

**ANÁLISE ESPAÇO-TEMPORAL DA
TRAJETÓRIA DE CARREIRA DE
PESQUISADORES BRASILEIROS**

CAIO ALVES FURTADO

**ANÁLISE ESPAÇO-TEMPORAL DA
TRAJETÓRIA DE CARREIRA DE
PESQUISADORES BRASILEIROS**

Dissertação apresentada ao Curso de Pós-Graduação
em Ciência da Computação do Instituto de Ciências
Exatas da Universidade Federal de Minas Gerais,
como requisito parcial à obtenção do grau de Mestre
em Ciência da Computação

ORIENTADOR: CLODOVEU AUGUSTO DAVIS JUNIOR

BELO HORIZONTE

MAIO 2015

© 2015, Caio Alves Furtado.
Todos os direitos reservados.

F992a Furtado, Caio Alves.
Análise espaço-temporal da trajetória de
carreira de pesquisadores brasileiros/Caio Alves
Furtado.
—Belo Horizonte, 2015.
xvii, 96f. : il. ; 29cm.

Dissertação (Mestrado) - Universidade Federal de
Minas Gerais – Departamento de Ciência da
Computação

Orientador: Clodoveu Augusto Davis Junior.

1. Computação - Teses. 2. Pesquisadores - Brasil.
Teses. 3. Curriculum vitae – Teses. 4. Sistemas de
informação geográfica - Teses. 5.I.Orientador. II.
Título.

519.6*72(043)

Dedicado a João José Alves

Agradecimentos

À minha mãe e irmã, pelo incentivo que sempre me deram.

À Juliana, pelo apoio.

Ao professor Clodoveu, por todos os anos de orientação.

Ao Douglas e à Thamara, pela grande ajuda durante o mestrado.

Aos colegas do CEFET-MG e da UFMG, pelo suporte e amizade.

*“I’ll come to meet you in Moscow, Idaho
I’ll drive down to you in Utah, specifically to Provo
I think it would be great
If you lived in a different state”
(Best Friends Forever – Tape Song)*

Resumo

O crescimento da produção científica brasileira nos últimos anos é notável, motivando uma investigação dos fatores, dentro e fora do país, que ajudaram a moldar esse rico ambiente de pesquisa. Este trabalho busca verificar a viabilidade de usar *curricula vitae* (CV) como fonte de dados, assim como um guia para usar visualizações geográficas com esses dados. Para esses fins, este trabalho fornece uma análise aprofundada da educação dos pesquisadores e da evolução dos principais grupos de pesquisa brasileiros, ao analisar a informação contida no CV de aproximadamente seis mil pesquisadores envolvidos nos Institutos Nacionais de Ciência e Tecnologia (INCTs). Dados sobre as etapas cursadas por cada pesquisador em sua formação, da graduação ao doutorado, incluindo também pós-doutorado e emprego, são extraídos do currículo Lattes dos pesquisadores. A localização e o tempo correspondentes a cada etapa da sua carreira são extraídos do CV e organizados para construir as trajetórias espaço-temporais de suas carreiras, que são então analisadas considerando fatores adicionais, incluindo a área do conhecimento dos INCTs aos quais os pesquisadores estão associados. As análises desses dados indicam aumento da prevalência de instituições brasileiras na formação desses pesquisadores, já que o número de doutorados obtidos no exterior diminui com o tempo. Pós-doutorados, por outro lado, são realizados em sua maioria na Europa e Estados Unidos. Obter um pós-doutorado internacional após completar toda sua formação no Brasil indica a busca, pelo pesquisador, por estágios mais avançados de experiência e cooperações de alto nível com grupos de pesquisa estrangeiros. Os resultados também mostram a tendência de pesquisadores brasileiros em buscar emprego perto da instituição em que obtiveram seu diploma de graduação, sugerindo baixa mobilidade dentro do país. São também discutidas as características particulares da visualização geográfica das trajetórias criadas. Mapas são capazes de mostrar padrões e exceções, mas quantidade de linhas exibidas pode atrapalhar a leitura. Filtros e transparência

facilitam a leitura do mapa, mas o uso desses recursos deve ser claro para o usuário. A capacidade da visualização de exibir muitos atributos é baixa, pois há um compromisso entre a legibilidade do mapa e formas de exibi-los. Elementos desse estudo podem ser usados na definição de políticas públicas para a correção de distorções na formação de pesquisadores. Além disso o trabalho permite verificar as limitações e potencialidades das técnicas de visualização na análise exploratória de dados espaço-temporais de currículos.

Palavras-chave: Trajetória de carreiras, Análise espaço-temporal, Curriculum Vitae, Ciência Brasileira, Visualização Geográfica

Abstract

The growth of the Brazilian scientific production in recent years is remarkable, which motivates an investigation on the factors, inside and outside the country, which helped shape this wealthy research environment. This work seeks to verify the viability to use curricula vitae data in spatio-temporal analysis, and provide a guide to use geographic visualizations with this data. Towards that end, this article provides a thorough analysis of the education and evolution of the main Brazilian research groups, by analyzing curriculum vitae data of about 6,000 researchers involved in the country's National Institutes of Science and Technology (INCT) initiative. Data on the steps taken by each researcher in her education, from the bachelor's degree to doctorate, including a possible postdoctoral experience, and employment, are extracted from an official curriculum vitae repository. The location and the time at which each career step was extracted from the CVs and are arranged to build spatiotemporal career trajectories, which are then analyzed considering additional factors, including the area of knowledge of the INCTs to which each researcher is associated. The analyses of said data indicate the increasing prevalence of Brazilian institutions in the education of Brazilian scientists, as the number of doctorates obtained abroad is diminishing over time. Postdoctoral stages, on the other hand, are primarily taken in Europe or in the United States. Taking an international postdoctoral position after a full education in Brazil indicates a drive towards seeking higher-level exchange and cooperation with foreign groups in a more advanced career stage. Results also point towards a tendency for Brazilian researchers to seek employment in regions that are close to the institutions at which they received their bachelor's degrees, suggesting low mobility within the country. We also discuss particular characteristics of the geographic visualization created. Maps are good to show patterns and outliers, but a big number of lines may hinder map reading. Filters and opacity levels help map reading, but their presence should be clear for the

user. The dimensionability is low, there's a compromise between map readability and how to show attributes. Elements from this study may be used in the definition of public policies for correcting distortions in the researchers' education. Moreover, it allows verify the limitations and potential of visualization techniques for exploratory analysis of spatiotemporal curricula data.

Keywords: Research career path, Spatio-temporal analysis, Curriculum vitae, Brazilian science, Geographic visualization

Lista de Figuras

Figura 3.1. Esquema do banco de dados.....	12
Figura 3.2. Exemplo de trajetória de pesquisador.....	14
Figura 3.3. Trajetórias do <i>dataset</i>	16
Figura 4.1. Distribuição cumulativa de pesquisadores para instituições (a) acadêmicas e (b) empregadoras.	20
Figura 4.2. Número de doutores graduados por década.....	22
Figura 4.3. Distribuição geográfica das instituições formadoras de PhDs (exceto Brasil).....	24
Figura 4.4. Proporção entre títulos brasileiros e internacionais	25
Figura 4.5. Regiões do Brasil.....	26
Figura 4.6. Títulos de PhD por região brasileira	27
Figura 4.7. Visualização geográfica das trajetórias dos pesquisadores da área de Agronegócio	28
Figura 4.8. Visualização geográfica das trajetórias dos pesquisadores da área de Energia	28
Figura 4.9. Visualização geográfica das trajetórias dos pesquisadores da área de Engenharia/TI.....	28
Figura 4.10. Visualização geográfica das trajetórias dos pesquisadores da área de Exatas	29
Figura 4.11. Visualização geográfica das trajetórias dos pesquisadores da área de Humanas/Sociais	29
Figura 4.12. Visualização geográfica das trajetórias dos pesquisadores da área de Meio Ambiente	29
Figura 4.13. Visualização geográfica das trajetórias dos pesquisadores da área de Nanotecnologia	30
Figura 4.14. Visualização geográfica das trajetórias dos pesquisadores da área de Saúde.....	30

Figura 4.15. PhDs concedidos por década.....	32
Figura 4.16. Títulos de PhD por década - Nacional.....	33
Figura 4.17. Títulos de PhD por década - Internacional.....	34
Figura 4.18. Títulos PhD por ano e país (a), estado (b), cidade (c)	35
Figura 4.19. Segmentos de trajetória: década de 1960	36
Figura 4.20. Segmentos de trajetória: década de 1970	36
Figura 4.21. Segmentos de trajetória: década de 1980	37
Figura 4.22. Segmentos de trajetória: década de 1990	37
Figura 4.23. Segmentos de trajetória: década de 2000	37
Figura 4.24. Segmentos de trajetória: década de 2010	38
Figura 4.25. Distribuição de doutores no Brasil e nos INCTs.....	40
Figura 4.26. Grafo de migração dos pesquisadores, do doutorado para o emprego	44
Figura 4.27. Trajetória dos pesquisadores no mestrado (a), doutorado (b) e pós-doutorado (c).....	47
Figura 4.28. Distância percorrida pelos pesquisadores até a instituição de doutorado.....	54
Figura 4.29. Distância percorrida pelos pesquisadores do primeiro ao último ponto da trajetória.....	55
Figura 5.1. Filtros presentes na ferramenta	58
Figura 5.2. Visualização geral dos dados de trajetória.....	60
Figura 5.3. Visualização dos dados de trajetória: estágios acadêmicos (a), pós-doutorado (b) e emprego (c).....	61
Figura 5.4. Trajetórias entre costa Leste e Oeste dos Estados Unidos.....	62
Figura 5.5. Visualização das trajetórias	63
Figura 5.6. Melhorias na leitura do mapa: zoom (a) e cor (b)	64
Figura 5.7. Mapa após agrupar as trajetórias por país (a) e por estado (b).....	66
Figura 5.8. Mapa exibindo 10.000 trajetórias.....	67
Figura 5.9. Mapa exibindo 5.000 trajetórias	67
Figura 5.10. Mapa exibindo 1.000 trajetórias.....	68
Figura 5.11. Mapa exibindo 500 trajetórias	68
Figura 5.12. Mapa exibindo 100 trajetórias	68
Figura 5.13. Mapa exibindo 50 trajetórias	69
Figura 5.14. Trajetórias de distância 0km (a), com menos de 5 mil km (b) e com mais de 5 mil km (c).....	70

Figura 5.15. Exemplos de trajetórias com nível de opacidade: 50% (a), 25% (b) e 1% (c).....	71
Figura 5.16. Exemplos de melhoria da leitura: Zoom e Opacidade de 25% (a) e 10% (b).....	73
Figura 5.17. Exemplos de mapas com uma (a), duas (b), quatro (c) e oito (d) cores ao mesmo tempo.....	76
Figura 5.18. Efeito da transparência na soma de cores.....	77
Figura 5.19. Exemplos de mapa com escala monocromática (a) e escala monocromática com opacidade (b)	78
Figura 5.20. Representação 1:1 de espessura e pesquisadores	79
Figura 5.21. Uso de espessura – Destaque em SP	80
Figura 5.22. Instituições onde pesquisadores estudaram duas ou mais etapas consecutivamente.....	81
Figura 5.23. Representação 1:1 do tamanho do ponto e pesquisadores.....	82
Figura 5.24. Legenda da simbologia usada para representar faixas de valor.....	82
Figura 5.25. Exemplo de mapa com espessura e símbolos em escala	83

Lista de Tabelas

Tabela 4.1. Distribuição de pesquisadores por área de pesquisa	21
Tabela 4.2. Mestres e pesquisadores com pós-doutorado por área de pesquisa	23
Tabela 4.3. Distribuição dos pesquisadores por região e estágio da trajetória	39
Tabela 4.4. Análise de Origem e Destino das trajetórias – da graduação para o emprego.	41
Tabela 4.5. Análise de Origem e Destino das trajetórias – da graduação para o doutorado.	42
Tabela 4.6. Análise de Origem e Destino das trajetórias – do doutorado para o emprego	43
Tabela 4.7. Estágios obtidos no exterior por modalidade	45
Tabela 4.8. Número de pesquisadores e entradas por estágio	49
Tabela 4.9. Padrões de modalidade acadêmica mais comuns.....	50
Tabela 4.10. Padrões de regiões mais comuns.....	51
Tabela 4.11. Padrões de regiões mais comuns – incluindo emprego.....	52
Tabela 4.12. Padrões genéricos de movimentação mais comuns – incluindo emprego.	53

Sumário

Resumo	vii
Abstract	ix
Lista de Figuras	xi
Lista de Tabelas	xv
1 Introdução	1
1.1 Objetivos	2
1.2 Objetivos específicos	3
1.3 Estrutura	4
2 Contextualização e Trabalhos Relacionados	5
3 Trajetória de Carreira de Pesquisadores	11
3.1 Descrição dos dados	11
3.2 Modelagem.....	12
3.3 Geocodificação	16
4 Análise dos Dados	19
4.1 Distribuição da afiliação em instituições acadêmicas e empregadoras.....	19
4.2 Distribuição de pesquisador por área de pesquisa	21
4.2.1 Mestrados e pós-doutorados por área de pesquisa	23
4.2.2 Distribuição geográfica das de acordo com a área de pesquisa	24
4.3 Distribuição espaço-temporal dos títulos de doutorado	31

4.4	Pontos de origem e destino das trajetórias	38
4.5	Internacionalização	45
4.6	Padrões de trajetória	49
5	Discussão das Técnicas Usadas	57
5.1	A Ferramenta	57
5.2	Tipo de Dado.....	59
5.3	Tipo de Tarefa.....	59
5.4	Escalabilidade.....	63
5.4.1	Filtros	65
5.4.2	Transparência	71
5.5	Dimensionabilidade.....	74
5.5.1	Cor	75
5.5.2	Espessura	78
5.6	Posição dos Atributos.....	83
6	Conclusão.....	85
6.1	Trabalhos Futuros.....	88
	Bibliografia	89

Capítulo 1

Introdução

A produção científica brasileira cresceu consideravelmente nos últimos quinze anos, chegando a uma taxa média de crescimento de 10,7% ao ano (Almeida et al., 2013). Com esse desempenho, o Brasil tem crescido num ritmo cinco vezes maior que a média mundial, e o país alcançou a 13^a posição no ranking internacional das nações mais produtivas (Almeida et al., 2013). O sistema acadêmico brasileiro, no entanto, pode ser considerado bem jovem quando comparado com aqueles de países da Europa, América do Norte, Ásia, Oceania ou mesmo da América do Sul. A primeira universidade brasileira foi fundada somente na década de 1910 e o principal “ecossistema” de pesquisa em universidades que existe atualmente, formado por universidades federais espalhadas por todos os estados do Brasil, data da década de 1950 (Souza, 1991).

Portanto, é interessante buscar entender como um ecossistema científico tão novo evoluiu para produzir resultados tão notáveis em um período de tempo tão curto, identificando as influências dentro e fora do país que ajudaram a formar esse ambiente de pesquisa produtivo. Um possível modo de analisar esse ecossistema é estudar os pesquisadores que fazem parte dele. Há vários aspectos na carreira do pesquisador que podem indicar características da comunidade científica na qual ele está inserido. Por exemplo, o número de artigos publicados e o tamanho médio da rede de coautoria são indícios da qualidade dos pesquisadores dessa comunidade. Local de formação e local de emprego, por outro lado, indicam a ocorrência de fenômenos como a migração de pesquisadores e endogamia na formação.

Existem vários meios para obter dados sobre os pesquisadores. Uma delas é o seu *curriculum vitae* (CV), que é uma fonte rica em informações dos cientistas brasileiros. No Brasil, existe o portal Lattes¹, uma aplicação online criada pelo Comitê Nacional de Desenvolvimento Científico e Tecnológico² (CNPq) para coletar e integrar informações do *curriculum vitae* da comunidade acadêmica brasileira como um todo. O uso de currículos não é livre de problemas: há falta de padronização entre currículos, dificuldade de se coletar uma quantidade grande de CV, e existem currículos com informações incompletas e desatualizados. No entanto, o sistema Lattes apresenta vantagens que superam vários desses problemas, por tornar disponível um conjunto de dados padronizado e de fácil acesso. Pesquisadores e estudantes são normalmente incentivados a informarem seus dados ao Lattes e a deixarem seu perfil sempre atualizado, já que o currículo é pré-requisito para a submissão de projetos de pesquisa, para o pedido de bolsas de estudo e de outras formas de financiamento. Toda a informação contida no portal fica disponível publicamente, e o pesquisador é o responsável pela veracidade e acurácia dos dados publicados. Atualmente, sua cobertura é de praticamente todos os pesquisadores, instituições e grupos de pesquisa brasileiros que estão ativos. Conseqüentemente, o Lattes é uma fonte de dados detalhados e atuais sobre a produção científica brasileira, formada a partir da perspectiva dos indivíduos envolvidos. Agências de fomento, fundações de apoio à ciência e tecnologia, assim como o próprio CNPq, usam a plataforma para gestão e planejamento.

1.1 Objetivos

O objetivo deste trabalho é analisar o comportamento espaço-temporal dos pesquisadores ao longo de suas carreiras, usando o seu CV Lattes como fonte de dados. Entre os diversos dados disponíveis no currículo Lattes estão as instituições onde o pesquisador obteve sua formação acadêmica e seu local de emprego. Cada nível de formação é associado aos anos de início e fim do curso, e

¹ Lattes: <http://lattes.cnpq.br>

² CNPq: <http://www.cnpq.br/>

cada instituição pode ser localizada espacialmente, ou seja, associada a um par de coordenadas geográficas representando sua sede. A sequência de locais e períodos de tempo associados, então, pode ser transformada em *trajetória espaço-temporal*. Esta dissertação apresenta detalhes sobre o processo de obtenção e tratamento dos dados utilizados, e discute como foram modelados para análise, em analogia à análise de objetos móveis em geoinformática. São apresentadas análises variadas, e discutidos os resultados encontrados. Adicionalmente, é feita uma discussão sobre o uso de técnicas de visualização aplicadas aos dados modelados, indicando o que foi possível descobrir com as análises, se a modelagem do dado em trajetória foi adequada para o problema, que tipos de visualização são mais úteis para encontrar padrões de comportamento, entre outras questões.

1.2 Objetivos específicos

Os seguintes objetivos específicos foram definidos, como etapas para a consecução do objetivo geral:

- Obter dados de currículos de um subconjunto representativo dos pesquisadores brasileiros; para este trabalho, são utilizados os currículos de todos os pesquisadores associados a algum Instituto Nacional de Ciência e Tecnologia³;
- Extrair atributos básicos e dados de formação (graduação, mestrado, doutorado, pós-doutorado) e emprego (afiliação institucional) dos pesquisadores;
- Localizar espacialmente as instituições citadas nos currículos de pesquisadores;

³ INCT é um programa criado pelo CNPq para impulsionar a pesquisa científica brasileira em certas áreas de interesse. A lista dos grupos apoiados pelo programa pode ser encontrada no site <http://inct.cnpq.br/institutos/>

O programa será discutido mais detalhadamente nas próximas seções.

- Formar um banco de dados contendo as trajetórias espaço-temporais dos pesquisadores e outros dados associados à sua área de atuação, em conexão com os INCTs;
- Desenvolvimento de uma ferramenta interativa para análises exploratórias dos dados. Avaliar as técnicas de visualização geográfica propostas, utilizando métricas existentes para compará-la com outras técnicas de visualização.

1.3 Estrutura

O trabalho está estruturado da seguinte forma. A Seção 2 contém trabalhos relacionados e uma contextualização do problema. A Seção 3 apresenta a modelagem do problema, a descrição da trajetória acadêmica, os dados utilizados e o tratamento aplicado neles. Na Seção 4 mostramos as análises feitas sobre os dados e, particularmente, sobre as trajetórias dos pesquisadores. Na Seção 5 são discutidas as vantagens e desvantagens das técnicas de visualização utilizadas. Por fim, a Seção 6 apresenta as conclusões obtidas e lista trabalhos futuros.

Capítulo 2

Contextualização e Trabalhos Relacionados

O estudo dos pesquisadores de uma comunidade científica é um tema constante de pesquisa. É de interesse do governo e da própria comunidade conhecer certos aspectos tais como (1) o local de formação dos doutores de um país (Arenas et al., 2001); (2) a internacionalização e migração que ocorreram com a população científica em um intervalo de tempo (Song, 1997); (3) a taxa de eugenia (contratação de cientistas que se formaram na própria instituição) dos pesquisadores (Padilha, 2007) e (4) as diferenças na carreira de pesquisadores de sexos diferentes (Schwabe, 2011). O presente trabalho envolve avaliar dados de um grupo de pesquisadores brasileiros, por meio da visualização espaço-temporal de dados contidos no currículo de cada um deles.

O uso de *curriculum vitae* como fonte de dados para análises já foi discutido pela comunidade científica. Gaughan & Bozeman [2002] usam dados obtidos a partir de CVs para verificar a influência do recebimento de investimentos do governo na produtividade científica de pesquisadores. Morzinski & Schubot [2000] também usam currículos para avaliar programas de desenvolvimento do corpo docente. Esses artigos, no entanto, citam a falta de padronização e dificuldade de obtenção dos currículos como desvantagem no uso desse documento como fonte de dados.

A plataforma Lattes é um sistema brasileiro, criado pelo CNPq, que funciona como repositório de currículos de pesquisadores. Ela possui dados

detalhados a respeito dos pesquisadores brasileiros, e premissa tais como livre acesso aos dados e estrutura padronizada fazem com que o uso dos seus currículos não apresente as desvantagens constatadas na literatura (Cañibano & Bozeman, 2009). Sua riqueza de dados faz com que a plataforma seja usada em vários estudos. Balancieri et al. [2005] fazem uma avaliação histórica de redes de colaboração e discutem como o Lattes ajuda no estudo dessas redes. Outro estudo da área foi feito por Mena-Chalco & Cesar Junior [2009], que coletam os currículos Lattes e fazem algumas análises desses dados, incluindo análises geográficas de orientação de pesquisadores. Entretanto, a variedade de análises disponíveis é pequena e voltada para um grupo limitado de pesquisadores. Entre os trabalhos que fazem uso das informações disponíveis no Lattes estão os de Augusta et al. [2008] e Kleber et al. [2012]. Os dois trabalhos concentram-se na análise do perfil de pesquisadores na área de saúde e estudam a distribuição dos pesquisadores pelo Brasil. Ambos os trabalhos, desenvolvidos independentemente um do outro, mostram a necessidade que os pesquisadores têm de uma ferramenta que os ajude a analisar os grupos científicos brasileiros.

O CNPq, junto com a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e outras fundações regionais de ciência, criou em 2008 um programa para fomentar e promover grupos de pesquisa brasileiros, chamado *Institutos Nacionais de Ciência e Tecnologia*⁴, ou INCT. Com um investimento substancial, foram criados 101 institutos espalhados em todas as regiões, cobrindo áreas consideradas estrategicamente importantes para o país. O programa não cobre todos os grupos de pesquisa brasileiros, mas inclui vários dos melhores grupos de pesquisa em atividade.

O Instituto Nacional de Ciência e Tecnologia para a Web⁵ (InWeb) é um dos institutos criados pelo programa INCT. O grupo tem como objetivo desenvolver modelos e algoritmos para melhorar a integração da Web à sociedade. Um dos projetos conduzidos pelo InWeb é o CiênciaBrasil (Laender et al., 2011). O projeto criou um portal⁶ onde os currículos Lattes são reestruturados de forma a configurar uma rede social baseada em pesquisadores.

⁴ INCT: http://estatico.cnpq.br/programas/inct/_apresentacao/apresentacao.html

⁵ InWeb: <http://www.inweb.org.br/>

⁶ CiênciaBrasil: <http://pbct.inweb.org.br>

Os relacionamentos nessa rede social são caracterizados como colaborações em publicações: se dois pesquisadores são coautores de um artigo, eles são considerados “amigos” nessa rede social. Essas redes permitem a investigação de vários aspectos da carreira dos pesquisadores, identificando práticas, comportamentos e características dos padrões obtidos da produção e colaboração científica. Atualmente, o portal inclui os dados de currículo dos pesquisadores membros de INCTs, que tenham pelo menos começado a cursar seu doutorado. No entanto, o conteúdo do CiênciaBrasil está constantemente em expansão.

Redes sociais acadêmicas, como o CiênciaBrasil, são úteis para estudar o perfil científico de um grupo. Foram propostas para analisar a mobilidade de estudantes e profissionais entre instituições participantes do Processo de Bolonha⁷ (Ferreira & Porfírio, 2010), um programa europeu feito para melhorar o ensino superior do continente. Redes sociais são também usadas na criação de planos de estudo (Jorge & Porfírio, 2010). Redes sociais acadêmicas podem ser usadas também para a recomendação de coautores para um pesquisador (Lopes et al., 2010; Oh & Jeng, 2011).

A vida acadêmica dos pesquisadores pode ser descrita como uma trajetória de objetos móveis (Etienne et al., 2012). A trajetória é definida como um conjunto de posições, que correspondem aos lugares em que o pesquisador obteve seus diplomas, associados a um período de tempo. Essas posições são compostas de uma marca temporal e uma coordenada espacial, correspondendo à posição da instituição onde o pesquisador estudou, e a ordem cronológica das posições forma a trajetória. Ao contrário de trajetórias de objetos móveis, a granularidade temporal da trajetória acadêmica é maior, podendo durar anos ou mesmo décadas, e ela possui um número menor de pontos, limitado ao número de graus de formação que o pesquisador obteve.

As trajetórias possuem características que devem ser consideradas para uma visualização. Spaccapietra [2007] define que uma trajetória é um conjunto de posições de um objeto que se move no espaço dado um intervalo de tempo. Spaccapietra também define que as trajetórias podem diferir na modelagem,

7 About the Bologna Process <http://www.ond.vlaanderen.be/hogeronderwijs/bologna/about/>

sendo metafóricas (movimento da academia para a indústria, por exemplo), *ingênuo-geográficas* (movimento de Minas Gerais para a Bahia), ou espaço-temporais (descrição do movimento de uma bola de boliche). Este trabalho irá tratar de casos *ingênuo-geográficos* e espaço-temporais.

Para obter as coordenadas espaciais das instituições é necessário geocodificá-las. Geocodificação é o ato de associar uma coordenada espacial a um endereço (Goldberg et al., 2007). Geocodificação costuma ser associada com a localização de endereços postais (Davis Jr. et al., 2013; Davis Jr. & Fonseca, 2011), mas tem sido adotado um conceito mais abrangente, considerando como fonte qualquer tipo de descrição textual (Goldberg et al., 2007; Goldberg, 2011), usando uma gama variada de dados de suporte como *gazetteers* (Machado et al., 2011) e *application programming interfaces* (APIs) de serviços de mapas online (Alencar & Davis, 2011).

Este trabalho utiliza visualizações geográficas para estudar padrões contidos nas trajetórias acadêmicas dos pesquisadores. A visualização de dados é um importante instrumento para a exploração de conjuntos de dados, bem como para melhorar a capacidade de compreensão e comunicação do resultado de análises. Algumas visualizações já são aplicadas no portal CiênciaBrasil, mas nenhuma delas é voltada para os dados geográficos. Neste trabalho propomos visualizações dos dados espaço-temporais contidos no Lattes. A vantagem de usar recursos de visualização de dados para descoberta de informação em bancos de dados é conhecida (Fayyad et al., 2001), sendo inclusive usada em outros portais de redes sociais acadêmicas (Ganev et al., 2009). Enkhsaikhan et al. [2008] discutem a eficiência de se usar visualização de dados para análise de redes sociais acadêmicas.

Câmara et al. [2004] consideram que a ênfase na análise de dados espaciais deve ser na medição de propriedades e relacionamentos, considerando explicitamente a localidade espacial como parte do fenômeno estudado. Na presente análise, nós nos concentramos nas instituições onde os pesquisadores dos INCTs estudaram/trabalham e como a trajetória entre elas se relaciona com certos aspectos do CV do pesquisador. Para tal, usamos técnicas de visualização geográfica de dados. A visualização geográfica dos dados também é útil para a descoberta de conhecimento (Gahegan et al., 2001). Ferramentas de visualização

são geralmente usadas como parte da análise de dados espaciais, especialmente na etapa exploratória. Ferramentas interativas de visualização de dados geográficos, como a que apresentamos neste trabalho, são meios de descobrir propriedades dos dados (Haining, 2003).

Alguns autores já propuseram formas de mapear dados cientométricos. Leydesdorff & Persson [2010] e Bornmann et al. [2011] mostram a viabilidade de usar mapas pontuais que relacionam os dados de produção científica e citação com a cidade em que os pesquisadores estão empregados. Bornmann & Waltman [2011] também usam *heatmaps* para mostrar dados de citação relacionados com o local de emprego. Em geral, encontramos visualizações clássicas da cartografia, como mapas pontuais, coropléticos e *heatmaps*. O trabalho de Börner et al. [2006], além de usar mapas coropléticos, utiliza uma visualização de trajetórias para estudar o fluxo da difusão de conhecimento nos Estados Unidos. No entanto, Börner não se aprofunda na discussão da utilidade dessa técnica para descoberta de conhecimento ao visualizar as trajetórias.

Os artigos acima possuem em comum o fato de não avaliarem as vantagens e desvantagens no uso dos mapas. Seria o mapa a melhor escolha para visualizar os dados ou opções mais tradicionais, como gráficos, irão exibir melhor o que o autor deseja? O livro *How to Lie with Maps* (Monmonier, 1996) discute como mapas podem distorcer a realidade, mesmo involuntariamente, se aspectos como legenda e escala não forem cuidadosamente escolhidos. Dias et al. [2012] discute como escolher a melhor técnica de visualização de um dado, considerando aspectos como escalabilidade da técnica, o tipo de dado e para qual finalidade a visualização será usada.

Embora o estudo de Dias envolva diversas técnicas de visualização, seu resultado não pode ser aplicado diretamente para visualizações geográficas. As visualizações apresentadas são: *Standard 1D-3D techniques*, *Icon-based techniques*, *Geometrically transformed displays*, *Pixel-oriented techniques*, e *Graph-based techniques*. Visualizações geográficas possuem algumas características em comum com as técnicas acima, mas o trabalho de Dias não leva em conta aspectos geográficos que visualizações geográficas, como mapas, precisam obedecer.

Neste trabalho serão avaliadas as visualizações geográficas geradas para dados de trajetórias de pesquisadores usando métricas propostas por Dias. Essas métricas definem os dados por tipo, dimensão e escala, além de medir as tarefas que podem ser exercidas ao analisar a visualização. As técnicas de visualização geográfica serão comparadas com técnicas mais clássicas utilizando as mesmas métricas. Usaremos o conjunto de dados do CiênciaBrasil como fonte de dados e, a partir das visualizações criadas, clássicas e geográficas, será feita uma avaliação das técnicas.

Capítulo 3

Trajетória de Carreira de Pesquisadores

O nosso conjunto de dados é composto do perfil Lattes de pesquisadores relacionados aos INCTs. Nesta seção serão discutidos os dados utilizados; o tratamento e a modelagem aplicados; e suas características.

3.1 Descrição dos dados

Os currículos dos pesquisadores vinculados a INCTs foram coletados do Lattes para o portal CiênciaBrasil (Laender et al., 2011). Os CVs estão publicamente disponíveis para download como arquivos XML, mas é necessária uma permissão especial do portal para baixar grandes volumes de dados. A coleta ocorreu entre dezembro de 2012 e agosto de 2013. Atualmente, o portal possui dados somente de pesquisadores de INCTs e que possuam um doutorado concluído ou em curso.

Nem todos os pesquisadores presentes no CiênciaBrasil possuem os dados necessários para as nossas análises. Removemos todos os pesquisadores para os quais não existe informação sobre local e ano de início do doutorado, resultando em 5.973 currículos válidos no total. Consideramos a informação dos currículos correta, devido ao uso dos dados para fomento de pesquisa e avaliação de produtividade. Além disso, verificamos as datas de modificação para verificar o quão atual são os dados. Descobrimos que 94,7% dos perfis foram modificados entre Janeiro de 2012 e agosto de 2013, e 76,8% dos perfis foram modificados entre janeiro de 2013 e agosto de 2013.

```

Researcher(LattesID, [degreeType, institutionID,
    startYear, endYear], workInstitutionID)
Institution(institutionID, institutionName, location)
INCT(inctID, inctName, inctArea)
INCTResearchers(inctID, LattesID)

```

Figura 3.1. Esquema do banco de dados

Não foram utilizados todos os dados disponíveis no currículo coletado. O esquema dos dados utilizados é apresentado na Figura 3.1. Neste trabalho usamos as informações de INCT associado, formação acadêmica e emprego do pesquisador. Um pesquisador pode ter diversas informações de formação, mas dispõe-se de somente do local de emprego mais atual. Os dados de formação incluem a modalidade cursada, ano de início e conclusão, além da instituição onde estudou. No caso do emprego, tem-se somente a instituição onde o pesquisador trabalha atualmente. Pesquisadores estão associados a diferentes INCTs. Um INCT possui no mínimo um pesquisador associado e cada pesquisador está associado a, no mínimo, um INCT. Podem existir casos (0,72% dos pesquisadores) onde um mesmo pesquisador está associado a mais de um INCT.

3.2 Modelagem

Foi considerado o conjunto de instituições em que o pesquisador se formou, junto com o intervalo de tempo em que ele esteve associado a essa instituição, como sua *trajetória acadêmica*. A trajetória envolve todos os eventos de sua formação, da graduação até o pós-doutorado e sua contratação. A vida acadêmica do pesquisador foi transformada em uma *trajetória acadêmica*, que inclui a formação da graduação até o pós-doutorado, terminando com o emprego. Cada evento (graduação, mestrado, etc.) é um ponto com dimensão espacial e temporal. Usamos como dimensão temporal os anos de início e fim de cada etapa de formação, e como dimensão espacial a posição encontrada pela geocodificação da instituição.

Os eventos são ligados por segmentos, ou *moves*, que representam o movimento do pesquisador ao fim de um evento para o início do próximo. Esses movimentos não possuem significado real: um pesquisador que terminou seu mestrado em São Paulo em 2002 e começou seu doutorado em Portugal em 2005 não passou três anos viajando de um país para outro. Também não podemos afirmar que durante o intervalo de 2002 e 2005 ele não esteve em outros países por motivos não relacionados à sua formação acadêmica. Os *moves* são uma ferramenta para auxiliar na visão de onde o pesquisador estava e para onde ele foi, considerando a informação acadêmica disponível no seu CV.

A trajetória esperada é *graduação* → *mestrado* → *doutorado* → *pós-doutorado* → *emprego*, mas essa ordem não é sempre verificada. Todo pesquisador possui pelo menos um doutorado, mas o número e a ordem das etapas de formação podem ser alterados. Um pesquisador pode ter dois doutorados e nenhum pós-doutorado, ou pode ter feito seu doutorado logo após terminar a graduação e em seguida obter a titulação de mestrado. A sequência é indicada pelo ano de conclusão de cada curso. No entanto, os dados obtidos possuem somente o local de emprego mais atual, sem as datas de contratação e conclusão indicadas. Sendo assim, consideramos que a etapa de emprego é sempre a última na trajetória.

Veja um exemplo de trajetória acadêmica na Figura 3.2. O pesquisador no exemplo fez seu bacharelado no Ceará, mudou-se para São Paulo para fazer seu mestrado, foi para Virgínia, Estados Unidos da América, para fazer seu doutorado, voltou para o Brasil, concluiu seu pós-doutorado em Minas Gerais e acabou sendo empregado nessa mesma instituição.

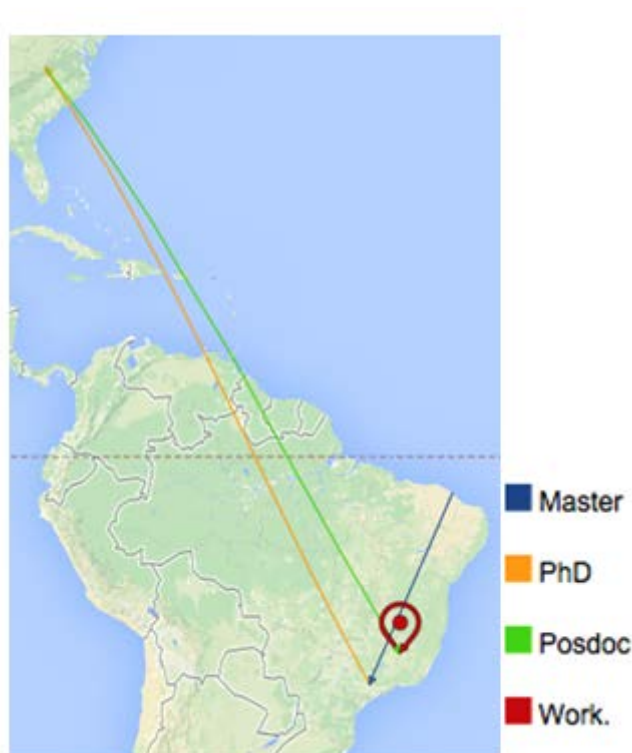


Figura 3.2. Exemplo de trajetória de pesquisador.

Como dito anteriormente, a trajetória acadêmica pode ser vista como uma trajetória de objetos móveis. A trajetória de um objeto móvel é definida como um conjunto de posições, onde cada posição possui uma localização geográfica e um *timestamp* (Etienne et al., 2012). A modelagem dos dados de CVs em trajetória é feita do seguinte modo. Os dados de formação e emprego definem as posições da trajetória do pesquisador. Os anos de início e fim de cada etapa formam o *timestamp* da posição. A instituição na qual o pesquisador obteve sua formação e/ou foi empregado é transformada em um ponto geográfico, servindo como a localização geográfica de uma etapa na trajetória. O conjunto de posições do pesquisador é ordenado cronologicamente pelo ano de conclusão de cada etapa. A posição do local de emprego é ordenada como o último ponto da trajetória, por representar o emprego mais atual e não possuir ano de conclusão.

Vale ressaltar que nem todo evento de formação acadêmica atende aos requisitos espaço-temporais necessários. O Lattes não exige o preenchimento obrigatório dos campos de ano de início e ano de conclusão, o que pode impedir

a inserção do evento na trajetória acadêmica. Algumas instituições também não foram geolocalizadas, não possuindo nenhuma coordenada geográfica relacionada a elas. Não consideramos esses casos na criação das trajetórias; um pesquisador que fez graduação, mestrado e doutorado, mas que não declarou o ano de conclusão do mestrado teria a trajetória *graduação* → *doutorado*. Da mesma forma, se a instituição declarada no seu bacharelado não pôde ser geolocalizada, sua trajetória seria *mestrado* → *doutorado*. No total, temos 27.565 eventos e 21.592 segmentos em nosso banco de dados. Removendo os eventos irregulares, temos 27.065 eventos e 21.092 segmentos. Foi criada uma ferramenta de visualização⁸ desses segmentos (Furtado et al., 2014), usada para enxergar a distribuição e encontrar possíveis padrões ou *outliers* nas trajetórias. A ferramenta mostra todas as trajetórias do banco, e é possível aplicar certos filtros para que seja exibida somente a informação desejada. Atualmente é possível filtrar as trajetórias por década de conclusão, área de pesquisa, atividade que exercia (formação acadêmica ou experiência profissional), além de poder agrupar as trajetórias por país e escolher trajetórias com certos destinos e origens. Também é possível procurar por trajetórias de pesquisadores específicos passando o identificador de seu currículo Lattes.

A Figura 3.3 mostra todas as trajetórias obtidas, com o tamanho do símbolo normalizado pelo número de vezes em que o segmento se repete no banco de dados. As linhas mostram movimentos entre duas instituições e os pontos mostram os segmentos de tamanho zero, que acontecem quando o pesquisador continua na mesma instituição em etapas consecutivas.

⁸ <http://aqui.io/trajectory>



Figura 3.3. Trajetórias do *dataset*

3.3 Geocodificação

O currículo Lattes possui vários atributos que indicam de forma direta ou indireta a posição geográfica do pesquisador em um dado instante do tempo. Nós usamos técnicas de geocodificação para estabelecer uma relação entre as instituições e sua real posição geográfica, usando o nome declarado pelo pesquisador para descobrir uma localidade associada a ela.

Uma das referências geográficas mais importantes contidas nos currículos são as instituições de afiliação. Para obter a trajetória acadêmica dos pesquisadores é necessário traduzir as instituições atendidas por eles para pontos no espaço e tempo. A informação temporal é dada pelo próprio pesquisador: ao declarar o local em que estudou também é informado os anos de início e conclusão. Para obter a posição geográfica da instituição, usamos técnicas de geocodificação para traduzir o nome em um ponto no mapa.

Nós geramos uma lista com referências únicas de instituições encontradas nos CVs. Essa lista foi geocodificada usando o Google Geocoding API⁹, passando o nome da instituição como endereço de entrada. A identificação do endereço depende da capacidade da API de reconhecer o nome de pontos de referência, como grandes construções ou instalações. No entanto, foram

⁹ <https://developers.google.com/maps/documentation/geocoding/>

detectadas variações na forma que os pesquisadores escreveram o nome das instituições no currículo. O mais comum é utilizar somente o nome da instituição em que trabalham ou estudaram, mas alguns pesquisadores também informaram o campus ou unidade. Essa informação extra nos ajudou a encontrar instituições que possuem campus em múltiplas cidades. Nos casos onde a informação de campus não foi incluída consideramos que o pesquisador estava localizado no campus sede da instituição.

No total, existem 3.478 instituições únicas no banco. Elas se dividem em *instituições de ensino* e *instituições de emprego*. Instituições de ensino são aquelas em que os pesquisadores estudaram e obtiveram sua formação. As instituições de emprego são aquelas onde os pesquisadores trabalham atualmente. A divisão é a seguinte: 2.858 instituições são exclusivamente de ensino, 399 são exclusivamente empregadoras e 221 instituições são instituições de ensino e de emprego ao mesmo tempo.

Para cada entrada única foi identificada a sua cidade, estado, país e coordenadas geográficas. Verificamos a validade do processo de geocodificação inspecionando manualmente um conjunto de instituições. Escolhemos instituições de emprego e instituições acadêmicas que concederam títulos de doutorado ou empregam cinco ou mais pesquisadores associados a elas. Obtivemos 84 instituições de ensino que atendem a essa condição, correspondendo a 80,59% (4.878) de todas as entradas referentes, e 106 instituições de emprego, correspondendo a 88,93% (5.207) de todas as entradas. O endereço declarado nos sites das instituições foi comparado com o endereço obtido pela API e contamos o número de casos incorretos.

Um total de 13 instituições acadêmicas foram geocodificadas incorretamente, correspondendo a 594 pesquisadores. Dez dessas instituições foram geocodificadas no estado e país corretos, mas na cidade errada. As três instituições restantes, correspondendo a 61 das 594 entradas, foram geolocalizadas em estados/países errados ou não foram encontradas. Do mesmo modo, um total de 22 instituições de emprego, correspondendo a 534 entradas, foram geolocalizadas incorretamente. Dessas, 6 instituições, correspondendo a 62 das 534 entradas, foram geocodificadas em estados/países errados ou não puderam ser geocodificadas. Dos erros encontrados, verificamos que a

geocodificação em cidades diferentes é causada normalmente por instituições com mais de um campus, onde a API não encontrou a cidade do campus sede (por exemplo, a USP). Os erros entre estados aconteceram em instituições com sedes em múltiplos estados, como o INPE. Erros no nível do país aconteceram com instituições diferentes com nomes similares, como o Observatório Nacional brasileiro e mexicano. Instituições que não foram geocodificadas foram aquelas com nomes incorretos ou genéricos, como “...” ou “Escola de Engenharia”.

Após a verificação, pudemos concluir que as instituições associadas com boa parte dos pesquisadores em nosso banco de dados estão corretamente geocodificadas. Além disso, como todos os erros encontrados na avaliação foram corrigidos. Uma taxa de erros tão baixa como a encontrada era esperada, pois o banco de dados é composto majoritariamente de universidades. Essas instituições geralmente ocupam grandes áreas em uma cidade, sendo assim importantes pontos de interesse para mapeamento urbano no Google Maps. Portanto, o processo de geocodificação simples que empregamos foi suficientemente correto para nossos experimentos.

Capítulo 4

Análise dos Dados

Nesta seção serão apresentadas as análises realizadas com os dados. Primeiro são descritas análises voltadas para entender a distribuição dos pesquisadores pelo ano e local de graduação, além da sua área de pesquisa. Em seguida, é analisada a evolução no tempo da educação dos pesquisadores, referenciando também sua área de pesquisa. Também é analisada a distribuição geográfica das instituições, no mundo e no Brasil, a fim de identificar trajetórias comuns. Finalmente, vamos analisar os tipos mais frequentes de trajetórias no conjunto de dados.

4.1 Distribuição da afiliação em instituições acadêmicas e empregadoras

Foi calculado o número de pesquisadores associados a instituições acadêmicas, considerando o nível de doutorado. A distribuição cumulativa dos pontos é exibida na Figura 4.1a. Existem 6.053 entradas únicas no gráfico (pesquisadores que estudaram em mais de uma instituição, por terem dois ou mais doutorados, são contados duas vezes). A distribuição apresenta um comportamento de cauda longa, que pode ser explicado por vários fatores. Em particular, acreditamos que esse resultado reflete dois fatos importantes: (1) existem inúmeras instituições acadêmicas fora do Brasil onde os títulos acadêmicos são obtidos, e (2) há várias instituições acadêmicas que raramente são associadas com pesquisa, como as faculdades regionais, refletindo a tendência nacional de que atividades de

pesquisa são conduzidas majoritariamente em universidades públicas¹⁰. Deve-se ressaltar que aproximadamente 50% dos pesquisadores estão concentrados em menos de 20 instituições, que podem ser consideradas como as principais condutoras de pesquisa no Brasil. Outro ponto é que o CV Lattes requer a indicação do nome da instituição, mas indicar o departamento ou instituto é opcional. Por exemplo, enquanto 1.206 pesquisadores indicaram que estudaram na “*Universidade de São Paulo*”, outros 113 indicaram subdivisões da USP (como institutos, campus ou unidades acadêmicas) como a instituição onde estudaram. Tais variações podem ter contribuído de certo modo para a distribuição observada na Figura 4.1a.

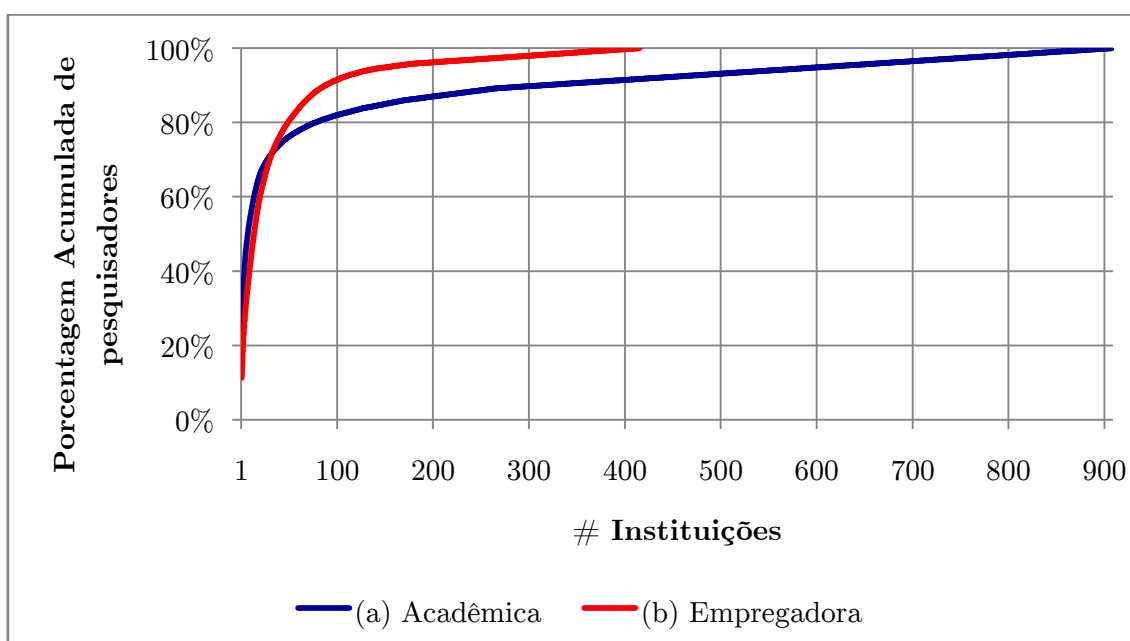


Figura 4.1. Distribuição cumulativa de pesquisadores para instituições (a) acadêmicas (nível doutorado) e (b) empregadoras.

A distribuição das instituições empregadoras segue um comportamento semelhante, como pode ser observado na Figura 4.1b: relativamente poucas instituições (34) empregam mais de 70% (4.199) dos pesquisadores. Novamente, as instituições que empregam a maior parte dos pesquisadores correspondem às

¹⁰http://agencia.fapesp.br/crise_economica_ainda_afeta_a_pesquisa_desenvolvimento_e_inovacao_em_escalas_global/20233/

melhores universidades brasileiras, assim como no caso anterior. A cauda é composta na sua maioria por instituições privadas.

4.2 Distribuição de pesquisador por área de pesquisa

A Tabela 4.1 mostra a distribuição dos pesquisadores de acordo com o tema de pesquisa do INCT ao qual são associados. Para cada área, são indicados o número de pesquisadores (86 pesquisadores estão associados a mais de um INCT), assim como o ano de graduação mais antigo e a mediana do ano em que esses pesquisadores receberam seu título de doutor. Saúde e Meio Ambiente são as áreas com a maior quantidade de pesquisadores: quase metade de todos os pesquisadores está associada a uma dessas duas áreas. A maior parte das áreas possui pelo menos um pesquisador que se formou entre o fim dos anos 50 e o início dos anos 60, com exceção de duas áreas: Agronegócios e Humanas/Sociais. Essas duas áreas possuem os pesquisadores seniores mais jovens e são mais jovens no conjunto. Por outro lado, Exatas, Saúde e Engenharia/TI possuem a maior parte dos pesquisadores mais antigos.

Tabela 4.1. Distribuição de pesquisadores por área de pesquisa

Área	Número de Pesquisadores	Ano de Formação mais Antigos (PhD)	Mediana do Ano de Formação (PhD)
Agronegócio	535	1972	2000
Energia	386	1960	1999
Engenharia/TI	691	1958	1999
Exatas	585	1958	1995
Humanas/Sociais	388	1969	2002
Meio Ambiente	897	1962	2000
Nanotecnologia	626	1961	1999
Saúde	1.908	1957	1999
Total	6.016	1957	1999

Inicialmente pode parecer estranho que áreas de pesquisa como Nanotecnologia possuam pesquisadores graduados há mais tempo que outras áreas mais tradicionais, como Agronegócios e Humanas/Sociais. Esse fato reflete a natureza multidisciplinar dos INCTs. Por exemplo, o pesquisador da área de

Nanotecnologia que se graduou em 1961 possui um doutorado em Química, e seu grupo também emprega biólogos, físicos e engenheiros¹¹.

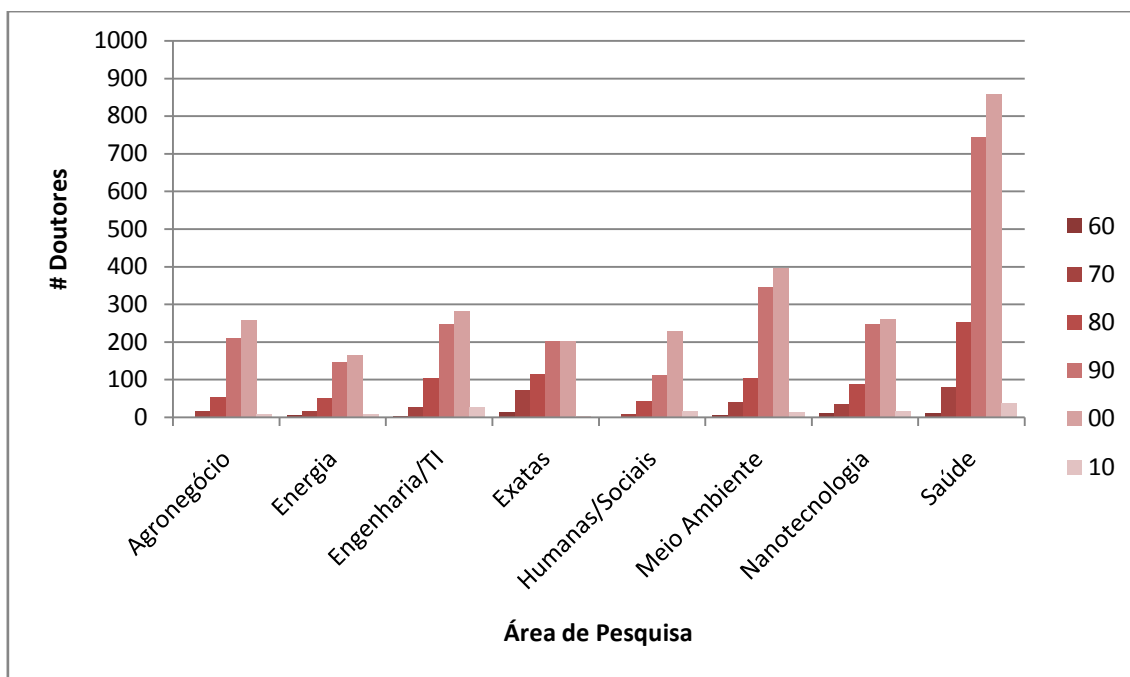


Figura 4.2. Número de doutores graduados por década

A Figura 4.2 mostra a distribuição da formação de PhDs ao longo das décadas para cada área. Agronegócio, Saúde e Engenharia/TI são algumas das áreas que mostram um grande crescimento no número de pesquisadores nas últimas décadas. Por outro lado, o crescimento é menor em áreas como Energia e Exatas.

O número relativamente maior de doutores na área de Saúde pode estar relacionado com o aumento na pesquisa relacionada a doenças típicas de países tropicais^{12,13}. Também existe um grande direcionamento estratégico dirigido à essa área, já que Saúde engloba o maior número de INCTs: 37 dos 101 institutos. Por outro lado, o número de pesquisadores cresce em todas as áreas, indicando o aumento das atividades científicas no Brasil nas duas últimas décadas.

¹¹ INCT-INAMI - <http://inct.cnpq.br/web/inct-inami/pesquisadores/>

¹² http://www.sbmt.org.br/site/corpo_texto/187

¹³ <http://www.fiocruz.br/rededengue/cgi/cgilua.exe/sys/start.htm?inoid=226&sid=3>

A respeito do pequeno número de pesquisadores graduados nas décadas de 1950 (não presente no gráfico; a década possui menos de 50 pesquisadores no total), 1960 e 1970 trata-se de um fenômeno esperado, por causa da aposentadoria. O conjunto de dados utilizado nesse estudo contém somente pesquisadores que ainda estão trabalhando. As leis brasileiras atuais permitem que homens se aposentem após 35 anos de contribuição e as mulheres, após 30 anos¹⁴. Um pesquisador que começou a trabalhar entre as décadas de 50 e 70 já teria acumulado tempo suficiente para se aposentar. Portanto, nossa análise reflete diretamente o *dataset* utilizado, composto de pesquisadores ativos, e não a situação do Brasil em geral.

4.2.1 Mestrados e pós-doutorados por área de pesquisa

Tendo analisado o doutorado, esta subseção foca nos outros estágios acadêmicos do pesquisador. Usando os dados do Lattes, analisamos como o número de pesquisadores com títulos de mestre e cargos de pós-doutorado varia conforme a área de pesquisa.

Tabela 4.2. Mestres e pesquisadores com pós-doutorado por área de pesquisa

Área	Nº		Nº total Postdoc	% (Total Postdoc)		
	Pesquisadores com Posdoc	% Postdoc/PhD		Nº MSc	% MSc/PhD	
Agronegócio	254	48,29%	348	66,16%	493	93,73%
Energia	146	38,52%	217	57,26%	342	90,24%
Engenharia/TI	317	47,60%	449	67,42%	643	96,55%
Exatas	436	74,53%	750	128,21%	508	86,84%
Humanas/Sociais	162	42,86%	213	56,35%	354	93,65%
Meio Ambiente	350	40,14%	476	54,59%	824	94,50%
Nanotecnologia	405	65,11%	604	97,11%	534	85,85%
Saúde	1004	53,38%	1.434	76,24%	1.525	81,07%

A Tabela 4.2 contrasta os grupos de pesquisa e o número de pesquisadores com pós-doutorado. A distribuição é desigual e em algumas áreas, como Exatas e Nanotecnologia, a fração de pesquisadores com PhD que também

¹⁴ <http://blog.previdencia.gov.br/?p=7270>

fizeram pós-doutorado (4^a coluna) é muito maior que em outras áreas. Além disso, se considerarmos todas as posições de pós-doutorado que os pesquisadores ocuparam, os pesquisadores de Exatas possuem mais períodos de pós-doutorado do que títulos de doutorado (6^a coluna).

A tabela também mostra a distribuição de títulos de mestrado (MSc, ou *Master of Science*) em cada área. A diferença entre as áreas não é tão expressiva como aquela observada para pós-doutorado. A tabela também mostra que a maior parte dos pesquisadores possui tanto um mestrado quanto um doutorado, embora a fração de pesquisadores com doutorado e *sem* mestrado seja um pouco menor nas áreas de Saúde, Nanotecnologia e Exatas.

4.2.2 Distribuição geográfica dos pesquisadores por área de pesquisa

Os pesquisadores participantes dos diversos INCTs obtiveram o título de doutorado em instituições de todo o mundo. A Figura 4.3 mostra os países de destino mais comuns para cada área, com exceção do Brasil (excluído para melhorar a leitura do gráfico). Em aproximadamente metade das áreas, a maior parte dos pesquisadores graduados em instituições dos Estados Unidos, enquanto os pesquisadores as áreas da Saúde e Engenharia/TI foram mais presentes nas instituições do Reino Unido e França.

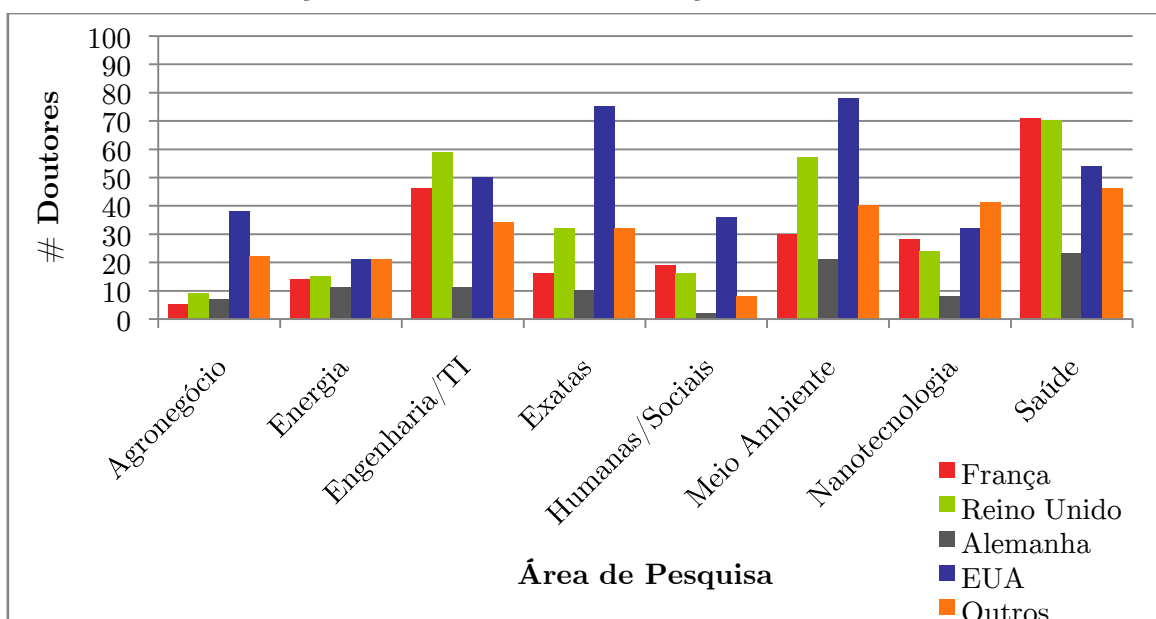


Figura 4.3. Distribuição geográfica das instituições formadoras de PhDs (exceto Brasil)

A Figura 4.4 mostra a proporção de doutorados obtidos em instituições brasileiras e estrangeiras. Embora a proporção do número de títulos obtidos fora varia de área para área, oscilando entre 15% e 35% do total, doutorados obtidos em instituições brasileiras são a maioria em todas as áreas. Essa distribuição mostra que a maior parte da educação dos pesquisadores brasileiros é feita dentro do país, mas que existe um número considerável de pesquisadores formados no exterior.

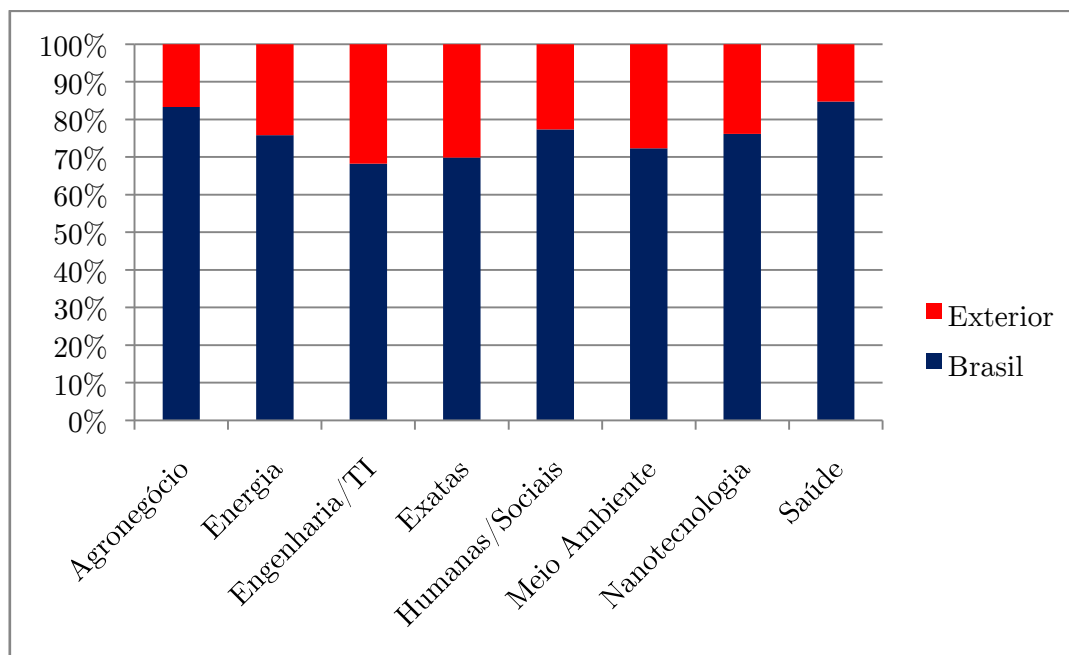


Figura 4.4. Proporção entre títulos brasileiros e internacionais

Analisamos mais detalhadamente os doutorados obtidos no Brasil ao olhar a distribuição das instituições ao redor do país. Foi considerado a divisão tradicional em regiões (Centro-Oeste, Norte, Nordeste, Sul, Sudeste), mas tratamos o estado de São Paulo separadamente da região Sudeste, como mostra a Figura 4.4. São Paulo sozinho é responsável pela formação de aproximadamente metade dos pesquisadores do nosso *dataset* que concluíram seu doutorado no Brasil, porém o estado concentra apenas cerca de 22% da população brasileira¹⁵. Por ter números tão expressivos, decidimos analisar o estado separadamente.

¹⁵A população de São Paulo era de 41 milhões de habitantes em 2010. A população brasileira total em 2010 é de aproximadamente 190 milhões.

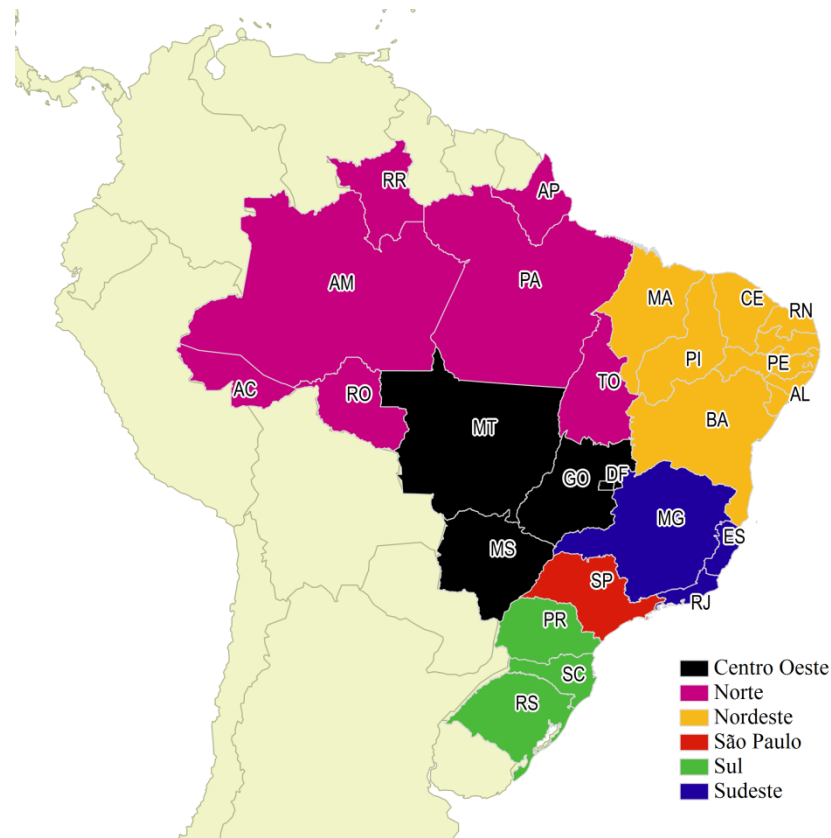


Figura 4.5. Regiões do Brasil

A Figura 4.6 mostra a distribuição do número de doutorados por área de pesquisa pelas regiões do Brasil. A maioria dos títulos obtidos na região Norte está relacionada às áreas de Energia e Meio Ambiente, região que contém a floresta Amazônica e recursos energéticos como a usina de Belo Monte. O Centro-Oeste se destaca com um número proporcionalmente alto de pesquisadores da área de Nanotecnologia. Também vale destacar que Agronegócio é a única área com mais pesquisadores formados em uma região diferente de São Paulo. As outras regiões seguem um padrão mais regular.

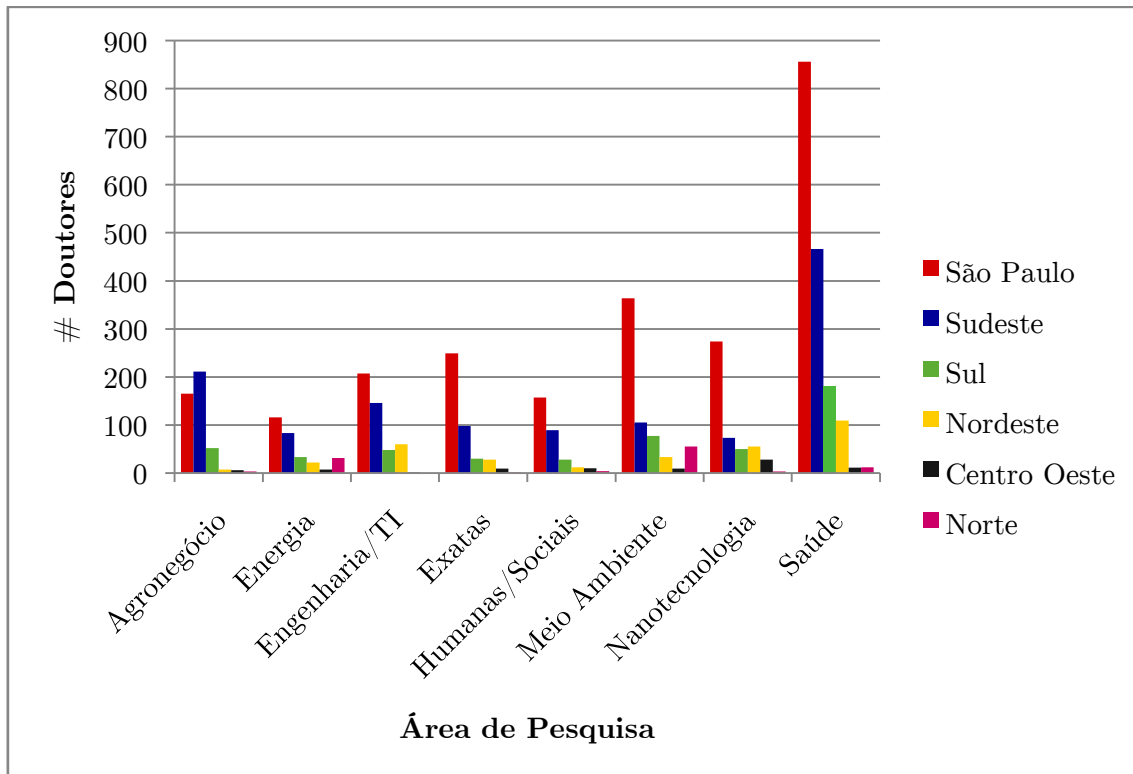


Figura 4.6. Títulos de PhD por região brasileira

As figuras 4.7 até 4.14 mostram geograficamente para onde os pesquisadores brasileiros foram para obter seu doutorado. As linhas mostram as trajetórias e a espessura mostra quantos pesquisadores possuem aquele segmento em sua trajetória. As figuras confirmam os destinos mais comuns mostrados nos gráficos anteriores, mas os mapas ajudam a entender melhor porque os pesquisadores de certas áreas escolhem destinos diferentes. Por exemplo, é possível ver que os pesquisadores relacionados a Agronegócio que estudam nos Estados Unidos escolheram instituições localizadas numa região conhecida pela sua produção agrícola, enquanto aqueles da Engenharia/TI escolheram instituições em regiões com desenvolvimento tecnológico grande, como a Califórnia.

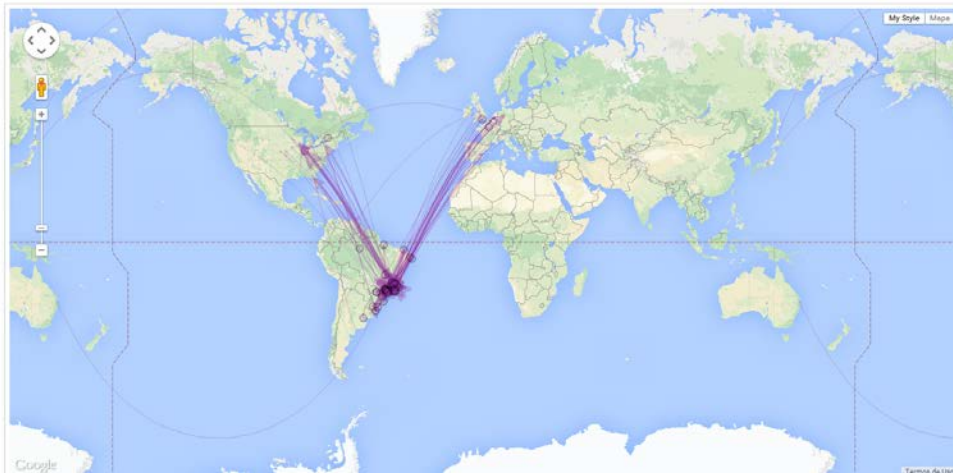


Figura 4.7. Visualização geográfica das trajetórias dos pesquisadores da área de Agronegócio



Figura 4.8. Visualização geográfica das trajetórias dos pesquisadores da área de Energia

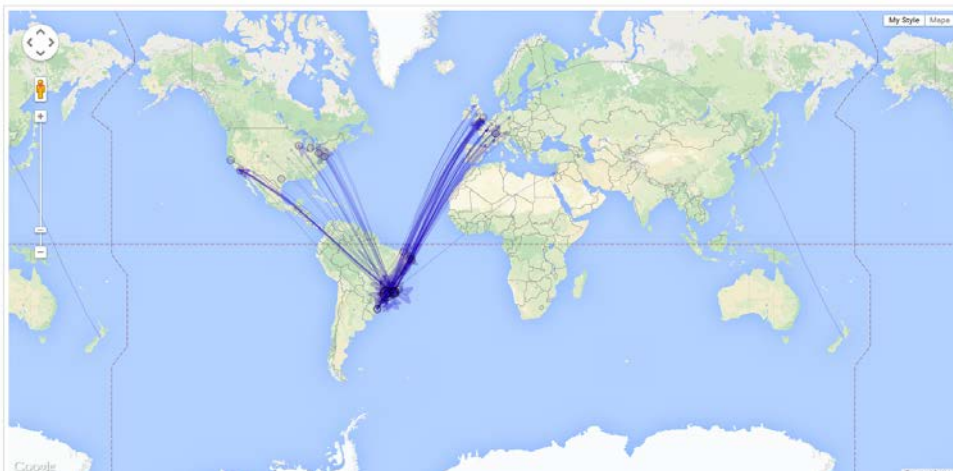


Figura 4.9. Visualização geográfica das trajetórias dos pesquisadores da área de Engenharia/TI

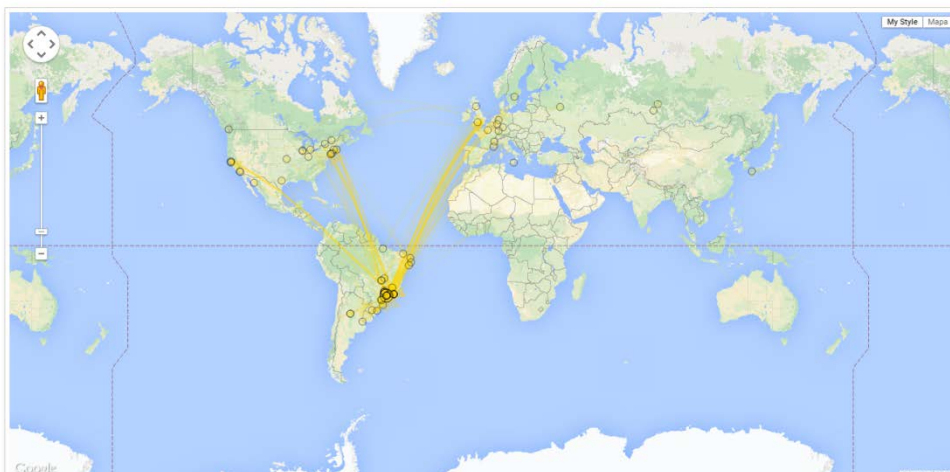


Figura 4.10. Visualização geográfica das trajetórias dos pesquisadores da área de Exatas

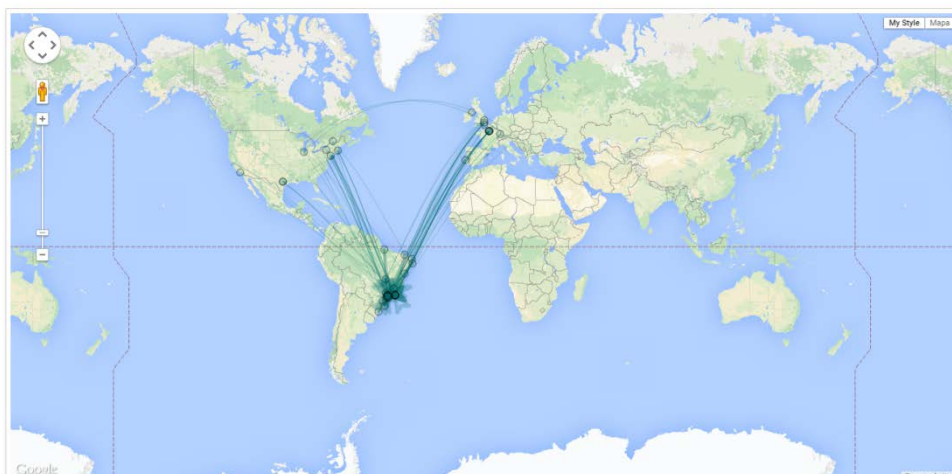


Figura 4.11. Visualização geográfica das trajetórias dos pesquisadores da área de Humanas/Sociais

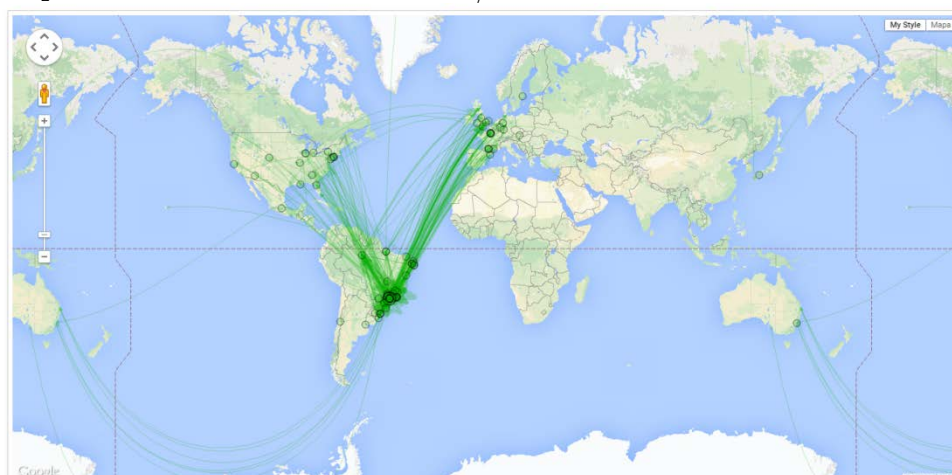


Figura 4.12. Visualização geográfica das trajetórias dos pesquisadores da área de Meio Ambiente

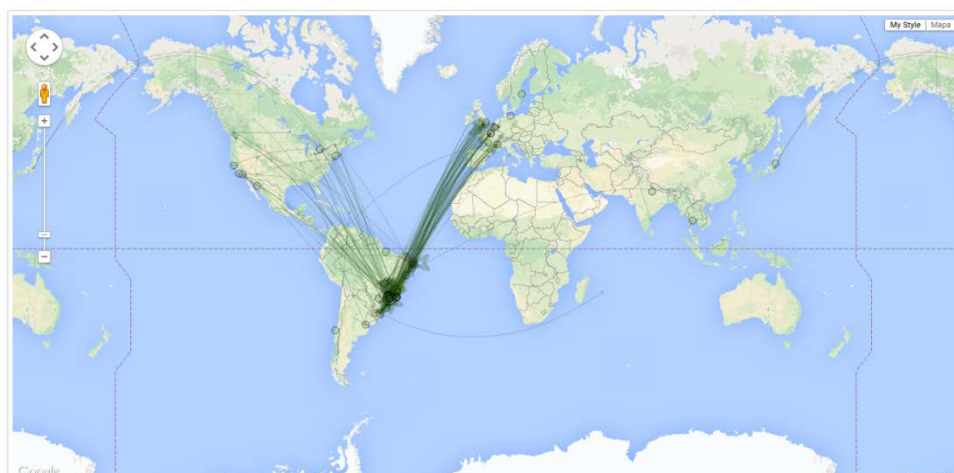


Figura 4.13. Visualização geográfica das trajetórias dos pesquisadores da área de Nanotecnologia

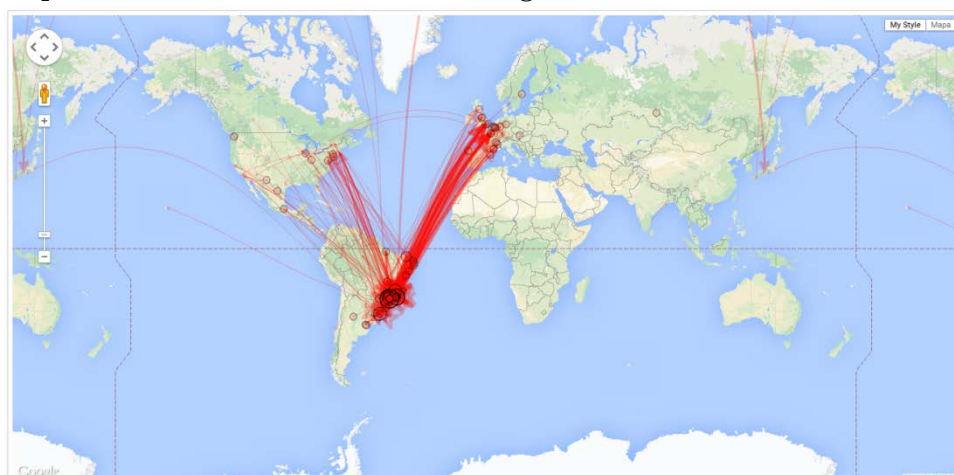


Figura 4.14. Visualização geográfica das trajetórias dos pesquisadores da área de Saúde

A seguir apresentamos possíveis explicações para os padrões observados na Figura 4.6. De acordo com o GeoCapes¹⁶ as regiões com o maior número de programas de pós-graduação são, respectivamente, o Sudeste, Sul, Nordeste, Centro-Oeste e Norte. A mesma ordem se repete na Figura 4.6, com São Paulo separado mas se destacando dos demais estados do Sudeste. De fato, o crescimento dos programas no Norte, Nordeste e Centro-Oeste é muito recente, de acordo com Avellar [2014], tendo obtido um crescimento expressivo de 50% entre 2000 e 2010. Isso pode explicar porque essas regiões ainda possuem uma

¹⁶ GeoCapes (<http://geocapes.capes.gov.br/geocapes2/>) é um sistema de informação georeferenciada mantido pela CAPES. O site oferece análises estatísticas por estado de certos dados sobre pós-graduação.

participação menor na formação da comunidade de pesquisa do Brasil. Uma exceção ao padrão geral é a área de Agronegócios, tendo São Paulo em segundo lugar. Isso pode ser explicado com a existência de um grande número de programas de pós-graduação no estado de Minas Gerais em comparação com São Paulo: um terço dos programas mineiros é dedicado a ciências agrícolas. Outra exceção é a área de Meio Ambiente, na qual a Região Norte possui um número expressivo de doutores e que deve estar relacionado à sua proximidade da floresta Amazônica.

A soberania de São Paulo na maioria das áreas pode ser explicada pelo fato de o estado abrigar algumas das melhores universidades brasileiras¹⁷, a maioria fundada pelo governo do estado de São Paulo. As melhores instituições brasileiras são, em sua maioria, fundadas pelo governo federal. Mesmo considerando o fato que cada estado possui pelo menos uma universidade federal, aquelas mais ativas em pesquisa estão concentradas nas regiões Sul e Sudeste, incluindo São Paulo¹⁸.

Em suma, o estado de São Paulo é de grande importância para a formação dos pesquisadores brasileiros, já que algumas das principais universidades do país estão localizadas nele. No entanto, como vamos discutir na próxima seção, a prevalência de São Paulo tem diminuído com o tempo. Outras regiões ainda estão bem atrás de São Paulo, mas algumas instituições nas regiões Norte e Nordeste são importantes tanto regionalmente quanto como centros de referências para assuntos específicos, como Meio Ambiente.

4.3 Distribuição espaço-temporal dos títulos de doutorado

Nesta seção é discutida a distribuição temporal de títulos de doutorado em diferentes regiões, considerando o ano de conclusão. A Figura 4.15 apresenta a distribuição dos títulos obtidos por década, considerando tanto títulos obtidos no Brasil quanto no exterior. A maioria dos pesquisadores se formou entre as

¹⁷ Três das instituições localizadas em São Paulo e financiadas pelo governo estadual aparecem no Top 10 do ranking QS de Universidades da América Latina de 2014 (<http://www.topuniversities.com/university-rankings/latin-american-university-rankings/2014>).

¹⁸ O mesmo ranking QS possui 17 instituições do Sul e Sudeste no Top 20 de Artigos por Corpo Docente, considerando somente instituições brasileiras.

décadas de 1990 e 2000. Perceba que a fração dos doutorados obtidos no exterior decresce de forma bastante expressiva a cada década. Essa porcentagem era de 63% na década de 1960, 48% na década de 1970, 39% na década de 1980, 27% na década de 1990, chegando a apenas 14% na década de 2000. Esses números mostram como a comunidade científica brasileira amadureceu com o tempo. A tendência observada está alinhada a programas e bolsas governamentais em vigor entre 1950 e 1980, que estimularam pesquisadores a estudarem no exterior. Em paralelo, nos anos 60, uma lei promovia a expansão do número de programas de pós-graduação a fim de empregar pesquisadores formados fora do país e assim atender às necessidades do país (Balbachevsky, 2011; Balbachevsky, 2013).

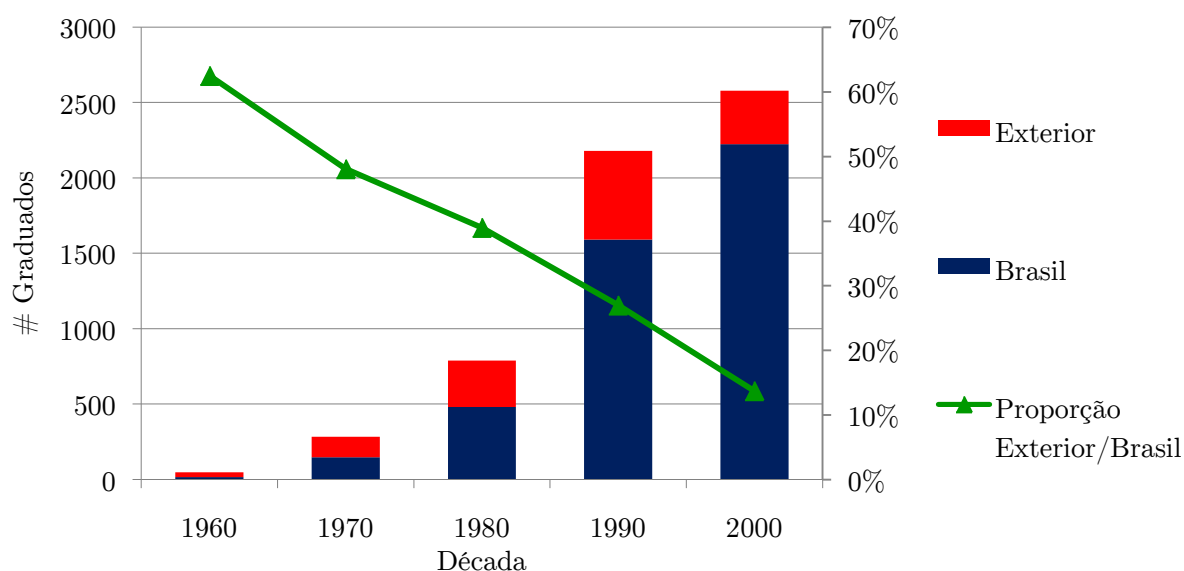


Figura 4.15. PhDs concedidos por década

Focando nas regiões do Brasil, a Figura 4.9 mostra que as instituições localizadas em São Paulo são responsáveis por uma grande porcentagem dos doutorados obtidos no Brasil nas últimas décadas. No entanto, a importância das outras regiões do Brasil vem aumentando recentemente. Considerando que nossos dados refletem somente os pesquisadores INCT, a figura mostra que a porcentagem de pesquisadores nos principais grupos de pesquisa brasileiros vindo de outras partes do Brasil vem aumentando ao longo dos anos. Considerando o crescimento absoluto do número de doutorados concedidos por

instituições brasileiras, a figura mostra na década de 1980 um crescimento na região Sudeste, e a partir de 1990 um crescimento nas instituições do Sul e Nordeste. Isso mostra a tendência de descentralização conforme as instituições distribuídas pelo país amadurecem. Os números das décadas de 1950 e 1960 são muito baixos, não chegando a 25 pesquisadores formados contando as duas décadas.

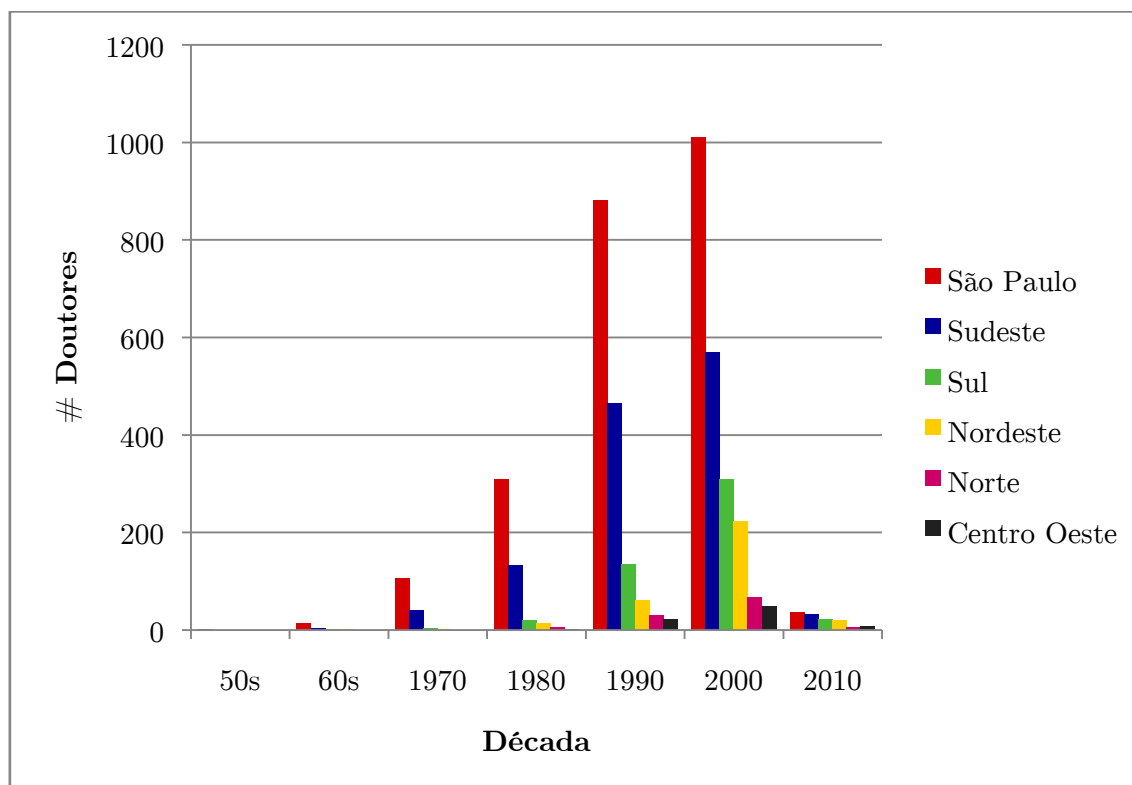


Figura 4.16. Títulos de PhD por década - Nacional

No total, existem 1.225 pesquisadores em nosso *conjunto de dados* que obtiveram seu doutorado no exterior. Como se pode perceber na Figura 4.10, a América do Norte (mais especificamente os Estados Unidos) possui uma participação grande, mas doutorados obtidos na Europa se tornaram predominantes a partir da década de 1980. É interessante notar o número pequeno de pesquisadores que estudaram em instituições da América Latina. Seria de se esperar que brasileiros fossem estudar em países cuja língua é próxima ao português ou que estão geograficamente próximos, mas esse não é o caso. Nem mesmo Portugal atrai muitos pesquisadores: o país formou apenas 6 dos 1.225 pesquisadores. Um fato similar também é observado em Portugal,

onde somente 4% dos projetos de pesquisa concedidos no país são atribuídos a pessoas originárias de países falantes de português (Heitor et al., 2014).

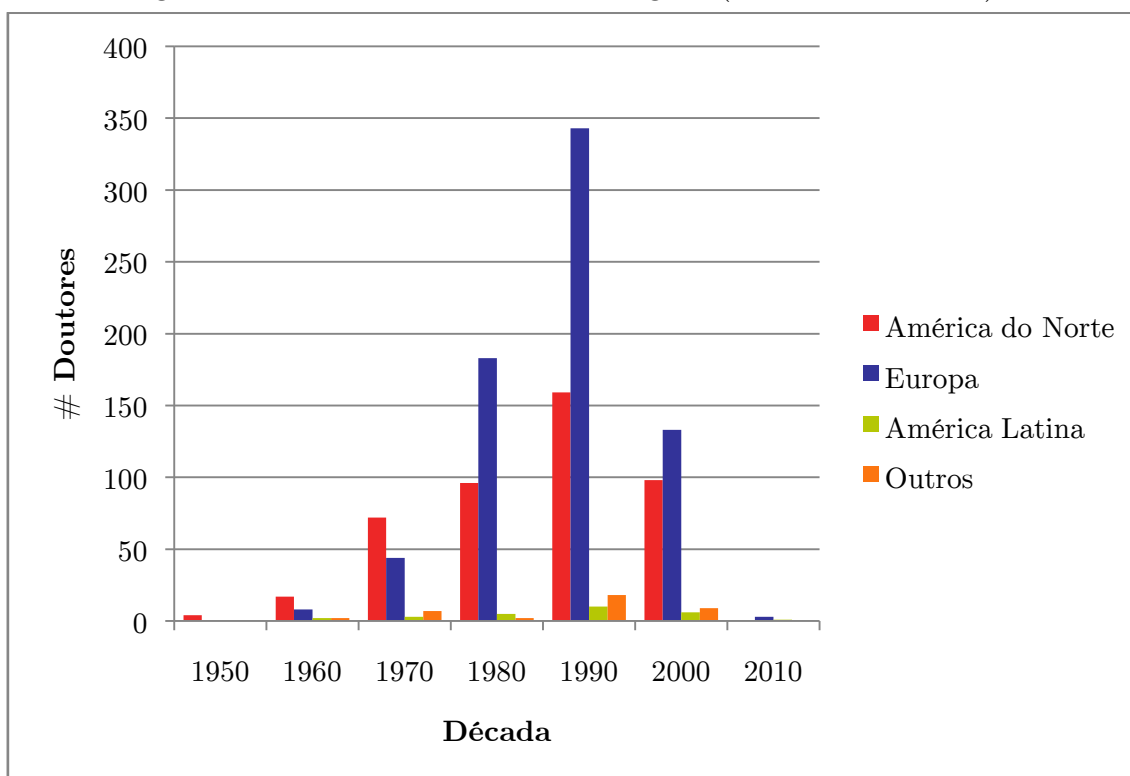
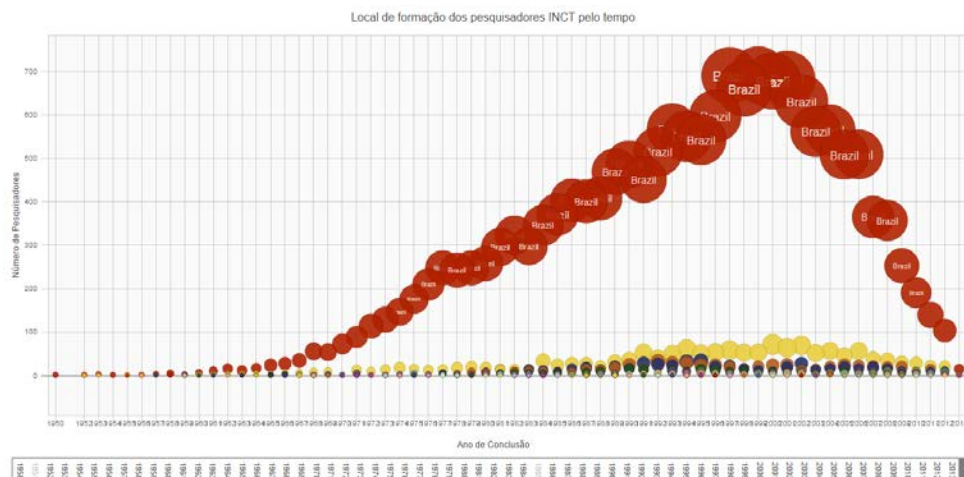


Figura 4.17. Títulos de PhD por década - Internacional

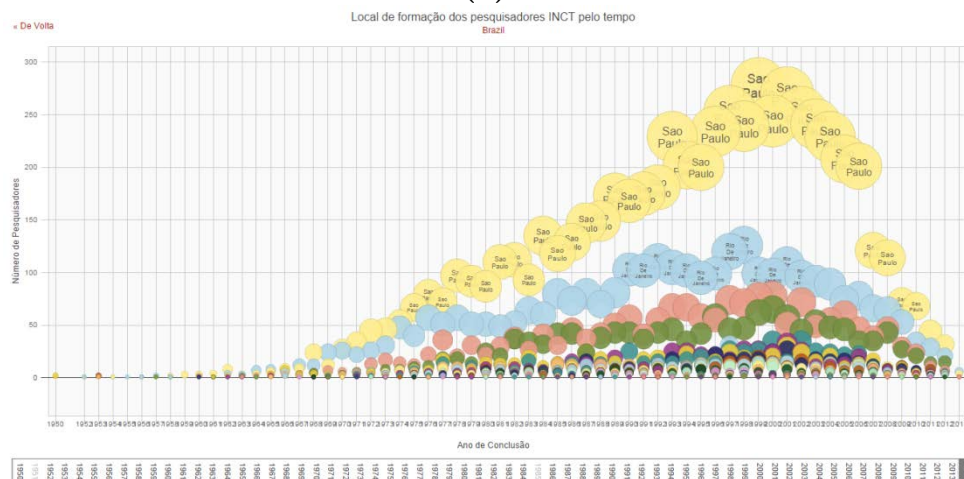
Um gráfico de distribuição interativo similar aos gráficos exibidos na Figura 4.9 e Figura 4.10, também foi desenvolvido como parte do trabalho e pode ser acessado pelo link¹⁹. Esse gráfico exibe os títulos de doutorado por ano e o usuário pode filtrar a informação por país, estado ou cidade. O que nota-se de diferente no gráfico interativo é que apesar de ocorrer uma descentralização com o tempo, a importância relativa de cada região praticamente não muda. Ou seja, ao longo de todos os anos São Paulo é o estado que forma mais pesquisadores, seguido do Rio de Janeiro, Minas Gerais e Rio Grande do Sul. O mesmo acontece ao analisar as cidades do estado de SP. A cidade de São Paulo é a que mais forma pesquisadores, com Campinas em segundo lugar. Acontece que o número de pesquisadores formados em cada lugar pode até mudar, mas o *ranking* dos locais que mais formam não muda, ou varia pouco. A Figura 4.11 mostra alguns exemplos de telas geradas pelo gráfico interativo.

¹⁹ <https://aqui.io/scatter>

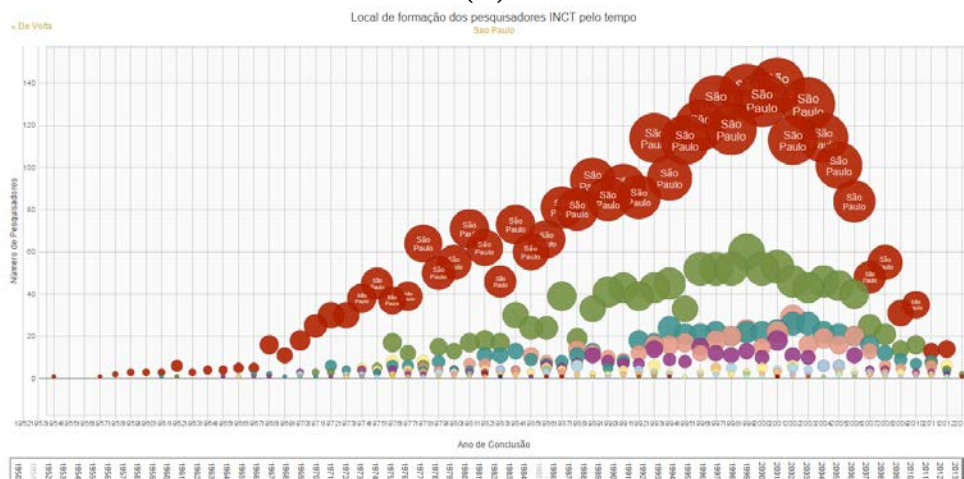
4.3 DISTRIBUIÇÃO ESPAÇO-TEMPORAL DOS TÍTULOS DE DOUTORADO ANÁLISES



(a)



(b)



(c)

Figura 4.18. Títulos PhD por ano e país (a), estado (b), cidade (c)

As figuras 4.19 até 4.24 mostram mapas que indicam o movimento dos pesquisadores a cada década. Se antes era possível ver facilmente o aumento do número de pesquisadores pelas décadas, agora passa a ser possível ver a

descentralização. Com o passar dos anos, o número e o local das instituições de destino aumenta. Embora essa variação possa ser atribuída simplesmente ao crescimento do número de pesquisadores, pode-se perceber que não é esse o caso. Comparando as décadas de 1970 e 2010, que possuem um número de pesquisadores similar, vemos como a década de 2010 possui uma variedade de instituições maior que na década de 1970.

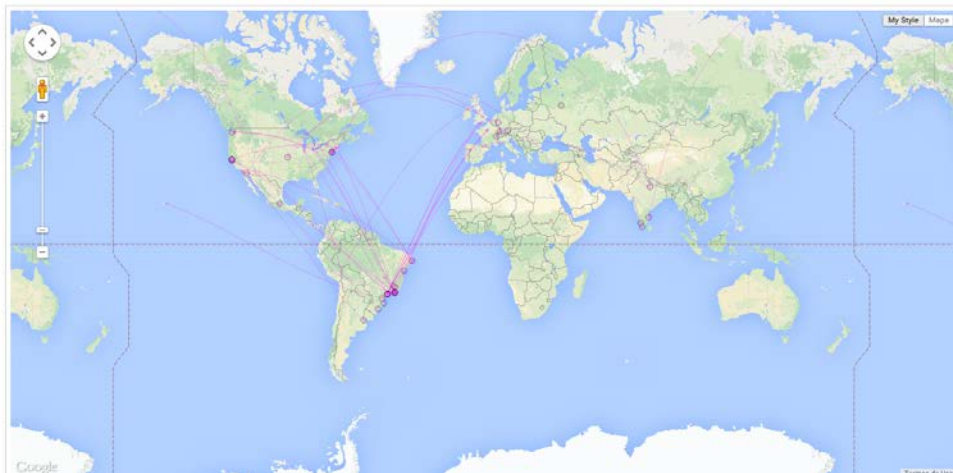


Figura 4.19. Segmentos de trajetória: década de 1960

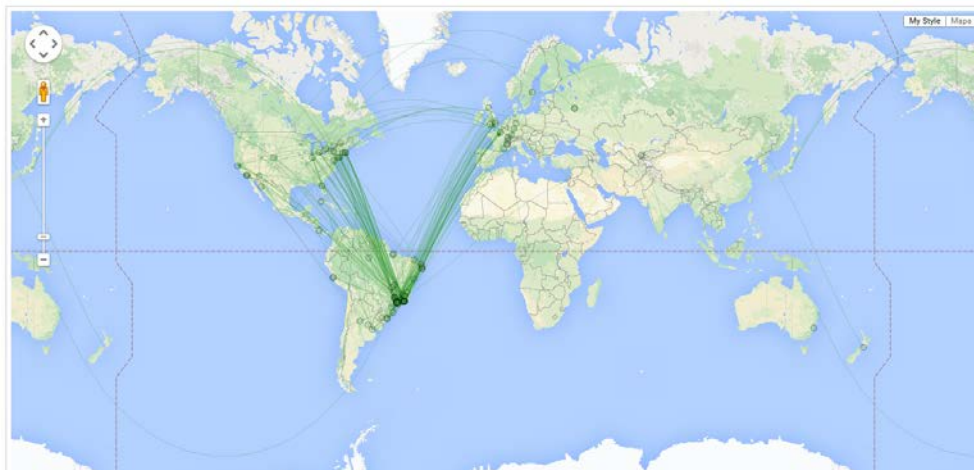


Figura 4.20. Segmentos de trajetória: década de 1970

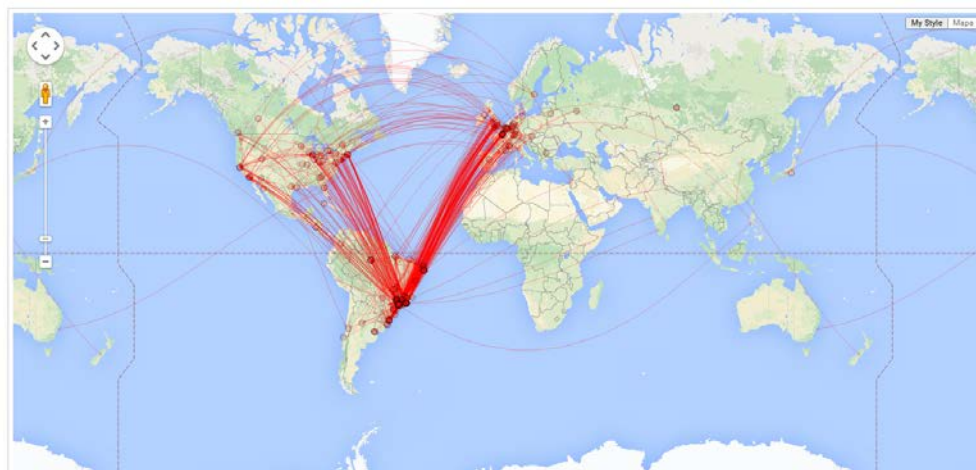


Figura 4.21. Segmentos de trajetória: década de 1980

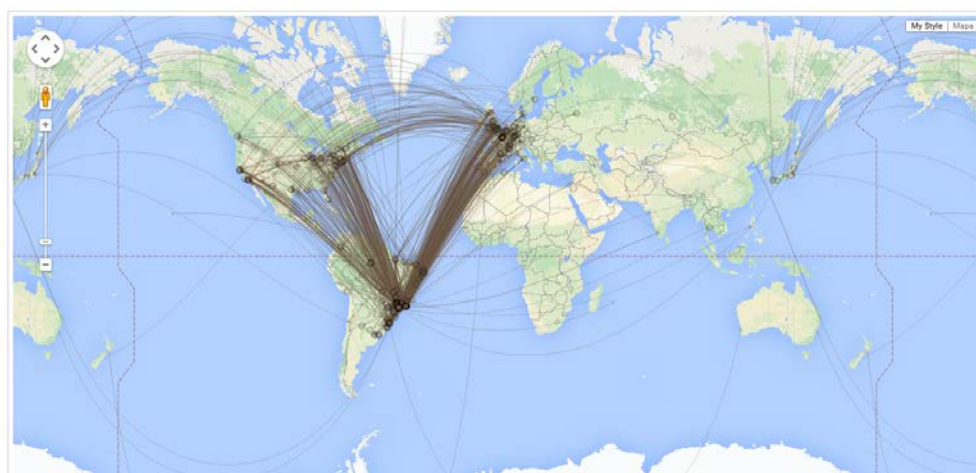


Figura 4.22. Segmentos de trajetória: década de 1990

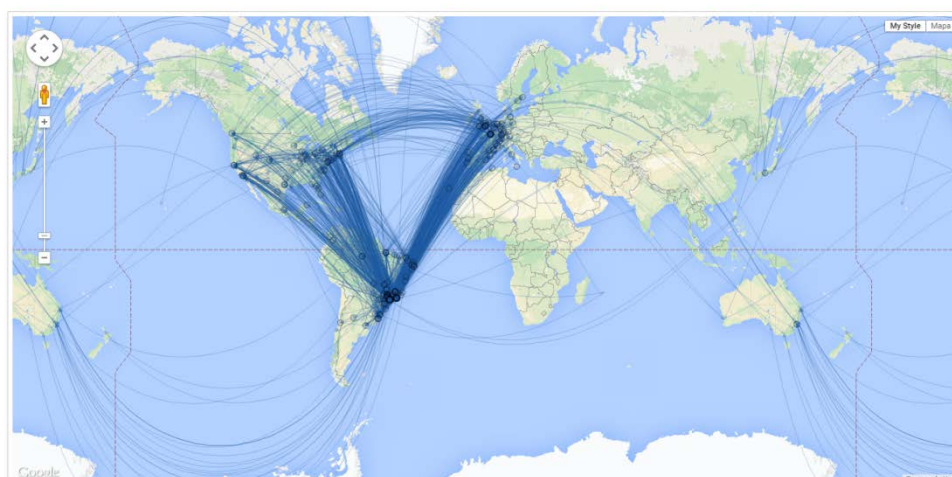


Figura 4.23. Segmentos de trajetória: década de 2000

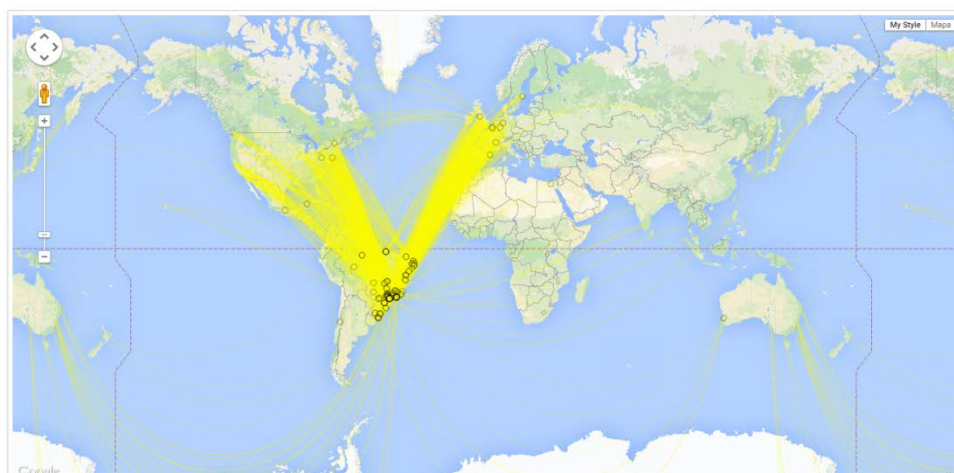


Figura 4.24. Segmentos de trajetória: década de 2010

Em suma, os dados do CiênciaBrasil mostram que a maioria dos pesquisadores membros de INCTs obtiveram seus doutorados no Brasil, e que a porcentagem de doutorados obtidos fora parece diminuir com o tempo. A Europa recebe a maior parte desses pesquisadores, mas uma grande parte também vai para os EUA.

4.4 Pontos de origem e destino das trajetórias

Nessa seção comparamos os pontos de origem e destino de certas etapas da trajetória acadêmica do pesquisador. Mais especificamente, iremos comparar o local de sua instituição de graduação, sua instituição de doutorado e seu local de trabalho. Consideramos o local da instituição de graduação como o local de origem do pesquisador. Os dados utilizados não possuem o local de nascimento do pesquisador, mas como a maioria dos alunos brasileiros não migra de estado ao escolher uma faculdade²⁰ nós consideramos essa aproximação confiável o bastante para as análises.

A Tabela 4.3 mostra a distribuição dos pesquisadores nas regiões do Brasil e exterior em três dos estágios da trajetória: graduação, doutorado e emprego. Consideramos somente a primeira entrada em caso de múltiplas entradas do mesmo estágio. A distribuição das instituições de graduação e de emprego é bem similar, enquanto o estado de São Paulo e as regiões da Europa

²⁰ <http://g1.globo.com/educacao/noticia/2013/05/13-dos-calouros-no-sisu-migram-de-estado-em-2013.html>

e América do Norte possuem uma concentração maior de instituições de doutoramento. Isso confirma a análise discutida na seção anterior: ao longo das décadas, esses foram os destinos mais procurados para estudos avançados. No entanto, a distribuição similar dos pontos iniciais e finais da trajetória parece indicar a tendência de pesquisadores voltarem à sua região de origem após realizarem seus estudos fora.

Tabela 4.3. Distribuição dos pesquisadores por região e estágio da trajetória (azul indica aumento e vermelho indica redução em relação à Graduação)

	Graduação		PhD		Emprego	
	Absoluto	Relativo	Absoluto	Relativo	Absoluto	Relativo
Brasil – Centro-Oeste	226	3,88%	78	1,33%	381	6,51%
Brasil – Nordeste	840	14,41%	310	5,28%	885	15,12%
Brasil – Norte	206	3,53%	106	1,81%	346	5,91%
Brasil – Sudeste	1.601	27,47%	1,24	21,14%	1.562	26,68%
Brasil – São Paulo	1.557	26,71%	2.335	39,81%	1.671	28,54%
Brasil – Sul	922	15,82%	483	8,23%	783	13,37%
Ásia e Oceania	18	0,31%	37	0,63%	2	0,03%
África	2	0,03%	2	0,03%	1	0,02%
Europa	109	1,87%	703	11,98%	17	0,29%
América Latina	138	2,37%	27	0,46%	1	0,02%
América do Norte	58	1,00%	429	7,31%	20	0,34%
Não Localizados	152	2,61%	116	1,98%	186	3,18%
Total	5.829	100%	5.866	100%	5.855	100%

Deve-se ressaltar que a concentração de estudos avançados em instituições paulistas e em outras regiões não é influência da concentração populacional. A Figura 4.25 compara a distribuição pelas regiões brasileiras do número de pessoas no Brasil com doutorado, de acordo com dados do CNPq²¹, e o número de pesquisadores de INCTs, do nosso conjunto de dados. Foi associada a região onde a instituição de trabalho está localizada, já que essa é a informação mais atual sobre a localização do pesquisador. Para fins de comparação, o gráfico inclui a porcentagem da população brasileira região, como informado pelo CNPq.

²¹ <http://www.memoria.cnpq.br/estatisticas/investimentos/regiao.htm>, ver Tabela 1.5.8

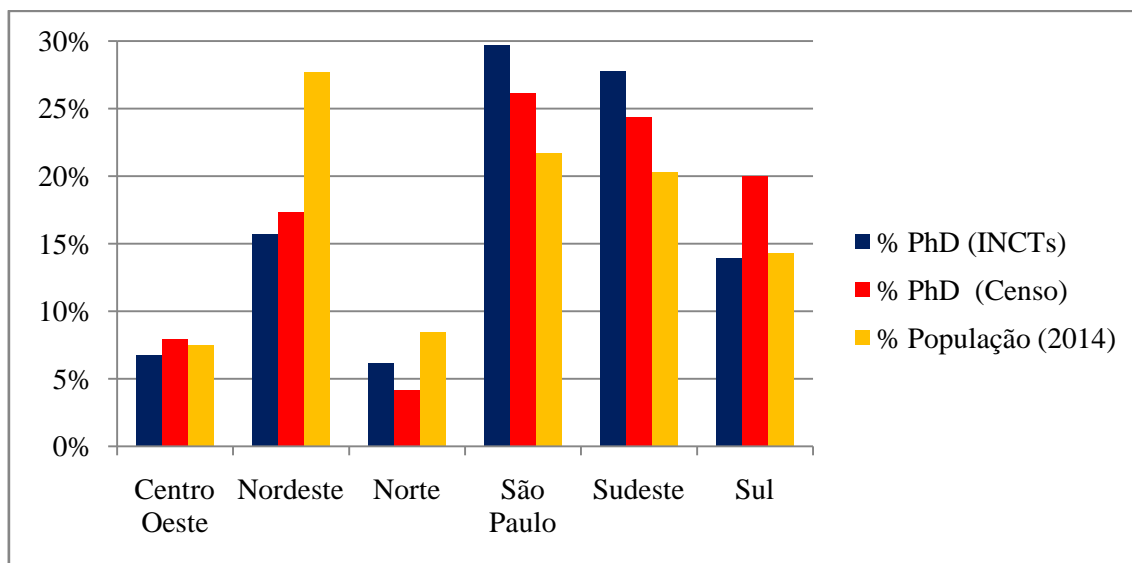


Figura 4.25. Distribuição de doutores no Brasil e nos INCTs

A distribuição é no geral similar, mas observa-se que as instituições de São Paulo e Sudeste juntas correspondem a mais da metade dos doutorados do país e a mais de 55% dos pesquisadores envolvidos com INCTs, porém correspondem a 42% da população total. Já que os INCTs representam os grupos de pesquisa mais avançados no Brasil, a grande concentração de pesquisadores nessas regiões (em relação à população) pode indicar instituições e grupos de pesquisa mais maduros em comparação àqueles localizados em outras regiões do país. O mesmo se aplica para o Sul, que é responsável por 14% dos pesquisadores INCT e a uma porcentagem similar da população total do país.

Por outro lado, a figura mostra que o programa INCT tem sido bem sucedido em envolver cientistas de todas as regiões, em uma proporção que é similar à capacidade delas (medida pelo número de doutores residentes). Os números absolutos, no entanto, indicam que ainda há um longo caminho a ser percorrido: enquanto o Sudeste, a região mais avançada do Brasil, possui 0,6 doutores por 1.000 habitantes, a proporção na Europa (considerando 27 países) e nos EUA está entre 6 a 8 doutores por 1.000 habitantes (IDEA Consult [2010]).

Tabela 4.4. Análise de Origem e Destino das trajetórias – da graduação para o emprego. (AA: Ásia; AF: África; BMI: Brasil Centro-Oeste; BNE: Brasil Nordeste; BSE: Brasil Sudeste (exceto São Paulo); BSP: Brasil São Paulo; BSU: Brasil Sul; EU: Europa; LA: América Latina; NA: América do Norte)

		Destino - Emprego											
		AA	AF	BMI	BNE	BNO	BSE	BSP	BSU	EU	LA	NA	Total
Origem - Graduação	AA	0	0	2	2	1	2	8	1	0	0	0	16
	AF	0	0	0	1	0	1	0	0	0	0	0	2
	BMI	0	0	113	11	14	27	41	9	0	0	1	216
	BNE	0	0	31	610	28	47	70	16	0	0	1	803
	BNO	0	0	10	9	152	8	11	4	1	1	0	196
	BSE	1	0	86	51	31	1.122	175	53	8	0	6	1.533
	BSP	0	0	68	82	35	131	1.076	70	0	0	8	1.470
	BSU	1	0	40	56	41	63	108	566	0	0	2	877
	EU	0	1	6	10	6	32	30	12	5	0	0	102
	LA	0	0	9	17	8	31	36	24	1	0	0	126
	NA	0	0	1	4	6	14	19	6	0	0	1	51
	Total	2	1	366	853	322	1.478	1.574	761	15	1	19	5.391

Foram analisados com mais detalhes os pontos de origem e destino na trajetória do pesquisador através da Tabela 4.4, Tabela 4.5 e Tabela 4.6, ignorando os casos onde os dados de origem e destino do pesquisador estão incompletos. A distribuição da Tabela 4.4 é calculada usando a posição da instituição de graduação como local de origem e a instituição de emprego como local de destino. A concentração na diagonal principal indica a tendência de não se mudar para outras partes do país em busca de emprego. A tendência das pessoas é preferir viver na mesma região em que completaram sua graduação. Podemos ver que instituições brasileiras empregam pessoas formadas em diversas regiões diferentes. No entanto, a maioria deles se formou em alguma instituição localizada na mesma região. De forma similar, a maioria dos pesquisadores que se estudaram em uma região tende a se manter nela. Uma exceção interessante é o Centro-Oeste: enquanto quase 70% dos pesquisadores trabalhando no Centro-Oeste não receberam seu título de graduação na região, aproximadamente 50% dos que lá se formaram trabalham própria na região.

Tabela 4.5. Análise de Origem e Destino das trajetórias – da graduação para o doutorado. (AA: Ásia; AF: África; BMI: Brasil Centro-Oeste; BNE: Brasil Nordeste; BSE: Brasil Sudeste (exceto São Paulo); BSP: Brasil São Paulo; BSU: Brasil Sul; EU: Europa; LA: América Latina; NA: América do Norte)

		Destino - PhD											
		AA	AF	BMI	BNE	BNO	BSE	BSP	BSU	EU	LA	NA	Total
Origem - Graduação	AA	5	0	1	0	1	2	2	0	2	0	3	16
	AF	0	0	0	0	0	1	0	0	0	0	0	1
	BMI	1	0	35	2	3	26	89	6	34	0	21	217
	BNE	4	0	6	254	6	96	259	16	114	1	42	798
	BNO	1	0	7	10	55	32	57	9	16	0	7	194
	BSE	7	0	11	11	12	871	292	22	179	2	146	1.553
	BSP	5	1	6	8	13	45	1.213	8	119	0	94	1.512
	BSU	11	1	6	7	8	53	241	390	131	1	46	895
	EU	0	0	0	0	0	14	19	3	58	1	6	101
	LA	0	0	1	3	2	30	47	7	7	22	12	131
	NA	1	0	0	1	1	6	6	3	4	0	34	56
	Total	1	2	73	296	101	1.176	2.225	464	664	27	411	5.440

A Tabela 4.5 compara a localização das instituições de graduação e de doutorado. Mais uma vez os números estão concentrados em sua maioria na diagonal principal, um padrão semelhante ao do emprego, ou seja, cursar o doutorado na mesma região em que o pesquisador fez sua graduação. Uma exceção notável é o número elevado de doutorados obtidos em São Paulo por pessoas que estudaram no Centro-Oeste, Nordeste e Sul, que era esperado considerando os resultados discutidos em seções anteriores.

Finalmente, a Tabela 4.6 cobre a transição da instituição de doutorado para a instituição onde o pesquisador trabalha. Naturalmente, considerando que o *dataset* contém pesquisadores de INCTs, a maioria dos pesquisadores trabalha no Brasil, normalmente em São Paulo, Sudeste e Sul. Perceba também que pesquisadores que obtiveram o doutorado em instituições paulistas estão espalhados por várias regiões do país, mas se concentram principalmente no próprio estado.

Tabela 4.6. Análise de Origem e Destino das trajetórias – do doutorado para o emprego. (AA: Ásia; AF: África; BMI: Brasil Centro-Oeste; BNE: Brasil Nordeste; BSE: Brasil Sudeste (exceto São Paulo); BSP: Brasil São Paulo; BSU: Brasil Sul; EU: Europa; LA: América Latina; NA: América do Norte)

		Destino - Emprego											
		AA	AF	BMI	BNE	BNO	BSE	BSP	BSU	EU	LA	NA	Total
Origem - PhD	AA	0	0	1	9	1	4	12	5	0	0	1	33
	AF	0	0	0	1	0	0	1	0	0	0	0	2
	BMI	0	0	44	5	10	6	2	3	0	0	1	71
	BNE	0	0	2	253	15	10	10	4	0	0	1	295
	BNO	0	0	5	4	87	3	2	1	0	0	0	102
	BSE	1	0	70	94	41	866	65	45	6	0	4	1.192
	BSP	0	0	151	273	99	253	1.230	194	2	1	7	2.210
	BSU	1	0	22	29	20	24	23	340	0	0	0	459
	EU	0	1	51	128	32	186	147	125	7	0	2	679
	LA	0	0	1	5	1	3	7	9	0	0	0	26
	NA	0	0	27	47	20	160	115	37	0	0	2	408
	Total	2	1	374	848	326	1.515	1.614	763	15	1	18	5.476

A Figura 4.26 contém mais uma representação do fluxo de migração dos pesquisadores, desta vez analisando a mudança do estado onde fez seu doutorado para o estado onde trabalha. O grafo mostra agora o fluxo de entrada e saída para cada estado. A cor do nó indica o seu grau: quanto mais vermelho, maior o número de arestas que entram e saem dele. As arestas representam o número de pesquisadores que migraram de um estado para outro. O sentido da aresta indica sua direção: caso a aresta esteja no sentido horário em relação ao nó ela representa uma aresta de saída; caso a aresta esteja no sentido anti-horário, ela representa uma aresta de entrada. Um círculo em cima do nodo representa as arestas que se ligam ao próprio nodo.

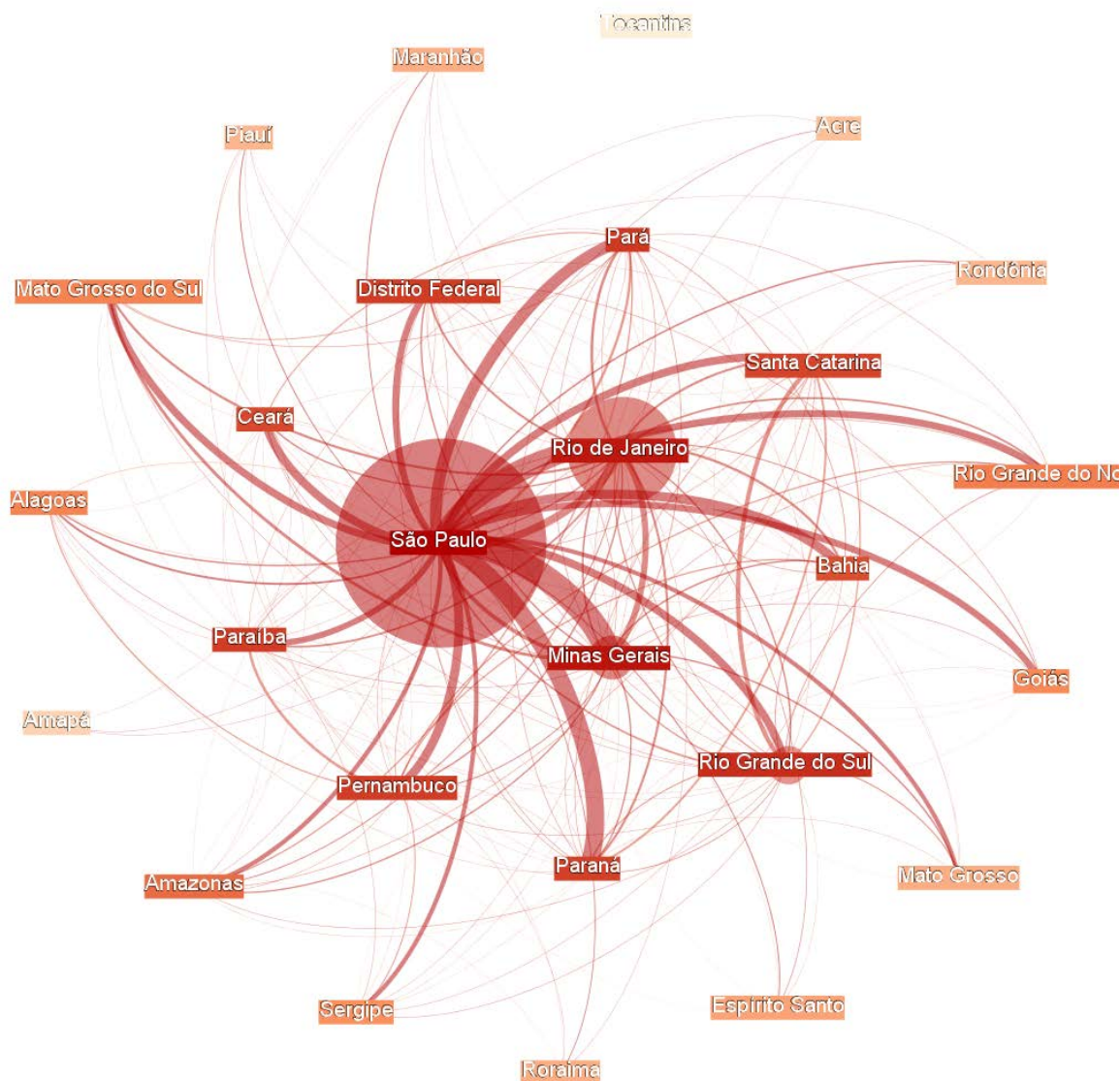


Figura 4.26. Grafo de migração dos pesquisadores, do doutorado para o emprego

O grafo mostra melhor como é a participação de cada estado nos números da Tabela 4.7, indicando de forma clara quais os estados que mais recebem e exportam pesquisadores para outros estados. Observa-se que os nós centrais são aqueles que possuem maior grau, e que o grau diminui dos nós centrais para os periféricos. Da mesma forma, os nós centrais são aqueles que possuem mais arestas saindo do que entrando, e a proporção vai se invertendo conforme os nós se afastam do centro, com os nós periféricos possuindo mais arestas de entrada do que de saída. Vê-se que mesmo dentro de uma região há estados mais atraentes que outros, que funcionam como centros regionais. O Distrito Federal

e o Pará são exemplos de estados assim. Também vale notar que o Espírito Santo, mesmo sendo parte da região Sudeste, possui comportamento parecido com estados do Norte e Nordeste, não possuindo participação expressiva nem na formação de pesquisadores nem como centro de pesquisa.

Concluindo, ao analisar as tabelas acima é possível notar que pesquisadores INCT possuem uma resistência em se mudar para outras partes do país. Os pesquisadores estudados são na maioria brasileiros que estudaram perto de casa e atualmente trabalha na mesma região. A proporção de pesquisadores que cursaram estudos avançados em outras regiões ou no exterior é de 32%. No entanto, o padrão difere de região para região. São Paulo e Sudeste são as únicas regiões onde a maioria dos pesquisadores é natural da própria região. O Centro-Oeste, Norte, Nordeste e em menor extensão o Sul mostram um padrão de migração temporário: uma porcentagem expressiva de pesquisadores emigra dessas regiões da graduação para cursar o doutorado, e da etapa de PhD para o emprego há um fluxo de pesquisadores de volta para a região. Essa observação contrasta fortemente com estudos de mobilidade a respeito de pesquisadores europeus (Van Bowel [2010]), que mostra que metade dos estudantes de doutorado que obtiveram doutorado em Economia nos Estados Unidos arranjaram um emprego naquele país. Da outra metade, um terço permaneceu no seu país de origem, enquanto os outros dois terços procuraram por empregos em outros países europeus. Também notamos que a participação de brasileiros trabalhando no exterior é pequena, e é uma evidência da necessidade de encorajar a diáspora científica brasileira a ter um papel mais importante na ciência brasileira (Balbachevsky & do Couto e Silva [2011]).

4.5 Internacionalização

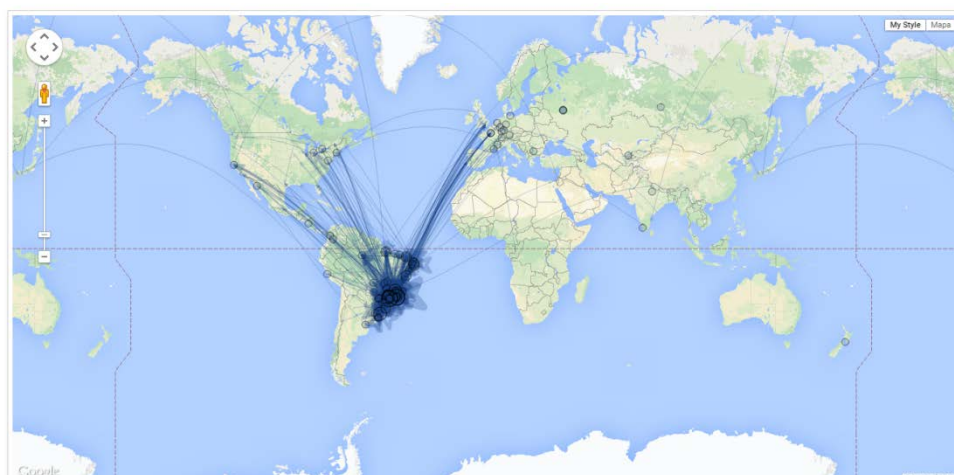
O conjunto de pontos nas trajetórias dos pesquisadores dos INCTs, embora concentrados em sua maioria no Brasil, inclui um número considerável em instituições estrangeiras. No total, 2.727 pesquisadores completaram pelo um estágio em instituições do exterior. A Tabela 4.8 mostra a porcentagem de estágios feitos no exterior por modalidade.

Tabela 4.7. Estágios obtidos no exterior por modalidade

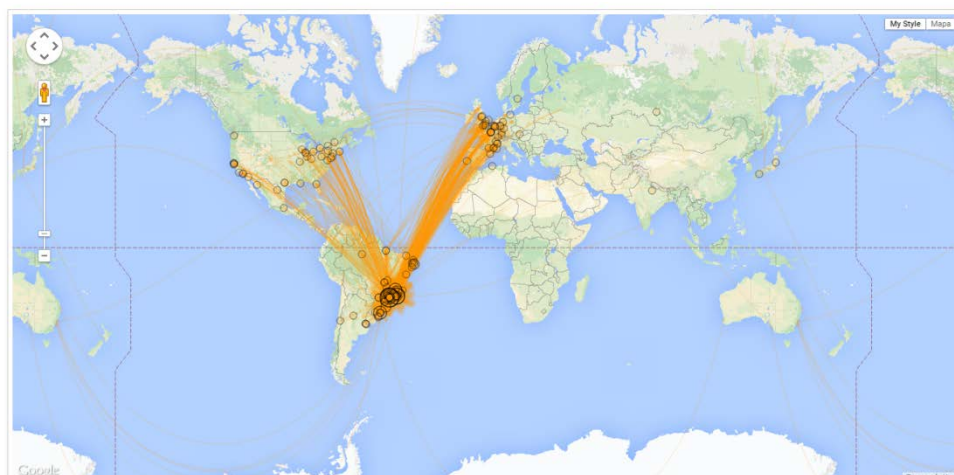
	% estágios obtidos no exterior
Graduação	5,78%
Mestrado	6,94%
Doutorado	20,58%
Pós-doutorado	62,46%

A porcentagem de títulos internacionais aumenta em estágios mais altos, chegando a quase 21% para o PhD. Se compararmos todos os estágios, descobrimos que o estágio mais comum fora do Brasil são os pós-doutorados. Em geral, vemos que quanto mais avançado o pesquisador está em sua trajetória, maior a chance dele fazer esse estágio em uma instituição fora do Brasil

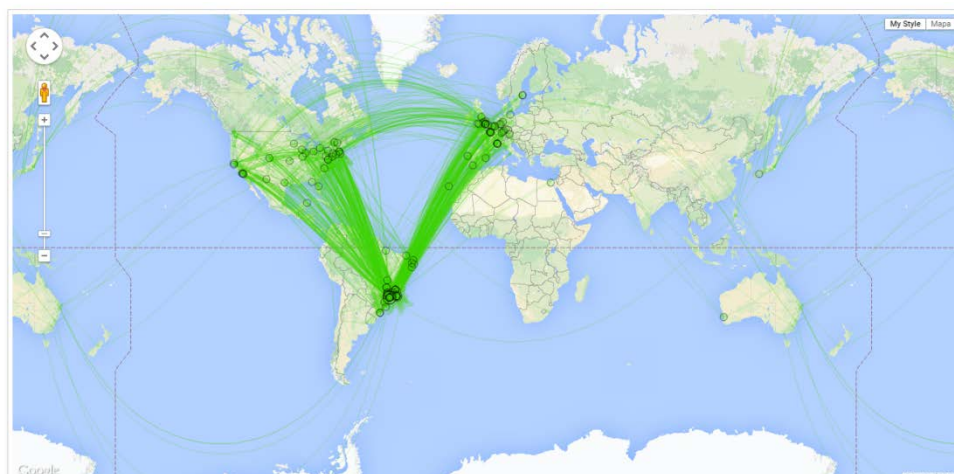
Por exemplo, mais de 60% dos pesquisadores dos INCTs que registram estágios de pós-doutorado o fizeram fora do Brasil. Além disso, um total de 1.262 pesquisadores (41% de todos os pesquisadores com pós-doutorado e 21% do total) possui o pós-doutorado como o único estágio feito fora do país. De fato, só 20% dos títulos de doutorado foram obtidos em instituições internacionais. Uma possível explicação para a diferença da porcentagem de PhDs e pós-doutorados em instituições do exterior pode ser que pesquisadores brasileiros que fizeram a maior parte ou toda sua educação no Brasil buscam posições de pós-doutorado como um meio de expandir sua experiência internacional e aumentar sua cooperação com grupos de pesquisa internacionais. Uma análise mais detalhada da intensidade de cooperação com cientistas internacionais após o pós-doutorado confirmaria essa hipótese. A lista de trabalhos publicados do pesquisador, assim como os autores dos artigos, pode ser obtida em seu currículo Lattes. No entanto, é de se esperar que cientistas que não são brasileiros não tenham seu currículo no sistema. Assim, deixamos essa análise para um trabalho futuro.



(a)



(b)



(c)

Figura 4.27. Trajetória dos pesquisadores no mestrado (a), doutorado (b) e pós-doutorado (c)

A Figura 4.27 exhibe como funciona a internacionalização no mestrado, doutorado e pós-doutorado. De forma geral, percebe-se que no mestrado existe uma grande concentração de trajetórias passando por instituições de São Paulo e Sudeste. A região serve como ponto de atração para pesquisadores das outras partes do Brasil. Também vemos que, daqueles que vão fazer seu mestrado fora do país, grande parte possui como origem instituições de São Paulo e Sudeste. No doutorado, São Paulo e Sudeste continuam a ser um ponto de atração, enquanto as instituições de outras partes do Brasil perdem importância e instituições estrangeiras ganham. O mesmo comportamento evolui no pós-doutorado. Percebe-se que a tendência dos pesquisadores que possuem estágios em instituições internacionais é estudar antes em uma instituição do Sudeste. Esse comportamento pode estar relacionado à busca contínua de instituições de referência na sua área de pesquisa.

Recentemente percebe-se uma demanda crescente para que posições de pós-doutorado no Reino Unido e EUA sejam ocupadas por profissionais estrangeiros (Cantwell [2011], Cantwell & Taylor [2013]). Esse fenômeno pode explicar (pelo menos parcialmente) a porcentagem alta de pesquisadores com posições de pós-doutorado fora do Brasil, mas ele não é o único fator. Ao contrário, observa-se um crescimento muito maior de pós-doutorados em instituições brasileiras nos últimos anos, se comparado com estágios realizados no exterior. Por exemplo, 1.016 pesquisadores em nossos dados fizeram pós-doutorado no exterior antes do ano 2000. O número de pesquisadores que ocuparam esses cargos a partir do ano 2000 é ligeiramente maior, 1.345. Por outro lado, enquanto só 216 pesquisadores fizeram pós-doutorado no Brasil antes de 2000, o número daqueles com pós-doutorado em instituições brasileiras a partir de 2000 aumentou em quase 5 vezes, chegando a 1.036. A internacionalização de pesquisadores na etapa de pós-doutorado parece ser um fenômeno recorrente, refletindo antigas práticas governamentais de incentivo de educação no exterior (Balbachevsky [2011]), aliado a uma demanda por profissionais estrangeiros em instituições fora do Brasil.

4.6 Padrões de trajetória

Nesta seção discutimos os padrões de trajetórias, considerando todas as etapas da formação assim como etapas de pós-doutorado e emprego dos pesquisadores dos INCT²². Nos casos onde o pesquisador possui mais de uma entrada no CV para uma mesma etapa (como possuir dois doutorados), nós consideramos o primeiro obtido. Esses casos são incomuns, mas observamos uma média de 1,48 pós-doutorados e 1,12 graduações por pesquisador (Tabela 4.9). Quase todo pesquisador possui pelo menos uma graduação e um doutorado válidos em seu currículo (mais de 97% nos dois casos). Também observamos uma grande proporção de pesquisadores com mestrados, e pouco mais de metade dos pesquisadores registram um pós-doutorado.

Tabela 4.8. Número de pesquisadores e entradas por estágio

	Nº Pesquisadores com pelo menos uma Entrada	% Pesquisadores com pelo menos uma Entrada	Nº Total Entradas	Entradas por Pesquisador
Graduação	5.811	97,29%	6.537	1,12
Mestrado	5.172	86,59%	5.405	1,05
Doutorado	5.866	98,21%	5.945	1,01
Pós-Doutorado	2.994	50,13%	4.439	1,48

Para cada pesquisador, trajetória foi definida como uma sequência de estágios acadêmicos e de emprego ao longo do tempo, tendo cada etapa relacionada a uma região geográfica, como definido na Seção 4.3. Foi feita uma análise de padrões frequentes em dois conjuntos de sequências: uma consistindo da série de etapas educacionais (da graduação ao doutorado) terminando com um possível pós-doutorado; e a outra consistindo da sequência de todas as etapas, incluindo o emprego. No primeiro caso, queremos encontrar os padrões geográficos frequentes na educação dos pesquisadores. No segundo caso,

²² Só serão incluídos títulos que já foram concluídos. Por exemplo, enquanto todos os pesquisadores do *dataset* registram pelo menos o início de um doutorado, nem todos eles concluíram de fato o curso. Portanto, só consideramos os estágios que possuem data de conclusão.

queremos ver como a região da instituição de emprego é influenciada pelas etapas anteriores. O algoritmo de análise de padrões frequentes simplesmente conta o número de ocorrências do padrão, e seleciona aqueles que se repetem em pelo menos $\theta\%$ dos casos. Em outras palavras, esses padrões possuem um *suporte mínimo* de θ . Todas as análises dessa seção consideram somente combinações com $\theta \geq 1\%$.

A Tabela 4.9 mostra os padrões mais comuns em relação às etapas acadêmicas, ou seja, a sequência de títulos assim como descrita no CV do pesquisador. A tabela corresponde a 84,9% de todos os pesquisadores INCT. Os dois padrões mais comuns são *Graduação → Mestrado → Doutorado* e *Graduação → Mestrado → Doutorado → Pós-Doutorado*. A mudança direta da graduação para o doutorado é muito menos comum, indicando que a educação dos cientistas brasileiros usualmente inclui um mestrado, o que é o padrão usual das instituições brasileiras.

Tabela 4.9. Padrões de modalidade acadêmica mais comuns. G: Graduação; M: Mestrado; D: Doutorado; P: Pós Doutorado.

Padrão	# Pesquisadores	% Suporte
GMD	2.090	34,99%
GMDP	1.454	24,34%
GMDPP	437	7,32%
GDP	257	4,30%
GD	248	4,15%
GGMD	187	3,13%
GMDPPP	133	2,23%
GDPP	101	1,69%
GM	84	1,41%
GGMDP	83	1,39%
Total	5.074	84,95%

A Tabela 4.10 mostra os padrões mais comuns em relação à sequência de regiões onde o pesquisador estudou em cada estágio de sua formação acadêmica. Os padrões foram obtidos usando a sequência *Graduação → Mestrado → Doutorado → Pós-Doutorado*, seguindo os padrões mais comuns encontrados na Tabela 4.10, usando *placeholders* no caso do pesquisador não possuir

informações daquele estágio. Em caso de múltiplas entradas para o mesmo estágio, foi considerado somente o primeiro estágio completo.

Tabela 4.10. Padrões de regiões mais comuns. (BNE: Brasil Nordeste; BSE: Brasil Sudeste (exceto São Paulo); BSP: São Paulo; BSU: Brasil Sul; EU: Europa; NA: América do Norte; ---: pesquisador não possui informação sobre a etapa)

Graduação	Padrão			# Pesquisadores	% Suporte
	Mestrado	Doutorado	Pós-Doutorado		
BSP	BSP	BSP	---	403	6,75%
BSE	BSE	BSE	---	370	6,19%
BSP	BSP	BSP	BSP	215	3,60%
BSU	BSU	BSU	---	178	2,98%
BNE	BNE	BNE	---	147	2,46%
BSP	BSP	BSP	NA	132	2,21%
BSE	BSE	BSE	BSE	126	2,11%
BSP	BSP	BSP	EU	122	2,04%
BSE	BSE	BSE	NA	95	1,59%
BSE	BSE	BSE	EU	76	1,27%
BSP	---	BSP	---	74	1,24%
Total				1.938	32,44%

Os padrões exibidos na Tabela 4.10 correspondem a 32,4% de todos os pesquisadores do *dataset*. O padrão mais comum consiste em completar todos os estágios no estado de São Paulo. Mais que isso, os 5 padrões mais comuns da Tabela 4.11 mostram trajetórias onde todos os estágios foram obtidos na mesma região, indicando uma baixa mobilidade dos pesquisadores brasileiros durante sua educação. Para padrões heterogêneos, ou seja, aqueles que possuem pelo menos um estágio fora de sua região natal, a situação mais comum é estudar em São Paulo e Sudeste, mas com uma experiência pós-doutoral no exterior. Finalmente, os padrões mais frequentes que exibem movimentação intranacional (mas que não são frequentes o bastante para aparecerem na tabela) consistem de pesquisadores que saíram de sua região original para fazer doutorado em São Paulo.

Quando o estágio correspondente à instituição empregadora é adicionado na tabela (Tabela 4.12), a tendência encontrada anteriormente é reforçada. Observa-se que praticamente os mesmos padrões exibidos na Tabela 4.11 reaparecem nessa análise. Em todos os casos da tabela, mesmo naqueles com

estágios no exterior, o pesquisador trabalha na mesma região onde concluiu sua graduação. O padrão mais frequente que mostra o pesquisador trabalhando em uma região diferente é o caso de pesquisadores que fizeram sua graduação, mestrado e doutorado no Sudeste e atualmente estão empregados no Centro-Oeste, que acontece em somente 0,50% dos casos.

Tabela 4.11. Padrões de regiões mais comuns – incluindo emprego. (BNE: Brasil Nordeste; BSE: Brasil Sudeste (exceto São Paulo); BSP: São Paulo; BSU: Brasil Sul; EU: Europa; NA: América do Norte; ---: pesquisador não possui informação sobre a etapa)

Graduação	Padrão				# Pesquisadores	% Suporte
	Mestrado	Doutorado	Pós- Doutorado	Emprego		
BSE	BSE	BSE	---	BSE	285	4,77%
BSP	BSP	BSP	---	BSP	276	4,62%
BSP	BSP	BSP	BSP	BSP	154	2,58%
BSU	BSU	BSU	---	BSU	148	2,48%
BNE	BNE	BNE	---	BNE	126	2,11%
BSE	BSE	BSE	BSE	BSE	108	1,81%
BSP	BSP	BSP	NA	BSP	107	1,79%
BSP	BSP	BSP	EU	BSP	98	1,64%
BSE	BSE	BSE	NA	BSE	77	1,29%
BSE	BSE	BSE	EU	BSE	69	1,16%
Total					1.448	24,25%

A Tabela 4.12 mostra uma análise de padrão similar, considerando uma divisão de regiões com granularidade mais grossa para distinguir melhor a diferença da mobilidade intra e internacional. Foi usado “A” para representar a região do Brasil onde o pesquisador concluiu sua graduação; “B” para representar as regiões brasileiras diferentes da região ‘A’ em cada caso; e “C” para indicar regiões fora do Brasil. Os padrões da tabela correspondem a 65,47% dos pesquisadores do *dataset*.

Tabela 4.12. Padrões genéricos de movimentação mais comuns – incluindo emprego. (A: instituição brasileira localizada na mesma região onde o pesquisador fez sua graduação; B: instituição brasileira localizada em uma região diferente de onde o pesquisador fez sua graduação; C: instituição está localizada fora do Brasil; Null: processo de geocodificação não identificou a região da instituição)

Padrão					# Pesquisadores	% Suporte
Graduação	Mestrado	Doutorado	Pós- Doutorado	Emprego		
A	A	A	---	A	876	14.67%
A	A	A	C	A	458	7.67%
A	A	A	A	A	317	5.31%
A	B	B	---	B	269	4.50%
A	A	C	---	A	202	3.38%
A	A	A	---	B	202	3.38%
A	A	C	C	A	180	3.01%
A	A	B	---	A	165	2.76%
A	B	B	---	A	164	2.75%
A	B	B	C	B	150	2.51%
A	A	B	---	B	112	1.88%
A	A	---	---	A	109	1.82%
A	---	A	---	A	107	1.79%
A	---	A	C	A	107	1.79%
A	B	B	B	B	101	1.69%
A	A	A	NULL	A	83	1.39%
A	A	A	C	B	82	1.37%
A	A	A	---	---	79	1.32%
A	A	A	A	B	78	1.31%
A	B	B	C	A	70	1.17%
Total					3.911	65,47%

Mais uma vez, os resultados exibidos na Tabela 4.12 confirmam as observações de que pesquisadores brasileiros tendem a ficar na mesma região na maior parte da sua vida acadêmica. Na maioria dos casos, pesquisadores que não trabalham na sua região original tendem a se mudar ao iniciar o mestrado, logo que concluem a graduação. Além disso, não encontramos influência clara na mobilidade dos pesquisadores que possuem algum estágio no exterior. Por exemplo, de acordo com a tabela, dos 1.249 pesquisadores que possuem algum estágio no exterior, 1.017 deles (81%) voltaram para a sua região de origem.

Também foi feita uma análise da distribuição da distância geográfica entre dois pontos da trajetória do pesquisador. No primeiro caso mostramos a distância entre a instituição onde o pesquisador fez seu doutorado e a instituição onde fez sua etapa anterior, normalmente a o mestrado ou a graduação.

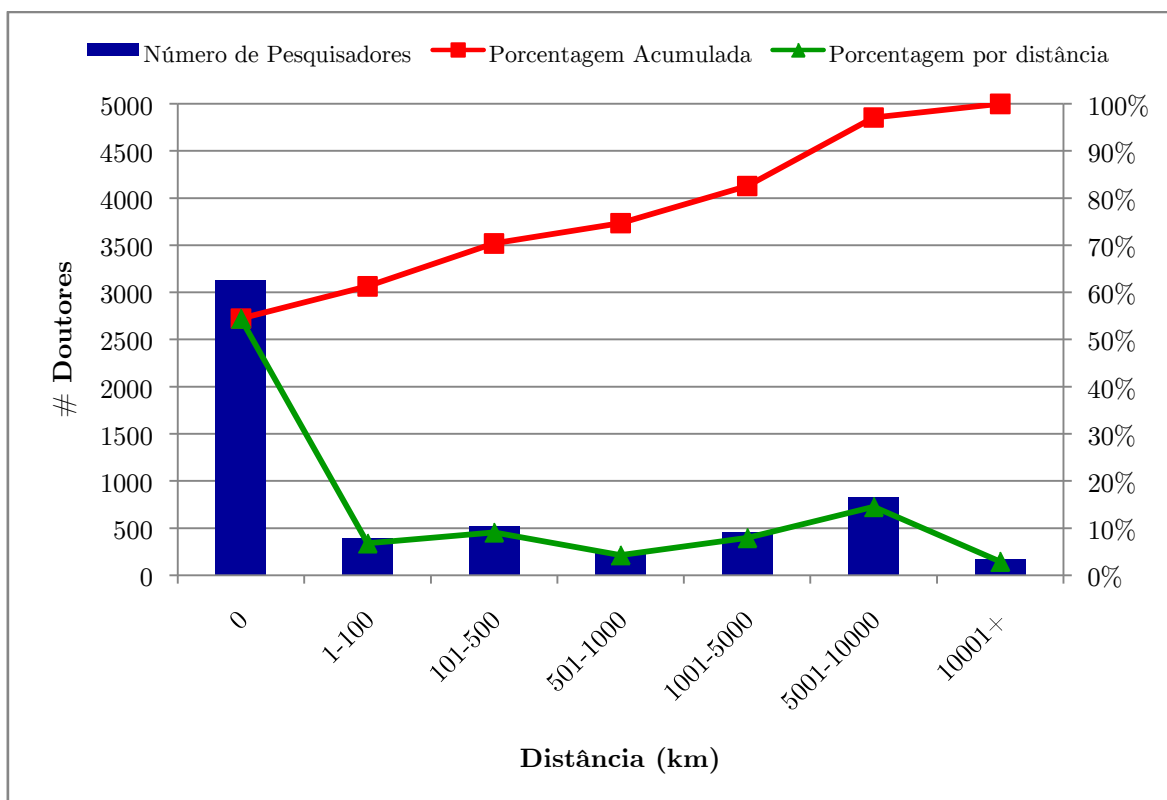


Figura 4.28. Distância percorrida pelos pesquisadores até a instituição de doutorado

A Figura 4.28 mostra a distribuição dos pesquisadores por certas faixas de valor correspondendo à distância entre os dois pontos analisados. Podemos dizer de forma aproximada que os pesquisadores que se moveram 1-500 km são aqueles que foram estudar em cidades próximas; os que moveram 501-1000 km são os que se moveram para estados próximos; entre 1001 e 5000 são os que se mudaram para países próximos ou regiões distantes no mesmo país; e aqueles que se moveram mais de 5001 km são tipicamente os pesquisadores que foram estudar em outros países.

O que vemos na Figura 4.15 é uma tendência dos pesquisadores obterem um PhD na mesma instituição onde obtiveram seu título anterior, seja esse uma

graduação, um mestrado ou mesmo outro doutorado. Analisando as faixas de valor, vemos que os pesquisadores que não se mudaram são a grande maioria, e os outros casos possuem distribuição similar. Aproximadamente 55% dos pesquisadores continuaram os estudos na mesma instituição. No total, pouco mais de 70% dos pesquisadores estudam em uma instituição a menos de 1000 km do ponto inicial de sua trajetória, confirmando mais uma vez a baixa mobilidade dos pesquisadores.

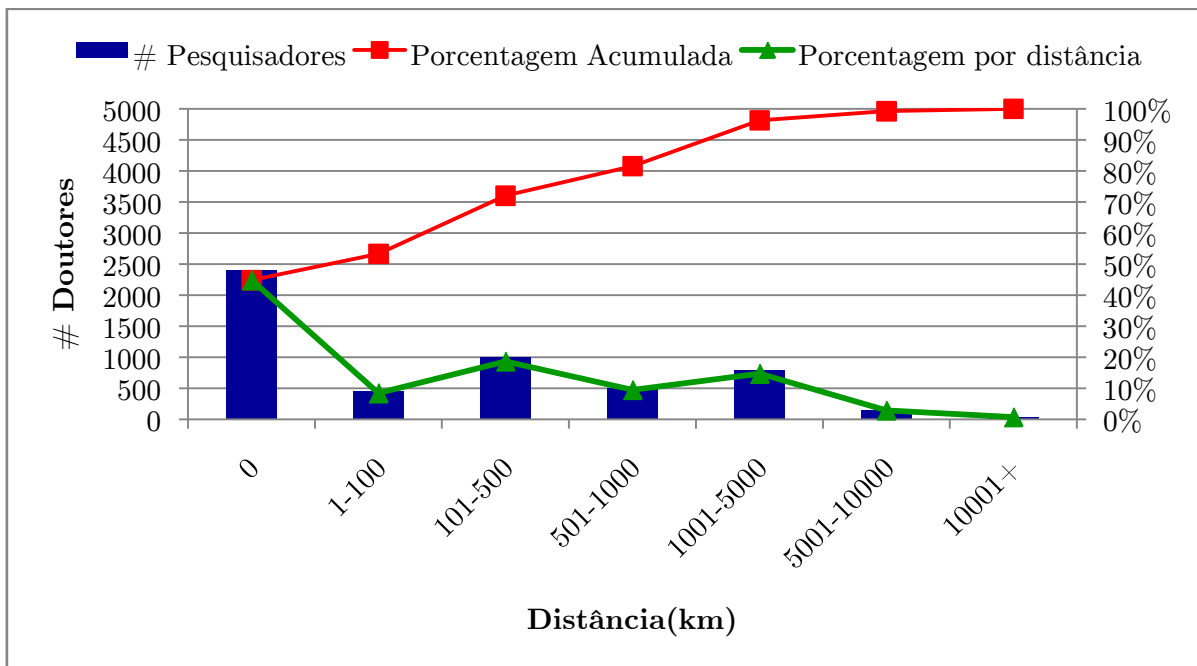


Figura 4.29. Distância percorrida pelos pesquisadores do primeiro ao último ponto da trajetória

No segundo caso mostramos a distância entre a primeira e última instituição da trajetória, usualmente a instituição de graduação e de emprego respectivamente. A Figura 4.29 mostra um padrão similar ao da Figura 4.28, mas com certas diferenças. A primeira figura mostra uma tendência de o pesquisador continuar os estudos na instituição em que estudava anteriormente, mas não é raro encontrar pesquisadores que foram estudar em instituições distantes. A segunda figura também mostra a tendência de não se mudar para longe, mas quando o pesquisador encontra emprego longe de sua região inicial este costuma ser em locais menos distantes. Na Figura 4.16 há uma grande concentração de pesquisadores que trabalham a menos de 100 km de distância

da sua primeira posição, mas a quantidade de pesquisadores é menor se comparada com a Figura 4.28.

Os padrões mais frequentes confirmam a tendência entre pesquisadores brasileiros de não se mudarem para outras partes do país. Uma grande porcentagem dos pesquisadores INCT conduziu todos os estágios de sua carreira na mesma região do país. Instituições estrangeiras só aparecem nos padrões frequentes em estágios mais avançados, como pós-doutorado. Essa análise também confirma a tendência da redução da demanda por doutorados obtidos no exterior em favor do aumento de doutorados obtidos em São Paulo. Se considerarmos que é desejável que o pesquisador busque uma variedade maior de instituições, grupos e centros de pesquisas durante sua trajetória educacional, os padrões encontrados indicam que algumas políticas poderiam ser aplicadas para promover mobilidade. Uma possível explicação para a mobilidade reduzida é a falta de incentivos apropriados ou falta de incentivo para se mudar, mas tais hipóteses não podem ser testadas usando nossos dados. Uma política nacional foi criada em 2010 para incentivar estudantes de graduação a cursar algumas matérias em instituições estrangeiras, com o programa Ciência Sem Fronteiras²³. Se esse tipo de programa vai influenciar a mobilidade das próximas gerações de pesquisadores ainda é uma incógnita. Não existem programas similares para incentivar a mobilidade dentro do país, exceto para cientistas seniores durante curtos períodos de tempo.

²³ <http://www.cienciasemfronteiras.gov.br/>

Capítulo 5

Discussão das Técnicas de Visualização Usadas

Nesta seção discutimos alguns aspectos do uso da ferramenta de visualizações das trajetórias acadêmicas. A ferramenta²⁴ foi criada para o trabalho como forma de verificar as trajetórias criadas e para auxiliar o entendimento do comportamento delas. Algumas telas da ferramenta foram usadas na seção anterior para ilustrar geograficamente os dados que eram discutidos.

Nesta seção discutiremos alguns parâmetros que devem ser levados em conta para indicar se determinada visualização é recomendada ou não para o cumprimento de certas finalidades de uso. A avaliação será realizada usando os cinco parâmetros usados por Dias et al. [2012]: tipo de dado; tipo de tarefa, escalabilidade, dimensionabilidade e posição dos atributos. Para cada parâmetro é apresentada uma discussão sobre a técnica de visualização e soluções adotadas para superar os problemas encontrados.

5.1 A Ferramenta

A ferramenta de visualização criada se trata de um mapa web, implementado em Javascript utilizando a API do Google Maps²⁵. O mapa exibe as trajetórias dos pesquisadores em forma de setas ou pontos. No caso onde os pontos de

²⁴ <http://aqui.io/trajectory>

²⁵ <https://developers.google.com/maps/>

origem e destino são diferentes, usa-se a linha direcionada; nos casos onde origem destino sejam a mesma instituição, usa-se o ponto.

A ferramenta permite que o usuário aplique diversos filtros aos dados, como pode ser visto na Figura 5.1. O usuário pode filtrar as trajetórias pelos atributos dos pesquisadores associados a elas. Por exemplo, ao aplicar os filtros “Exatas”, “PhD” e “Trajetorias Finalizadas na década de 1980”, o mapa exibirá as trajetórias dos pesquisadores da área de Exatas, considerando somente os segmentos que o pesquisador obteve um doutorado na década de 1980. Também é possível filtrar as trajetórias pela distância dos segmentos ou pela origem e destino.

Legenda

- Todas as áreas.
- Engenharia E Tecnologia Da Informação
- Saúde
- Nanotecnologia
- Ecologia E Meio Ambiente
- Ciências Humanas E Sociais Aplicadas
- Energia
- Ciências Agrárias E Agronegócio
- Exatas

- Todas as modalidades B (A->B).
- Undergraduate
- Master
- PhD
- Posdoc
- Work.

- Todas as trajetórias (por década).
- Trajetórias Finalizadas na década de 1950.
- Trajetórias Finalizadas na década de 1960.
- Trajetórias Finalizadas na década de 1970.
- Trajetórias Finalizadas na década de 1980.
- Trajetórias Finalizadas na década de 1990.
- Trajetórias Finalizadas na década de 2000.
- Trajetórias Finalizadas na década de 2010.

Opacidade:

- 100%
- 75%
- 50%
- 25%
- 10%
- 5%

Modo:

- Usar dados de instituições.
- Agrupar instituições por países.

Escolha qual tema de cores usar:

- Usar preto (default)
- Área
- Modalidade
- Década

Trajetoias nulas (Que vão de um ponto para ele mesmo):

- Ignorar trajetórias nulas
- Considerar apenas trajetórias nulas

Trajetoias unitárias (Que apenas um único pesquisador percorreu):

- Remove trajetórias unitárias.

A grossura da linha e tamanho do ícone são proporcionais a quantidade de pesquisadores que realizaram aquele trajeto:

- Ignorar quantidade de pesquisadores.

Origens e Destinos:

- Apenas Trajetórias com origem no Brasil
- Apenas Trajetórias com destino no Brasil
- Apenas Trajetórias com origem fora do Brasil
- Apenas Trajetórias com destino fora do Brasil

Comprimento máximo e mínimo das trajetórias: Escolha o tamanho máximo e mínimo (em quilômetros) das trajetórias que serão exibidas

Mínimo Máximo

Visualizar trajetórias de um pesquisador específico (segundo seu identificador): Clique sobre a trajetória para mais informações.

Figura 5.1. Filtros presentes na ferramenta

Os outros filtros estão relacionados a como os dados são exibidos. É possível alterar a opacidade das linhas; mudar sua espessura de acordo com o número de pesquisadores que seguiram aquela trajetória; agrupar as linhas por países; e remover as linhas pertencentes à só um pesquisador. Por último, é possível procurar a trajetória de um pesquisador específico, buscando-o pelo identificador único do seu currículo Lattes.

5.2 Tipo de Dado

O tipo de dado é um fator determinante para a escolha de qual técnica de visualização utilizar. Shneiderman [1996] classifica os dados de acordo com o número de atributos (dimensão) e também quanto à natureza, que pode ser quantitativa (dados numéricos) ou qualitativa (dados categóricos). Assim como Dias, será considerada nessa seção somente a natureza, deixando a dimensão dos dados para outra discussão.

A definição de mapas temáticos²⁶ diz que eles são capazes de mostrar tanto dados qualitativos (tipo de vegetação de uma área) quanto quantitativos (densidade populacional em uma área) através de cores e símbolos, entre outras técnicas. Portanto, caso a trajetória possua dados de natureza quantitativa e qualitativa, a visualização geográfica de trajetórias é capaz de mostrá-los.

Vale ressaltar que a visualização discutida neste trabalho não é capaz de exibir todos os tipos de *trajetórias*. Pela definição de Spaccapietra [2007] as trajetórias podem ser metafóricas, ingênuo-geográficas e espaço-temporais. Como a posição tratada nas trajetórias metafóricas é imaginária, não há como localizar os pontos espacialmente. Embora uma trajetória acadêmica possa ser vista como metafórica (o pesquisador se moveu da graduação para o doutorado, daí para o pós-doutorado e finalmente para o emprego), o fato de cada etapa possuir uma instituição geolocalizável associada a ela permite visualizar essas trajetórias.

5.3 Tipo de Tarefa

Tipo de tarefa refere-se às atividades que o usuário pode fazer de acordo com os objetivos do uso de uma representação gráfica (Kleim [2002], Shneiderman [1996]). As tarefas consideradas por Dias et al. [2012] são:

- **Visão Geral dos Dados:** capacidade de mostrar todos os dados;
- **Correlação de Atributos:** o grau de relacionamento entre as variáveis pode revelar padrões de comportamento ou tendências;
- **Identificação de padrões,** normas e características importantes;

²⁶ <http://www12.statcan.gc.ca/census-recensement/2011/ref/dict/geo047-eng.cfm>

- **Identificação de Agrupamento:** atributos com comportamentos parecidos;
- **Identificação de Anomalia:** dados com comportamento atípico em comparação ao resto.

A visualização geográfica das trajetórias permite a visualização geral dos dados, mas sem filtragem dos dados a execução das outras tarefas é prejudicada.

A Figura 5.2 mostra um exemplo.

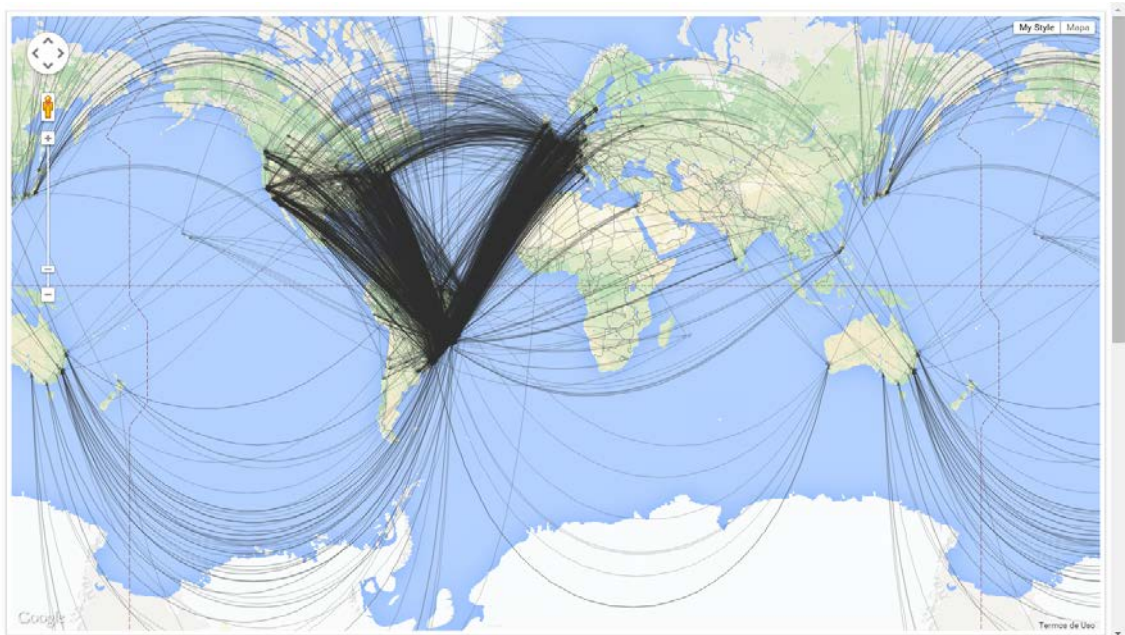
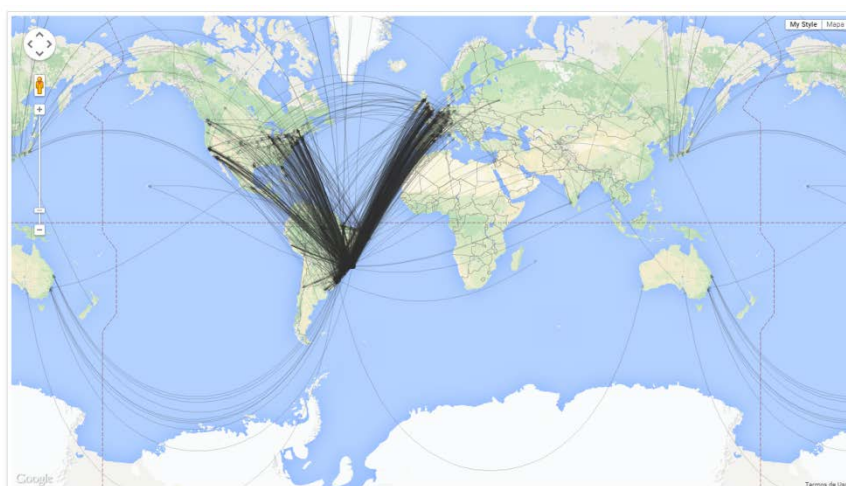
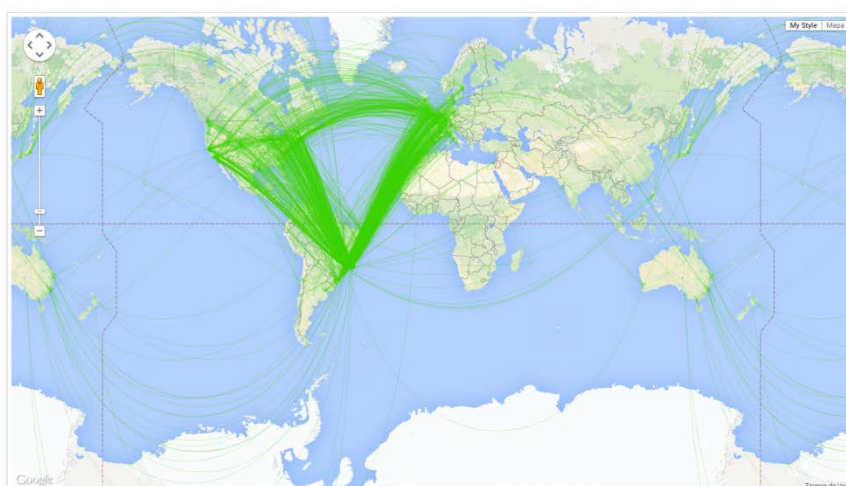


Figura 5.2. Visualização geral dos dados de trajetória.

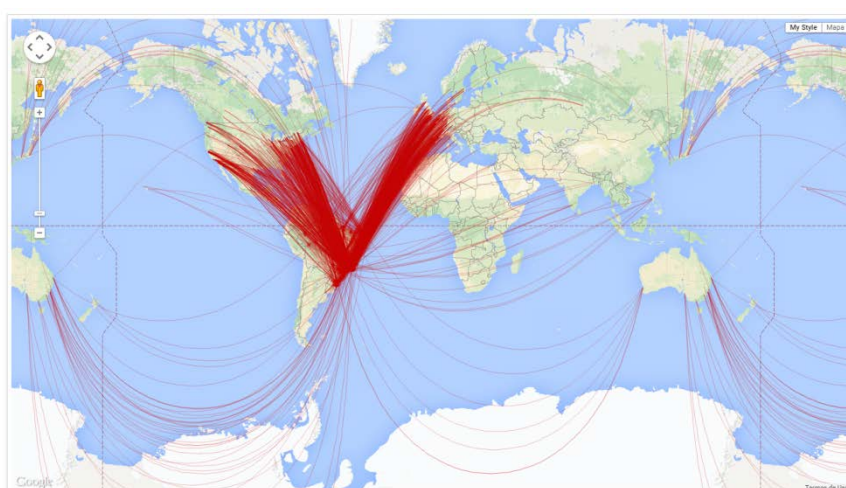
Na Figura 5.2 são apresentadas todas as trajetórias do nosso banco de dados. É possível ver que existem três padrões mais comuns de trajetórias: Brasil ↔ Europa, Brasil ↔ EUA e Europa ↔ EUA. Outras regiões comuns são a Austrália e o Japão. Além disso, é incomum que a trajetória não tenha como origem ou destino o Brasil. Análises mais detalhadas são difíceis de fazer visualmente devido ao número de linhas no mapa, por isso é necessário usar filtros para permitir a execução das tarefas listadas por Dias.



(a)



(b)



(c)

Figura 5.3. Visualização dos dados de trajetória: estágios acadêmicos (a), pós-doutorado (b) e emprego (c).

A Figura 5.3 mostra as mesmas trajetórias da Figura 5.2, mas agora separadas pelo motivo que originou aquela movimentação: se a trajetória está relacionada à obtenção de um diploma de graduação, mestrado ou doutorado (a), se o pesquisador foi participar de pós-doutorado (b) ou se o pesquisador foi trabalhar (c). Ao filtrar os dados é possível ver que as modalidades influenciam os padrões e podemos separar o comportamento habitual das exceções. Por exemplo, o movimento entre EUA e Europa é um padrão comum para a etapa de pós-doutorado, ou seja, muitos pesquisadores que cursaram o doutorado em um continente vão ao outro para o pós-doutorado. No entanto, esse padrão é uma exceção em outras etapas.

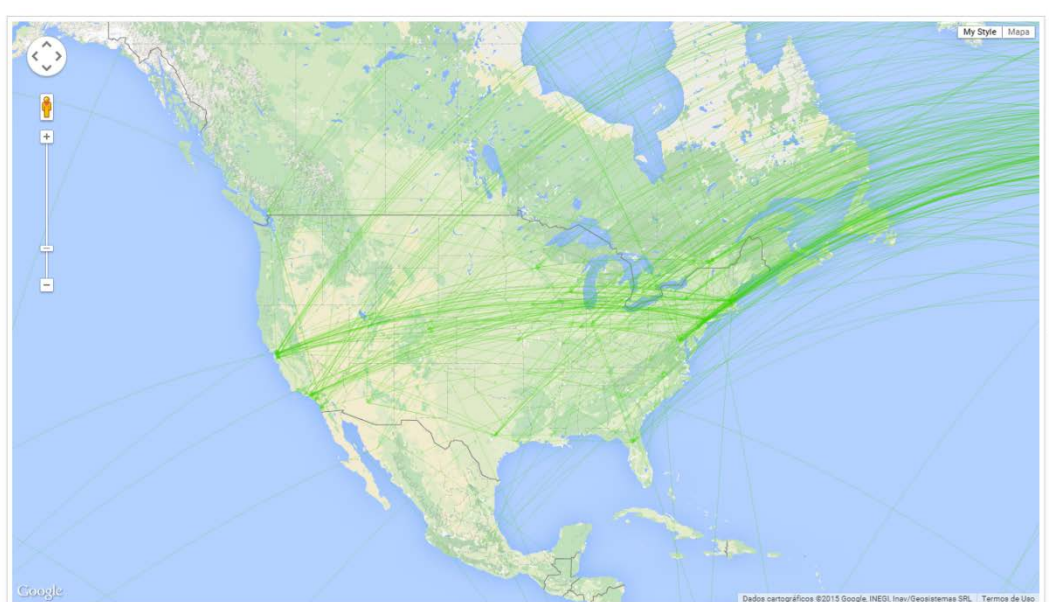


Figura 5.4. Trajetórias entre costa Leste e Oeste dos EUA

A visualização proposta não é capaz de mostrar agrupamento dos dados. É possível ver trajetórias parecidas, mas a ausência de métricas para medir o quanto parecido são duas trajetórias impede de classificar os grupos de fato.

Embora a visualização geográfica permita realizar quase todas as tarefas descritas, mas nem sempre elas são facilmente realizadas. O usuário da visualização deve sempre ter em mente que podem existir comportamentos difíceis de serem encontradas mesmo depois de filtrar dados indesejados. A Figura 5.4 mostra as trajetórias de pós-doutorado nos EUA, removendo aquelas que possuem como destino ou origem o Brasil. Analisando somente a Figura 5.3.b, as trajetórias entre as duas costas dos Estados Unidos poderiam passar

despercebidos por um usuário sem atenção ou que não esperava esse comportamento.

5.4 Escalabilidade

Escalabilidade é uma característica a ser observada no dado, que reflete o tamanho do *dataset* a ser visualizado. Os dados podem ser classificados pela ordem de grandeza como volume pequeno (10^1 até 10^2 entradas diferentes), volume médio (10^3 até 10^5) e volume grande (10^6 até 10^7). O conjunto de dados usado possui 21.092 trajetórias e 8.533 trajetórias únicas, sendo classificado como volume médio.

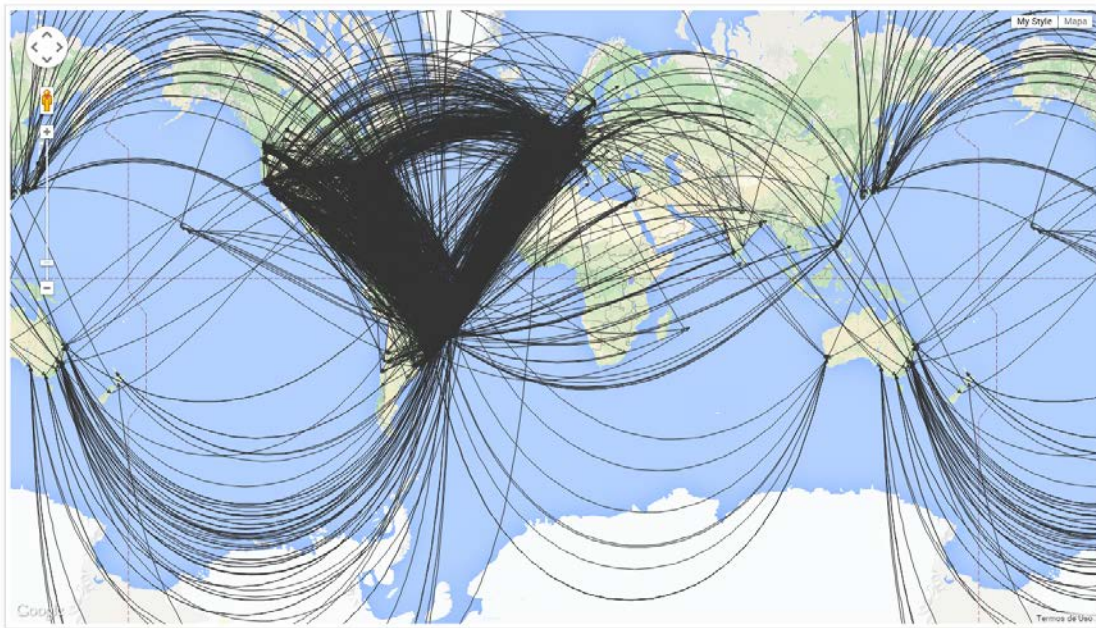
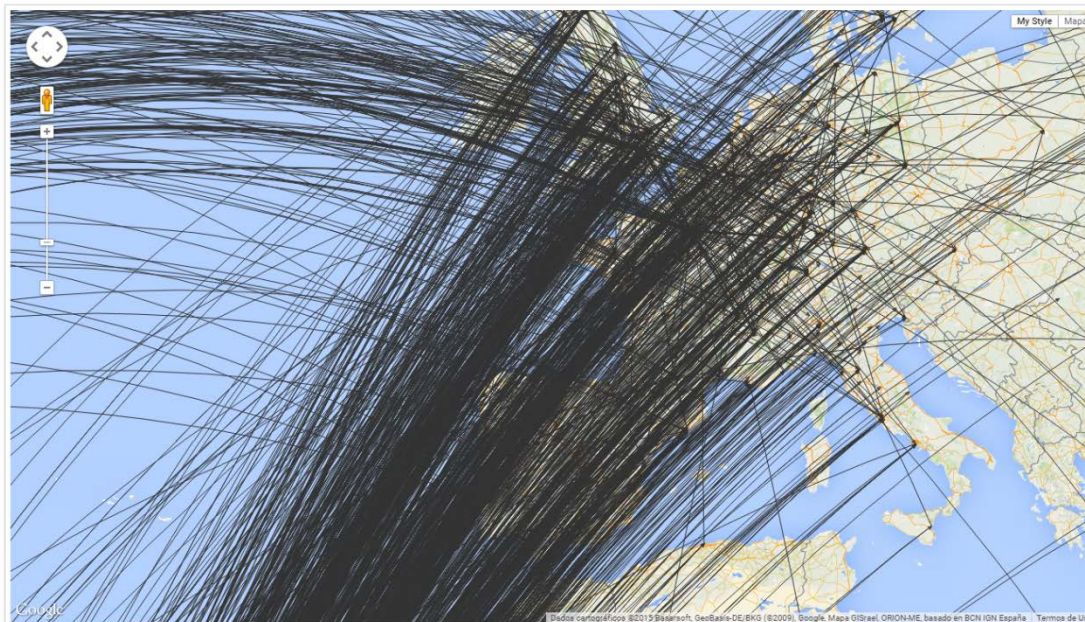


Figura 5.5. Visualização das trajetórias

A visualização proposta possui problemas ao apresentar um grande número de dados. A Figura 5.5 mostra como fica o mapa se tentarmos ver todas as trajetórias. A concentração de linhas no Brasil, EUA e Europa é muito grande e elas se sobrepõem umas sobre as outras. Isso nos trás dois problemas: a maioria dos casos não é possível identificar origem ou destino, somente nas trajetórias *outliers* é possível identificar um desses dois pontos. Além disso, as linhas mais curtas são dificilmente identificadas e nas figuras acima não seria

possível ver trajetórias regionais, por exemplo, entre estados brasileiros. O mapa, neste caso, serve para ver onde estão os padrões mais comuns e os *outliers*, mas não é possível ver informações mais detalhadas.



(a)



(b)

Figura 5.6. Melhorias na leitura do mapa: zoom (a) e cor (b)

Seria de se esperar que o problema fosse causado pela falta de diferenciação entre as linhas, mas não é o caso. A Figura 5.5 mostra que o zoom e a colorização não são soluções para o problema. Aumentar o zoom do mapa ajuda a encontrar os pontos de chegada e saída das trajetórias, mas ainda há muita sobreposição entre as linhas e é difícil ver os pontos que estão debaixo das grandes concentrações como, por exemplo, a cidade de Paris. As cores diferentes ajudam a mostrar que existem várias linhas diferentes, mas ainda não é possível ver de onde elas surgem e para onde elas vão.

5.4.1 Filtros

Como a visualização não possui boa capacidade de mostrar todos os dados do nosso *dataset* foi necessário encontrar meios de melhorar sua leitura. Duas opções foram encontradas. Uma é diminuir o número de trajetórias mostradas ao mesmo tempo, a outra é facilitar a leitura do mapa sem diminuir o número de dados mostrados. Para o primeiro caso adotamos o uso de filtros para remover dados indesejados ou agrupar as trajetórias. Para o segundo, usamos linhas com graus de opacidade diferentes.

A Figura 5.6 mostra duas opções encontradas para agrupar as trajetórias em menos linhas. As trajetórias no modo normal da visualização possuem como pontos de origem e destino as coordenadas geográficas das instituições, obtidas na etapa de geocodificação. Ao agrupar as instituições por país ou estado temos um ganho na facilidade de leitura ao custo da perda de detalhes: as linhas são mais fáceis de serem lidas e o mapa mostra uma visão geral das trajetórias, mas não é possível identificar padrões regionais.



(a)



(b)

Figura 5.7. Mapa após agrupar as trajetórias por país (a) e por estado (b)

Na Figura 5.7a é possível ver que existe a trajetória Brasil \leftrightarrow EUA mas, ao contrário da Figura 5.4, não é possível ver exatamente de onde os pesquisadores estão vindo e para onde estão indo. Isso pode influenciar o usuário a pensar que a distribuição geográfica das instituições é homogênea, o que não é

o caso. Outro problema do mapa é que não é possível ver as trajetórias intranacionais e intra-estaduais, o que também pode levar o usuário a tirar conclusões erradas. Essas trajetórias não aparecem no mapa porque o mapa mostra o movimento dos pesquisadores, mas não mostra aqueles que continuaram no mesmo ponto em etapas subsequentes. Iremos discutir mais sobre trajetórias com origem e destino iguais na seção 5.5.



Figura 5.8. Mapa exibindo 10.000 trajetórias



Figura 5.9. Mapa exibindo 5.000 trajetórias



Figura 5.10. Mapa exibindo 1.000 trajetórias



Figura 5.11. Mapa exibindo 500 trajetórias



Figura 5.12. Mapa exibindo 100 trajetórias



Figura 5.13. Mapa exibindo 50 trajetórias

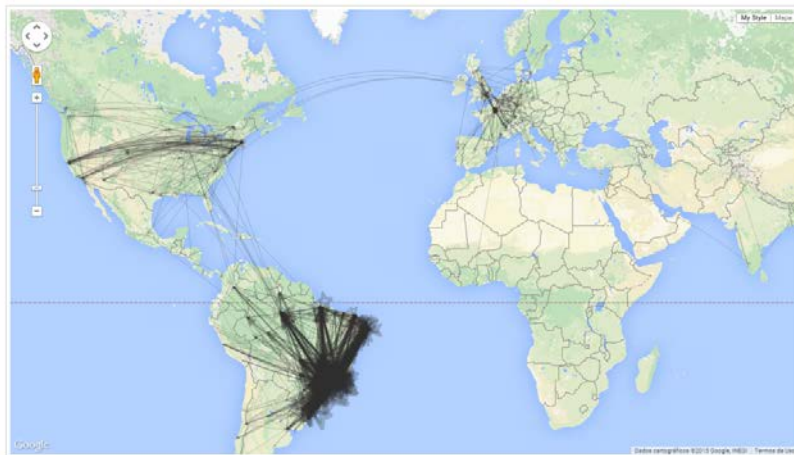
As Figuras 5.8 até 5.13 mostram como fica a visualização ao exibir menos dados. As três primeiras figuras mostram o que seria um volume médio de dados, enquanto as três últimas mostram um volume de dados baixo. Na nossa avaliação, o mapa permite a apresentação de conjuntos de dados de volume aproximado de 1.000 entradas. Com essa quantidade, os mapas mostram as linhas de forma visível e já é possível ver os pontos de entrada e saída, embora ainda seja difícil vê-los quando há muita concentração de linhas sobre os pontos. Nas figuras com mais de mil entradas ainda há muita sobreposição entre as linhas e a redução não facilita a leitura do mapa. Já nas figuras com poucas linhas a leitura é fácil e é possível identificar facilmente as trajetórias, mas não é necessário reduzir tanto o volume de dados para que a leitura seja clara. Assim como no caso anterior, reduzir o número de linhas exibido pode esconder do usuário comportamento dos dados. No caso desta ferramenta, é dada ao usuário a oportunidade de filtrar as linhas pelas variáveis da trajetória como, por exemplo, filtrar os pesquisadores que se formaram na década de 80 ou filtrar todos os pesquisadores da área de Exatas pelo grau de formação.

Outra opção criada foi o filtro de distâncias. Como visto anteriormente, a mobilidade dos pesquisadores é baixa. No entanto, as linhas mais compridas chamam mais a atenção, cobrindo as linhas mais curtas. Para resolver o problema, foi criado um filtro onde o usuário pode escolher o comprimento mínimo e máximo das trajetórias exibidas. A Figura 5.14 mostra o resultado do filtro. Aproximadamente, a figura (a) mostra 55% dos pesquisadores, a figura

(b) mostra 35% e a figura (c) mostra 20%. Percebe-se que Figura 5.14c aparenta ter mais linhas do que (a) ou (b), e também é muito similar à Figura 5.2. Ao filtrar trajetórias longas evita-se esconder os padrões das trajetórias curtas.



(a)



(b)



(c)

Figura 5.14. Segmentos de trajetória de distância 0km (a), com menos de 5 mil km (b) e com mais de 5 mil km (c)

5.4.2 Transparência

Nos dois casos anteriores foi visto que reduzir o número de trajetórias exibidas pode esconder padrões visuais do usuário, o que não é desejável para a ferramenta. Para evitar que seja necessário remover linhas que podem ser de interesse do usuário, acrescentamos às linhas um nível de opacidade. Com as linhas transparentes é possível enxergar o que há por baixo, justamente um dos problemas da nossa visualização. A cor das linhas possui propriedade aditiva, ou seja, se duas linhas se sobrepõem a cor resultante terá intensidade maior, como se possuíssem um nível de opacidade duas vezes maior que o original. Essa propriedade apresentou uma vantagem que não era esperada, destacando grupos de linhas mais comuns ao aumentar a intensidade delas.



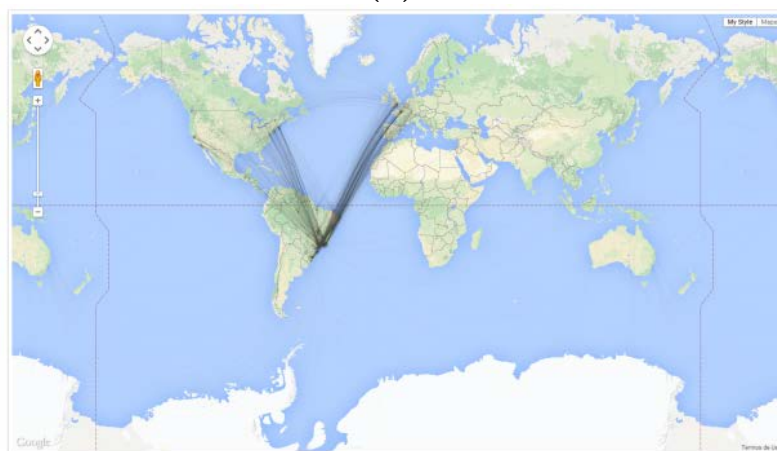
(a)



(c)



(b)



(d)

Figura 5.15. Exemplos de trajetórias com nível de opacidade: 50% (a), 25% (b), 10% (c) e 1% (d)

A Figura 5.15 mostra exemplos de níveis de opacidade que podem ser adotados no mapa. Níveis de transparência de 50% ou mais, como visto na Figura 5.15a, não apresentam muita diferença quando comparamos com o mapa da Figura 5.5. Por outro lado, níveis de transparência de 1% ou menos, como na Figura 5.15d, mostram as linhas quase invisíveis e somente os padrões muito comuns ganham destaque. Os outros dois níveis de transparência, mostrados na figuras 5.15b e 5.15c, amenizam o problema de sobreposição das linhas e apresenta boa visibilidade dos pontos de início e fim. A Figura 5.16 mostra com detalhes a região da Europa mostrada na Figura 5.6b com níveis de transparência. Nos dois casos é possível perceber a melhoria na leitura. Além disso, a Figura 5.16a retira o destaque das anomalias, enquanto na Figura 5.9b somente os padrões mais comuns se destacam.



(a)



(b)

Figura 5.16. Exemplos de melhoria da leitura: Zoom e Opacidade de 25% (a) e 10% (b)

Foi possível constatar que os níveis de transparência melhoram a leitura e que dependendo da intensidade a opacidade pode facilitar a realização de certas tarefas, como encontrar os padrões ou os *outliers*. No caso de uma ferramenta de

análise exploratória é importante dar o controle do nível de opacidade para o usuário, assim ele pode variar o nível de opacidade em busca da visualização mais apropriada para ajudar uma realizar a tarefa ou destacar um padrão.

5.5 Dimensionabilidade

Dimensionabilidade é outra característica relacionada ao dado a ser visualizado, e é relacionada ao número de atributos que o dado possui. Dias define dados com até quatro atributos como baixa dimensionalidade, entre cinco e nove são como média dimensionalidade, e dados com dez ou mais atributos como alta dimensionalidade. Cada ponto da trajetória do *dataset* possui os seguintes atributos:

- Pesquisador;
- Área de pesquisa do pesquisador;
- Nome da Instituição, com cidade, estado, país e coordenada geográfica;
- Atividade que o pesquisador fazia na instituição (se ele estava no mestrado, doutorado, emprego...);
- Ano de início e conclusão da atividade.

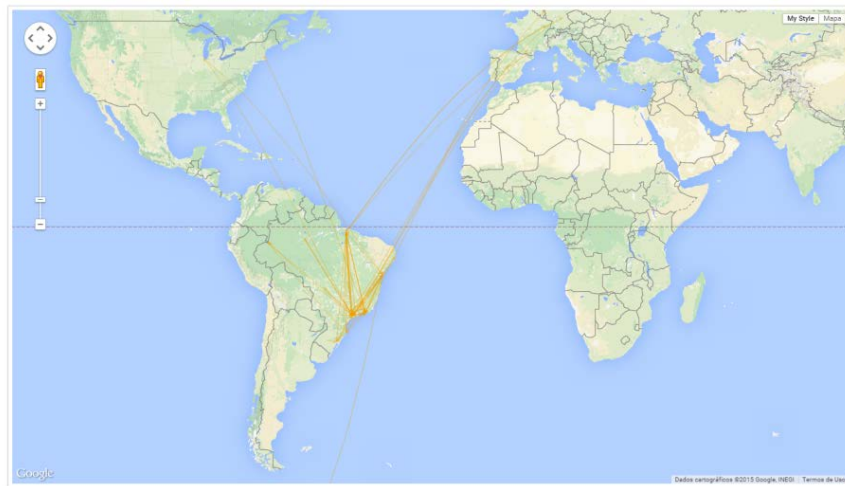
Cada segmento da trajetória é formado por um ponto de origem e um ponto de destino, com uma seta indicando a direção origem/destino. A trajetória completa do pesquisador é formada por todos os segmentos unidos.

Além do traço e dos pontos de início e fim, usamos cores, transparência, espessura e filtros para representar os atributos. Usamos as cores para representar classes de atributos qualitativos e espessura para representar as faixas numéricas dos dados quantitativos. A transparência é usada para representar trajetórias comuns. Outras opções para exibir múltiplos atributos ao mesmo tempo são o uso de degradê, animação (Schich et al. [2014]) e visualização 3D (Tominski et al. [2012]), que não são abordados nesse trabalho.

Este trabalho aborda somente uma das possibilidades de representar atributos com cores e tamanhos. É possível usar escala de cor para representar categorias ou a espessura para representar elementos quantitativos. No entanto, a discussão a seguir se refere somente às decisões tomadas para este trabalho.

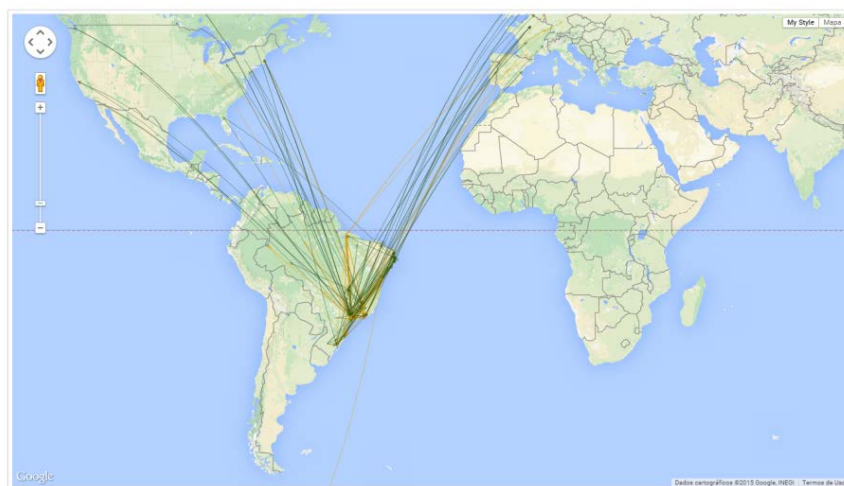
Cor

A Figura 5.10 mostra exemplos de telas onde usamos uma ou mais cores para representar classes diferentes de trajetórias. Nesse caso, mostramos trajetórias comuns de diversas áreas de pesquisa. Podemos ver que as cores são um bom método de diferenciação, mas o número de classes pode limitar sua utilidade. São dois os motivos dessa limitação. Um é que a partir de quatro cores é difícil criar uma paleta onde cada cor seja facilmente distinguível²⁷. A outra é a transparência usada nas linhas, que cria novas cores que podem confundir o usuário. Como visto na seção anterior, em conjuntos de dados com muitas linhas se cruzando é necessário usar transparência para que seja possível ler o mapa sem dificuldades. Quando duas ou mais linhas se cruzam, a visualização exibe uma linha cuja cor é igual à soma da cor das linhas.

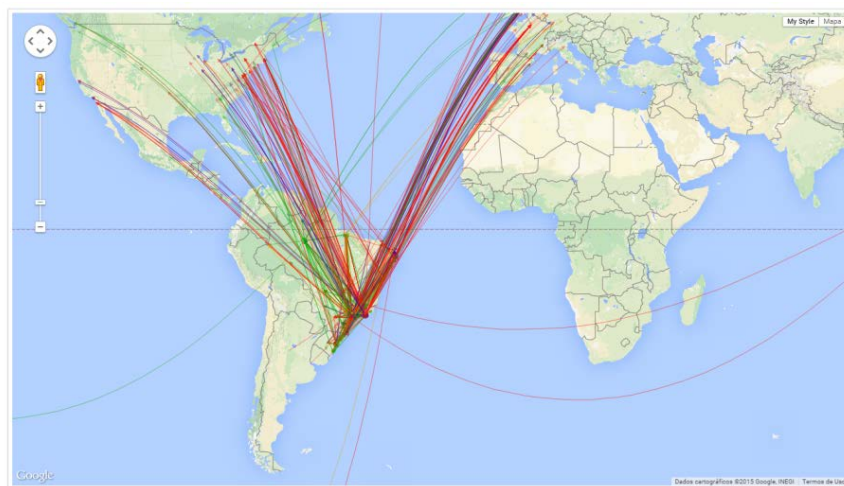


(a)

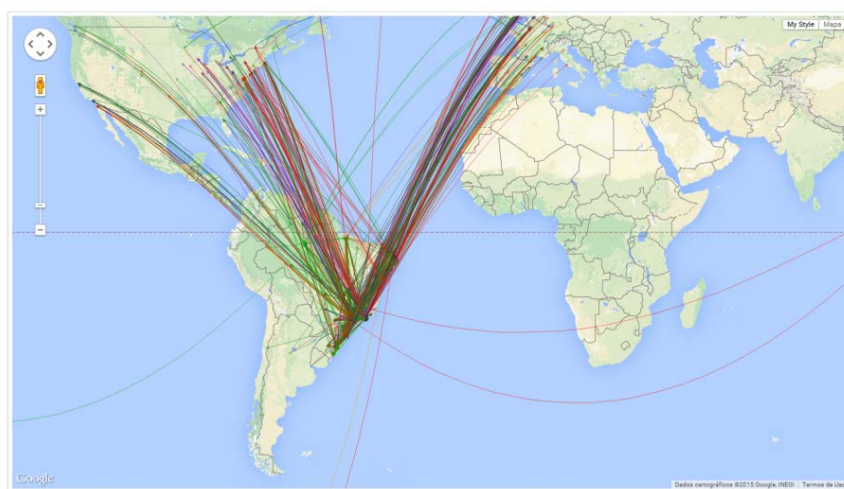
²⁷ http://www.tigercolor.com/color-lab/color-theory/color-theory-intro.htm#color_harmonies



(b)



(c)



(d)

Figura 5.17. Exemplos de mapas com uma (a), duas (b), quatro (c) e oito (d) cores ao mesmo tempo

A Figura 5.18 exibe as linhas com espessura maior para mostrar melhor como a transparência afeta as cores. As linhas da imagem podem ser azuis, amarelas ou verdes, representando respectivamente mestrado, doutorado e pós-doutorado. No entanto, vemos que há mais cores do que só essas três. É possível ver que existem tons mais escuros das três cores, além de linhas que possuem tons parecidos com cinza.

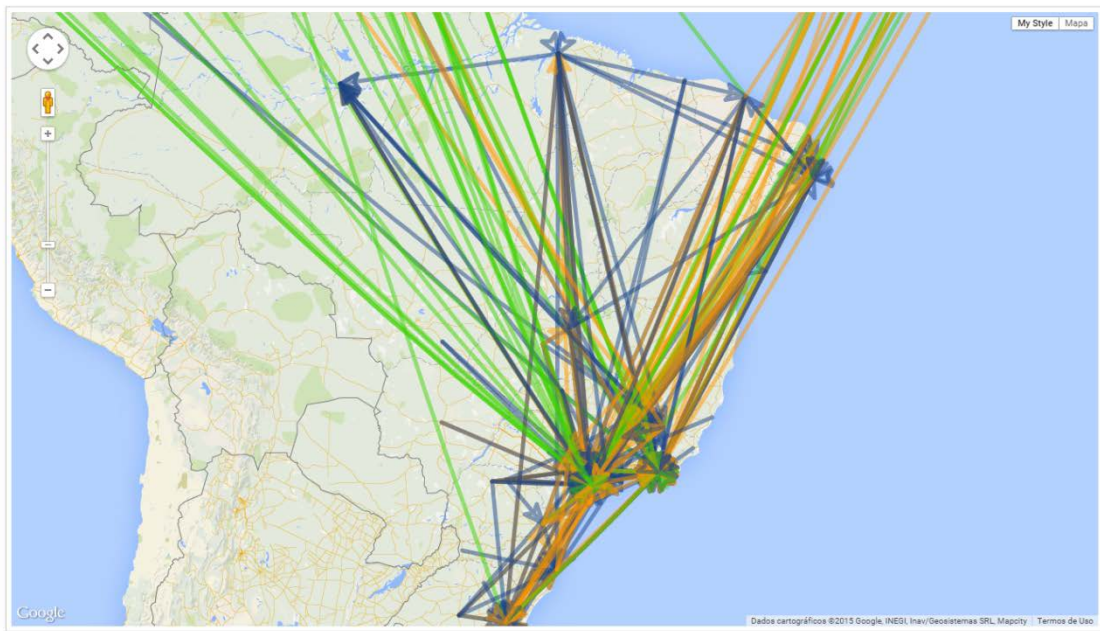
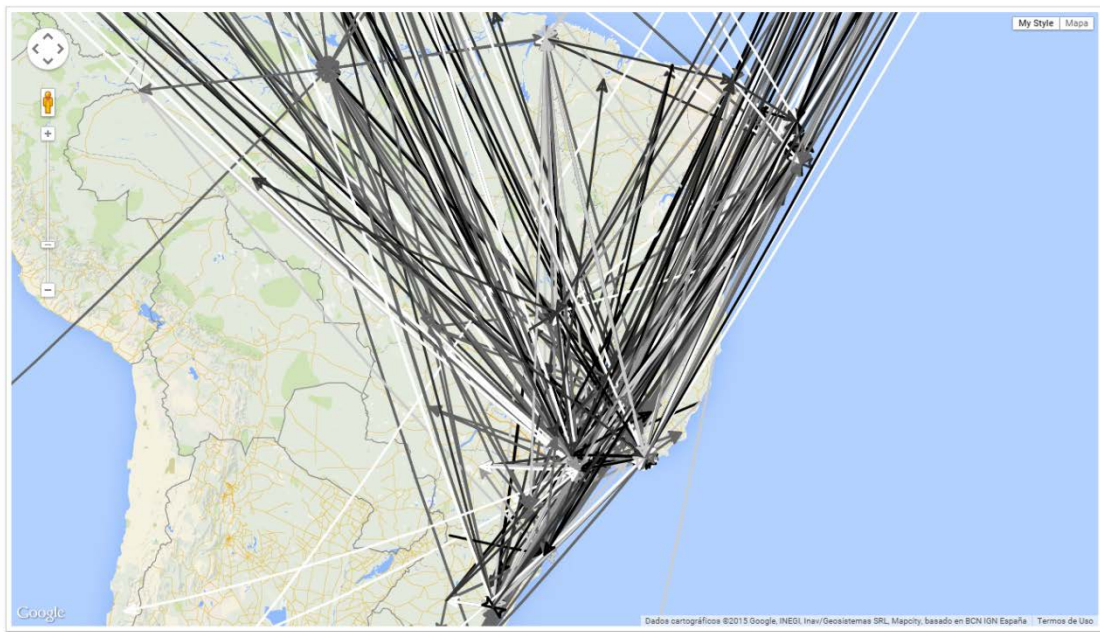
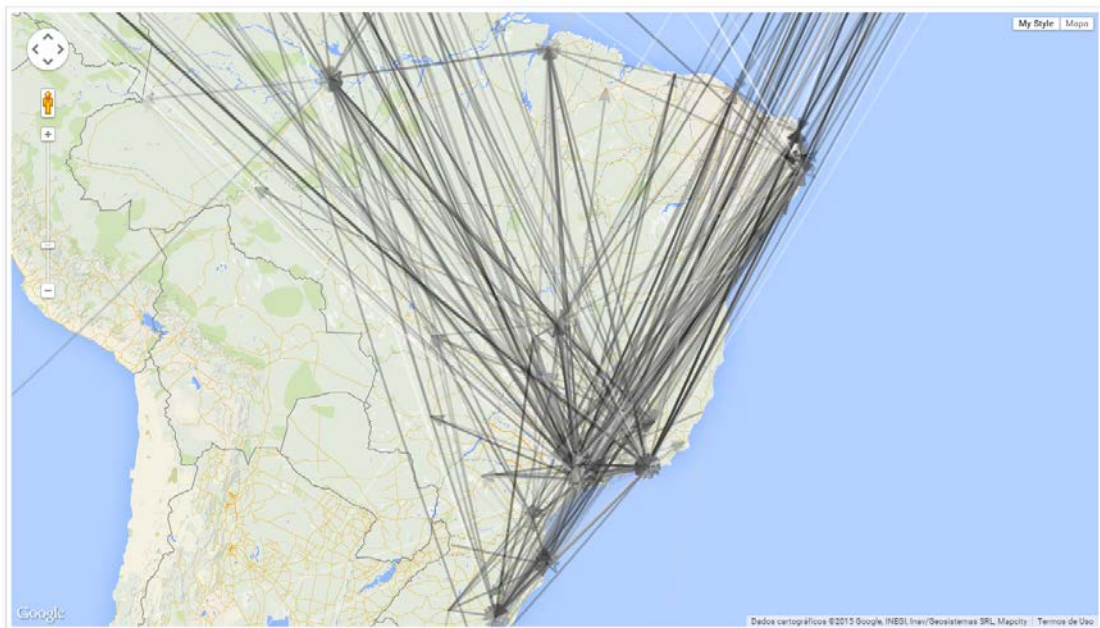


Figura 5.18. Efeito da transparência na soma de cores

Usar uma escala monocromática para diferenciar as linhas é uma possibilidade, mas que não foi usada neste trabalho porque o uso de opacidade impede a interpretação correta dos níveis de cor. Como pode ser visto na Figura 5.12b, a transparência não permite discernir o tom de cinza das linhas.



(a)



(b)

Figura 5.19. Exemplos de mapa com escala monocromática (a) e escala monocromática com opacidade (b)

5.5.1 Espessura

A espessura foi outro aspecto da linha que usamos para representar atributos. Cada linha representa um pesquisador que se formou na instituição de origem e começou uma nova etapa de ensino na instituição de destino. A espessura

representa quantos pesquisadores fizeram o mesmo movimento em sua carreira acadêmica. Alguns comentários devem ser feitos a respeito do uso de espessura. Primeiro é que quando o foco da visualização é mostrar o conjunto de dados completo, a espessura da linha não pode prejudicar a leitura do mapa. Além disso, a ferramenta utilizada não consegue desenhar linhas muito espessas. A espessura que escolhermos deve ser tal que obedeça a esses dois limites.

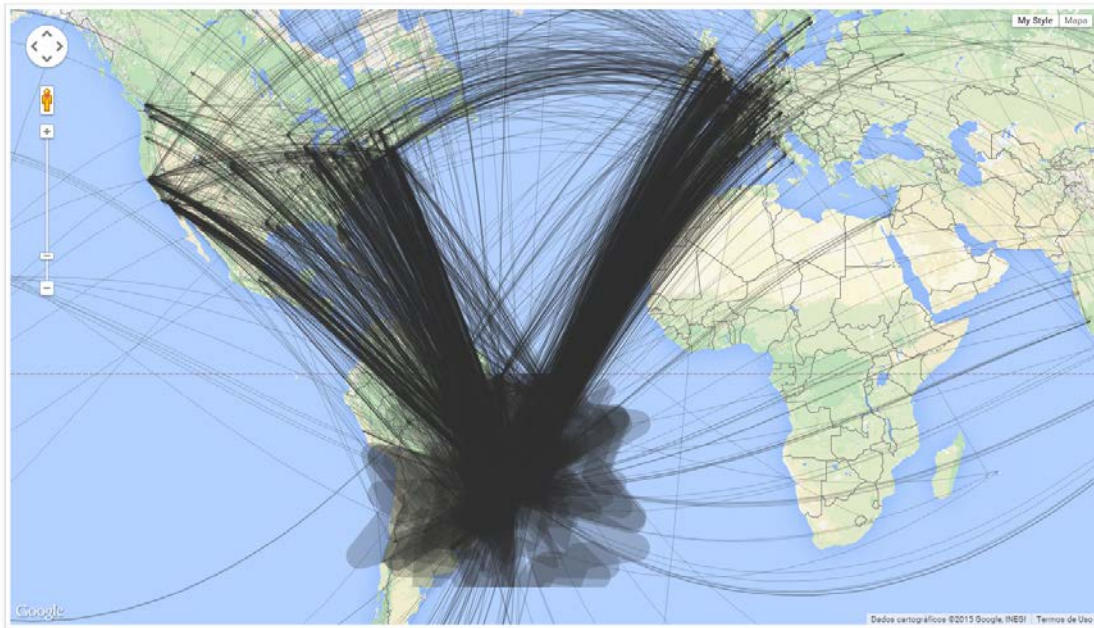


Figura 5.20. Representação 1:1 de espessura e pesquisadores

A Figura 5.13 mostra um mapa dos segmentos mais populares no *dataset*. A espessura da linha tem tamanho, em pixels, igual ao número de vezes que aquele segmento aparece nas trajetórias²⁸. Ou seja, se existem dez pessoas que saíram de A para B, a linha $A \rightarrow B$ terá espessura de 10 pixels. No entanto, essa escolha não é boa. O mapa mostra que a movimentação dentro do Brasil é maior que para o exterior, mas não é possível tirar conclusões mais detalhadas. Algumas linhas ficaram espessas demais, com as setas indicando o destino com dimensão maior do que a da própria linha.

²⁸ Um pesquisador pode aparecer mais de uma vez nessa contagem caso o padrão se repita na sua trajetória. Um pesquisador cuja trajetória acadêmica é $UFMG \rightarrow USP \rightarrow UFRJ$ será contado duas vezes na trajetória $UFMG \rightarrow USP$, uma vez na trajetória $USP \rightarrow UFRJ$ e uma vez na trajetória $UFMG \rightarrow UFRJ$.

Muitos pesquisadores se movimentam distâncias curtas, entre as instituições dentro do país. Ao mostrar o conjunto inteiro de dados teremos linhas curtas, porém muito espessas, e linhas compridas, mas finas. A Figura 5.14 é um exemplo, mostrando que algumas das trajetórias mais populares são também curtas.

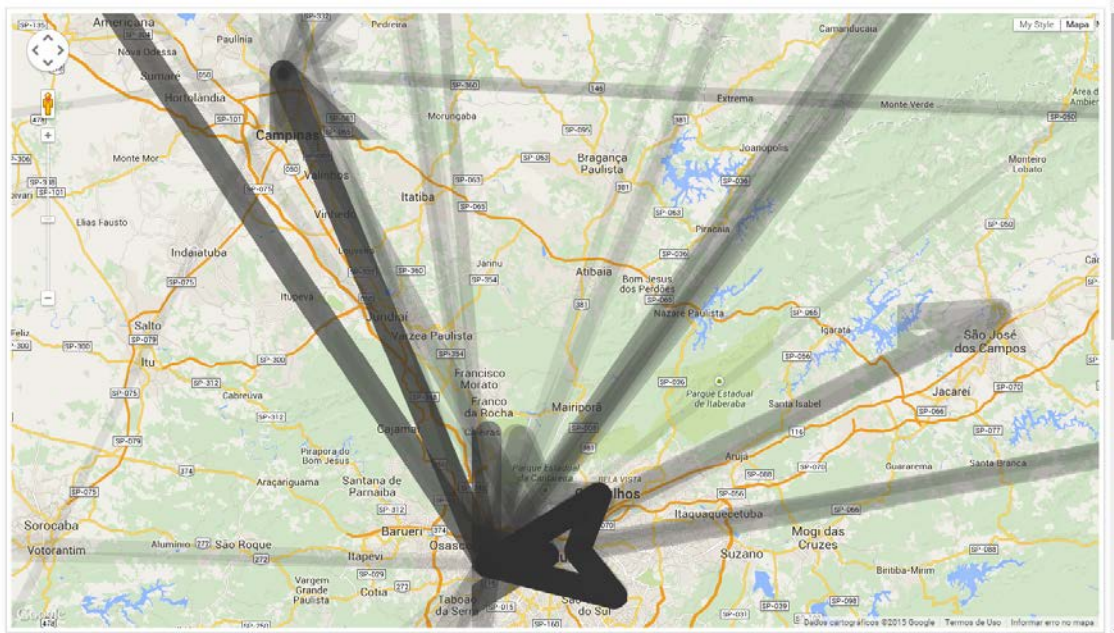


Figura 5.21. Uso de espessura – Destaque em SP

Além do desbalanceamento entre o número de pesquisadores e as trajetórias, há um desbalanceamento entre o número de pesquisadores que se move entre etapas acadêmicas e o número que continua na mesma instituição. Até agora só foi mostrado nos mapas a movimentação dos pesquisadores, quando eles saem de uma instituição para estudar em outra. Aqueles pesquisadores que continuam na mesma instituição não aparecem no mapa. Algumas análises de trajetória ignoram os pontos que não se movem, mas para o nosso caso é importante saber onde estão e quantos são os pesquisadores que não se movimentam.

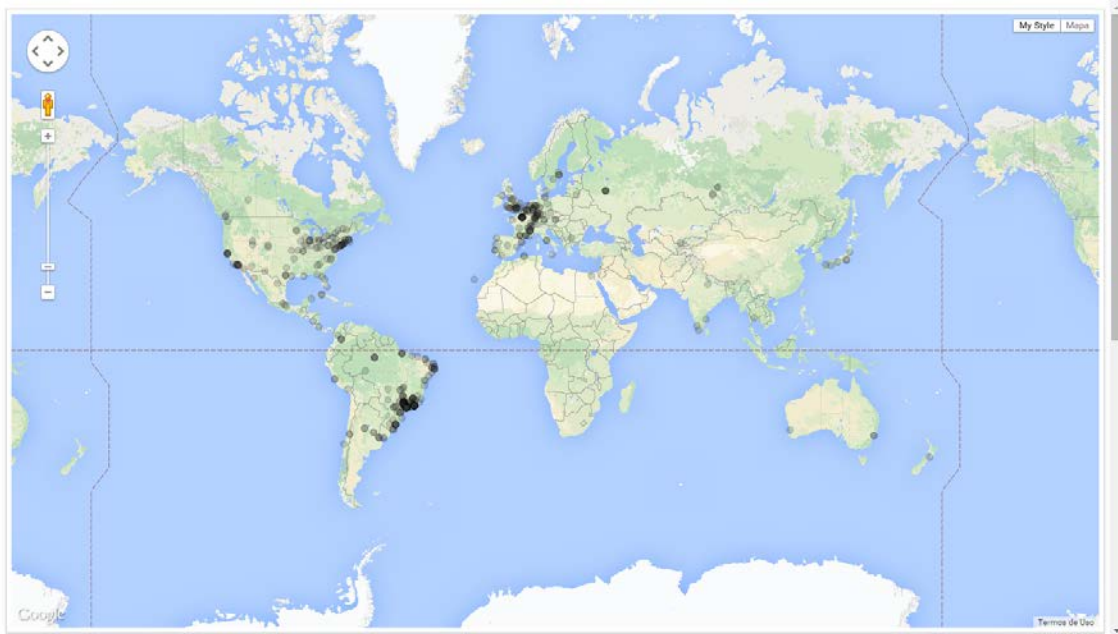


Figura 5.22. Instituições onde pesquisadores estudaram duas ou mais etapas consecutivamente

A Figura 5.22 mostra as instituições que tiveram pesquisadores estudando nelas em duas ou mais etapas consecutivas. Essas instituições aparecem nos mapas (como o da Figura 5.1) como pontos de origem e destino das linhas. O mapa reflete a tendência já observada anteriormente: os pesquisadores se concentram nas regiões Sul e Sudeste do Brasil, além do litoral do Nordeste; no exterior, as instituições estão localizadas principalmente na Europa e Estados Unidos.



Figura 5.23. Representação 1:1 do tamanho do ponto e pesquisadores

Se alterarmos o tamanho dos pontos de acordo com o número de vezes que pesquisadores fizeram estudos consecutivos, vemos como a discrepância é enorme. A Figura 5.23 mostra como o mapa fica ao colocar o número de vezes que aquele segmento aparece nas trajetórias como o diâmetro em pixels do ponto. Os pontos ficam tão grandes que o maior deles não aparece na tela. Analisando os números vemos que a movimentação entre instituições mais comum é da *Universidade Estadual Paulista “Júlio de Mesquita Filho”* para a *Universidade de São Paulo*, que se repete 94 vezes no conjunto de dados. Já a instituição com maior quantidade de etapas consecutivas é a *Universidade de São Paulo*, que aparece 1634 vezes.







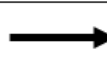
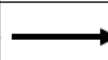
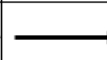
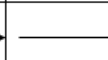
Icons					Arrows				
Greater than 500	Greater than 250	Greater than 100	Greater than 50	Less than 50	Greater than 50	Greater than 20	Greater than 5	Greater than 2	Equals 1
									

Figura 5.24. Legenda da simbologia usada para representar faixas de valor

Com uma diferença tão grande entre as linhas e os pontos, é necessário criar um jeito de facilitar a visualização e permitir que o usuário consiga comparar linhas e pontos. A solução foi criar faixas de valores, usando um símbolo graduado para representar cada faixa. A Figura 5.24 mostra a legenda com as faixas de valores escolhidas. Não houve uma análise estatística para escolher essas faixas; elas foram definidas com base nos valores máximos e foram experimentados valores diferentes em diversos filtros. O resultado final pode ser visto na Figura 5.25.



Figura 5.25. Exemplo de mapa com espessura e símbolos em escala

5.6 Posição dos Atributos

O método de visualização proposto, por ser um mapa 2D, faz parte da classe dos métodos de visualizações cuja posição dos atributos não influencia na interpretação (Shneiderman, 1996). Portanto, não há necessidade de analisar esse parâmetro com a nossa visualização. A posição dos atributos está mais relacionada aos métodos iconográficos (Dias et al., 2012).

5.7 Avaliação final

A ferramenta de visualização criada permite que o usuário realize diversas tarefas, como visualização geral dos dados, análises de padrões e de *outliers*. A visualização é específica para dados de trajetórias e deve ser usada quando os pontos de origem e destino são importantes para as análises e o usuário deseja

visualizar outros atributos, qualitativos ou quantitativos, relacionados às trajetórias.

A visualização apresentada possui baixa dimensionalidade e escalabilidade. A quantidade de dados pode prejudicar a legibilidade do mapa, mas o formato das trajetórias e comprimento das linhas são características que influenciam mais diretamente a legibilidade. Outro aspecto, a escalabilidade, foi afetada pela biblioteca utilizada para implementar a visualização. Existem diversas maneiras de mostrar diferentes atributos em uma linha (cor, formato, espessura...), mas nem todas elas puderam ser utilizadas. Também se deve levar em conta que a forma como mostrar atributos pode afetar a legibilidade do mapa.

Capítulo 6

Conclusão

Este trabalho teve como objetivo estudar as trajetórias de carreira de um conjunto de pesquisadores utilizando os dados informados por eles próprios em seus currículos Lattes. Os dados de currículo são transformados em trajetórias geográficas, utilizando as instituições e etapas de formação acadêmica e de emprego como pontos no espaço e no tempo. As trajetórias foram usadas em diversas análises para descobrir padrões e comportamentos dos pesquisadores brasileiros durante sua formação acadêmica. Também foi criada uma ferramenta de visualização das trajetórias, além de discutir as vantagens e dificuldades que foram encontradas ao usar essa ferramenta durante as análises.

O trabalho obteve sucesso em seus objetivos. No entanto, ressaltamos a seguir alguns problemas encontrados durante sua execução. O currículo Lattes foi uma fonte de dados completa, mas alguns aspectos de como o currículo foi organizado pelo CNPq poderiam facilitar esse tipo de pesquisa. As formas como o usuário do portal listam a instituição onde está trabalhando e as instituições onde ele estudou é diferente: só no primeiro caso há a opção de listar o nome, a seção/campus/instituto e o endereço de fato da instituição, no outro caso só pode ser passado o nome da instituição. No entanto, o campo do nome é de preenchimento livre do usuário. O CNPq possui um Diretório de Instituições²⁹ e durante o preenchimento do currículo o site sugere instituições que já estão cadastradas no diretório, mas o usuário não é obrigado a aceitar as sugestões. Isso leva a casos em que um pesquisador informa o nome como sugerido pelo

²⁹ <http://di.cnpq.br/di/index.jsp>

site (Universidade de São Paulo); um nome detalhado da instituição (Escola de Enfermagem da USP, campus Ribeirão Preto); um nome genérico, que não fornece informação de localização (Escola de Enfermagem); ou até mesmo informar um nome incorreto (Um dos pesquisadores informou “...” como nome da instituição). O nome sugerido é sempre o que possui mais entradas associadas, mas os outros casos se repetem de forma significativa no banco. Uma validação mais rigorosa da informação poderia diminuir os problemas, como ambiguidade, causado por esses nomes.

Com a ausência de endereço para todas as instituições, foram usadas ferramentas de geocodificação para encontrar a localização delas usando exclusivamente o nome informado pelo usuário. O resultado da geocodificação foi preciso o bastante para nossos dados, mas a informação disponível para este fim contém imprecisões e incertezas. Por exemplo, aproximadamente 1.200 pessoas informaram como a instituição que obtiveram seu doutorado a “Universidade de São Paulo”. Ao geocodificar a instituição pelo nome informado, obtemos o endereço do campus central, na cidade de São Paulo. No entanto, não sabemos dizer se todos os 1.200 pesquisadores realmente estudaram no campus de São Paulo ou se eles estudaram em outras cidades. Pudemos afirmar com certeza a cidade da instituição somente nos casos em que as instituições possuem uma única unidade ou nos casos em que o pesquisador informou dados adicionais junto ao nome da instituição. A disponibilidade dos dados do Diretório de Instituições para coleta poderia ajudar na identificação da posição exata do pesquisador em cada etapa da trajetória.

As análises foram feitas levando em consideração a incerteza no nível municipal da localização. Mesmo assim, obtivemos diversas conclusões que acreditamos ser de grande importância para o estudo da comunidade científica brasileira. Encontramos diversos indícios de que os pesquisadores brasileiros tendem a possuir mobilidade reduzida em relação aos números encontrados em estudos na Europa e Estados Unidos (IDEA, 2010; Van Bowell, 2010). Observamos que o estado de São Paulo é responsável pela formação de uma parcela significativa de pesquisadores. É comum pessoas de várias regiões do Brasil saírem de seus estados para fazer doutorado ou pós-doutorado em São Paulo, e no fim do curso voltarem para sua região. Também encontramos uma

tendência de descentralização, com mais pesquisadores estudando em instituições fora do eixo São Paulo-Sudeste, mas a ordem relativa de cada região nos *rankings* criados continua inalterada com o passar do tempo.

A ferramenta de visualização geográfica de trajetórias, criada para este trabalho, foi muito útil como ferramenta exploratória dos dados. Nas etapas iniciais a representação visual ajudou a esclarecer alguns dos fenômenos observados e também a encontrar padrões que não eram esperados. O mapa é grande ajuda para enxergar os padrões de trajetória mais comuns e encontrar os casos mais incomuns, mas não é bom para estimar quantidades absolutas. Como discutido anteriormente, o número de pesquisadores que faz parte de uma trajetória é representado pela espessura da linha ou tamanho do ponto, que é uma representação abstrata da quantidade. Além disso, as trajetórias com mais pesquisadores (mais espessas) costumam ser também as mais curtas. As linhas mais compridas, mesmo que finas, muitas vezes chamam mais atenção visualmente do que as linhas espessas, mas curtas. Não foi possível implementar outras formas de representar quantidade usando a API do Google Maps.

O mapa de trajetórias oferece ao usuário a oportunidade de realizar diversas tarefas, mas não permite mostrar com clareza muitos dados ao mesmo tempo. Esse é um problema causado pelo formato das trajetórias, já que há muito cruzamento entre as linhas do mapa. O número alto de linhas no mapa prejudica a leitura e dificulta a análise visual dos dados. Algumas soluções foram adotadas para melhorar a escalabilidade, como redução no número de linhas exibidas ao mesmo tempo e uso de transparência nas linhas. Reduzir o volume de dados exibido ao mesmo tempo melhora a leitura, mas pode esconder padrões do usuário e levá-lo a tirar conclusões erradas sobre os dados. A transparência, no entanto, ajuda na leitura de mapas com grande volume de dados sem nenhuma desvantagem aparente. É importante deixar claro para o usuário qual nível de transparência está sendo usado, para evitar esconder informação dele.

O mapa proposto também apresentou baixa dimensionalidade de atributos visualizáveis ao mesmo tempo, mas o problema foi causado em parte pela escolha da biblioteca utilizada na implementação. Ferramentas especificamente voltadas para visualização de dados devem ser consideradas em

futuras implementações. Houve um compromisso entre escalabilidade e dimensionabilidade. Ao usar a transparência, diminuimos as opções de cores para representar atributos. Remover a transparência aumenta as possibilidades, mas dificulta a leitura do mapa.

O presente trabalho gerou as seguintes publicações:

- Furtado, C.A.; Andrade, T.K. & Davis Jr., C.A. (2014) *Geovisualization of The Academic Trajectories of Brazilian Researchers*. GeoInfo 2014, pp. 83-94;
- *A Spatiotemporal Analysis of Brazilian Science from the Perspective of Researchers' Career Trajectories* (submetido em periódico internacional; em estágio final de avaliação)

6.1 Trabalhos Futuros

Como trabalho futuro, uma direção prioritária consiste em estender o conjunto de dados utilizado, usar novos tipos de dados e melhorar a ferramenta de visualização de trajetórias. A ferramenta foi desenvolvida usando Javascript e a API do Google Maps, mas sem usar nenhuma biblioteca específica de visualização de dados. Com isso, a implementação do código tornou-se mais difícil; algumas funcionalidades estéticas, como linhas com bordas ou dégradé, não foram tratadas devido a limitações da biblioteca gráfica usada; e a usabilidade da ferramenta em si é baixa. A própria definição do arquivo de entrada impediu a realização de certas atividades. O arquivo de entrada fracionava os dados em segmentos, que foram pré-calculados baseando-se na ordem cronológica que cada estágio foi concluído. Dessa forma não foi possível criar visualizações análogas às tabelas da seção 4.4.

A extensão do banco de dados permitiria estender análises confirmar algumas de nossas hipóteses validadas para o grupo de pesquisadores dos INCTs. Esse grupo de pesquisadores usado é bastante específico: são pesquisadores ligados a um INCT que possuem ou que estão cursando um doutorado. Seria interessante verificar como os pesquisadores ausentes diferem do conjunto que estudamos, que pode ser considerado como sendo composto por

um extrato superior dos pesquisadores do Brasil³⁰. Os comportamentos encontrados são nítidos e ajudam a entender a comunidade científica brasileira, mas os resultados podem ter um viés, positivo ou negativo, por causa do conjunto menor. A expansão do conjunto de dados para todos os pesquisadores que possuem um currículo Lattes é possível, mas mesmo assim os dados não cobririam todos os pesquisadores brasileiros. Não é garantido que a diáspora científica brasileira possui perfis no Lattes, e as estatísticas dos brasileiros que trabalham no exterior são necessárias para as análises.

Também seria interessante usar novos conjuntos de dados de trajetória, como migração interna em um país ou movimentação casa-emprego em cidades grandes. Esses dados apresentam granularidades variadas, tanto na distância quanto no tempo. A diferença na estrutura dos dados exigiria análises e visualizações diferentes. Um conjunto de dados diferente ajudaria principalmente a abstrair as decisões tomadas para todos os tipos de dados. Assim, o mapa de trajetórias e as decisões tomadas para a exibição dos dados poderiam ser adaptados para as necessidades específicas do conjunto de dados.

³⁰ <http://inct.cnpq.br/sobre/>

Bibliografia

- Arenas, J. L.; Castaños-Lomnitz, H.; Valles, J.; González, E. & Arenas-Licea, J. (2001) *Mexican scientific brain drain: causes and impact*. Research Evaluation 10(2): 115-119.
- Alencar, R.O. & Davis Jr, C. A. (2011) *Evaluation of the quality of an online geocoding resource in the context of a large Brazilian city*. Transactions in GIS 15: 851-868.
- Almeida, E.; Chaves, E.; Guimarães, J. A. (2013) *Brazil's growing production of scientific articles: how are we doing with review articles and other qualitative indicators?* Scientometrics 97(2):287-315.
- Avellar, S. O. C. (2014) *Migração interna de mestres e doutores no Brasil: algumas considerações*. Revista Brasileira de Pós Graduação 11(24):20.
- Balancieri, R.; Bovo, A. B.; Kern, V. M.; Pacheco, R. C. S. & Barcia, R. M. (2005) *A análise de redes de colaboração científica sob as novas tecnologias de informação e comunicação: um estudo na Plataforma Lattes*. Ciência da Informação [online] 34(1):64-77.
- Balachevsky E. (2011) *Academic Careers in Brazil: The Legacy of the Past*. The Journal of the Professoriate 4(2):27.
- Balachevsky E. (2013) *Higher Education in Brazil: Different Worlds and Diverse Beliefs*. Comparative & International Higher Education 5(3):4.

- Balbachevsky, E. & do Couto e Silva, E. (2011) *A diáspora científica brasileira: perspectivas para sua articulação em favor da ciência brasileira*. Revista Parcerias Estratégicas 16(33):14.
- Börner, K.; Penumarthy, S.; Meiss, M. & Ke, W. (2006) *Mapping the diffusion of scholarly knowledge among major U.S. research institutions*. Scientometrics 68(3):415-426.
- Bornmann, L.; Leydesdorff, L.; Walch-Solimena, C. & Ettl, C. (2011) *Mapping excellence in the geography of science: An approach based on Scopus data*. Journal of Informetrics 5(4): 537–546.
- Bornmann, L. & Waltman, L. (2011) *The detection of “hot regions” in the geography of science - A visualization approach by using density maps*. Journal of Informetrics, 5(4), 547–553.
- Canibaño, C. & Bozeman, B. (2009) *Curriculum vitae method in science policy and research evaluation: the state-of-the-art*. Research Evaluation 18 (2): 86-94.
- Cantwell, B. (2011) *Academic in-sourcing: international postdoctoral employment and new modes of academic production*. Journal of Higher Education Policy and Management 33(2):14.
- Cantwell, B. & Taylor, B.J. (2013) *Internationalization of the postdoctorate in the United States: analyzing the demand for international postdoc labor*. Higher Education: The International Journal of Higher Education and Educational Planning 66(5):17.

- Cavalcante, R. A.; Barbosa, D. R.; Bonan, P. R. F.; Pires, M. B. O. & Martelli-Júnior, H. (2008) *Perfil dos pesquisadores da área de odontologia no Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)*. Revista Brasileira de Epidemiologia 11(1):106-113.
- Davis Jr., C.A.; Fonseca, F.T.; & Borges, K. A. V. (2003) *A flexible addressing system for approximate urban geocoding*. GeoInfo 2013. ISBN 85-17-00026-9.
- Davis Jr., C.A. & Fonseca, F.T. (2007) *Assessing the Certainty of Locations Produced by an Address Geocoding System*. Geoinformatica 11:103-129.
- Dias, M. M.; Yamaguchi, J. K.; Rabelo, E. & Franc, C. (2012) *Visualization Techniques: Which is the Most Appropriate in the Process of Knowledge Discovery in Data Base?* Advances in Data Mining Knowledge Discovery and Applications. ISBN: 978-953-51-0748-4.
- Enkhsaikhan, M.; Liu, W. & Reynolds, M. (2008) *Geographical and Temporal Visualisation of Social Relationships*. PACIS 2008 Proceedings 1: 433-443.
- Etienne, L.; T. Devogele & A. Boujou (2012). *Spatio-temporal analysis of mobile objects following the same itinerary*. Advances in Geo-Spatial Information Science 10: 47-57.
- Fayyad, U. M.; Grinstein, G. G. & Wierse, A. (2002) *Information Visualization in Data Mining and Knowledge Discovery*. San Francisco: MK/Morgan Kaufmann. ISBN:1-55860-689-0.
- Ferreira, J. & Porfrio, F. P. (2009) *Towards an Academic Social Network for Bologna Process*. Technological Developments in Networking, Education and Automation, pp. 151-156.

- Furtado, C. A.; Andrade, T.K. & Davis Jr., C.A. (2014) *Geovisualization of The Academic Trajectories of Brazilian Researchers*. GeoInfo 2014, pp. 83-94.
- Gahegan, M.; Wachowicz, M.; Harrower, M. & Rhyne, T. (2001) *The Integration of Geographic Visualization with Knowledge Discovery in Databases and Geocomputation*. Cartography and Geographic Information Science. 28(1), pp 29-44.
- Ganev, V.; Guo, Z.; Serrano, D.; Tansey, B.; Barbosa, D. & Stroulia, E. (2009) *An environment for building, exploring and querying academic social networks*. MEDES '09 Proceedings of the International Conference on Management of Emergent Digital EcoSystems, pp.282-289.
- Gaughan, M. & Bozeman, B. (2002) *Using curriculum vitae to compare some impacts of NSF research grants with research Center funding*. Research Evaluation 11(1):17-26.
- Goldberg, D. W.; Wilson, J.P. & Knoblock, C.A. (2007) *From text to geographic coordinates: the current state of geocoding*. URISA Journal 19:33-46.
- Goldberg, D. W. (2011) *Advances in geocoding research and practice*. Transactions in GIS 15:727-733.
- Heitor, M.; Horta, H. & Mendonça, J. (2014) *Developing human capital and research capacity: Science policies promoting brain gain*. Technological Forecasting & Social Change 82:17.
- IDEA Consult. (2010) *Study on mobility patterns and career paths of EU researchers*. Brussels, Belgium: European Commission, Research Directorate-General.

- Jorge, A. L. & Porfírio, F. P. (2010) *Building an Academic Social Network for Bologna Mobility*. SNS '10 Proceedings of the 3rd Workshop on Social Network Systems, ISBN: 978-1-4503-0080-3.
- Laender, A. H. F.; Moro, M. M.; Silva, A. S.; Davis Jr., C. A.; Gonçalves, M. A.; Galante, R.; Silva, A. J. C.; Bigonha, C. A. S.; Dalip, D. H.; Barbosa, E. M.; Borges, E. N.; Cortez, E.; Procópio Jr., P.; Alencar, R. O.; Cardoso, T. N. C. & Salles, T. (2011) *Ciência Brasil - the Brazilian portal of science and technology*. Seminário Integrado de Software e Hardware.
- Leydesdorff, L. & Persson, O. (2010) *Mapping the Geography of Science: Distribution Patterns and Networks of Relations among Cities and Institutes*. ArXiv e-prints.
- Lopes, G. R.; Moro, M. M.; Wives, L. K. & Oliveira, J. P. M. (2010) *Collaboration Recommendation on Academic Social Networks*. Advances in Conceptual Modeling – Applications and Challenges 6413:190-199.
- Machado, I. M. R.; Alencar, R. O.; Campos Jr., R. O. & Davis Jr, C. A. (2011) *An ontological gazetter and its application for place name disambiguation in text*. Journal of the Brazilian Computer Society 17: 267-279.
- Monmonier, M. (1996) *How to Lie with Maps*. University Of Chicago Press, 2^a edição. ISBN: 978-0226534213
- Morzinski, J.A. & Schubot, D. B. (2000) *Evaluating Faculty Development Outcomes by Using Curriculum Vitae Analysis*. Fam Med. 32(3):185-9.

- Oh, J. S. & Jeng, W. (2011) *Groups in Academic Social Networking Services: An Exploration of Their Potential as a Platform for Multi-disciplinary Collaboration*. Privacy, Security, Risk and Trust (PASSAT), IEEE Third International Conference. ISBN: 978-1-4577-1931-8.
- Mena-Chalco, J. P. & Cesar Jr., R. M. (2009) *scriptLattes: an open-source knowledge extraction system from the Lattes platform*. Journal of the Brazilian Computer Society 15(4):31-39.
- Padilha, L. E. (2008) *How has Mexican faculty been trained? A national perspective and a case study*. Higher Education: The International Journal of Higher Education and Educational Planning, 56(2):167-183.
- Schich, M.; Song, C.; Ahn, Y.; Mirsky, A.; Martino, M.; Barabási, A. & Helbing, D. (2014). *A network framework of cultural history*. Science 345(6196):558-562.
- Schuab, M. (2011) *The Career Paths of Doctoral Graduates in Austria*. European Journal of Education European Journal of Education 46(1):153–168.
- Shneiderman, B. (1996) *The eyes have it: A task by data type taxonomy for information visualizations*. VL'96: Proceedings of the 1996 IEEE Symposium on Visual Languages, pp. 336-343.
- Song, H. (1997) *From Brain Drain to Reverse Brain Drain: Three Decades of Korean Experience*. Science Technology & Society 2: 317-345.
- Souza, P. N. P. (1991) *Estrutura e funcionamento do ensino superior brasileiro*. Editora Pioneira Ciências Sociais.

Spaccapietra, S.; Parent, C.; Damiani, M. L.; de Macedo, J. A.; Porto, F. & Vangenot, C. (2007) *A conceptual view on trajectories*. Data & Knowledge Engineering 65(1):126-146.

Teixeira, R. K. C.; Gonçalves, T. B. & Botelho, N. M. (2012) *Análise quantitativa de pós-graduandos em ciências de saúde no Brasil: perfil por estados*. UNOPAR Científica Ciências Biológicas e da Saúde 14(3).

Tominski, C.; Schumann, H.; Andrienko, G. & Andrienko, N. (2012) *Stacking-Based Visualization of Trajectory Attribute Data*. Visualization and Computer Graphics 18(12):2565-2574.

Van Bouwel, L. A. C. (2010) *International mobility patterns of researchers and their determinants*. “Sumer” Conference 2010 on Opening Up Innovation: Strategy, Organization and Technology, pp. 1-26.