos foram salvos, extraindo seu identificador único e a data da última atualização do mesmo. Para esse fim, somente o cabeçalho de cada currículo em que estão presentes estas informações necessitam ser carregados. Tais dados serão armazenados em um arquivo de identificação

contendo código, identificador, data da última atualização no CNPq e data da atualização do currículo armazenado localmente.

Para a atualização do repositório de currículos é necessário o arquivo de identificação, por se tratar da base para extração do mesmo. Durante a atualização o primeiro componente do processo de extração é executado, resultando na extração de todos os códigos cadastrados na plataforma. Os códigos já registrados no arquivo de identificação são ignorados, e no final do arquivo são adicionados aqueles novos códigos, representantes dos novos currículos ainda não extraídos.

Com o uso dos códigos, são acessados os cabeçalhos de cada um dos currículos sendo extraídos códigos identificadores e as datas de atualização junto à Plataforma Lattes, tanto para currículos já extraídos, como para os novos currículos, ocorrendo a atualização do arquivo de identificação a cada nova extração. O acesso ao cabeçalho torna mais ágil a extração de dados, por não haver a necessidade de esperar todo o currículo ser gerado.

Por fim, acontece a extração de currículos, sendo o extrator o responsável por verificar se possuem currículos cuja data de atualização local é diferente da data de atualização junto ao CNPq; caso seja divergente, o currículo da Plataforma é extraído e substitui o currículo local, modificando a data de atualização. Quando não, o currículo permanece o mesmo sem alterações. Por fim são extraídos novos currículos cadastrados, para serem inseridos ao final do arquivo base. Inicialmente esses novos currículos não possuem data de atualização por terem sido extraídos pela primeira vez, sendo assim, é inserido sua data de atualização local.

Finalmente, todos os currículos são armazenados em XML (*eXtensible Markup Language*), não necessitando de um novo repositório de dados feito por bancos de dados relacionais, buscando um menor custo computacional.

Vale ressaltar que com todos os currículos armazenados localmente, torna-se mais fácil e flexível a manipulação de dados coletados a partir da Plataforma Lattes.

3.1 COMPONENTES DESENVOLVIDOS PARA TRATAMENTO DOS DADOS

O *LattesDataXplorer* foi utilizado especificamente para a coleta e seleção dos dados curriculares da Plataforma Lattes, no qual obteve-se o todo o Repositório de Currículos em formato XML.

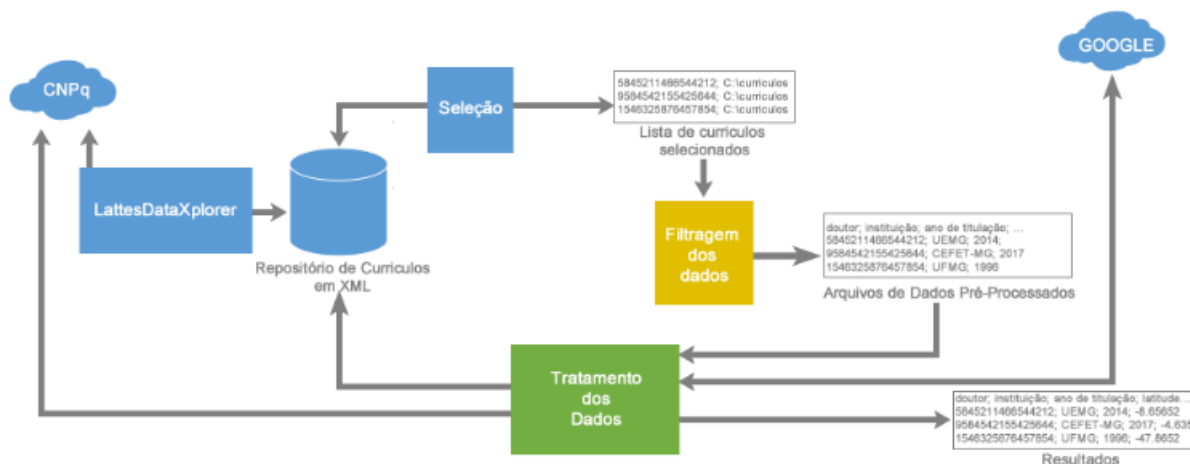
A "Seleção" do conjunto de dados a ser analisado utiliza a linguagem de consulta *XPath* (*XML Path Language*) para pesquisa e posterior geração dos subgrupos a serem analisados. A linguagem *XPath* possibilita a construção de expressões que vão processar e percorrer um documento XML de forma similar ao uso de expressões regulares. Portanto, possibilita o agrupamento de um conjunto de currículos com parâmetros desejados. Assim sendo, em busca pelos parâmetros em cada um dos currículos, independentemente ou não de qual seção ele(s) seja(m) encontrado(s), tais currículos são selecionados e formam um grupo para análises. A partir de então os dados dos currículos são organizados em uma lista de currículos que foram selecionados.

A lista armazena os identificadores de cada currículo e o caminho que ele está armazenado localmente, sendo assim, será possível analisar somente os currículos selecionados.

Diante do exposto foram coletados somente currículos de indivíduos com doutorado concluído, por se tratar do grupo com o maior nível de formação acadêmica; por se tratar de currículos que são frequentemente atualizados e grande parte dos parâmetros necessários para o presente trabalho estarem registrados em seus currículos.

A fim de mapear o êxodo de indivíduos brasileiros cadastrados na Plataforma Lattes, foi efetuada a mineração de dados para filtrar os dados relevantes para esta pesquisa, logo após os dados serem filtrados ocorreu um tratamento com o intuito de enriquecê-los para as análises a serem realizadas. A Figura 1 apresenta um aspecto geral do conjunto de componentes que foram desenvolvidos objetivando obter as análises desejadas.

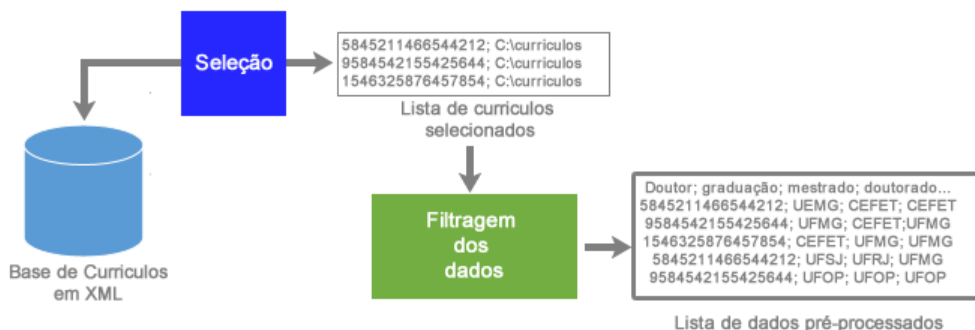
Figura 1 - Aspecto geral do conjunto de componentes utilizados



Fonte: Autores.

Após a Seleção do conjunto a ser analisado, ocorre o módulo “Filtragem dos dados” (Figura 2). Esta fase é responsável por analisar os arquivos XML com o intuito de obter informações relevantes à pesquisa, armazenando-as em um estrato de dados formatados (Arquivos de dados pré-processados). As informações dos currículos presentes no arquivo possuem: identificador do currículo; estado e cidade de nascimento; grande área; área; código, identificador, nome e CEP (Código de Endereçamento Postal) do vínculo atual de atuação do indivíduo, além do código de identificação, e nome da instituição, início e fim de cada nível de formação acadêmica concluída, juntamente com a instituição em que foi realizada a formação considerando desde a graduação até o doutorado.

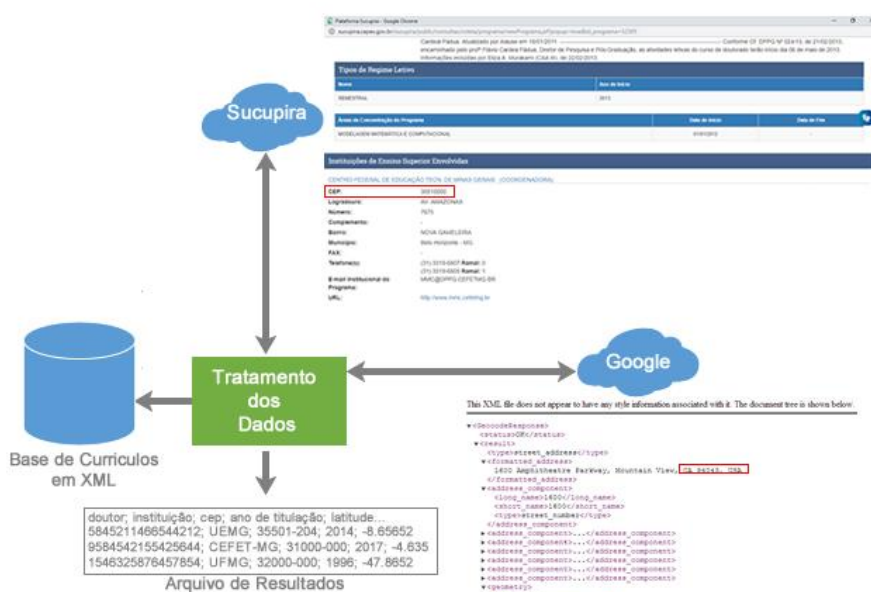
Figura 2 - Filtragem dos dados



Fonte: Autores.

Posteriormente é executado o Módulo de "Tratamento dos Dados" (Figura 3) que tem o intuito de processar os dados dos doutores coletados, tratá-los e caracterizá-los, resultando em outros arquivos, a fim de facilitar as análises dos dados. Nesse processo são realizadas quatro etapas: Obtenção de CEP da instituição; Busca pela localização geográfica; Limpeza e agrupamento de dados e Normalização dos dados.

Figura 3 - Tratamento dos dados



Fonte: Autores.

- A primeira etapa efetuada, é a "Obtenção de CEP da instituição" em que a partir do código da instituição recuperado no currículo do indivíduo, o mesmo é consultado no diretório de instituições da Plataforma Lattes no intuito de obter os dados da instituição e dessa forma, recuperar da seção de endereço o CEP da instituição.
- A etapa de "Busca pela localização geográfica" é uma tarefa a ser realizada com a finalidade de obter coordenadas geográficas de uma instituição. Em que ao acessar a API (*Application Programming Interface*) de geolocalização do Google, será enviado o CEP da instituição, para posteriormente ter como retorno a localização geográfica da instituição.
- Na etapa de "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* nos nomes das cidades; 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 com a ausência de dados, como por exemplo dados de geolocalização ausência de algarismos.

Posteriormente é gerado o arquivo de "Resultados" representando uma sumarização de todos os dados obtidos nos currículos de doutores brasileiros, não necessitando de consultas dos arquivos XML dos currículos extraídos, possuindo todos os dados específicos para a realização das análises desta pesquisa.

Logo, após toda a execução das etapas descritas anteriormente, e com os arquivos formatados de tal forma a facilitar as análises, diversas métricas são aplicadas para compreensão sobre como tem ocorrido a mobilidade dos doutores brasileiros ao longo do seu processo de formação acadêmica.

4 RESULTADOS

Como resultados iniciais, a partir do conjunto de dados extraído, foi possível mensurar as distâncias percorridas pelos indivíduos ao longo do processo de formação acadêmica (Tabela 1), relatando todos os níveis de formação, desde o nascimento até o doutorado, nesta tabela ignorou-se a atuação profissional, e foram consideradas todas as formações acadêmicas independentemente do ano que ocorreu.

Tabela 1 - Distância média percorrida em Km entre os níveis de formação

Distância(Km)	Graduação	Mestrado	Doutorado
Nascimento	300,71	857,49	1.595,63
Graduação	-	722,56	1.510
Mestrado	-	-	1.053,72

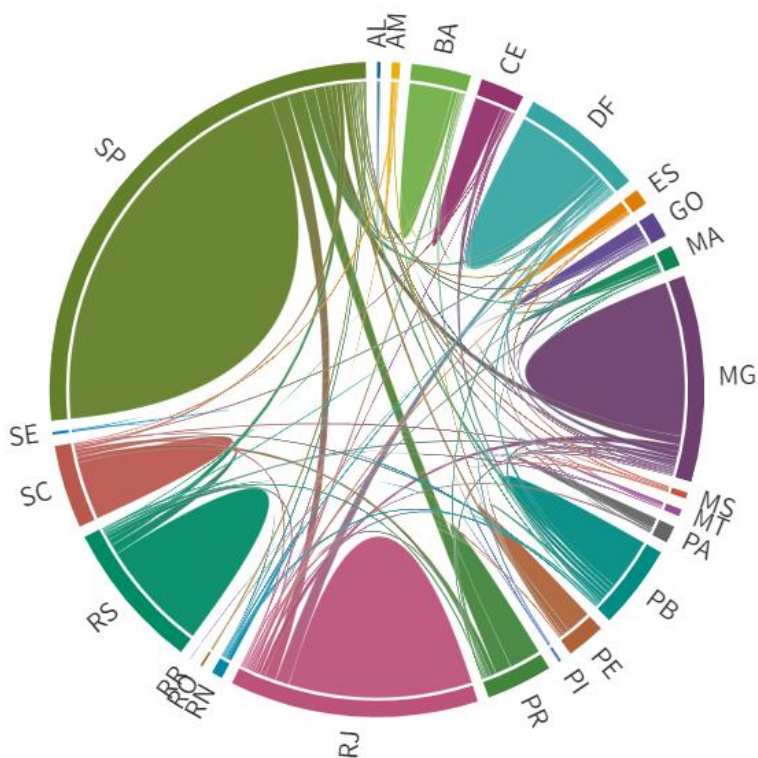
Fonte: Autores.

Percebe-se que a distância média entre as etapas tem uma variação significativa. Analisando a distância média do local de nascimento para graduação (primeiro registro sobre graduação concluída), observa-se que esta é a menor distância média calculada, podendo ser influenciada pelo fato de que grande parte das cidades brasileiras possuem instituições que proporcionam ao estudante cursos de graduação, e aquelas que não possuem na maioria das vezes, ficam próximas a outras cidades que detêm os cursos neste nível de capacitação de interesse dos estudantes. Já as maiores distâncias estão entre o local de nascimento e de capacitação é a nível de doutorado, seguido da graduação/doutorado respectivamente, possuindo as maiores distâncias percorridas pelos indivíduos entre estes níveis de formação. Destaca-se que o valor médio da díade graduação-doutorado pode ser influenciado por um quantitativo de indivíduos que realizam seus doutorados no exterior, cujas distâncias são mais representativas. Foi possível efetuar a distribuição entre os níveis considerando a mediana das distâncias percorridas no processo de formação acadêmica dos indivíduos analisados.

Como era de se esperar, os menores valores de mediana são entre mestrado-doutorado, graduação-mestrado respectivamente, uma vez que os estudantes em grande parte decidem seguir na mesma instituição durante essas transições em seus processos formativos. A maior mediana observada é a entre nascimento-doutorado, por se tratar de o local de nascimento do indivíduo até o mais alto nível de formação resultando em uma representação de toda a sua vida acadêmica.

Na Figura 4 foi possível representar o deslocamento dos indivíduos entre os estados brasileiros. Levando em conta desde o nascimento, todos os níveis de formação e atuação profissional.

Figura 4 - Fluxos a nível estadual percorridos pelos doutores



Fonte: Autores.

Inicialmente, destaca-se a ausência de alguns estados brasileiros, como por exemplo: Acre, Amapá e Tocantins, podendo ser justificado pela não habitação dos doutores nestes estados em algum momento de suas carreiras acadêmicas, ou porque os indivíduos nestas regiões não informaram em seus currículos.

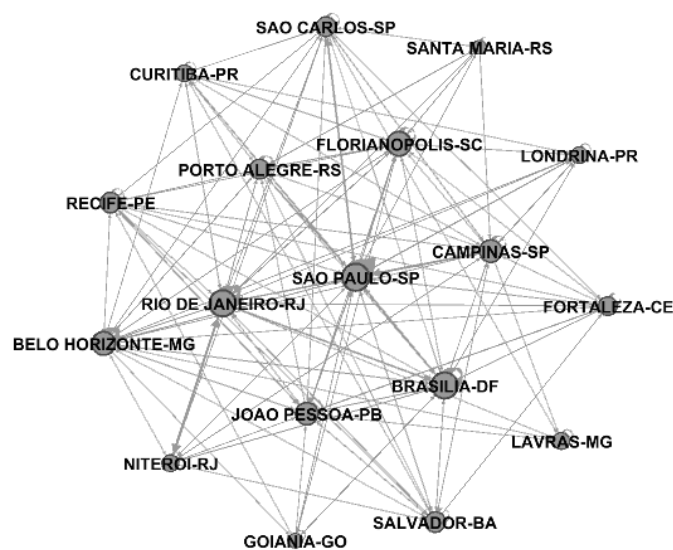
É possível observar estados como São Paulo, Rio de Janeiro, Minas Gerais e Rio Grande do Sul e o Distrito Federal se destacam por possuir mais caminhos percorridos dentro de seus próprios territórios em detrimento de outros estados brasileiros.

Destaca-se que os estados com maiores quantidades interestaduais se interagem, principalmente na emigração de Rio de Janeiro, Minas Gerais e Paraná para a imigração no estado de São Paulo que representa o estado com maior número de doutores atuando.

Ressalta-se também que todos os estados possuem vínculo com os outros estados do país, apesar de alguns em menores quantidades, como Roraima e Rondônia. Sendo esses os estados menos representativos, dos

presentes no gráfico, já que também se destacam por possuir quantidades menos representativas de indivíduos que nasceram neles. Na Figura 5 foi possível caracterizar uma rede representando onde os indivíduos mais optaram por se localizar em algum momento de sua formação acadêmica, levando em consideração os graus de entrada, filtrando-se para que o grau de entrada esteja entre 10 e 86. Ou seja, aqueles que possuíam entre 10 e 86 conexões entre os outros estados.

Figura 5 - Rede contendo as cidades com as maiores quantidades de vínculos.



Fonte: Autores.

Identificou-se que grande parte das cidades presentes no gráfico correspondem a capitais brasileiras, ou àquelas que possuem universidades públicas de grande relevância na educação brasileira.

Os nós de cada cidade são representados pelos círculos, em que quanto maior o diâmetro, maior será a quantidade de grau de entrada de um determinado nó (cidade), percebendo assim que a cidade de São Paulo corresponde a cidade com maior quantidade de vínculos.

Percebe-se também que cidades como São Paulo-SP, Rio de Janeiro-RJ e Belo Horizonte-MG somente não possuem um vínculo com uma cidade representada na rede, sendo assim, são bastante influentes na mobilidade acadêmica na área de Ciência da Informação.

5 CONCLUSÕES

Percebe-se que os doutores da área de Ciência da Informação, tendem a percorrer baixas distâncias ao longo do processo de formação, principalmente se tratando de um nível de formação para o seu superior, como por exemplo de mestrado para doutorado, pois grande parte dos indivíduos optam por permanecer na mesma instituição. Observa-se também que a localização geográfica possui forte influência sobre o processo migratório para capacitação, em que, os doutores da Ciência da Informação tendem a se habitar principalmente em estados da região Sudeste, Sul e no Distrito Federal. Se tratando das cidades com maiores quantidades de grau de entrada nas redes de vínculos, percebe-se o quanto a cidade de São Paulo é mais influente comparada às demais.

REFERÊNCIAS

ALBARRÁN, P.; CARRASCO, R.; RUIZ-CASTILLO, J. Geographic mobility and research productivity in a selection of top world economics departments. **Scientometrics**, [S. l.], v. 111, n. 1, p. 241-265, abr. 2017.

ALMEIDA, G.; ZANLORENSSI, G. **Fluxos migratórios**: a distribuição da população de cada estado pelo país. [S. l.]: Nexo, 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>. Acesso em: 19 jul. 2021.

AVEIRO, T. M. M. O programa Ciência sem Fronteiras como ferramenta de acesso à mobilidade internacional. **#Tear: Revista de Educação, Ciência e Tecnologia**, [S. l.], v. 3, n. 2, dez. 2014.

CAÑIBANO, C.; OTAMENDI, F. J.; SOLÍS, F. International temporary mobility of researchers: a cross-discipline study. **Scientometrics**, [S. l.], v. 89, n. 2, p. 653-675, nov. 2011.

DEMARTINI, M. **Falta de oportunidades mantém cientistas brasileiros no exterior**. [S. l.]: Exame, 2017. Disponível em: <https://exame.com/ciencia/falta-de-oportunidades-mantem-cientistas-brasileiros-no-exterior/>. Acesso em: 14 abr. 2021.

DIAS, T. M. R. **Um estudo da produção científica brasileira a partir de dados da Plataforma Lattes**. 2016. 181 f. Tese (Doutorado em Modelagem

Matemática e Computacional) - Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte, 2016.

DUBOIS, P.; ROCHET, J. C.; SCHLENKER, J. M. Productivity and mobility in academic research: evidence from mathematicians. **Scientometrics**, [S. l.], v. 98, n. 3, p. 1669-1701, mar. 2014.

KOSZTYÁN, Z.; FEHÉRVÖLGYI, B.; CSIZMADIA, T.; KEREDES, K. Investigating collaborative and mobility networks: reflections on the core missions of universities. **Scientometrics**, [S. l.], v. 126, p. 6 3551-3564, fev. 2021.

LANE, J. Let's make science metrics more scientific. **Nature**, v. 464, n. 7288, p. 488-489, mar. 2010.

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

MADEIRA, R. M.; MARENCO, A. Os desafios da internacionalização: mapeando dinâmicas e rotas da circulação internacional. **Revista Brasileira de Ciência Política**, [S. l.], p. 47-74, abr. 2016.

MARQUES, F. Registros valiosos. **Pesquisa FAPESP**, São Paulo, SP, v. 233, p. 34-37, jul. 2015. Disponível em: <https://revistapesquisa.fapesp.br/registros-valiosos/>. Acesso em: 21 jul. 2021.

RAMOS, M.; VELHO, L. Formação de doutores no Brasil e no exterior: impactos na propensão a migrar. **Educação & Sociedade**, [S. l.], v. 32, p. 933-951, dez. 2011.

SACHINI, E.; KARAMPEKIOS, N.; BRUTTI, P.; SIOUMALAS-CHRISTODOULOU, K. Should I stay or should I go? Using bibliometrics to identify the international mobility of highly educated Greek manpower. **Scientometrics**, [S. l.], v. 125, n. 1, p. 641-663, out. 2020.

SILVA, F. M.; SMIT, J. W. Organização da informação em sistemas eletrônicos abertos de Informação Científica & Tecnológica: análise da Plataforma Lattes. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 14, p. 77-98, abr. 2009.

VERGINER, L.; RICCABONI, M. Cities and countries in the global scientist mobility network. **Applied Network Science**, [S. l.], v. 5, n. 1, p. 38, jul. 2020.

AN ANALYSIS OF ACADEMIC MOBILITY BASED ON DATA FROM DOCTORS IN THE INFORMATION SCIENCE AREA

ABSTRACT

Objective: To analyze how the scientific exodus occurs in Brazil based on doctors in the area of Information Science, in which individuals from different locations in the country migrate in search of better academic training. **Methodology:** the Lattes Platform was used as a data source, extracting all the curricula with the LattesDataXplorer tool and from then on, all data processing was carried out for further analysis. **Results:** calculations were made of the average distances covered by the Doctors of Information Science, and it was also possible to highlight the flows of states throughout their academic training, as well as the main cities inhabited by these individuals. **Conclusions:** it was observed that individuals tend to travel short distances throughout the training process, geographic location has a strong influence on the migratory process for training and that the city of São Paulo is preferred by individuals in the training process.

Descriptors: Academic mobility. Information Science. Data analysis.

UN ANÁLISIS DE LA MOVILIDAD ACADÉMICA A PARTIR DE DATOS DE MÉDICOS DEL ÁREA DE CIENCIAS DE LA INFORMACIÓN

RESUMEN

Objetivo: Analizar cómo ocurre el éxodo científico en Brasil a partir de doctores en el área de Ciencias de la Información, en el que individuos de diferentes lugares del país migran en busca de una mejor formación académica. **Metodología:** se utilizó como fuente de datos la Plataforma Lattes, extrayendo todos los planes de estudio con la herramienta LattesDataXplorer y a partir de ahí se realizó todo el procesamiento de datos para su posterior análisis. **Resultados:** se realizaron cálculos de las distancias promedio recorridas por los Doctores en Ciencias de la Información, y también fue posible destacar los flujos de estados a lo largo de su formación académica, así como las principales ciudades habitadas por estos individuos. **Conclusiones:** se observó que los individuos tienden a viajar distancias cortas a lo largo del proceso de formación, la ubicación geográfica tiene una fuerte influencia en el proceso migratorio para la formación y que la ciudad de São Paulo es preferida por los individuos en el proceso de formación.

Descriptores: Movilidad académica. Ciencias de la Información. Análisis de datos.

Recebido em: 19.12.2022

Aceito em: 22.03.2023