Journal of the Brazilian Computer Society, 2022, doi: 10.5753/jbcs.2022.2887

A Geographic Analysis of the Interdisciplinary Collaborations in the Brazilian Scientific Community

Geraldo J. Pessoa Junior [**Universidade Federal de Viçosa** | ***geraldo.pessoa@ufv.br***]
Thiago M. R. Dias [**Centro Federal de Educação Tecnológica de Minas Gerais** | ***thiago@div.cefetmg.br***]
Thiago H. P. Silva [**Universidade Tecnológica Federal do Paraná** | ***thiagosilva@utfpr.edu.br***]
Alberto H. F. Laender [**Universidade Federal de Minas Gerais** | ***laender@dcc.ufmg.br***]

Abstract
Interdisciplinary collaborations have recently attracted the attention of scholars, since they help bridging academic relationships and contribute to make scientific collaboration networks even stronger. However, previous works on this subject have mainly focused on characterizing such interdisciplinary collaborations in specific research groups or scientific communities. In this article, we start from a previous work in which we characterized the interdisciplinary collaborations within the entire Brazilian scientific community, as defined according to the upper level of the knowledge area classification scheme proposed by CNPq, the Brazilian National Council for Scientific and Technological Development, considering the following eight major areas: *Agrarian Sciences*, *Applied and Social Sciences*, *Biological Sciences*, *Engineering*, *Exact and Earth Sciences*, *Health Sciences*, *Humanities*, and *Linguistics, Letters and Arts.* Based on this interdisciplinary collaboration network, we conducted a geographic analysis that characterizes how these collaborations have been spread across the Brazilian geographic regions. Overall, our results show strong collaborative ties involving the triad formed by the three main Brazilian geographic regions (Southeast, South and Northeast) for all major areas. Besides, three of the eight major areas (*Agrarian Sciences*, *Biological Sciences*, and *Health Sciences*) show a massive participation in interdisciplinary collaborations across all regions. Despite that, geographic proximity is an important factor, since the proportion of interdisciplinary collaborations involving researchers from the same region is high. Finally, we analyze the patterns of interdisciplinary collaboration by regions and by major areas, thus showing that the Brazilian interdisciplinary network is highly connected.

Keywords: *Interdisciplinary Collaborations, Coauthorship Networks, Scientific Communities, Geographic Analysis, Lattes Platform*

1 Introduction

In recent years, interdisciplinary collaborations have attracted the attention of the scientific community as a means to aggregate knowledge from different domains to better understand and address real problems. For instance, in a pioneering effort researchers from different knowledge fields and associated with institutions from the nine states of Brazil's Northeast region team up together with the aim to promote scientific solutions to attenuate the global spread of the COVID-19[1]. Coordinated by the physicist Sérgio M. Rezende and the neuroscientist Miguel Nicolelis, this effort is a reference in terms of interdisciplinary collaborations, since it involves researchers from distinct areas such as infectology, biodiversity, sociodemography, biomathematics, among others. It also emphasizes the importance of observing specific problems from distinct perspectives, as well as of applying new techniques already known from other contexts to solve them (Haythornthwaite, 2006). Another example that illustrates this kind of effort in a distinct scenario is the use of drones for smart agriculture automation (Kulbacki et al., 2018). Initially idealized for defense projects, this kind of technology aggregates distinct research efforts ranging from chemistry to machine learning.

Due to an enormous academic mobility and the rapid growth of the Internet, global scientific collaborations have become highly connected. As reinforced by the Royal Society (Wilsdon et al., 2011), interdisciplinary collaborations bring several benefits to researchers such as (i) the opportunity to apply new theories that allow them to view problems from a new perspective, (ii) access to funding available exclusively for interdisciplinary research and (iii) the expansion of their respective collaboration networks. Indeed, Adams et al. (2005) have shown that research teams from US universities that have collaborations with other departments, companies and foreign institutions tend to have higher productivity indicators, as well as that researchers who earned prestigious awards usually participate in large teams.

Previous studies regarding interdisciplinary collaborations in scientific coautorship networks (Abramo et al., 2018; Iglič et al., 2017; Jones et al., 2008; Mena-Chalco et al., 2014; Shi et al., 2018) have focused on external collaborations (e.g., cooperation among research groups (Freire and Figueiredo, 2011), knowledge exchange (Haythornthwaite, 2006), role of geography (Hoekman et al., 2010; Pan et al., 2012), migration of researchers (Kato and Ando, 2013) and influence from distinct research areas (Lima et al., 2013)) as being an important factor in the evolution of scientific communities. In this article, instead of characterizing specific research groups or knowledge areas, we address a complete and robust scientific community. Particularly, we study the Brazilian scientific collaboration network based on the upper

[1]The Scientific Committee of the Northeast Consortium: `https://www.comitecientifico-ne.com.br/`.

level of the knowledge area classification scheme proposed by CNPq[2], the Brazilian National Council for Scientific and Technological Development, which considers the following eight major areas: *Agrarian Sciences*, *Applied and Social Sciences*, *Biological Sciences*, *Engineering*, *Exact and Earth Sciences*, *Health Sciences*, *Humanities*, and *Linguistics, Letters and Arts*.

In this regard, in a previous work (Pessoa Junior et al., 2019, 2020), we provided a detailed analysis of the interdisciplinary collaborations in the Brazilian scientific scenario. For this, we collected from the CNPq's Lattes Platform[3] data related to 263,264 Brazilian researchers holding a PhD degree and generated two academic social networks: one considering all collaborations involving the researchers and another one considering only interdisciplinary collaborations among them. As our main contribution in that work, we investigated the intensity of the interdisciplinary collaborations across each one of the major areas considered, thus emphasizing their role for strengthening the entire network. We also analyzed the interdisciplinary collaborations along the time by focusing on the researchers' academic age and on the temporal evolution of the collaboration rates per major area (Pessoa Junior et al., 2020).

In this article, we focus on a more specific analysis and consider the geographic dimension as an additional layer of complexity. Although communication technologies have diminished spatial distances, Pan et al. (2012) argue that geography remains an important factor that affects the dynamics of science. Specifically, we investigate how collaborative are the major research areas among the five Brazilian geographic regions. In summary, the main contribution of this article is a geographic analysis of the interdisciplinary collaborations in the Brazilian scientific community as defined by the eight major areas of the CNPq knowledge area classification scheme.

Thus, the remaining of this article is organized as follows. Firstly, Section 2 reviews the current literature, whereas Section 3 describes how we acquired and prepared the data we use in our analysis, thus providing some preliminary figures about them. Then, Section 4 presents a detailed discussion of the results of our analysis, highlighting the interdisciplinary collaboration patterns in the coauthorship networks by considering the eight major areas defined by the CNPq knowledge area classification scheme and the five Brazilian geographic regions. Finally, Section 5 presents our conclusions and provides some insights for future work.

2 Related work

Various studies have explored different features when characterizing academic collaborations. For example, Silva et al. (2017) addressed this issue by considering the following five scientometric dimensions: (i) *focus of the study* (e.g., individual scholars or research groups), (ii) *geographic coverage* (e.g., a country or a specific region), (iii) *adopted metrics* (e.g., citation counts and venues' quality), (iv) *scope of the analysis* (e.g., network properties, collaboration patterns or ranking assessment), and (v) *researchers' profile* (e.g., mentored students, career length or gender). Yet, Morillo et al. (2003) presented a bibliometric methodology that aimed to provide a general overview of several scientific disciplines, with special attention to their interrelation. In such a work, the authors established a tentative typology of disciplines and research areas according to their degree of interdisciplinarity. Here, we focus on an analysis of interdisciplinary collaborations, considering the geographic factor as an additional layer of complexity.

As discussed by Sonnenwald (2007), scientific collaboration is the interaction between scientists with the purpose of sharing activities in order to achieve common goals as a final result. In this context, complex network analyses have been carried out to understand the meaning of such collaborations in different scenarios (Garay et al., 2016; Hua and Haughton, 2012; Jiang et al., 2017; Liu et al., 2005; Mooney et al., 2013; Morillo et al., 2003). By analyzing successful collaborations, Wagner et al. (2002) listed motivating factors for understanding why researchers collaborate, such as to gain access to foreign structures, access information in real time and increase research creativity. In fact, by working with new collaborators, researchers can enhance the quality of their work (Adams et al., 2005; Jones et al., 2008; Silva et al., 2016). Specifically, with respect to the Brazilian computer science community, Silva et al. (2016) investigated how academic mobility affects productivity, showing that researchers with foreign ties tend to publish in more prestigious publication venues. Likewise, Kato and Ando (2013) concluded that international collaboration improves the overall research performance in Chemistry.

Regarding the geography of the scientific collaborations, Sidone et al. (2017) studied the geographic patterns of Brazilian scholarly publications and scientific collaborations over six triennia (1992–2009). They revealed that such publications and collaborations have an intense spatial heterogeneity across the country, which is concentrated in the Southeast and South regions. In addition, they showed that geographic proximity plays an important role in determining interregional collaboration in Brazil (e.g., an increase of 100 km between two researchers reduces the likelihood of collaboration by an average of 16%). Considering research areas, Mena-Chalco et al. (2014) carried out a pioneering study within the Brazilian scientific community. In their study, they characterized the eight major areas that we address in this article, but considering only topological metrics. As a result, they contributed to a better understanding of how Brazilian researchers collaborate to each other.

Similarly to us, other initiatives have analyzed the importance of interdisciplinarity. For example, Porter et al. (2008) report the results obtained by a team responsible for evaluating a program for supporting interdisciplinary research in the United States, the NAKFI[4]. In such a work, they proposed and tested two quantitative metrics, *integration* (dispersion among the areas of the publications cited by a research) and *specialization* (dissemination of the areas in which the re-

[2]This classification scheme is organized into four levels: major area (e.g., Exact and Earth Sciences), area (e.g., Computer Science), subarea (e.g., Theory of Computation) and specialty (e.g., Formal Languages and Automata). For more details, refer to de Siqueira et al. (2020).

[3]Lattes Platform: http://lattes.cnpq.br

[4]National Academies Keck Futures Initiative. More information at https://www.keckfutures.org.

search is published), to evaluate the research results of the NAKFI program. By using these metrics, they concluded that the degree of research interdisciplinarity (i.e., integration) in the evaluated program was surprisingly high and that specialization reflected the concentration of a researcher's publications on a specific topic or research area. On the other hand, Huang and Chang (2011) carried out a study to investigate interdisciplinary changes in the area of information science between 1978-2007 based on direct citations and coauthorships. For this, they used the Brillouin's index (Peet, 1975) to measure the degree of interdisciplinarity in that area for each year of the period of study. Their results show that the degree of interdisciplinarity in direct citations and coauthorships grew over the years, with coauthorships having a more evident growth rate than direct citations. Based on data from the Journal Citation Reports[5], Silva et al. (2013) also show quantitatively that science fields are becoming increasingly interdisciplinary.

More specifically, Abramo et al. (2018) conducted a study aiming at verifying the influence of multidisciplinarity among the researchers involved in a same research project and how scientific production can be affected by this type of collaboration. In their study, they analyzed the publications of all Italian faculty members during a period of five years, from 2004 to 2008. Like us, they also adopted a classification scheme and stratified the Italian researchers according to their specialties in eight major areas of knowledge (Biology, Earth Sciences and Space, Engineering, Physics, Mathematics, Clinical Medicine, Biomedical Research and Chemistry). The results obtained in such a study show that scientific publications on specific topics are produced by researchers from the same knowledge area, while the most diversified ones come from multidisciplinary groups. Another aspect addressed by the authors is how diverse a publication is, since the greater the number of areas involved, the greater the diversity of that specific scientific production. In another work, Shi et al. (2018) carried out a study aimed at evaluating whether the degree of reputation of a scientific institution influences interdisciplinarity in the knowledge production process. To measure the prestige of an institution, they took into consideration its scientific production. A metric based on the IDR (Irreproducible Discovery Rate) approach (Li et al., 2011), combined with the *P-rank* indicator (Yan et al., 2011), was used to analyze the level of importance of an institution's scientific output. The results showed that more prestigious institutions provide a greater interdisciplinarity to the knowledge flow.

Finally, looking closer at regional aspects, Chiarini et al. (2014) investigated bibliographic production by Brazilian states from 2000 to 2010, considering a total of 147,638 researchers registered in the CNPq Directory of Research Groups. They identified three states in the southeast region (São Paulo, Rio de Janeiro and Minas Gerais) and one in the south region (Rio Grande do Sul) as those with the highest productivity. In addition, they showed that the researchers in these states focused mainly on the major areas of *Agrarian Sciences*, *Biological Sciences* and *Health Sciences*.

[5]`https://clarivate.com/webofsciencegroup/solutions/journal-citation-reports`

3 Background

In this section, we first introduce our dataset, which was used in our previous work to characterize the interdisciplinary collaborations in the Brazilian scientific community (Pessoa Junior et al., 2019, 2020). Then, we present an overview of such interdisciplinary collaborations, upon which we carry out our geographic analysis.

3.1 Our dataset

Our dataset covers 263,264 Brazilian researchers and more than 10 million publications. In addition, the researchers' expertise are identified in terms of the eight major areas derived from the CNPq knowledge area classification scheme (de Siqueira et al., 2020). In order to identify researchers in different collaborations, we adopted a strategy proposed by Dias and Moita (2015), which generates a kind of collaboration identifier. This strategy first removes all special characters and stop words from the publication titles. Then, it concatenates all generated tokens and adds the publication year to the final string, thus creating a hash key for each publication. Finally, the hash keys are used as an identifier for the authors' collaborations, i.e., there is a collaboration involving two or more researchers if they share the same hash key. As reported by their authors, this strategy is very effective, providing average precision and recall values for distinct scientific areas of 100% and 98.08%, respectively, thus making it very suitable for our purposes. We also notice that, in the analyses that we carry out in this article, we only consider those collaborations that correspond to coauthorships identified from journal articles, since this type of publication is the most common and the most relevant for the majority of the research areas.

Tables 1 and 2 list the eight major areas considered and shows, respectively, the distribution of researchers and publications in each one of them. As we can see in Table 1, the number of researchers in each major area (second column) is quite uneven and basically reflects the popularity of each one of them. However, looking at the third column, it shows that a large number of researchers (151,664, 57.6% of the total) has at least one interdisciplinary collaboration, i.e., such researchers have published at least one article in coauthorship with one or more researchers from another major area, which clearly indicates that the Brazilian researchers are quite open to this kind of collaboration.

Regarding publications by major area (see Table 2 where figures between square brackets correspond to percentages with respect to the overall numbers in the second column), we can see that the number of interdisciplinary publications[6] per major area varies from 1.3% and 5.5% (*Linguistics, Letters and Arts*) to 24.1% and 24.5% (Biological Sciences), when we consider, respectively, only the interdisciplinary collaborations (figures between parentheses) and all publications together, i.e., non-interdisciplinary and interdisciplinary ones (figures between brackets). Looking closer to such figures, although in general the number of interdisciplinary collaborations in each major area is related to its total number of

[6]Interdisciplinary publications are those that have at least two coauthors from distinct major areas.

Table 1. Distribution of Researchers by Major Area.

Major Area	Researchers	Interdisciplinary Researchers
Agrarian Sciences	26,953 (10.2%)	19,328 (12.7%)
Applied and Social Sciences	29,146 (11.1%)	13,289 (8.8%)
Biological Sciences	36,356 (13.8%)	25,095 (16.5%)
Engineering	24,746 (9.4%)	13,925 (9.2%)
Exact and Earth Sciences	39,231 (14.9%)	21,414 (14.1%)
Health Sciences	45,990 (17.5%)	28,844 (19.0%)
Humanities	44,743 (17.0%)	23,294 (15.4%)
Linguistics, Letters and Arts	16,099 (6.1%)	6,475 (4.3%)
Total	263,264 (100.0%)	151,664 (100.0%)

Table 2. Distribution of Publications by Major Area.

Major Area	Publications	Interdisciplinary Publications
Agrarian Sciences	1,416,249 (14.0%)	252,616 (15.6%) [17.8%]
Applied and Social Sciences	768,179 (7.6%)	83,431 (5.1%) [10.9%]
Biological Sciences	1,598,303 (15.8%)	391,000 (24.1%) [24.5%]
Engineering	958,899 (9.5%)	185,772 (11.5%) [19.4%]
Exact and Earth Sciences	1,395,113 (13.8%)	271,423 (16.7%) [19.5%]
Health Sciences	2,297,864 (22.7%)	285,359 (17.6%) [12.4%]
Humanities	1,306,579 (12.9%)	131,625 (8.1%) [10.1%]
Linguistics, Letters and Arts	386,728 (3.8%)	21.161 (1.3%) [5.5%]
Total	10,127,914 (100.0%)	1,622,387(100.0%) [16.0%]

publications, there are cases where this pattern does not apply as, for example, in the major areas of *Engineering* and *Health Science*. Moreover, *Health Sciences* is the major area with the largest number of publications (2,297,864), but it appears only in the fifth place when we consider the interdisciplinary publications (285,359).

3.2 Interdisciplinary collaborations

Since interdisciplinary collaborations play an important role in the Brazilian scientific community (Mena-Chalco et al., 2014; Pessoa Junior et al., 2019, 2020), we now present an overview of the structure of its two networks. The first one, called Global, includes a node for each one of the 263,264 researchers in our dataset, with an edge being created between two nodes whenever the respective researchers have at least one scientific collaboration. The second network, called Interdisciplinary, also includes a node for each researcher, but its edges represent only interdisciplinary collaborations, which are identified according to the first major area indicated by the researchers as the closest related to their research activities. Figure 1 shows the evolution of the two collaboration networks over the years. As we can see, after the year 2000 both networks present an intensive growth in the number of scientific collaborations involving PhD researchers, but the intensity of such collaborations starts to decelerate from 2015 onwards, tending to some stability from 2017 onwards.

Table 3 shows the figures of some usual metrics and statistics for the Global and Interdisciplinary networks. First, we highlight that 35.2% of the edges (900,992 out of 2,563,017) correspond to interdisciplinary collaborations and 57.6% of the nodes (151,664 out of 263,264) correspond to researchers who have at least one interdisciplinary collaboration. These are very expressive numbers, since they refer to the network of the entire scientific community from a whole country. In addition, the fact that a large number of the edges from the giant component[7] of the Global network, which includes 95.5% of all edges, are kept in the Interdisciplinary network is also very relevant, since 87.8% of such edges belong to its giant component. Moreover, this reinforces how important such collaborations are in an academic context.

As expected, the number of edges and the average degree of the nodes is smaller in the Interdisciplinary network, whereas the total number of isolated components is higher. Despite that, the diameters and the average path lengths in both networks are very close. Regarding the density of the networks (ratio between the number of existing edges and the number of possible edges in the complete graph), the Interdisciplinary one naturally tends to be less dense. Finally, we also calculated the Assortative Mixing Coefficient (Newman, 2003) of the Global network. The value of this metric (0.8375) evidentiates a high level of connectivity involving the nodes of the Global network, thus corroborating the importance of the interdisciplinary collaborations in the context of the Brazilian scientific community.

4 Interdisciplinary collaborations across geographic regions

Why should we analyze interdisciplinary collaborations across geographic regions? Since Brazil is a continental coun-

[7] Maximal subgraph that includes a path connecting each pair of nodes of a network.

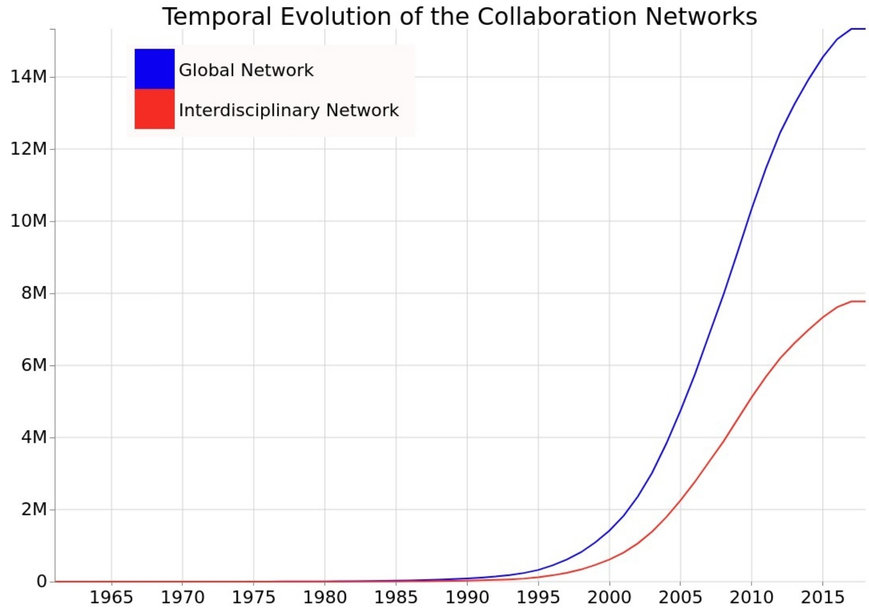

Figure 1. Temporal Evolution of the Collaboration Networks.

Table 3. Metric Figures of Both Networks.

Metric	Global	Interdisciplinary
Number of Nodes	263,264	263,264
Number of Connected Nodes	211,594	151,664
Number of Isolated Nodes	51,670	111,600
Number of Edges	2,563,017	900,992
Average Degree	11.9	6.8
Size of the Giant Component	207,583	145,255
% of Nodes in the Giant Component	78.8%	55.2%
% of Edges in the Giant Component	95.5%	87.8%
Average Path Length	5.2	7.1
Network Diameter	15	18
Network Density	3.98E-5	1.30E-5

try, addressing this issue is fundamental to understand how scientific collaborations occur inside and outside the Brazilian regions when considering the major areas involved. Doing so, we are able to reveal, for example, what are the main collaboration patterns in our community and evidentiate the role played by geography in this regard.

Thus, given that there is an expressive number of interdisciplinary collaborations involving researchers of all major areas, in this section we investigate how such collaborations occur across the five Brazilian geographic regions[8]. Brazil is composed of 26 states and a Federal District that locates its capital, Brasília, which means that characterizing such scientific collaborations in terms of its geographic regions is very important to better understand how scientific knowledge disseminates across the country.

For better understanding the importance of each one these regions, Figure 2 presents some demographic data about each one of them, as provided by IBGE, the Brazilian Institute of Geography and Statistics, according to its lasted official report[9]. As we can see, there is a large gap in the gross product among the geographic regions, being the respective per capita values much smaller in the North and Northeast regions (US$3,256 and US$4,714, respectively) than in the Central-West, Southeast and South ones (US$ 11,333, US$11,234 and US$10,659, respectively). Thus, these uneven figures by region affect how investments in science and technology are carried out across the country, thus impacting the collaborations among research groups and, therefore, the scientific growth and development in Brazil. To illustrate this, we can observe the discrepancy in the number of researchers in the Southest and North regions, which have, respectively, 856 and 103 researchers per million inhabitants. Moreover, according to data from CAPES[10], the Ministry of Education Graduate Studies Agency, 75.3% of the graduate programs with the highest grade of excellence, i.e., graduate programs graded 7[11], are located in the Southeast region, whereas in the North region there is not a single program with such a grade (see NGP7 values in Figure 2).

Next, we divide our analysis into two parts. First, in Subsection 4.1, we look at interdisciplinary collaborations by considering a more general context and highlighting how

[8]See: https://en.wikipedia.org/wiki/Regions_of_Brazil
[9]IBGE: https://www.ibge.gov.br/en/home-eng.html
[10]http://www.capes.gov.br
[11]The CAPES evaluation system rates the graduate programs in Brazil from grades 3 to 7, being grades 6 and 7 assigned to programs of high excellence.

Figure 2. The Brazilian Geographic Regions (data according to the lasted official report by IBGE as of 2018).

such collaborations occur between geographic regions and major areas. Then, in Subsection 4.2, we analyze separately the scenarios per region and per major area, thus providing a characterization of the interdisciplinary collaboration patterns.

4.1 Interdisciplinary collaborations per geographic region

Figure 3 presents the interdisciplinary collaboration rates involving the researchers of each two regions. It is worth mentioning that, although the Interdisciplinary network involves 151,664 researchers, for this specific analysis only those 131,936 researchers that are actually associated with an academic or research institution in Brazil, according to information in their own curricula, were considered. As can be seen, the Southeast region (in purple) is the one with the highest rates of interdisciplinary collaboration in all scenarios, reaching more than 46% in each one of them. A first explanation for these figures is the fact that such a region is the richest and most populated one in the country (see Figure 2) and, therefore, the one with the greatest access to resources, both governmental and private, for research funding. In fact, there is an intrinsic relation between the wealth and the size of the population of the regions when considering the distribution of interdisciplinary collaborations. In addition, it should be noticed that in all scenarios the rates of interdisciplinary collaborations involving researchers from their own regions is quite relevant, reinforcing that geographic proximity is also an important factor in this regard.

When we look at the pairwise comparisons of the collaboration percentages across the regions (Table 4), the contrast becomes even more evident. For instance, the proportional participation of collaborations involving researchers from the North region (total of 3.7%, first four lines) is almost 21 times lower than that involving researchers from the Southeast region (total of 77.3%), despite having a population 4.8 times smaller and a gross income 16.4 times lower, as shown in Figure 2. On the other hand, such discrepancy in the Northeast region (total of 44.7%) in relation to the Southeast region is smoother with a proportional share that is 1.7 times smaller, thus better reflecting the size of its population and its per capita income, which are, respectively, 1.6 and 3.7 smaller. Again, as can be seen by such figures, the proportional participation by region is strongly related to the respective number of programs of high excellence (see Figure 2).

Table 4. Percentage of Interdisciplinary Collaborations Between Brazilian Regions.

Regions	% Collaborations
Central-West - North	0.3%
North - Northest	0.7%
North - South	0.8%
North - Southest	1.9%
Central-West - Northest	3.8%
Central-West - South	4.9%
Central-West - Southest	11.7%
Northest - South	12.2%
Northest - Southest	28.0%
Southest - South	35.7%
Total	100.0%

In order to present a more detailed view of the interdisci-

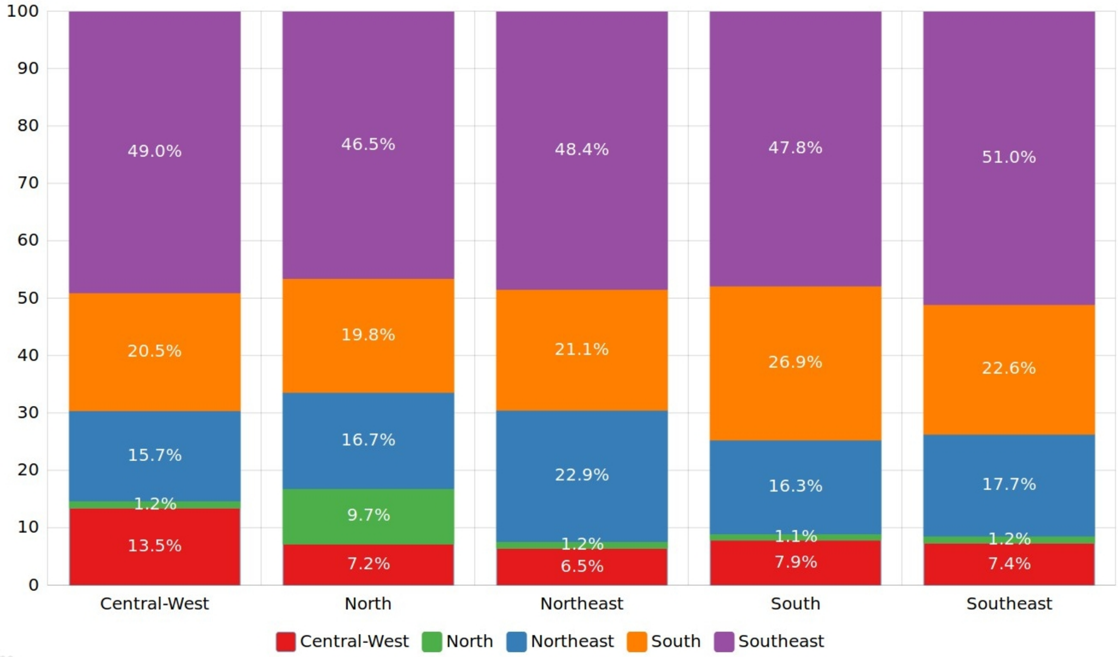

Figure 3. Distribution of the Interdisciplinary Collaborations per Geographic Region.

plinary collaboration patterns, Figure 4 provides a visualization of how they correlate across the country. In that figure, there is a heat map with the proportional correlation matrix of interdisciplinary collaborations according to their normalized values, considering both regions and major areas. In this way, bluer cells inform those regions and their respective areas that have a strong collaboration compared to the others, while lighter tones indicate low collaboration ones. Thus, Figure 4 describes the intensity of the collaboration patterns reflected by the distribution of the researchers per major area in their respective regions. Also, note that the values are globally normalized, so we can compare cell by cell, as well as region by region across the quadrants. Moreover, by definition, there are completely blank quadrants (i.e., when there is no correlation) as, for example, the first eight rows and columns that refer to the North region, since collaborations within a same region are not considered.

Regarding the most relevant collaboration patterns in terms of the number of researchers involved, there is a strong collaboration between *Biological Sciences* researchers from the Southeast region (row 27) and *Health Sciences* researchers from the South region (row 38). Moreover, the reciprocal is also true as showed by rows 30 and 35, i.e., even having access to other researchers due to geographic proximity (which, as already shown, facilitates collaborations), there is an intensive search for knowledge from other domains. Similarly, the *Biological Sciences* and the *Exact and Earth Sciences* major areas have a considerable reciprocity involving researchers from the Southeast and South regions (cells 27 and 37, and 29 and 35). This evidence reinforces that Brazilian researchers also look for establishing new scientific collaborations, making the Brazilian collaboration network more cohesive and the access to information more widely available.

Figure 4 also describes the intensity of the collaboration patterns with respect to the total number of researchers from each region. In accordance with the percentage of interdisciplinary collaborations by regions (see Table 4), we can observe stronger tones in the quadrants that refer to collaborations that involve the following regions: Northeast (quadrant 9 to 16), Southeast (quadrant 25 to 32) and South (quadrant 33 to 40). On the other hand, we can also observe that the lightest tones in the figure are those that refer to the North region (quadrant 1 to 8) and the Central-West region (quadrant 17 to 24), the two regions with the smallest GDP rates in the country.

Looking now more closely at each major area in Figure 4, we notice a quite uneven correlation among them. Although this reflects the popularity of each major area, the chart enables us to understand the intensity of the interdisciplinary collaborations across the country. For instance, considering the major areas of *Engineering* (rows 4, 12, 20, 28 and 36) and *Applied and Social Sciences* (rows 2, 10, 18, 26 and 34), which are two comparable communities in terms of the number of researchers (24,746 and 29,146, respectively), they have quite different collaboration patterns. More specifically, researchers from *Engineering* present strong ties with those from *Exact and Earth Sciences* (rows 5, 13, 21, 29 and 37), whereas the collaborations between researchers from *Applied and Social Sciences* and *Humanities* (rows 7, 15, 23, 31 and 39) are less intensive. It is also worth noticing that *Biological Sciences* (rows 3, 11, 19, 27 and 35) is the major area most involved in interdisciplinary collaborations. Although it is only the third largest major area in terms of number of researchers (see Table 1), it has a remarkable presence in all scenarios. Additionally, we can also see that *Agrarian Sciences* (rows 1, 9, 17, 25 and 33), *Biological Sciences* (rows 3, 11, 19, 27 and 35) and *Health Sciences* (rows 6, 14, 22, 30 and 38) are among the major areas most involved in interdisciplinary collaborations across the regions (respective arcs are much wider). In contrast, the rows and columns that refer to the *Linguistics, Letters and Arts* major area (rows 8,

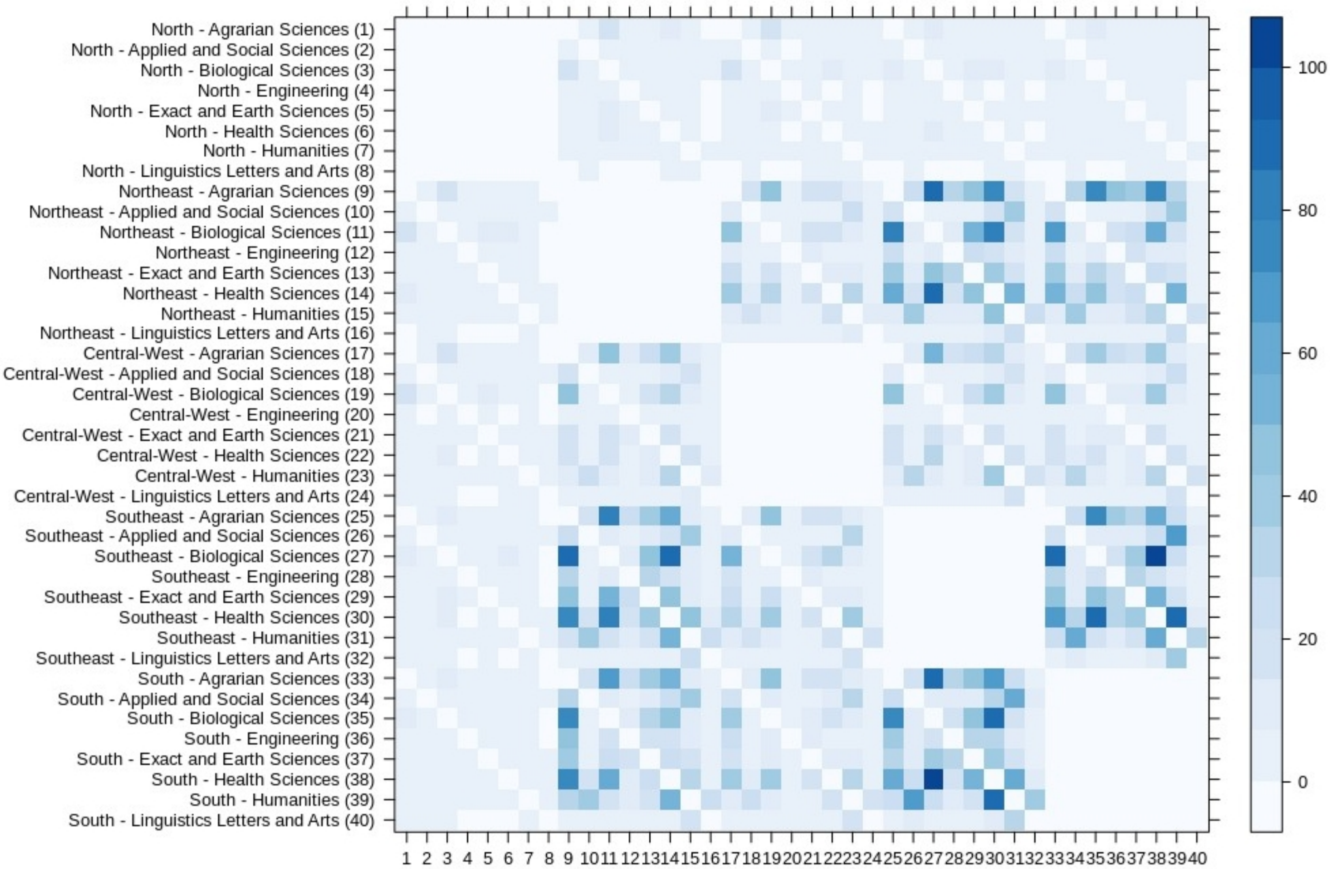

Figure 4. Normalized Interdisciplinary Collaboration Patterns Between the Brazilian Geographic Regions.

16, 24, 32 and 40), the smallest major area in terms of researchers (see Table 1), appears more isolated in the figure with more milder tones.

Table 5 summarizes the ten most relevant collaboration patterns in terms of the number of collaborations involved. For example, the first line shows a strong collaboration between *Biological Sciences* researchers from the Southeast region and *Health Sciences* researchers from the South region. In addition, the second line confirms that the reciprocal is also true, i.e., even having access to other researchers due to geographic proximity (which, as already shown, facilitates collaborations), there is an intense search for knowledge from other domains. Likewise, the major areas of *Biological Sciences* and *Agrarian Sciences* have considerable reciprocity involving the Southeast and Northeast regions. This evidence reinforces that Brazilian researchers also seek to establish new scientific partnerships, making the Brazilian collaboration network more cohesive and the access to information more widely available.

4.2 Interdisciplinary collaboration patterns

Having characterized the interdisciplinary collaborations across the geographic regions and major areas, now we take a closer look at them by considering two distinct perspectives: regions subdivided by major areas and major areas subdivided by regions. Doing so, we aim to provide a better understanding of how such collaborations occur across the country.

Interdisciplinary collaboration patterns between major areas per region. Figure 5 shows the general collaboration patterns involving the major areas for each one of the five Brazilian geographic regions. The size of the nodes represent the percentage of researchers from each major area involved in interdisciplinary collaborations and the thickness of the edges expresses the percentage of collaborations between researchers from the major areas connected by them. The graphs show that the major area most involved in interdisciplinary collaborations in all regions is *Biological Sciences* (black nodes), corroborating the previous finding that this is the most collaborative major area (see Tables 1 and 2). In contrast, *Applied and Social Sciences* (orange nodes) and *Linguistics, Letters and Arts* (beige nodes) appear among the three less collaborative major areas in all regions, thus corroborating the fact that they are among the major areas with the smallest rates of collaboration overall, as already shown in Tables 1 and 2. Looking more closely at Figure 5, we note that *Health Sciences* (purple) has a great appeal in the three wealthy and most populous regions of the country (Northeast, Southeast and South), whereas *Agrarian Sciences* (red) appears as the second most popular major area in the regions with the largest territorial extension (Central-West and North).

With respect to the diversity of the collaborations, we note that the Northeast and South regions are the most democratic ones (largest nodes with more homogeneous sizes), i.e., in these two regions there are more major areas involved in interdisciplinary collaborations (*Biological Sciences*, *Health Sciences*, *Agrarian Sciences* and *Exact and Earth Sciences*).

Table 5. Top Ten Patterns of Interdisciplinary Collaborations Between Brazilian Regions

Major Area (Region)	Major Area (Region)	Nr. of Collaborations (%)
Biological Sciences (Southeast)	Health Sciences (South)	6,217 (2.07%)
Biological Sciences (South)	Health Sciences (Southeast)	5,667 (1.89%)
Humanities (South)	Health Sciences (Southeast)	5,662 (1.89%)
Biological Sciences (Southeast)	Agrarian Sciences (South)	5,467 (1.82%)
Biological Sciences (Southeast)	Health Sciences (Northeast)	5,299 (1.77%)
Biological Sciences (Southeast)	Agrarian Sciences (Northeast)	5,069 (1.69%)
Biological Sciences (Northeast)	Health Sciences (Southeast)	4,909 (1.64%)
Biological Sciences (Northeast)	Agrarian Sciences (Southeast)	4,749 (1.58%)
Biological Sciences (South)	Agrarian Sciences (Southeast)	4,726 (1.57%)
Agrarian Sciences (Northeast)	Health Sciences (Southeast)	4,396 (1.46%)

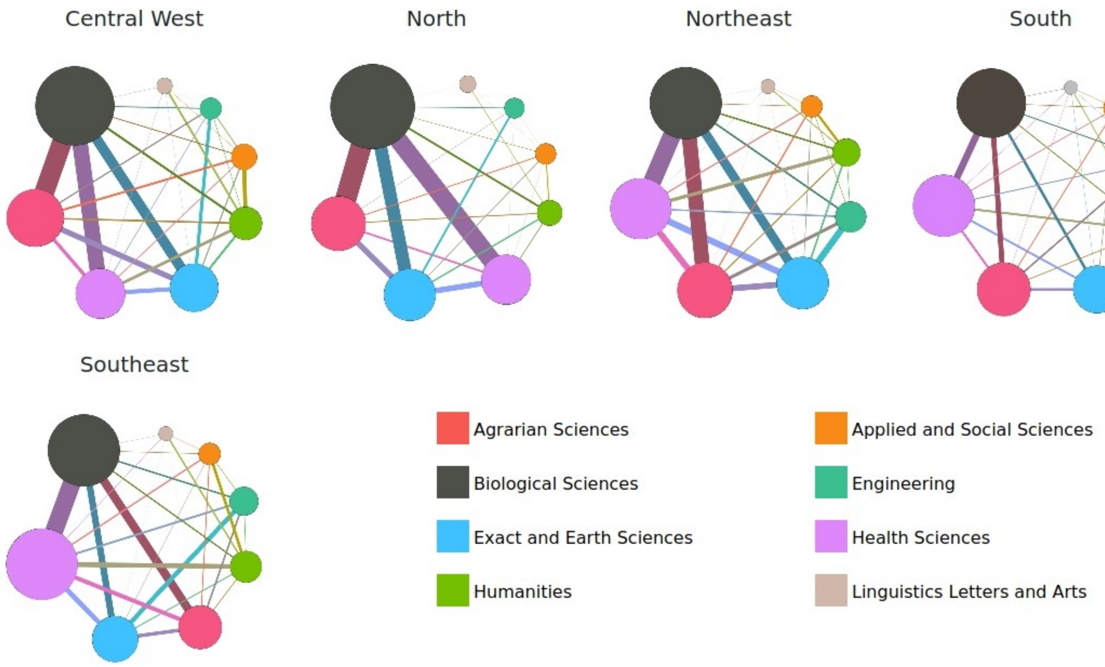

Figure 5. Normalized Interdisciplinary Collaboration Patterns Between Major Areas per Region.

In contrast, the Central-West, North and South regions concentrate most of their interdisciplinary collaborations on just three major areas: *Biological Sciences*, *Agrarian Sciences* and *Health Sciences.*

Finally, the edge thickness in the graphs of Figure 5 show that the collaborations between the major areas of *Biological Sciences* (black) and *Agrarian Sciences* (red), as well as those between the major areas of *Biological Sciences* (black) and *Health Sciences* (purple), are relevant in all scenarios. In contrast, the collaborations between the major areas of *Exact and Earth Sciences* (blue) and *Engineering* (sea green) are more expressive in the Northeast, Southeast and South regions, the three richest regions in Brazil. We also note that in the Northeast and South regions the thickness of the incident edges coming from the major area of *Exact and Earth Sciences* (blue) into the major areas of *Biological Sciences* (black), *Health Sciences* (purple), *Engineering* (sea green) and *Agrarian Sciences* (red) are very alike, showing a similar collaboration pattern of their researchers in both regions.

Interdisciplinary collaboration patterns between regions per major area. Looking now at Figure 6, it shows well-defined collaboration patterns for each one of the major areas, which includes an expressive Southeast-South-Northeast triad with few variations of the South and Northeast regions appearance as the second more relevant one. In contrast, the North region has a very low proportion of collaborations (small size nodes) and also very few strong relationships (thin edges) in all scenarios.

We can also highlight which major areas are the most important ones according to each region. For instance, the Central-West region has more relevant relationships in the major areas of *Agrarian Sciences* and *Linguistics, Letters and Arts* with the Southeast and South ones, whereas the Northeast region shows very thick edges with the Southeast one for the major areas of *Biological Sciences*, *Engineering*, and *Exact and Earth Sciences*. Such numbers are in line with the highest grade of excellence given by CAPES to the respective graduate programs in these major areas. For example, in the Northeast region, out of the nine graduate programs with the highest CAPES grade, four are from *Exact and Earth Sciences*, four from *Engineering* and one from *Biological Sciences*.

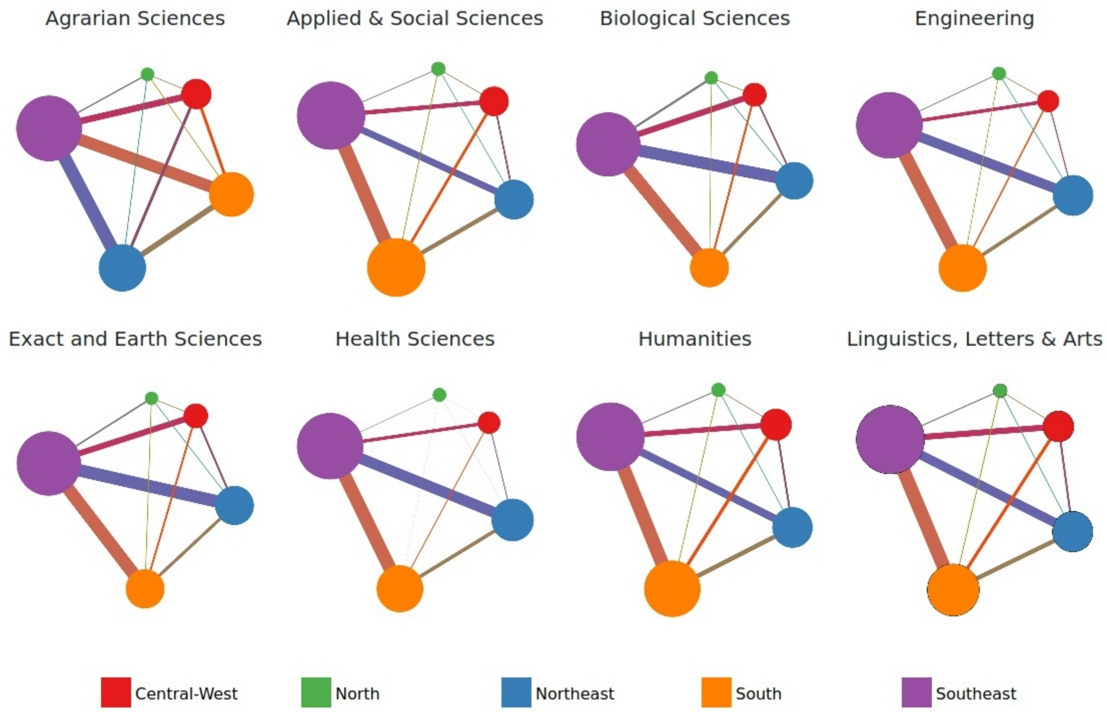

Figure 6. Normalized Interdisciplinary Collaboration Patterns Between Regions per Major Area.

5 Conclusions and future work

In this article, we provided a geographic analysis of the interdisciplinary collaborations in the Brazilian scientific community. For this, we used data extracted from the curricula vitae of 263,264 researchers holding a PhD degree collected from the Lattes platform and grouped them in terms of their contributions according to the eight major areas defined by the CNPq knowledge area classification scheme. As a result, we characterized how such collaborations have been spread across the Brazilian geographic regions. Particularly, 35.2% of all collaborations are interdisciplinary and 57.6% of the researchers have at least one scientific collaboration of this kind.

By considering the interdisciplinary collaborations across geographic regions, we concluded that geography plays a key role in favoring richer regions. Specifically, the total number of high excellence programs and the rate of researchers are quite unequal per million inhabitants (see Figure 2). Such issues have resulted in a predominance in the distribution of interdisciplinary collaborations, with the *Southeast* region being responsible for practically half of all such collaborations in each one of the regions. Although there are uneven patterns between the regions, there is a strong Southeast-South-Northeast triad for all major areas. On the other hand, by exploring interdisciplinarity in terms of the major areas, we observe distinct patterns that demonstrate the dynamics and peculiarities of each one of them in different regions. In general, *Biological Sciences*, the third most populous major area, stands out as the most democratic in terms of interdisciplinary diversity, whereas the major area of *Linguistics, Letters and Arts*, the less populous one, presents the lowest rates of interdisciplinary collaborations overall.

As future work, we intend to extend our current results by analyzing the researchers' academic mobility across the Brazilian geographic regions, thus characterizing their trajectories over their academic education (Furtado et al., 2015; Leydesdorff and Persson, 2010; Silva et al., 2016). Particularly, an interesting issue to address is to verify whether geographical location and interdisciplinarity are related to each other and what is the influence of the network structure in this specific case. Another perspective is to make explicit the group of researchers that make the coauthorship networks more integrated by transferring additional knowledge (Leão et al., 2018; Silva et al., 2019), therefore providing new indicators in terms of both interdisciplinarity and inter-regional collaborations (Dornbusch et al., 2013).

Acknowledgements

This work was supported by the authors' individual grants from CAPES and CNPq (308528/2019-0). The authors would like to thank the reviewers for their fruitful comments, which helped improving the final version of this article.

References

Abramo, G., D'Angelo, C. A., and Di Costa, F. (2018). The effect of multidisciplinary collaborations on research diversification. *Scientometrics*, 116(1):423–433.

Adams, J. D., Black, G. C., Clemmons, J. R., and Stephan, P. E. (2005). Scientific teams and institutional collaborations: Evidence from U.S. universities, 1981–1999. *Research Policy*, 34(3):259–285.

Chiarini, T., Oliveira, V. P., and do Couto e Silva Neto, F. C. (2014). Spatial distribution of scientific activities: An exploratory analysis of Brazil, 2000–10. *Science and Public Policy*, 41(5):625–640.

de Siqueira, G. O., Canuto, S. D., Gonçalves, M. A., and Laender, A. H. F. (2020). A pragmatic approach to hierarchical categorization of research expertise in the presence of scarce information. *International Journal on Digital Libraries*, 21(1):61–73.

Dias, T. M. R. and Moita, G. F. (2015). A method for the identification of collaboration in large scientific databases. *Em Questão*, 21(2).

Dornbusch, F., von Proff, S., and Brenner, T. (2013). The organizational and regional determinants of inter-regional collaborations - Academic inventors as bridging agents. Technical report, Working Papers on Innovation and Space.

Freire, V. P. and Figueiredo, D. R. (2011). Ranking in collaboration networks using a group based metric. *Journal of the Brazilian Computer Society*, 17(4):255–266.

Furtado, C. A., Davis Jr, C. A., Gonçalves, M. A., and de Almeida, J. M. (2015). A Spatiotemporal Analysis of Brazilian Science from the Perspective of Researchers' Career Trajectories. *PloS One*, 10(10):e0141528.

Garay, Y., Akbar, M., and Gates, A. Q. (2016). Towards Identifying Potential Research Collaborations from Scientific Research Networks using Scholarly Data. In *Proceedings of the 16th ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL 2016, Newark, NJ, USA, June 19 - 23, 2016*, pages 217–218.

Haythornthwaite, C. (2006). Learning and knowledge networks in interdisciplinary collaborations. *Journal of the American Society for Information Science and Technology*, 57(8):1079–1092.

Hoekman, J., Frenken, K., and Tijssen, R. J. (2010). Research collaboration at a distance: Changing patial patterns of scientific collaboration within Europe. *Research Policy*, 39(5):662–673.

Hua, G. and Haughton, D. (2012). A network analysis of an online expertise sharing community. *Social Network Analysis and Mining*, 2(4):291–303.

Huang, M.-H. and Chang, Y.-W. (2011). A study of interdisciplinarity in information science: Using direct citation and co-authorship analysis. *Journal of Information Science*, 37(4):369–378.

Iglič, H., Doreian, P., Kronegger, L., and Ferligoj, A. (2017). With whom do researchers collaborate and why? *Scientometrics*, 112(1):153–174.

Jiang, J., Shi, P., An, B., Yu, J., and Wang, C. (2017). Measuring the social influences of scientist groups based on multiple types of collaboration relations. *Information Processing & Management*, 53(1):1–20.

Jones, B. F., Wuchty, S., and Uzzi, B. (2008). Multi-University Research Teams: Shifting Impact, Geography, and Stratification in Science. *Science*, 322(5905):1259–1262.

Kato, M. and Ando, A. (2013). The relationship between research performance and international collaboration in chemistry. *Scientometrics*, 97(3):535–553.

Kulbacki, M., Segen, J., Knieć, W., Klempous, R., Kluwak, K., Nikodem, J., Kulbacka, J., and Serester, A. (2018). Survey of Drones for Agriculture Automation from Planting to Harvest. In *2018 IEEE 22nd International Conference on Intelligent Engineering Systems (INES)*, pages 353–358. IEEE.

Leão, J. C., Brandão, M. A., de Melo, P. O. S. V., and Laender, A. H. F. (2018). Who is really in my social circle? - Mining social relationships to improve detection of real communities. *Journal of Internet Services and Applications*, 9(1):20:1–20:17.

Leydesdorff, L. and Persson, O. (2010). Mapping the geography of science: Distribution patterns and networks of relations among cities and institutes. *Journal of the American Society for information Science and Technology*, 61(8):1622–1634.

Li, Q., Brown, J. B., Huang, H., Bickel, P. J., et al. (2011). Measuring Reproducibility of High-throughput Experiments. *The Annals of Applied Statistics*, 5(3):1752–1779.

Lima et al., H. (2013). Aggregating Productivity Indices for Ranking Researchers Across Multiple Areas. In *Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL'13, Indianapolis, IN, USA, July 22 - 26, 2013*, pages 97–106.

Liu, X., Bollen, J., Nelson, M. L., and Van de Sompel, H. (2005). Co-authorship networks in the digital library research community. *Information Processing and Management*, 41(6):1462–1480.

Mena-Chalco, J. P., Digiampietri, L. A., Lopes, F. M., and Cesar, R. M. (2014). Brazilian bibliometric coauthorship networks. *Journal of the Association for Information Science and Technology*, 65(7):1424–1445.

Mooney, H. A., Duraiappah, A., and Larigauderie, A. (2013). Evolution of natural and social science interactions in global change research programs. *Proceedings of the National Academy of Sciences*, 110(Supplement 1):3665–3672.

Morillo, F., Bordons, M., and Gómez, I. (2003). Interdisciplinarity in science: A tentative typology of disciplines and research areas. *Journal of the American Society for Information Science and Technology*, 54(13):1237–1249.

Newman, M. (2003). Mixing patterns in networks. *Physical Review E*, 67(026126).

Pan, R. K., Kaski, K., and Fortunato, S. (2012). World citation and collaboration networks: uncovering the role of geography in science. *Scientific reports*, 2:902.

Peet, R. K. (1975). Relative Diversity Indices. *Ecology*, 56(2):496–498.

Pessoa Junior, G. J., Dias, T. M., Silva, T. H., and Laender, A. H. F. (2020). On interdisciplinary collaborations in scientific coauthorship networks: the case of the Brazilian community. *Scientometrics*, 124(3):2341–2360.

Pessoa Junior, G. J., Dias, T. M. R., Silva, T. H. P., and Laender, A. H. F. (2019). Interdisciplinary Collaborations in the Brazilian Scientific Community. In *Digital Libraries for Open Knowledge - 23rd International Conference on Theory and Practice of Digital Libraries, TPDL 2019, Oslo, Norway, Sept. 9-12, 2019, Proceedings*, pages 145–153.

Porter, A. L., Roessner, D. J., and Heberger, A. E. (2008). How interdisciplinary is a given body of research? *Re-*

search evaluation, 17(4):273–282.

Shi, S., Zhang, W., Zhang, S., and Chen, J. (2018). Does prestige dimension influence the interdisciplinary performance of scientific entities in knowledge flow? Evidence from the e-government field. *Scientometrics*, 117(2):1237–1264.

Sidone, O. J. G., Haddad, E. A., and Mena-Chalco, J. P. (2017). Scholarly publication and collaboration in Brazil: The role of geography. *Journal of the Association for Information Science and Technology*, 68(1):243–258.

Silva, F. N., Rodrigues, F. A., Jr., O. N. O., and da F. Costa, L. (2013). Quantifying the interdisciplinarity of scientific journals and fields. *J. Informetrics*, 7(2):469–477.

Silva, T. H. P., Laender, A. H. F., Davis, C. A., da Silva, A. P. C., and Moro, M. M. (2017). A profile analysis of the top Brazilian Computer Science graduate programs. *Scientometrics*, 113(1):237–255.

Silva, T. H. P., Laender, A. H. F., Davis Jr, C. A., da Silva, A. P. C., and Moro, M. M. (2016). The Impact of Academic Mobility on the Quality of Graduate Programs. *D-Lib Magazine*, 22(9/10).

Silva, T. H. P., Laender, A. H. F., and Vaz de Melo, P. O. S. (2019). Characterizing Knowledge-Transfer Relationships in Dynamic Attributed Networks. In *IEEE/ACM 2019 International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2019, Vancouver, Canada, August 28-31, 2019*, pages 234–241.

Sonnenwald, D. H. (2007). Scientific collaboration. *Annual Review of Information Science and Technology*, 41(1):643–681.

Wagner, C. S., Staheli, L., Silberglitt, R., Wong, A., and Kadtke, J. (2002). *Linking Effectively: Learning Lessons from Successful Collaboration in Science and Technology*. RAND's Science & Technology Policy Institute, Arlington, VA.

Wilsdon, J. et al. (2011). *Knowledge, networks and nations: Global scientific collaboration in the 21st century*. The Royal Society, London, UK.

Yan, E., Ding, Y., and Sugimoto, C. R. (2011). P-rank: An indicator measuring prestige in heterogeneous scholarly networks. *Journal of the American Society for Information Science and Technology*, 62(3):467–477.