

# Geovisualization of the Academic Trajectories of Brazilian Researchers

Caio Alves Furtado, Thamara Karen Andrade, Clodoveu A. Davis Jr.

Departamento de Ciência da Computação – Universidade Federal de Minas Gerais  
(UFMG) – Belo Horizonte – MG – Brazil

{caioaf, thamara, clodoveu}@dcc.ufmg.br

**Abstract.** *People seeking academic careers usually pursue degrees in various institutions, looking for the best combinations regarding their intellectual interests, their personal means and the opportunities that arise in their lives. This can be perceived by looking at the stages in the education of current researchers, as recorded in their curricula vitae. We define the academic trajectories of researchers as the sequence of degrees obtained by a researcher, culminating with a work position. After geolocating academic and work institutions, and using the time period associated with each degree in the vitae data, the trajectories become spatiotemporal objects. We present an analysis of the academic trajectories of nearly 6,000 researchers associated with CNPq's National Institutes of Science and Technology program, based on their curriculum vitae data. An interactive visualization tool has been developed. Analyses include the variety of trajectories, preferred destinations, variations along time and different patterns related to research area.*

## 1 Introduction

Lattes<sup>1</sup> is a Web-based application created by CNPq (the Brazilian National Council for Scientific and Technological Development) to collect and integrate a wide range of information about the Brazilian academic community. One of its services is *Currículo Lattes*, a curriculum vitae Web system for researchers and students. All information in Lattes is publicly available, and currently covers the vast majority of active Brazilian researchers, groups, and institutions. Lattes is a rich database on Brazilian scientific research, built from the perspective of the individuals involved in it.

CNPq, along with CAPES (the Brazilian Ministry of Education's organization for graduate courses and curricula) and regional foundations, also created a program to foster and promote Brazilian research groups, called *Institutos Nacionais de Ciência e Tecnologia*<sup>2</sup> (National Institutes of Science and Technology, or INCT). The program created 101 institutes, covering thematic areas deemed strategically important for the country. While the INCT program does not cover every research group in the country, it includes many of the best Brazilian research groups.

CiênciaBrasil<sup>3</sup> is one of the research projects conducted within one of the INCTs. The project built a portal that uses Lattes curriculum vitae (CV) data to

---

<sup>1</sup> Lattes: <http://lattes.cnpq.br>

<sup>2</sup> INCT: [http://estatico.cnpq.br/programas/inct/\\_apresentacao/apresentacao.html](http://estatico.cnpq.br/programas/inct/_apresentacao/apresentacao.html)

<sup>3</sup> CiênciaBrasil: <http://pbct.inweb.org.br>

configure a research-based social network, in which relationships are characterized by collaboration in publications. This allows CiênciaBrasil to study many aspects of the researchers' careers and activities, seeking patterns from practices and behaviors captured from scientific production and collaboration. Currently, the portal includes CVs from INCT researchers that hold a Ph.D., but contents are constantly expanding.

Our work involves studying and characterizing patterns of placement and mobility of Brazilian researchers through time. We characterize events in the researchers' careers as points in space and in time, thus defining *academic trajectories*, comparable to trajectories of moving objects. In both cases trajectories are described by a set of positions collected through a time period, but the average academic trajectory has fewer positions and spans a much larger time interval, covering years. In this paper, we explore alternatives for using geographic visualization techniques to provide tools for exploratory analysis of such academic trajectory data.

This paper is organized as follows. Section 2 discusses some related work. Section 3 describes our dataset and the steps taken in its preparation. Section 4 presents several analyses, with a variety of visualization alternatives, exploring multiple aspects of the problem. Finally, Section 5 presents a final discussion and indicates future work.

## **2 Related Works**

This paper explores the spatiotemporal academic trajectories of INCT researchers, including stages such as graduation, graduate studies, employment in a university or research institution, and aims to discover and discuss the patterns in such trajectories. We focus in understanding the diversity in the researchers' formation, analyzing the trajectory patterns and their changes through time.

CiênciaBrasil is an INCT project that studies Brazilian researchers and research groups (Laender, Moro et al. 2011). It uses Lattes CV data and focuses on each researcher's publications and collaborations, which define a co-authorship network. This work uses CiênciaBrasil data, and explores the researchers' academic historic and their spatio-temporal aspect, also expanding the analysis to institutions.

Some works study the mobility of researchers. Van Bouwel (2010) discusses the international mobility decisions of researchers that obtained a Ph.D. in Economics in the United States. The report verifies the final work destinations of researchers after getting their Ph.D.s and explores the factors that drive their movement, showing that about 50% of them remain in the USA after getting their degrees, while only one third of the remaining half returns to their home country. A report from the European Commission (IDEA Consult 2010) details the mobility patterns of European researchers and discusses the influencing factors and impacting effects of their mobility. Our study focuses a smaller group, most of which employed in Brazil, but with a variety of academic trajectories up to their participation in an INCT.

Schich et al. (2014) developed a study on the mobility of a large number of people through history. Gathering birth and death dates and places from FreeBase biographies and other sources, the authors reconstructed simple two-point trajectories of more than 150,000 notable individuals. Visualization techniques, including animation, were used to show patterns in each period in human history, by identifying sources and destinations of the lives of so many people.

Câmara et al. (2004) consider that the emphasis of spatial data analysis should be on measuring properties and relationships, explicitly considering the spatial location as part of the studied phenomenon. In our analysis, we concentrate on the institutions in which the current INCT researchers studied and worked, and show how the trajectory between them can be influenced by certain aspects of the researcher’s CV. Visualization tools are often used as part of spatial data analysis, especially in the exploratory phase. Interactive scientific data visualization, as presented in this work, provides means to discover data properties (Haining 2003).

An academic trajectory is similar to a mobile object trajectory (Etienne, Devogele et al. 2012). The trajectory can be defined as a set of positions that correspond to places where the researcher obtained a degree in his education, associated with a time period. These positions are composed of a timestamp and a spatial coordinate, corresponding to the location of the institution at which the degree was obtained, and the chronological ordering of the positions forms the trajectory. The temporal granularity of an academic is measured in years, and time periods potentially cover several decades. A researcher will rarely have more than five positions in his trajectory.

Spaccapietra et al (Spaccapietra, Parent et al. 2008) discuss the semantics of trajectories in more detail. They describe three types of trajectories: *metaphorical* (steps in an evolutionary process), *naïve geographical* (major geographic points along a path), or *spatiotemporal* (spatial coordinates associated to timestamps). According to this classification, academic trajectories are metaphorical, since stops correspond to stages in a researcher’s education, but naïve geographical concepts appear when analyzing the location of the affiliation institutions along the career. Furthermore, Spaccapietra et al. define trajectories as collections of stops and moves. In the case of academic trajectories as defined here, we are mostly interested in the moves.

### 3 Dataset

Our dataset is composed of a set of researchers’ Lattes CVs, or profiles. The profiles are publicly available for download as XML files, and were collected from December 2012 through August 2013 as part of the CiênciaBrasil project. For now, the CiênciaBrasil portal covers only researchers associated with an INCT. Researchers are required to have and to keep up-to-date their vitae in Lattes, as a precondition to applying for grants and other forms of financing. Considering that 94.7% of the profiles were modified between January 2012 and August 2013, we consider the information provided by the profiles is correct and current enough for our studies.

The Lattes profile has academic information on the user, such as his name, workplace, and a list with his graduation degrees. Each degree has an institution associated with it, indicating where said researcher concluded his graduation, and the starting and ending years. We consider the set of institutions attended by a person, along with the time periods in which she attended, to be his *academic trajectory*. The trajectory is chronologically ordered. The expected trajectory is *bachelor* → *master* → *doctorate* → *pos-doctorate*, but this order isn’t always verified. Every researcher in the database has at least one Ph.D. degree, but the number of degrees and the order in which he completed them may change. A researcher can have two doctorates and no pos-doctorate, or can have his doctorate course right after his bachelor, and get a master’s degree after that, as indicated by the years associated with each degree.

Figure 1 shows a simplified schema for the data, after collection and geocoding. The block corresponding to each degree in the researcher's data is multivalued. Additional data on the researcher can be obtained using the Lattes ID, since the entire contents of the CV are stored elsewhere in CiênciaBrasil. INCTs are classified into 8 research areas, defined by CNPq: agriculture, environment, energy, exact sciences, humanities, nanotechnology, health/medical, and engineering/information technology (IT). Researchers can be associated with more than one INCT.

The database is composed of 5,973 researchers, all of which associated with at least one INCT. There are 3,478 unique institutions, including academic institutions and employment institutions. From these, 2,858 institutions appear exclusively as places of education, 399 institutions that appear exclusively as employers, and 221 institutions that are both academic and employers.

```
Researcher(LattesID, [degreeType, institutionID,
startYear, endYear], workInstitutionID)

Institution(institutionID, institutionName, location)

INCT(inctID, inctName, inctArea)

INCTResearchers(inctID, LattesID)
```

**Figure 1. Data schema**

The distribution of researchers throughout the institutions is uneven. More than 50% of the researchers work in only 15 of the 620 employment institutions. This behavior is also found in academic entries, in a characteristic long-tail distribution. Analyzing individual trajectory segments, there are 21,092 segments for the 5,973 researchers, about 3.5 segments per researcher. Of those, 7,930 segments connect an institution to itself, thus representing a continuation of studies in the same institution.

To materialize the academic trajectory as a spatiotemporal object, the institutions need to be geographically located, but Lattes does not provide the institutions' addresses or coordinates. We geocoded the institutions based on their names, as supplied in Lattes. The institution's name is processed as given by the user, thus errors and typos may occur. In the case of institutions with multiple campuses, Lattes usually does not define a specific one. Therefore, we considered the institution's main campus as a default. We selected a set of institutions to check the geocoding process. We picked every institution with at least 5 researchers associated with it (84 institutions), corresponding to 80.59% (4,878) of the recorded number of PhD degrees, and manually checked their location, correcting it when needed. Only three institutions could not be geocoded, due to faulty information, corresponding to 61 researchers.

## 4 Analysis

We created an interactive Web map application<sup>4</sup> to show researcher trajectories. We used a Google Maps layer as a background, and a Javascript module to draw the trajectories. There are filters on some dimensions, such as INCT area, type of degree, and trajectory segment final year. Some controls on the appearance of trajectories are also available, such as varying the thickness of lines according to the number of

---

<sup>4</sup> <http://aqui.io/trajectory/>

equivalent segments, and hiding segments that connect Brazil and foreign institutions. In the next subsections, we show steps in an exploratory analysis of the described dataset, using geovisualization in an interactive environment.

#### 4.1 Individual trajectory

The trajectory of any given individual can be shown in the interactive map, by supplying the researcher's Lattes ID. Figure 2 shows an example, in which the researcher starter his career at Fortaleza, in the Brazilian northeast, where he received his bachelor's degree. Then he moved to Campinas, to get a master's degree, next to the USA, for a Ph.D. Getting back to Brazil, he had a post-doctorate stage at Belo Horizonte, after which he was hired also in Belo Horizonte as a professor in a university. The last segment is a null link, i.e., a new stage in the researcher's career, but at the same location.

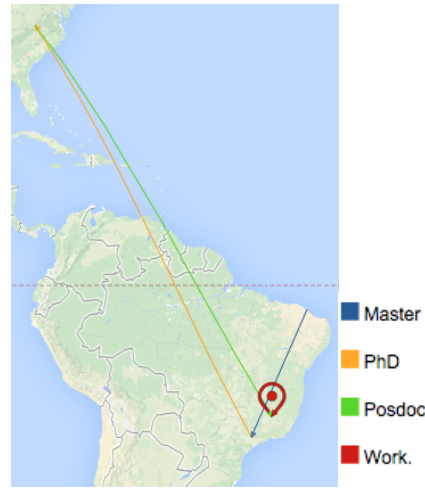


Figure 2. An individual trajectory

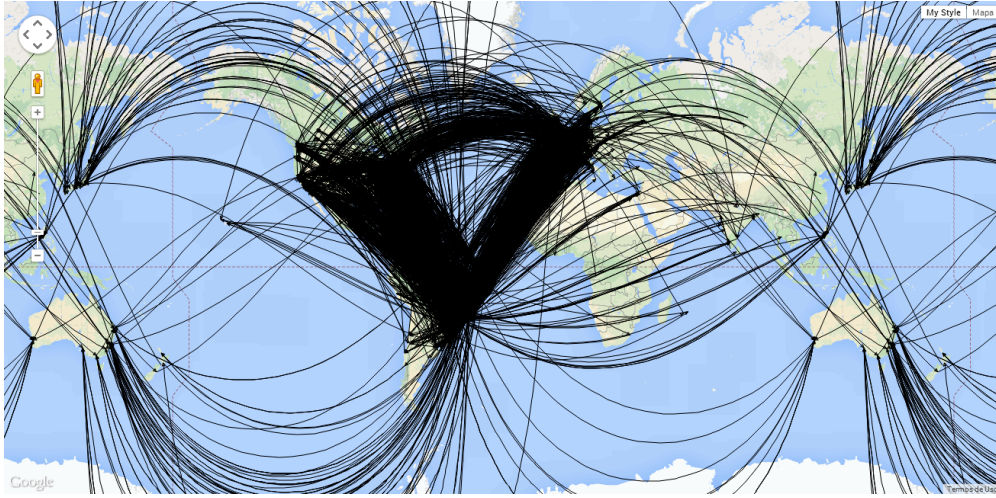
#### 4.2 Frequent destinations

Figure 3 shows all trajectories simultaneously. It is apparent that most trajectories link Brazil to the United States or Europe, but most links are null.

Table 1 shows the total number of segments grouped into the most frequent origins and destinations. Naturally, most destinations are within Brazil, but the number of segments that involve destinations in the United States or Europe is expressive, while segments including Africa, Asia or Oceania are infrequent.

Table 1. Start and end points of trajectory segments

Start \ End	Brazil	USA	Europe	Other
Brazil	14,373	1,026	1,282	424
USA	976	385	134	47
Europe	1,244	83	579	50
Other	254	41	57	137

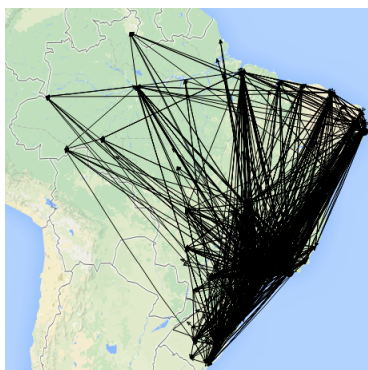


**Figure 3. Existing trajectories worldwide**

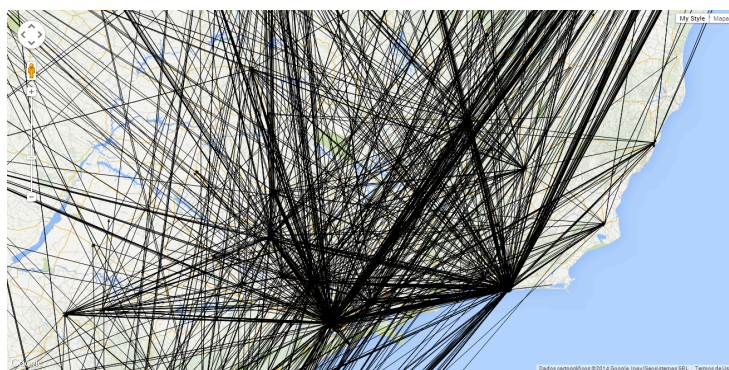
Figure 4 shows trajectory segments with both ends in Brazilian institutions, indicating successive stages in the researcher's education that were carried out in Brazilian institutions. Notice (Figure 4b) the concentration of trajectory segments in southeastern institutions, especially in São Paulo and Rio de Janeiro. Table 2 shows that these states, plus Minas Gerais, Rio Grande do Sul and Pernambuco, concentrate 78% of the destinations within Brazil. Notice also that most segments start and end in the same state. Next section explores this tendency further.

**Table 2. Most common start and end points within Brazil**

Start \ End	SP	RJ	MG	RS	PE	Others
São Paulo (SP)	4,553	169	163	57	56	620
Rio de Janeiro (RJ)	227	2,070	54	21	13	171
Minas Gerais (MG)	225	102	1,035	17	8	147
Rio Grande do Sul (RS)	144	45	14	883	9	120
Pernambuco (PE)	89	30	10	9	292	69
Others	612	184	98	73	77	1,907



**(a)**



**(b)**

**Figure 4. Existing trajectories (a) in Brazil and (b) Southeast region close-up**

### 4.3 Null trajectory segments

Many researchers fulfill consecutive stages in their education at the same institution. As a result, there are many null trajectory segments. Figure 5 shows concentrations of null segments as graduated symbols in a map, focusing on the Brazilian south and southeast

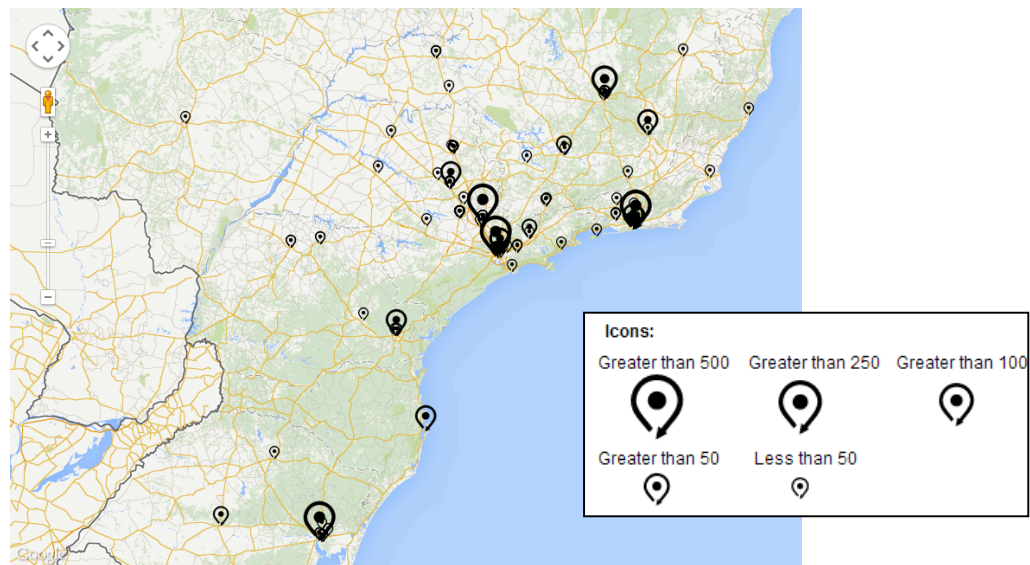


regions. Naturally, the sites of major Brazilian universities concentrate null segments, in cities such as São Paulo, Rio de Janeiro, Porto Alegre, Belo Horizonte and Campinas. This result does not imply that the researchers tend to work close to their origins; an analysis on that is presented in Section 4.7.

In the next subsection, we show the classification of segments according to the educational stages in the researcher's trajectory.

#### 4.4 Education degree

Looking now at the spatial distribution of segments according to education degree, we observe that the highest incidence of stages outside Brazil occurs in post-doctorate work (Table 3). Naturally, as the level of specialization increases, the more people go abroad for their studies. This trend is shown in Figure 6. For the master's degree, the destinations concentrate in the American East coast and in Western Europe. Ph.D.-related destinations are more varied, and more institutions serve as destination. In the INCT database, doctoral studies abroad are 3.6 times more frequent than master's degrees obtained abroad. Post-doctoral trajectories are the most internationalized, and the map shows that post-doctoral stages are held in an even greater variety of institutions. The number of post-docs held abroad is about twice the number of Ph.D.s obtained abroad. Post-docs in Brazil comprise 38% of the total, whereas almost 80% of the Ph.D. work is conducted in Brazilian institutions. This is to be expected, since post-doctoral stages are meant as an opportunity for the researcher to expand on his cooperation network and to interact with top research groups in his area, wherever they are located. The high concentration of Ph.D. degrees obtained in Brazil shows that Brazilian institutions are capable of educating most of the country's researchers, at least from the INCT point of view.



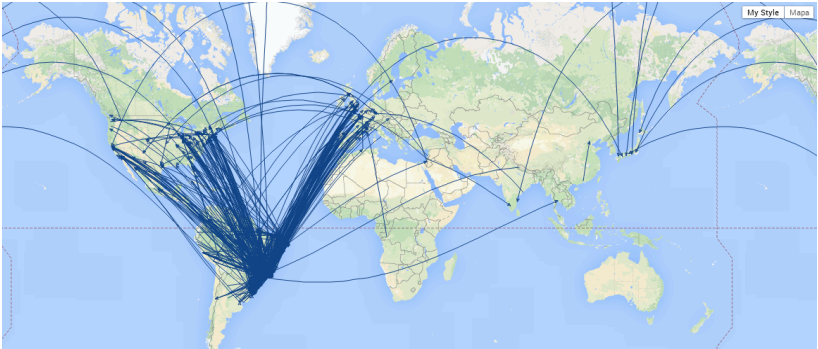
**Figure 5. Null trajectory segments**

Within Brazil, similar patterns can be seen. Figure 7 shows displacements for master's, Ph.D. and post-doc stages, indicating a concentration of destinations in southeastern institutions. These maps concentrate all links to or from a state in the state's centroid, so that the most important flows can be perceived. Notice the important

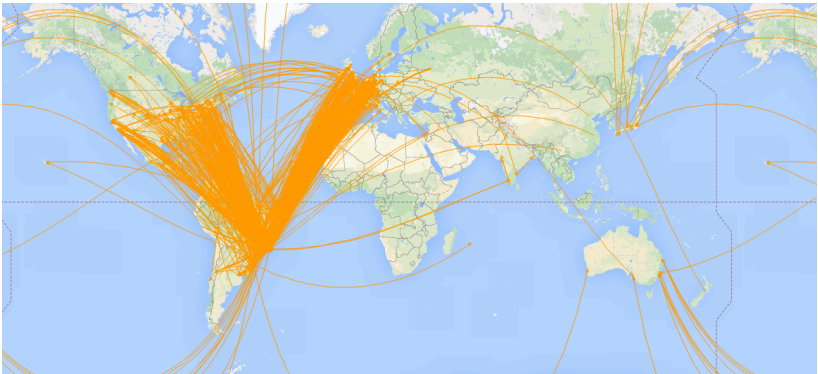
link in the master's degree map, towards Pará. These trajectories indicate a concentrated interest for the environmental researchers.

**Table 3. Distribution of segment destinations per degree**

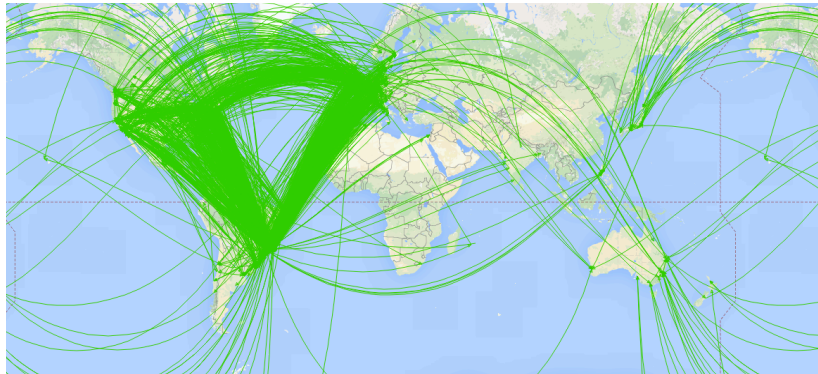
	Brazil	USA	Europe	Other
<b>Bachelors</b>	514	4	22	14
<b>Masters</b>	4,784	117	139	68
<b>Ph.D.</b>	4,570	351	697	119
<b>Post-doc</b>	1,486	1,056	1,081	279



**(a)**



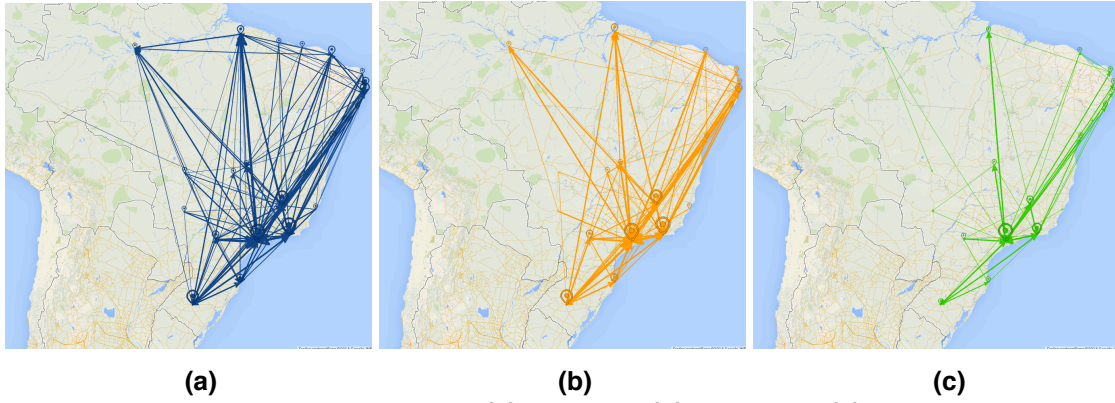
**(b)**



**(c)**

**Figure 6. Trajectory segments for (a) master's, (b) Ph.D., and (c) post-doc degrees**

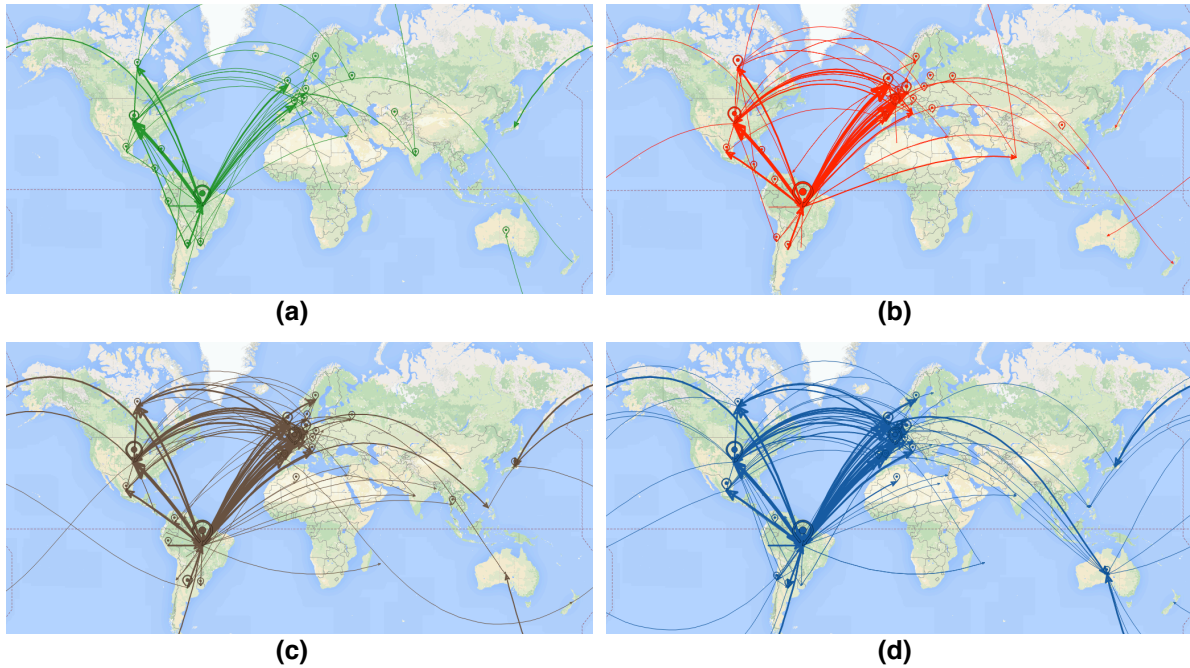




**Figure 7. Trajectory segments for (a) master's, (b) Ph.D., and (c) post-doc in Brazil**

#### 4.5 Variations along time

Figure 8 shows the mobility filtered by decades. In these maps, all incoming or outgoing links for a country are concentrated in a single node within that country, and indicating the number of coincident trajectories using line thickness. Icons and icon sizes indicate the number of null trajectory segments. We observe that, although the general pattern apparently does not change along time (i.e., the attractive destinations continue to be attractive), the intensity (reflected in the thickness of the lines) varies significantly, reflecting the growth in the number of researchers along time (Table 4). The decades from 1980 to 2000 concentrate most of the INCT researchers. Researchers that studied in the 1970s are scarcer, since many of them would be near the age for retirement by now.



**Figure 8. Trajectory segments along time: (a) 1970s, (b) 1980s, (c) 1990s, (d) 2000s**

We can see an evolution in the number of links leading to Europe. The trajectories involving countries in Asia and Oceania also increases, but that can be

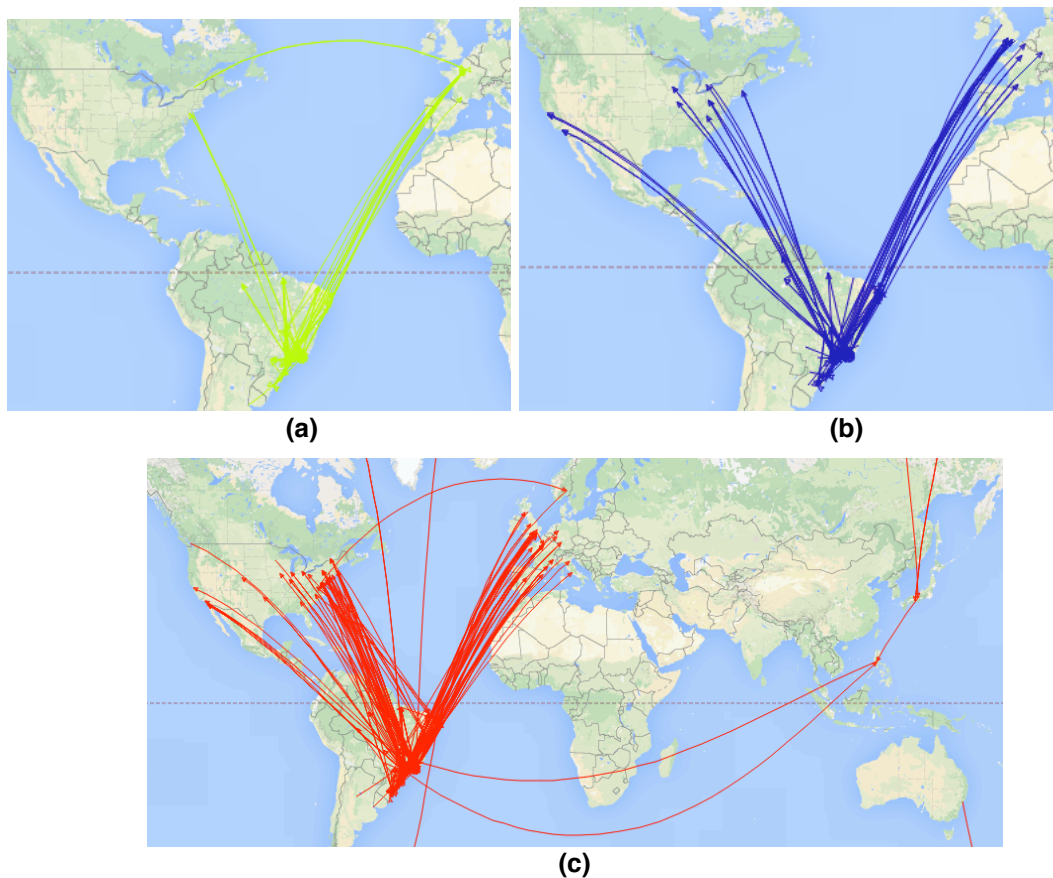
associated with the increase in the number of researchers. The trajectories between USA and Europe appear throughout all the period, and show that this behavior is not recent, even though most links are related to post-doctoral work.

**Table 4. Number of trajectory segments per decade (end year)**

Decade	# Segments
1950	12
1960	119
1970	871
1980	2338
1990	5205
2000	6004
2010	752

#### 4.6 Variations according to INCT research area

The visualization tool can also filter trajectories based on the research area associated with each INCT. This allows us to verify the preferred destinations for academic degrees according to the field of knowledge. Figure 9 shows three peculiar areas in this respect. With maps that exclude trajectories traversed by a single researcher. Humanities researchers (Figure 9a) seem to prefer European destinations, France in particular. Engineering and IT (Figure 9b) seek the USA and Europe, with a major drive towards London. Health/medical researchers (Figure 9c) have a wider variety of destinations, with an emphasis on the USA's east coast, various locations throughout Europe and a number of Asian destinations.



**Figure 9. International destinations according to INCT research area:**  
**(a) humanities, (b) engineering/IT, and (c) health/medical**

#### 4.7 Start and end points of trajectories

Table 5 shows mobility data among Brazilian states. Column *From* indicates the number of researchers whose first trajectory node is in that state. *i.e.*, the state in which they obtained their first degree, which we assume to be close to home. Column *Stay* indicates how many of the first group ended up working in the same state. Columns *Out* and *In* respectively indicate the number of researchers that have gone to work elsewhere, and the number of researchers that move in to the state. The final column, *Work total*, indicates the number of INCT researchers currently employed by institutions in that state. Overall, 41% of the researchers end up working in a state that is different from the one in which they obtained their bachelor's degree. Table 5 also shows that some states notably attract researchers whose trajectories start elsewhere, especially Mato Grosso do Sul (91% of the researchers come from somewhere else), Sergipe (86%), and the Federal District (73%), excluding the states with a very small number of researchers. Other states are notable for exporting their graduates for work in other states, as in the case of Espírito Santo (69% of the graduates of Espírito Santo work in other states), Federal District (63%) and Piauí (56%). On the other hand, relatively few researchers from Amazonas, Pará and Rio de Janeiro leave the state (16%, 26% and 27%, respectively).

**Table 5. Mobility of researchers among Brazilian states**

State	From	Stay	Out	In	Total
Acre	2	1	1	8	9
Alagoas	18	12	6	27	39
Amapá	-	-	-	5	5
Amazonas	50	42	8	106	148
Bahia	142	91	51	86	177
Ceará	159	110	49	40	150
Distrito Federal	117	43	74	115	158
Espírito Santo	29	9	20	11	20
Goiás	52	23	29	34	57
Maranhão	17	8	9	10	18
Mato Grosso	30	17	13	39	56
Mato Grosso do Sul	25	11	14	107	118
Minas Gerais	654	389	265	178	567
Pará	179	133	46	174	307
Paraíba	134	65	69	44	109
Paraná	259	126	133	102	228
Pernambuco	231	125	106	80	205
Piauí	18	8	10	10	18
Rio de Janeiro	949	692	257	269	961
Rio Grande do Norte	94	58	36	71	129
Rio Grande do Sul	540	314	226	67	381
Rondônia	1	-	1	16	16
Roraima	2	2	-	16	18
Santa Catarina	118	68	50	123	191
São Paulo	1,588	1,004	584	587	1,591
Sergipe	15	7	8	42	49
Tocantins	1	1	-	1	2
Outside Brazil	367	9	358	55	64
<b>Total</b>	<b>5,791</b>	<b>3,368</b>	<b>2,423</b>	<b>2,423</b>	<b>5,791</b>

## 5 Discussion and Future Work

This work shows visualization techniques applied to academic mobility data, extracted from the records of the researchers' Lattes CVs. The data source is highly reliable, being supplied and curated by the researchers themselves, but the process of obtaining and preparing such data for visual analysis is not yet straightforward. We created a Web application in which users can interactively manipulate visualization parameters, data

filters and other controls to analyze the mobility of researchers according to many different perspectives, from the individual trajectory to regional grouping.

In our visually-enhanced exploratory analysis of the data, we observed the prevalence of Brazilian destinations in the various steps of the researchers' education, and showed preferences as to international stages. Regarding the degree of education, we observed a tendency towards seeking foreign positions in the latter stages, especially for post-doc. Brazilian institutions handle most of the degrees, with some mobility between states and a concentration in the Brazilian southeast region, especially in São Paulo state. We found distinct mobility pattern variations between the various research areas, and further study could help determining more about the preferred institutions. We also observed concentrations of researchers of a given research area in some regions of the country – the most notable example is the environmental field, in which many researchers move to institutions closer to the Amazon. In the time-based analysis, we noticed that the prevalence of education degrees within Brazil is growing, but the participation of international stages is important.

Future work includes expanding the database to a larger and more variety set of researchers, and improving the visualization tool as to increase the variety of visual parameters that can be controlled by the user. We also emphasize the possibility of using such tools for other kinds of data, such as students in the Brazilian government's Science without Borders program, or migration demographics from the decennial census or from the Ministry of Education's student census.

## Acknowledgments

Authors acknowledge the support of CNPq (308678/2012-5), FAPEMIG (CEX-PPM-00518/13), and CAPES, Brazilian agencies in charge of fostering research initiatives.

## References

- Câmara, G., A. M. V. Monteiro, S. Druck and M. S. Carvalho (2004). Análise espacial e geoprocessamento. Análise espacial de dados geográficos. S. Druck, M. S. Carvalho, G. Câmara and A. M. V. Monteiro. Brasília (DF), EMBRAPA.
- Etienne, L., T. Devogele and A. Boujou (2012). Spatio-temporal analysis of mobile objects following the same itinerary. Advances in Geo-Spatial Information Science, CRC Press. **10**: 47-57.
- Haining, R. (2003). Spatial data analysis: theory and practice, Cambridge University Press.
- IDEA Consult (2010). Study on mobility patterns and career paths of EU researchers. Brussels, Belgium, European Commission, Research Directorate-General.
- Laender, A. H. F., M. M. Moro, A. S. Silva, C. A. Davis Jr, M. A. Gonçalves, R. Galante, A. J. C. Silva, C. A. S. Bigonha, D. H. Dalip, E. M. Barbosa, E. N. Borges, E. Cortez, P. Procópio Jr., R. O. Alencar, T. N. C. Cardoso and T. Salles (2011). CiênciaBrasil - the Brazilian portal of science and technology. Seminário Integrado de Software e Hardware (SEMISH). Natal (RN), Brazil, Sociedade Brasileira de Computação (SBC): 1366-1379.
- Schich, M., Song, C., Ahn, Y.-Y., Mirsky, A., Martino, M., Barabási, A.-L. Helbing, D. (2014). "A network framework of cultural history." Science **345**(6196): 558-562.
- Spaccapietra, S., C. Parent, M. L. Damiani, J. A. Macedo, F. Porto and C. Vangenot (2008). "A conceptual view on trajectories." Data & Knowledge Engineering **65**(2008): 126-146.
- 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. Imperial College London Business School: 1-26.