

ARTIGO

Recebido em:  
11/02/2016

Aceito em:  
10/11/2016

*Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, v. 22, n.48, p. 59-72, jan./abr., 2017. ISSN 1518-2924. DOI: 10.5007/1518-2924.2017v22n48p59

## Um processo para caracterização e análise de redes de colaboração científica institucional

*A process for characterization and analysis of networks scientific collaboration institutional*

---

Thiago Magela Rodrigues DIAS ([thiagomagela@gmail.com](mailto:thiagomagela@gmail.com))\* - Roberth Santos GOMES ([rsg\\_roberth@yahoo.com.br](mailto:rsg_roberth@yahoo.com.br))\*\* - Jhonatan Fernando OLIVEIRA ([jhonatan.oliveira@uemg.br](mailto:jhonatan.oliveira@uemg.br))\*\*\* - Gray Farias MOITA ([gray@dpp.cefetmg.br](mailto:gray@dpp.cefetmg.br))\*

\* Professor do Centro Federal de Educação Tecnológica de Minas Gerais

\*\* Mestrando em Modelagem Matemática e Computacional da Universidade do Estado de Minas Gerais

\*\*\* Professor da Universidade do Estado de Minas Gerais

---

### Resumo

O conhecimento é nos dias atuais fator primordial na definição e análise do desenvolvimento de um país e de sua população como um todo. A construção do conhecimento ocorre inevitavelmente através de colaborações, onde pesquisadores se reúnem em prol de objetivos em comum, formando as redes de colaboração científica. Tais redes são caracterizadas quando pesquisadores realizam e publicam seus trabalhos de forma conjunta, independentemente da instituição ou região geográfica em que estejam localizados. A colaboração científica entre instituições distintas, possibilita o intercâmbio de conhecimento e consequentemente efetiva troca de experiências, impulsionando a evolução da ciência. Nesse contexto, analisar tais redes pode proporcionar conhecimento sobre como vem ocorrendo a colaboração científica entre instituições de pesquisa e o quanto estas colaborações tem contribuído para a excelência das pesquisas realizadas. Este trabalho tem como objetivo analisar as redes de colaborações institucionais a partir de dados dos currículos Lattes. As redes são caracterizadas pelas colaborações entre pesquisadores de diferentes instituições, após a caracterização das redes, técnicas de visualização de grande volume de dados e métricas de análise de redes são aplicadas para verificar como a colaboração científica nacional tem ocorrido.

**Palavras-chave:** Redes de colaboração científica. Extração de dados. Recuperação da informação.

### Abstract

Knowledge is, nowadays, a key factor in defining and analyzing the development of a country and its population as a whole. The construction of knowledge inevitably occurs through collaborations, where researchers gather towards common goals, forming the scientific collaboration networks. Such networks are characterized when researchers perform and publish their work jointly, regardless of the institution or geographic region in which they are located. The scientific collaboration between different institutions enables the exchange of knowledge and consequently effective exchange of experiences, promoting the evolution of science. In this context, analyzing such networks can provide knowledge as it has the scientific collaboration between research institutions and how these collaborations have contributed to the excellence of the research conducted. This job aims to analyze the network of institutional collaborations from data from Lattes Curriculum. Networks are characterized by collaboration between researchers from different institutions after the characterization of networks, large data visualization techniques and metrics of network analysis are applied to see how the national scientific collaboration has occurred.

**Keywords:** Networks scientific collaboration. Data extraction. Recovery information.



v. 22, n. 48, 2017.  
p. 59-72  
ISSN 1518-2924



Esta obra está licenciada sob uma [Licença Creative Commons](https://creativecommons.org/licenses/by/4.0/).

## **1 INTRODUÇÃO**

A colaboração científica, objeto de estudos de vários trabalhos vem evoluindo ao longo dos anos. Os autores Beaver e Rosen (1978), em seu trabalho, afirmam que o primeiro artigo científico escrito de forma colaborativa entre diferentes pesquisadores, surgiu em 1678. Com o passar dos anos as redes de colaboração científica foram se consolidando e passaram a ter relação direta com o crescimento do conhecimento científico e das comunidades de pesquisa como um todo.

Nudelman e Landers (1972), sugerem que a credibilidade concedida pela comunidade científica a artigos de autoria conjunta é consideravelmente maior do que a creditada a artigos de autoria única, o que posteriormente é comprovado por Goffman e Warren (1980), ao mostrarem que as pesquisas desenvolvidas por grandes grupos tendem a ter mais influência no meio científico.

Portanto, visualizar a colaboração ocorrida na construção do conhecimento passa a ser algo relevante para que se possa mensurar o grau de importância de cada região ao longo deste processo, podendo auxiliar instituições, agências de fomento ou órgãos governamentais no direcionamento mais assertivo de seus investimentos.

No contexto deste trabalho, para a modelagem e caracterização de redes de institucionais, em especial, de instituições brasileiras, a Plataforma Lattes surge como excelente fonte de dados. A Plataforma Lattes tem como objetivo representar a experiência do CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) na integração de bases de dados de currículos, grupos de pesquisa e de instituições em um único sistema de informações. Cadastrar um currículo na plataforma e mantê-lo atualizado, atualmente, é indicado e está disponível a todos aqueles inseridos no contexto acadêmico e/ou científico. Isto se deve em função de sua adoção pela maior parte das instituições de fomento, universidades e institutos de pesquisa do país. A riqueza de informações providas pela plataforma em função de seu constante crescimento, abrangência e confiabilidade, torna-a indispensável e compulsória na análise meritória a pleitos de financiamentos de pesquisas científicas e tecnológicas. Lane (2010) destaca a Plataforma Lattes como excelente fonte para medição das produções científicas, em seu artigo publicado na revista Nature.

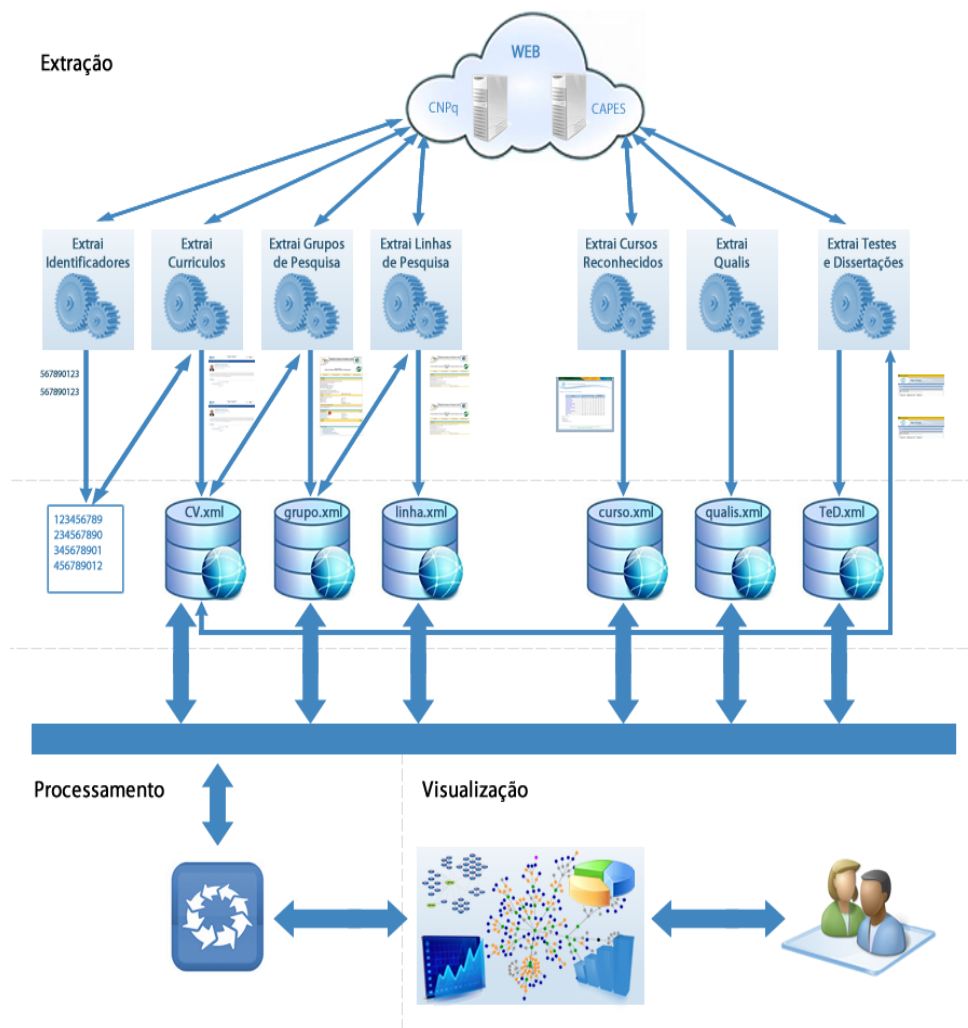
Em fevereiro de 2016, a plataforma Lattes atingiu a marca de 4.480.000 currículos cadastrados. Estes currículos estão disponíveis para consulta, porém, a consulta do currículo de forma isolada fragmenta o conhecimento disposto no conjunto de currículos, além do fato que a ferramenta de consulta disponibilizada pela própria plataforma Lattes não permite a identificação de colaborações científicas, por este motivo, o processamento e análise do grande volume de dados torna-se essencial para transformação destes currículos em conjunto de dados que possam auxiliar na compreensão de como ocorre a colaboração científica institucional no Brasil e ainda nas tomadas de decisões sobre a distribuição de recursos e investimentos.

Atualmente não há ferramentas que possibilitem a visualização de mapas de colaboração entre as instituições de pesquisa no cenário nacional. Essa ausência de ferramenta corrobora para inexistência de trabalhos compostos por embasamento visual que demonstre o grau de colaboração entre instituições de pesquisa no que tange ao conhecimento provido em determinada área do conhecimento.

Portanto, o objetivo deste trabalho consiste na exploração da plataforma Lattes como principal fonte de informação, a fim de caracterizar redes de colaboração e aplicar técnicas de visualização de grande volume de dados através da geração de mapas de colaboração institucionais.

## **2 MATERIAIS E MÉTODOS**

Os dados dos currículos Lattes utilizados para a caracterização das redes foram obtidos pela plataforma de extração e integração de dados proposta por Dias e Moita (2014). Na plataforma desenvolvida pelos autores, é apresentado que o processo de extração e integração dos dados foi dividido em três partes principais denominadas: Extração, Processamento e Visualização. Porém, neste trabalho somente foram utilizados os resultados da etapa de extração dos currículos, tendo em vista, que para a identificação das colaborações utiliza-se dos dados das publicações cadastradas em cada um dos artigos, como título e autores. Figura 1.



**Figura 1:** Arquitetura da Plataforma de Extração e Integração.  
**Fonte:** Dias e Moita (2014).

Todos os currículos extraídos são armazenados em disco em formato XML (*eXtensible Markup Language*). Na etapa de extração, várias falhas estruturais que estão nos currículos são corrigidas. Erros estes que dificultam o processo de análise dos dados. Importante ressaltar que neste caso, tratamentos de exceções foram desenvolvidos para contornar estes problemas e permitir que os arquivos sejam salvos normalmente e com o máximo de consistência possível. Os autores destacam que, diante do conjunto de currículos extraídos é possível realizar diversos tipos de análises bibliométricas, já que todos os dados dos currículos estão armazenados em arquivos estruturados. Tendo em vista a grande quantidade de dados extraídos, um grande desafio para a compreensão de como acontece a colaboração científica nacional passa a ser a identificação das colaborações. Para isso, o processo de identificação proposto por Dias e Moita (2015) é utilizado. Tal processo permite identificar colaborações científicas com base nos títulos das publicações de trabalhos em conjuntos com grande quantidade de dados, apresentando resultados satisfatórios com baixo custo computacional.

De posse de todas as colaborações devidamente identificadas entre pares de pesquisadores, o nome das instituições e o respectivo endereço profissional de cada pesquisador são considerados e dessa forma é possível identificar a colaboração entre as instituições, nesse caso, caracterizada pelo trabalho publicado por pesquisadores de instituições distintas, sendo utilizado a instituição de pesquisa atual para a caracterização das redes institucionais.

### 3 MODELAGEM E CARACTERIZAÇÃO

Para análise e visualização das informações, foi necessário o desenvolvimento de uma plataforma utilizando a linguagem Python. Os principais motivos pela escolha da linguagem, consistem: na praticidade; otimização dos códigos e no alto poder de processamento de grandes volumes de dados textuais, todos obtidos através de bibliotecas nativas da linguagem ou desenvolvidas por terceiros.

A plataforma desenvolvida, permite processar e analisar os dados do repositório Lattes, podendo ser parametrizada para trabalhar com áreas específicas do conhecimento como, por exemplo, as nove grandes áreas: Ciências Exatas e da Terra; Ciências da Saúde; Linguística, Letras e Artes; Ciências Biológicas; Engenharias; Ciências Humanas; Ciências Agrárias; Ciências Sociais Aplicadas e Outras. Áreas estas extraídas do currículo de cada pesquisador.

Diante os dados referente as colaborações que foram identificadas, as manipulações e análises são realizadas pela plataforma desenvolvida. As etapas descritas abaixo relatam o funcionamento da plataforma:

- **Controle:** como o foco do trabalho é nas instituições de pesquisa, se faz necessário uma base que contenha essas informações. Para este fim, um arquivo CSV (*Comma Separated Values*) foi gerado tendo como referência as instituições informadas nos currículos dos pesquisadores, acrescidas de um ID (identificador único), sigla, endereço completo e as coordenadas geográficas que foram obtidas através do endereço completo da instituição;

- **Seleção de Conjuntos:** consiste na definição de um filtro para os dados que deseja trabalhar, por exemplo, uma grande área ou determinada região. Através desta definição, novos arquivos de pesquisadores e colaborações são gerados;

- **Identificação de Colaboração:**

1. Arquivo de pesquisadores: ocorre uma comparação entre o arquivo de pesquisadores e o arquivo de instituições, onde o ID dos pesquisadores na plataforma Lattes é substituído pelo ID da instituição a qual estão vinculados. As coordenadas geográficas das respectivas instituições dos pesquisadores, também são acrescidas ao arquivo.

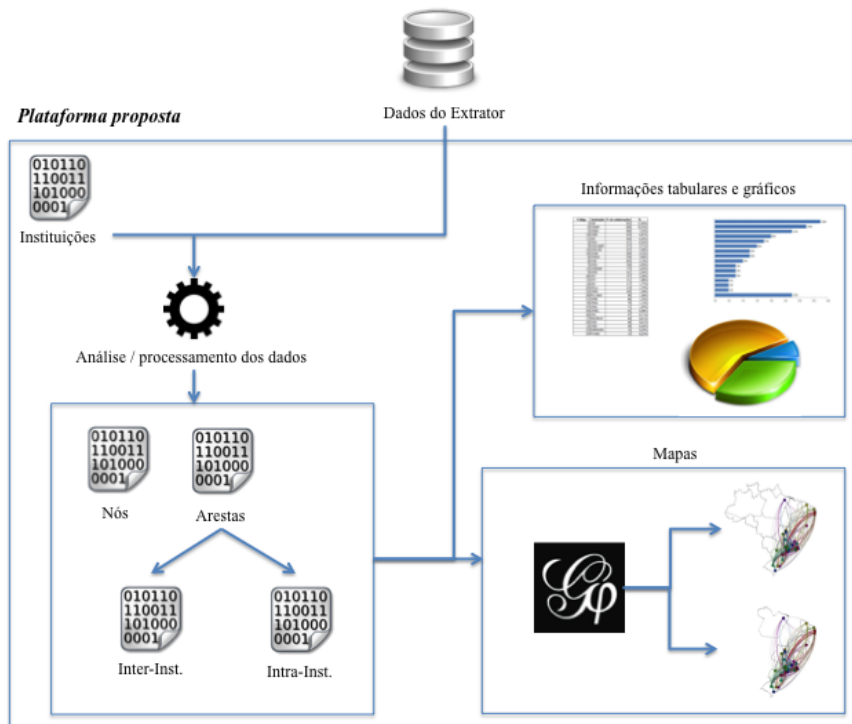
2. Arquivo de memória: o arquivo é gerado com base na comparação anterior e contém: ID antigo (da plataforma Lattes) e o novo ID (da instituição de pesquisa).

3. Arquivo de colaborações: tendo como referência o arquivo de memória, os ID's do arquivo de colaborações são substituídos pelos ID's das instituições de pesquisa correspondente dos pesquisadores. Posteriormente, há uma separação dos arquivos de colaborações, gerando dois arquivos, sendo: um arquivo contendo as colaborações entre pesquisadores de instituições distintas e outro contendo as colaborações dos pesquisadores das mesmas instituições.

- **Cálculo de Métricas:** nesta etapa ocorre a geração das informações tabulares (ranking's e matriz de colaboração) e dos arquivos a serem manipulados para a visualização das redes, sendo um arquivo de nós, relativo aos pesquisadores do presente estudo e um arquivo de arestas não direcionadas (colaborações);

- **Mapa de colaborações / gráfico de representatividade:** com a adoção da ferramenta Gephi, ocorre a geração do grafo de colaborações entre as instituições de pesquisa e do gráfico correspondente a representatividade de cada instituição.

A figura 2 descreve o funcionamento da plataforma proposta, nela as etapas: seleção de dados, manipulação dos arquivos e processamento de dados, descritas anteriormente, estão representadas por análise / processamento dos dados.



**Figura 2:** Funcionamento da plataforma proposta.  
**Fonte:** Os autores.

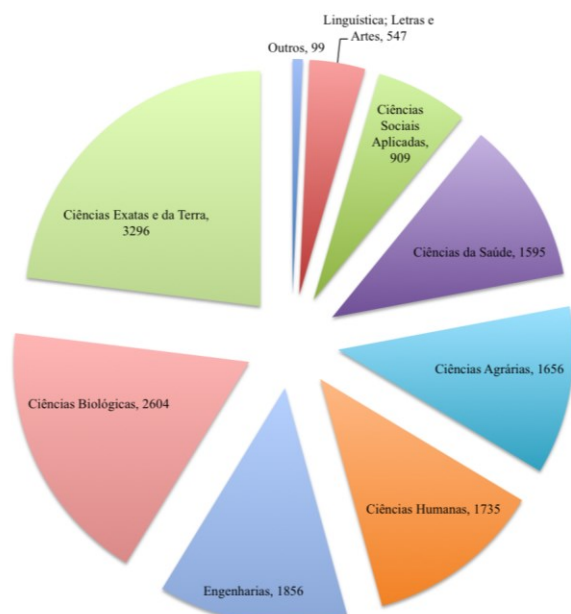
Os mapas foram gerados, utilizando como parâmetro a instituição atual informada no currículo dos pesquisadores, logo, todas as colaborações realizadas (independente do período) foram acrescidas a instituição atual a qual estão vinculados. Quando o pesquisador migra de uma instituição, a atual recebe todas as suas produções realizadas nas instituições anteriores, já que não neste trabalho não é possível vincular a colaboração a instituição da época.

Em função do objetivo do presente trabalho (gerar mapas de colaboração institucionais), os currículos que não possuem instituição de atuação registrado foram descartados, pois a ausência das informações de sua instituição, impossibilita a manipulação das informações nos mapas.

#### 4 RESULTADOS OBTIDOS

Para a análise e processamento dos dados aqui apresentados, utilizou-se como amostra os bolsistas de produtividade em pesquisa do CNPq, bem como suas respectivas colaborações. A escolha dos bolsistas justifica-se por serem os pesquisadores de excelência do país e tendo em vista que um dos critérios para receber a bolsa é manter o currículo Lattes atualizado, logo estes se caracterizam como um excelente conjunto para análise.

Além da definição da amostra, utilizou-se como filtro, a grande área Ciências Exatas e da Terra, por ser a com o maior número de bolsistas, conforme representado no gráfico 1. Os bolsistas dessa grande área, correspondem a 23% do total de bolsistas do país, sendo todos eles doutores.



**Gráfico 1:** Bolsistas por grande área.

**Fonte:** Os autores.

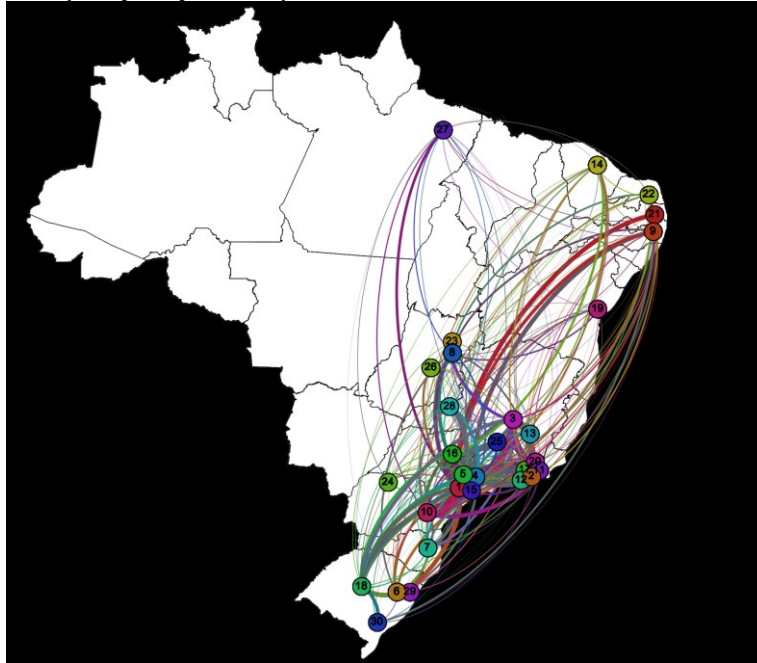
Para melhor representação visual e apresentação tabular dos resultados, apenas as colaborações entre 30 instituições de pesquisa com o maior número de bolsistas foram consideradas. O critério para definição das instituições foi o de bolsistas a elas vinculados. O ranking das 30 instituições está representado na tabela 1.

**Tabela 1:** Instituições de pesquisa selecionadas.

Código	Instituição	Sigla	Nº. de bolsistas
1	Universidade de São Paulo	USP	499
2	Universidade Federal do Rio de Janeiro	UFRJ	213
3	Universidade Federal de Minas Gerais	UFMG	153
4	Universidade Estadual Paulista Júlio de Mesquita Filho	UNESP	144
5	Universidade Estadual de Campinas	UNICAMP	200
6	Universidade Federal do Rio Grande do Sul	UFRGS	160
7	Universidade Federal de Santa Catarina	UFSC	71
8	Universidade de Brasília	UNB	79
9	Universidade Federal de Pernambuco	UFPE	89
10	Universidade Federal do Paraná	UFPR	75
11	Universidade Federal Fluminense	UFF	97
12	Universidade do Estado do Rio de Janeiro	UERJ	37
13	Universidade Federal de Viçosa	UFV	27
14	Universidade Federal do Ceará	UFC	67
15	Universidade Federal de São Paulo	UNIFESP	21
16	Universidade Federal de São Carlos	UFSCAR	64
17	Fundação Oswaldo Cruz	FIOCRUZ	6
18	Universidade Federal de Santa Maria	UFSM	41
19	Universidade Federal da Bahia	UFBA	51
20	Pontifícia Universidade Católica do Rio de Janeiro	PUC-RIO	57
21	Universidade Federal da Paraíba	UFPB	51
22	Universidade Federal do Rio Grande do Norte	UFRN	48
23	Empresa Brasileira de Pesquisa Agropecuária	EMBRAPA	7
24	Universidade Estadual de Maringá	UEM	32
25	Universidade Federal de Lavras	UFLA	12
26	Universidade Federal de Goiás	UFG	34
27	Universidade Federal do Pará	UFPA	28
28	Universidade Federal de Uberlândia	UFU	36
29	Pontifícia Universidade Católica do Rio Grande do Sul	PUCRS	14
30	Universidade Federal de Pelotas	UFPEL	20

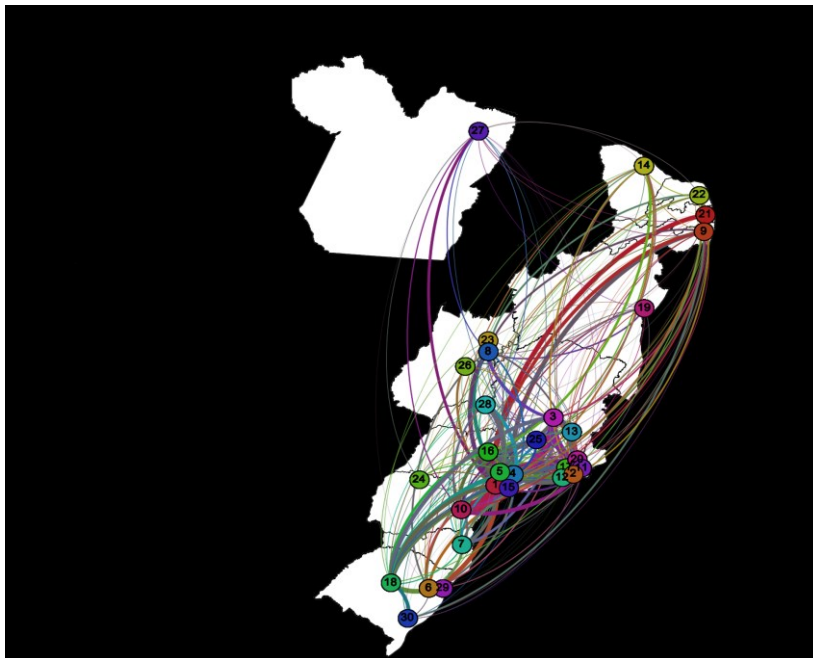
**Fonte:** Os autores.

Os mapas de colaboração, demonstrados nas figuras 3 e 4, representam todo processamento e análise realizada sobre os dados da grande área Ciências Exatas e da Terra. Neles constam apenas as colaborações interinstituições (entre pesquisadores de instituições distintas). Informações relevantes podem ser extraídas através da sua análise visual, como por exemplo, a maior concentração das publicações nas regiões sudeste do país, intensificadas em sua grande maioria por instituições federais, como: Universidade de São Paulo, Universidade Federal do Rio de Janeiro e Universidade Federal de Minas Gerais. Em contra partida, temos a imaturidade da rede, tendo em vista a relação direta de dependência de instituições para proliferação do conhecimento.



**Figura 3:** Mapa de colaboração interinstituições de todo o país.

**Fonte:** Os autores.



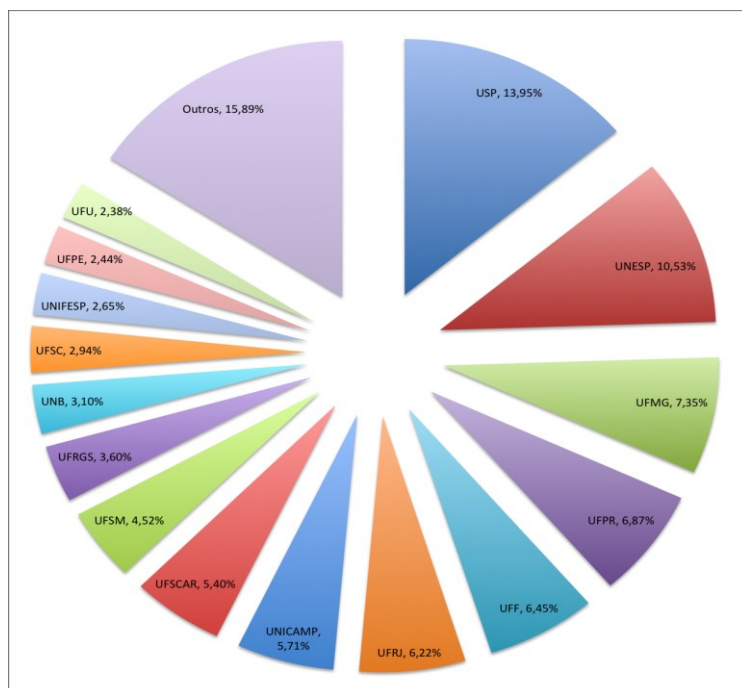
**Figura 4:** Mapa dos estados com colaborações.

**Fonte:** Os autores.

O gráfico 2 demonstra a representatividade das principais instituições de pesquisa no processo de proliferação do conhecimento. Nele, para melhor visualização, as instituições que possuem um percentual de representatividade inferior a 2%, estão agrupadas em outros.



Em sequência, o ranking das colaborações interinstituições, representado na tabela 3, complementa as informações apresentadas no gráfico.



**Gráfico 2:** Representatividade das instituições nas colaborações interinstituições.

**Fonte:** Os autores.

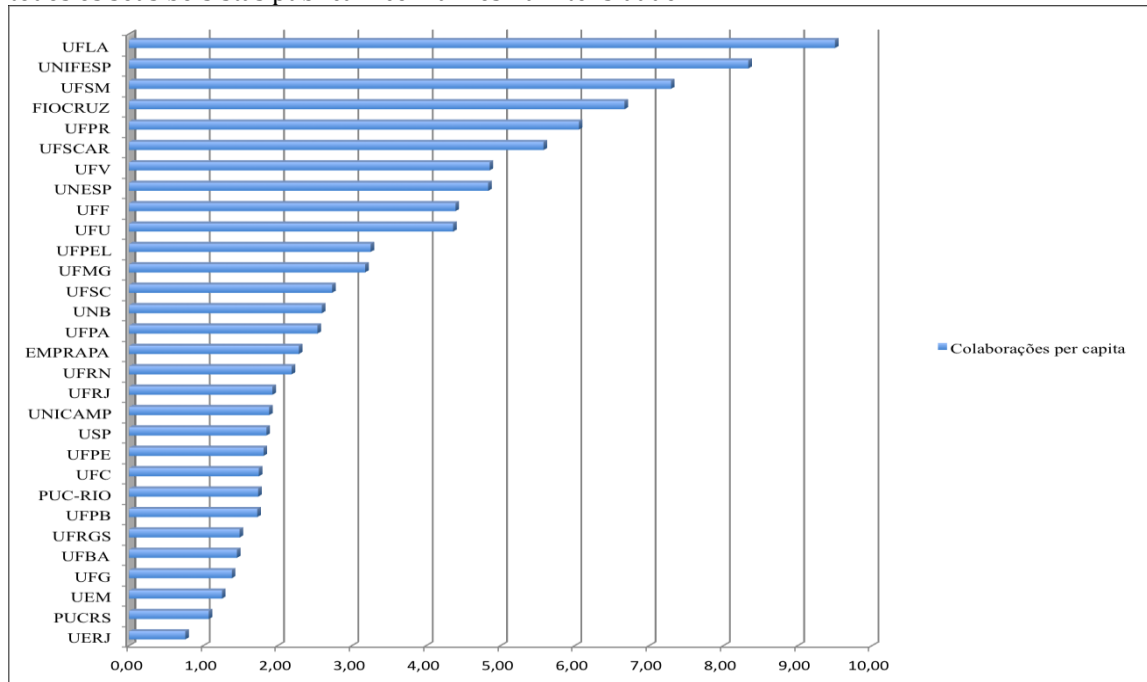
**Tabela 2:** Ranking de colaborações interinstituições.

Instituição	N. de colaborações	%
USP	922	13,95%
UNESP	696	10,53%
UFMG	486	7,35%
UFPR	454	6,87%
UFF	426	6,45%
UFRJ	411	6,22%
UNICAMP	377	5,71%
UFSCAR	357	5,40%
UFSM	299	4,52%
UFRGS	238	3,60%
UNB	205	3,10%
UFSC	194	2,94%
UNIFESP	175	2,65%
UFPE	161	2,44%
UFU	157	2,38%
UFV	131	1,98%
UFC	117	1,77%
UFLA	114	1,73%
UFRN	105	1,59%
PUC-RIO	99	1,50%
UFPB	88	1,33%
UFBA	74	1,12%
UFPA	71	1,07%
UFPEL	65	0,98%
UFG	47	0,71%
FIOCRUZ	40	0,61%
UEM	40	0,61%
UERJ	28	0,42%
EMPRAPA	16	0,24%
PUCRS	15	0,23%

**Fonte:** Os autores.



Pelo gráfico 3, verifica-se que quando a análise das colaborações é realizada de forma proporcional ao número de bolsistas que cada instituição possui, há modificação no ranking. Como exemplo, pode se mencionar a USP, que tem um número considerável de colaborações em função da quantidade de bolsistas que possui, pois quando a análise é realizada de forma per capita, sua representatividade diminui o que comprova que nem todos os seus bolsistas publicam com a mesma intensidade.



**Gráfico 3:** Colaborações per capita.

**Fonte:** Os autores.

Importante mencionar que a proximidade das instituições de pesquisa em alguns estados, prejudica a visibilidade de suas respectivas colaborações no mapa apresentado. Como exemplo, as instituições de pesquisa do estado do Rio de Janeiro. O estado possui cinco instituições dentre o conjunto selecionado, e destas, quatro estão em sua capital. Portanto, ao gerar o grafo utilizando as coordenadas geográficas, em função da escala do mapa político do Brasil, as instituições ficam praticamente sobrepostas. Quanto as colaborações, a UFRJ e UERJ, por exemplo, possuem 20 entre si, no mapa tal informação não pode ser visualizada. Situação semelhante também ocorre no estado de São Paulo.

Esta situação de difícil solução, contribui com a ausência de trabalhos correlatos que tratem instituições de pesquisa dentro de um contexto nacional. Em virtude da dificuldade mencionada e para complementar à informação visual apresentada no mapa, uma matriz de colaborações foi gerada e está sendo representada nas figuras 5 e 6. Na matriz é possível verificar com quais instituições de pesquisa houve as colaborações com suas respectivas quantidades.

	USP	UFRJ	UFMG	UNESP	UNICAMP	UFRGS	UFSC	UNB	UFPE	UFPR	UFF	UERJ	UFV	UFC	UNIFESP	UFSCAR
USP	488	37	58	126	93	32	26	30	25	92	56	2	16	13	44	84
UFRJ	37	161	38	33	26	20	13	13	5	19	63	20	6	11	7	14
UFMG	58	38	235	51	31	10	19	24	12	33	56	0	19	8	12	29
UNESP	126	33	51	217	42	16	16	13	32	75	52	0	19	9	42	57
UNICAMP	93	26	31	42	159	9	9	26	11	31	14	0	4	10	3	11
UFRGS	32	20	10	16	9	276	19	13	3	17	8	2	2	3	3	5
UFSC	26	13	19	16	9	19	60	6	2	15	12	0	2	1	3	5
UNB	30	13	24	13	26	13	6	171	15	1	4	0	1	1	0	0
UFPE	25	5	12	32	11	3	2	15	115	11	7	0	2	1	3	5
UFPR	92	19	33	75	31	17	15	1	11	118	31	0	11	8	10	24
UFF	56	63	56	52	14	8	12	4	7	31	79	2	15	6	12	26
UERJ	2	20	0	0	0	2	0	0	0	0	2	51	0	0	0	0
UFV	16	6	19	19	4	2	2	1	2	11	15	0	47	4	4	8
UFC	13	11	8	9	10	3	1	1	1	8	6	0	4	117	2	18
UNIFESP	44	7	12	42	3	3	3	0	3	10	12	0	4	2	11	10
UFSCAR	84	14	29	57	11	5	5	0	5	24	26	0	8	18	10	100
FIOCRUZ	0	7	7	2	5	0	2	5	0	2	6	0	2	0	0	0
UFSM	28	13	29	34	4	35	9	0	6	23	29	0	8	4	8	19
UFBA	7	6	3	6	6	7	8	12	0	1	0	1	0	0	0	1
PUC-RIO	8	23	10	8	4	2	4	1	5	6	8	0	2	1	2	4
UFPB	33	3	4	1	2	2	8	5	5	1	0	0	0	2	2	8
UFRN	14	11	3	6	1	1	3	14	4	7	0	0	0	7	0	1
EMPRAPA	3	0	0	0	0	0	0	7	0	0	1	0	0	0	0	4
UEM	8	1	5	7	8	0	0	0	0	0	0	0	0	0	2	6
UFLA	10	8	12	16	14	3	3	0	3	13	7	0	4	1	2	6
UFG	16	7	2	4	5	2	0	1	0	0	0	0	0	0	2	0
UFPA	20	3	2	2	5	2	0	8	1	9	2	1	0	3	0	0
UFU	38	3	9	26	3	3	7	5	3	14	6	0	2	2	2	7
PUCRS	0	0	0	0	0	13	2	0	0	0	0	0	0	0	0	0
UFPEL	3	1	0	1	0	6	0	0	0	0	3	0	0	2	0	5

**Figura 5:** Matriz de colaborações.

**Fonte:** Os autores.

	FIOCRUZ	UFSM	UFBA	PUC-RIO	UFPB	UFRN	EMPRAPA	UEM	UFLA	UFG	UFPA	UFU	PUCRS	UFPEL
USP	0	28	7	8	33	14	3	8	10	16	20	38	0	3
UFRJ	7	13	6	23	3	11	0	1	8	7	3	3	0	1
UFMG	7	29	3	10	4	3	0	5	12	2	2	9	0	0
UNESP	2	34	6	8	1	6	0	7	16	4	2	26	0	1
UNICAMP	5	4	6	4	2	1	0	8	14	5	5	3	0	0
UFRGS	0	35	7	2	2	1	0	0	3	2	2	3	13	6
UFSC	2	9	8	4	8	3	0	0	3	0	0	7	2	0
UNB	5	0	12	1	5	14	7	0	0	1	8	5	0	0
UFPE	0	6	0	5	5	4	0	0	3	0	1	3	0	0
UFPR	2	23	1	6	1	7	0	0	13	0	9	14	0	0
UFF	6	29	0	8	0	0	1	0	7	0	2	6	0	3
UERJ	0	0	1	0	0	0	0	0	0	0	1	0	0	0
UFV	2	8	0	2	0	0	0	0	4	0	0	2	0	0
UFC	0	4	0	1	2	7	0	0	1	0	3	2	0	2
UNIFESP	0	8	0	2	2	0	0	2	2	2	0	2	0	0
UFSCAR	0	19	1	4	8	1	4	6	6	0	0	7	0	5
FIOCRUZ	9	0	0	0	0	0	0	0	0	0	0	0	0	2
UFSM	0	127	4	4	0	7	1	0	4	0	0	5	0	25
UFBA	0	4	51	0	0	0	0	0	0	4	4	2	0	2
PUC-RIO	0	4	0	24	0	1	0	0	1	1	0	4	0	0
UFPB	0	0	0	0	66	6	0	0	0	2	1	2	0	1
UFRN	0	7	0	1	6	44	0	0	2	0	1	2	0	14
EMPRAPA	0	1	0	0	0	0	0	0	0	0	0	0	0	0
UEM	0	0	0	0	0	0	0	69	0	1	1	1	0	0
UFLA	0	4	0	1	0	2	0	0	18	0	0	5	0	0
UFG	0	0	4	1	2	0	0	1	0	25	0	0	0	0
UFPA	0	0	4	0	1	1	0	1	0	0	10	6	0	0
UFU	0	5	2	4	2	2	0	1	5	0	6	57	0	0
PUCRS	0	0	0	0	0	0	0	0	0	0	0	0	7	0
UFPEL	2	25	2	0	1	14	0	0	0	0	0	0	0	59

**Figura 6:** Continuação da matriz de colaborações.

**Fonte:** Os autores.

Como exemplo de possibilidade de informações que podem ser obtidas através da exploração qualitativa da matriz apresentada, realizou-se uma análise, com a finalidade de identificar porque a maior parte das colaborações realizadas pela PUCRS ocorreram com a UFRGS. Os fatores primordiais foram: pesquisadores terem pelo menos uma formação acadêmica em conjunto, ou por publicarem com seus orientadores.

A tabela 3 representa o ranking das colaborações intra-instituição (entre pesquisadores da mesma instituição). Embora estas colaborações não sejam o foco principal do trabalho, sua análise foi fundamental para posterior comparação entre as colaborações interinstituições, possibilitando verificar não apenas a representatividade de cada instituição, mas também como ocorreram as suas colaborações.

**Tabela 3:** Ranking de colaborações intra-instituição.

Instituição	N. de colaborações	%
USP	488	16,43%
UFRGS	276	9,29%
UFMG	235	7,91%
UNESP	217	7,30%
UNB	171	5,76%
UFRJ	161	5,42%
UNICAMP	159	5,35%
UFSM	127	4,27%
UFPR	118	3,97%
UFC	117	3,94%
UFPE	115	3,87%
UFSCAR	100	3,37%
UFF	79	2,66%
UEM	69	2,32%
UFPB	66	2,22%
UFSC	60	2,02%
UFPEL	59	1,99%
UFU	57	1,92%
UERJ	51	1,72%
UFBA	51	1,72%
UFV	47	1,58%
UFRN	44	1,48%
UFG	25	0,84%
PUC-RIO	24	0,81%
UFLA	18	0,61%
UNIFESP	11	0,37%
UFPA	10	0,34%
FIOCRUZ	9	0,30%
PUCRS	7	0,24%
EMPRAPA	0	0,00%

**Fonte:** Os autores.

As colaborações das instituições selecionadas, correspondem a um total de 6275. Destas, 3304 (52,65%) são interinstituições e 2971 (47,35%) intra-instituição, ou seja, a diferença apresentada é relativamente pequena.

A tabela 4 detalha como as instituições de pesquisa contribuíram através das suas colaborações. Nesta é possível verificar que as instituições de pesquisa UFRGS, UERJ e UEM; possuem menos de 50% das suas publicações com instituições distintas, ou seja, a maior parte do conhecimento gerado nas instituições ocorre entre seus próprios pesquisadores. Em contra partida, EMPRAPA (100%), UNIFESP (94%), UFPA (88%) e FIOCRUZ (82%) tem grande parte do conhecimento provido realizado de forma conjunta com outras instituições.

Analisando de forma isolada a instituição EMPRAPA, verifica-se que grande parte dos seus bolsistas selecionados, são da área de química e têm pelo menos uma de suas formações acadêmicas na USP, porém as datas de formação são distintas, assim como a data de ingresso a instituição, o que possivelmente justifica o fato de todas as colaborações serem realizadas com outras instituições de pesquisa.

As instituições USP, UNESP e UFMG que estão no topo do ranking de colaborações interinstituições, possuem respectivamente os seguintes percentuais: 65%, 76% e 67%; ou seja, mesmo tendo uma intensidade considerável nas colaborações com outras instituições, ainda possuem um percentual relevante de publicações intra-instituição.

**Tabela 4:** Detalhamento das colaborações.

Instituição	T. de colaborações	Intra-Instituição	%	Interinstituições	%
USP	1410	488	35%	922	65%
UFRJ	572	161	28%	411	72%
UFMG	721	235	33%	486	67%
UNESP	913	217	24%	696	76%
UNICAMP	536	159	30%	377	70%
UFRGS	514	276	54%	238	46%
UFSC	254	60	24%	194	76%
UNB	376	171	45%	205	55%
UFPE	276	115	42%	161	58%
UFPR	572	118	21%	454	79%
UFF	505	79	16%	426	84%
UERJ	79	51	65%	28	35%
UFV	178	47	26%	131	74%
UFC	234	117	50%	117	50%
UNIFESP	186	11	6%	175	94%
UFSCAR	457	100	22%	357	78%
FIOCRUZ	49	9	18%	40	82%
UFSM	426	127	30%	299	70%
UFBA	125	51	41%	74	59%
PUC-RIO	123	24	20%	99	80%
UFPB	154	66	43%	88	57%
UFRN	149	44	30%	105	70%
EMPRAPA	16	0	0%	16	100%
UEM	109	69	63%	40	37%
UFLA	132	18	14%	114	86%
UFG	72	25	35%	47	65%
UFPA	81	10	12%	71	88%
UFU	214	57	27%	157	73%
PUCRS	22	7	32%	15	68%
UFPEL	124	59	48%	65	52%

**Fonte:** Os autores.

## 5 CONCLUSÕES

Através das técnicas aplicadas, foi possível analisar um grande volume de dados e visualizar a representatividade dos principais institutos de pesquisa do país, no processo evolutivo da ciência referente a grande área Ciências Exatas e da Terra. Demonstrando dessa forma a viabilidade da adoção do processo de análise de redes de colaborações institucionais, proposto neste trabalho.

Pelo estudo realizado, conclui-se que embora a comunidade científica brasileira esteja cada vez mais receptiva as análises de redes de colaboração, o percentual de publicações entre pesquisadores da mesma instituição ainda é relativamente alto, contribuindo com o baixo intercâmbio de informações, dificultando a proliferação do conhecimento e conseqüentemente a evolução científica. Há ainda uma relação direta de dependência de algumas instituições na consolidação da rede, caracterizando-as como redes pouco conectadas.

## REFERÊNCIAS

BEAVER, Donald B.; ROSEN, Robert. Studies in Scientific Collaboration: part I: the professional origins of scientific co-authorship. **Scientometrics**, Budapeste, v. 1, n. 1, p. 65-84, 1978.

DIAS, Thiago M. R.; MOITA, Gray F.; DIAS, Patrícia M.; MOREIRA, Tales H. J. Identificação e Caracterização de Redes Científicas de Dados Curriculares. **iSys: Revista Brasileira de Sistemas de Informação**, Rio de Janeiro, v. 7, n. 3, p. 5-18, 2014.

DIAS, Thiago M. R.; MOITA, Gray F. Method for Identification of Collaborations in Large Scientific Databases. **Em Questão**, Porto Alegre, v. 21, n. 2, p. 140-161, 2015.

GOFFMAN, William; WARREN, Kenneth S. **Scientific information systems and the principle of selectivity**. New York: Praeger, 1980. p. 127.

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

NUDELMAN, Arthur. E.; LANDERS, Clifford. E. The failure of 100 divided by 3 to equal 33 1/3. **The American Sociologist**, v. 7, n. 9, p. 9-11, 1972.

SIDONE, Otávio. J. G. Análise Espacial do Conhecimento no Brasil: Parte II - Redes de Colaboração Científica. **Boletim de Informações FIPE - Temas de Economia Aplicada**, n. 400, p. 19-27, 2014.

Editores do artigo: Adilson Luiz Pinto, Rafaela Paula Schmitz e Enrique Muriel-Torrado