

Análise da Rede de Relacionamentos dos Doutores Brasileiros

Luciano A. Digiampietri¹, Caio M. Alves¹, Caio C. Trucolo¹,
Karina V. Delgado¹, Rogério Mugnaini¹

¹Escola de Artes, Ciências e Humanidades – Universidade de São Paulo (USP)
Av. Arlindo Béttio, Ermelino Matarazzo – 03828-000 – São Paulo – SP – Brasil

digiampietri@usp.br

Abstract. *The study of Brazilian academic characteristics is very important task for the understanding of national scientific production and the creation of effective science policy. It is a quite challenging activity due to the enormous cultural and geographical diversity of the country. This work aims to study the academic network of PhDs working in Brazil in order to identify characteristics of both national and state networks.*

Resumo. *O estudo das características acadêmicas brasileiras é uma atividade muito importante para o entendimento da produção científica nacional e para a criação de políticas científicas efetivas, porém é uma atividade bastante desafiadora devido à diversidade cultural e geográfica do país. Este trabalho visa a estudar a rede acadêmica dos doutores que atuam no Brasil a fim de se identificar características tanto da rede nacional como das redes estaduais.*

1. Introdução

Estudar qualquer característica que envolva o Brasil como um todo é uma atividade bastante desafiadora por causa de sua dimensão continental e suas enormes diversidades cultural e social. Em termos acadêmico-científicos, diversos esforços têm sido feitos para tentar potencializar o desenvolvimento nas regiões que possuem uma menor quantidade de doutores ou centros universitários ou de pesquisa ainda não consolidados¹.

Alguns estudos sugerem que tão importante (ou talvez ainda mais importante) do que facilitar o financiamento da ciência em centros ainda em consolidação é inserir (de maneira real e efetiva) estes centros nas redes acadêmicas e para isto, visitas de média ou longa duração de pesquisadores de destaque podem ser extremamente eficientes.

Porém, antes da definição das políticas científicas a serem adotadas, é fundamental um profundo entendimento do cenário acadêmico nacional. Neste sentido, este artigo visa a contribuir para este entendimento no que tange as redes acadêmicas formadas pelos doutores que atuam no Brasil.

Neste artigo são criadas e analisadas redes sociais acadêmicas a partir dos relacionamentos presentes nos currículos Lattes dos doutores brasileiros que atuam no Brasil. Além disso, técnicas de mineração de texto são aplicadas para identificar quais são os assuntos que se destacam nas publicações destes doutores.

¹Por exemplo, através da reserva de parte das verbas de editais nacionais de financiamento de projetos de pesquisa para estados ou regiões específicos ou mesmo a criação de editais que exijam a colaboração de centros de pesquisa menores com centros de pesquisa consolidados.

2. Metodologia

Este trabalho foi dividido em quatro atividades: obtenção dos dados e seleção da amostra; seleção das informações de interesse; cálculo de métricas; e análise dos resultados.

Obtenção dos dados e seleção da amostra. Todos os dados utilizados neste projeto foram obtidos de Currículos Lattes (<http://lattes.cnpq.br/>). Em julho de 2013 foi realizada uma consulta na plataforma Lattes a fim de se identificar todos os currículos cadastrados. Este processo resultou em cerca de 3,2 milhões de currículos, cujos arquivos XML foram baixados. Foram então selecionados os currículos que atendessem aos seguintes critérios: possuir doutorado na formação e possuir como endereço profissional atual um endereço no Brasil. Assim, os currículos de 156.278 doutores foram selecionados.

Seleção das informações de interesse. Para as análises realizadas neste artigo, as seguintes informações foram consideradas: endereço profissional; título dos artigos publicados em periódicos e em anais de eventos; ligações explícitas entre os currículos (cada currículo pode possuir diversas referências a outros currículos: coautores; coparticipantes de projetos; orientadores; orientandos e coparticipantes em bancas e comissões julgadoras). Todas as ligações foram consideradas para a montagem das redes sociais.

Cálculo de métricas. Dois tipos de métricas foram utilizadas neste artigo: métricas oriundas da análise de redes sociais/teoria dos grafos (Tabela 1) e métricas de mineração de textos (para identificar as expressões mais usadas nas publicações). Todas as métricas foram calculadas para 28 grupos: um composto pela rede social de todos dos 156.278 doutores; e 27 conjuntos correspondendo aos estados e ao Distrito Federal.

Análise dos resultados. As redes sociais acadêmicas construídas foram analisadas de maneira comparativa, conforme será discutido na próxima seção.

Tabela 1. Métricas Oriundas da Teoria dos Grafos Utilizadas

Métrica	Descrição
Nós	Número de nós presente na rede atual.
Arestas	Número de arestas presente na rede atual.
Nós no Componente Gigante	Número de nós no componente gigante (maior componente conexo).
Porcentagem de Nós no Componente Gigante	Porcentagem dos nós no componente gigante em relação a todos os nós da rede atual.
Densidade	Número de arestas do grafo atual em relação ao número máximo possível.
Grau Médio	Grau médio dos nós da rede atual.
Coefficiente de Clusterização	Quantidade de cliques de tamanho três dividida pela quantidade de três nós conectados.
Assortatividade de Grau	Métrica que calcula a tendência de nós de se conectarem a nós de mesmo grau (1 indica que todos os nós se conectam apenas a nós de mesmo grau e -1 indica que todos os nós se conectam a nós de grau diferentes dos seus).
Centralização de Grau	Métrica derivada da centralidade de grau que mede o quão central o nó mais central é em relação a todos os outros nós da rede, baseada no grau dos nós.
Centralização de Proximidade	Métrica derivada da centralidade de proximidade (<i>closeness</i>) que mede o quão central o nó mais central é em relação a todos os outros nós da rede, baseada na distância existente entre todos os pares de nós.
Diâmetro	Diâmetro da rede (grafo) atual.
Tamanho da Clique Máxima	Tamanho da maior clique do grafo atual, isto é, tamanho do subconjunto máximo de nós no qual todos os elementos do conjunto estão ligados uns aos outros.
Média dos Caminhos Mínimos	Média dos caminhos mínimos entre todos os pares de nós no componente gigante.

3. Análise dos Resultados

As redes sociais da Figura 1 têm como nós cada uma das cidades que possuem ao menos um doutor. Na rede da esquerda são apresentadas apenas as arestas que ligam duas cidades do mesmo estado e indicam que há doutores nestas cidades que se relacionam. É possível observar que os estados das regiões Sul e Sudeste brasileiros são aqueles que possuem doutores atuando em praticamente todo o seu território enquanto que no Norte e Centro-Oeste existem poucas (e dispersas) cidades onde há atuação de doutores. Na rede da direita são apresentadas todas as arestas entre as cidades (em cinza estão as arestas entre cidades de estados diferentes).

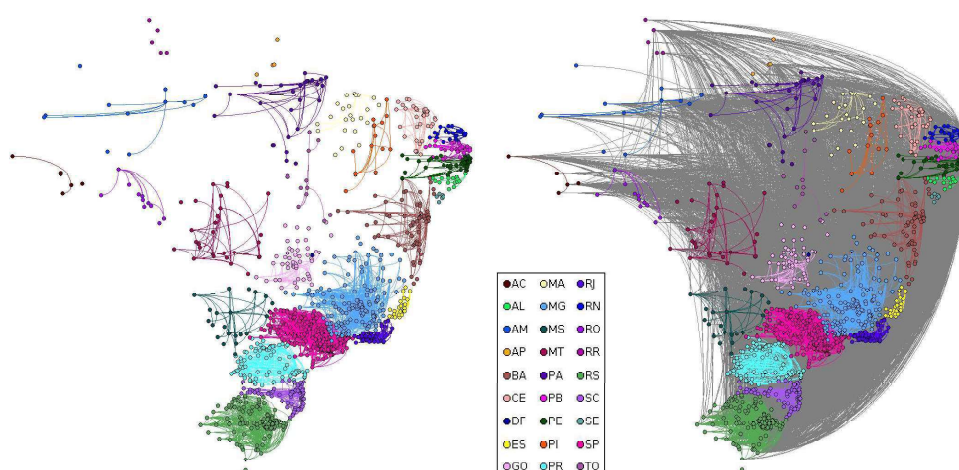


Figura 1. Rede Social dos Doutores - Cidades

Nas redes sociais da Figura 2, cada um dos nós representa um dos 156.278 doutores, os quais estão posicionados nas proximidades de sua cidade de atuação (de forma a minimizar a sobreposição entre doutores que atuam numa mesma cidade). Estas redes destacam a grande quantidade de doutores nas regiões Sudeste, Sul e parte do Nordeste, mas também mostram polos de doutores espalhados por todos os estados brasileiros, a maioria deles encontram-se nas capitais dos estados.

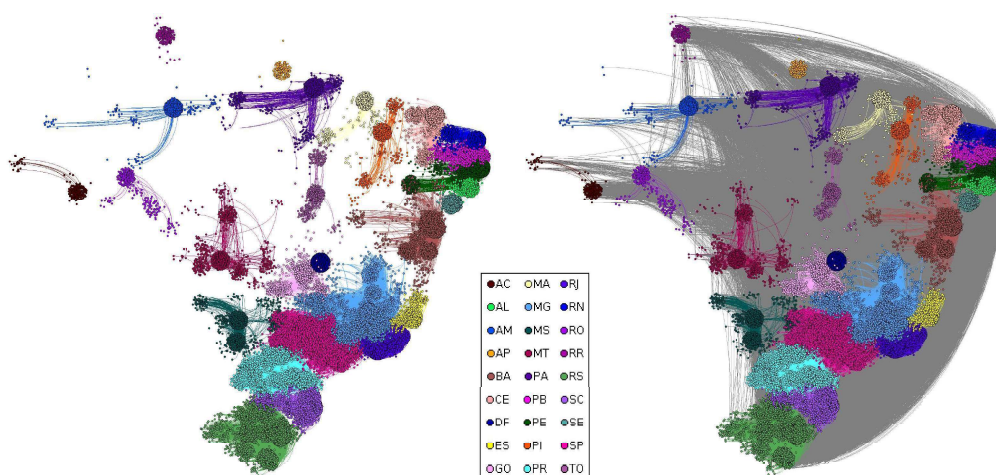


Figura 2. Rede Social dos Doutores que Atuam no Brasil

A Tabela 2 apresenta as métricas da rede nacional e das 27 redes (estaduais e do Distrito Federal). A primeira métrica apresenta o número de doutores (nós da rede) e a segunda o número de arestas em cada uma das redes. A maioria dos doutores atuando no país (53,2%) estão em São Paulo, Rio de Janeiro e Minas Gerais. Ao se adicionar também os doutores do Rio Grande do Sul, Paraná, Pernambuco e Santa Catarina temos quase 75% do total. Por outro lado, ao se somar as porcentagens de doutores do Amapá, Roraima, Acre, Rondônia e Tocantins tem-se menos de 1% do total.

A métrica *porcentagem de nós no componente gigante* contém a porcentagem dos nós que pertencem ao maior componente conexo de cada rede. Nota-se que 75,94% dos nós estão na componente gigante da rede nacional (118.678 nós). Apenas na rede do Rio

Tabela 2. Métricas Calculadas para cada Rede Social Produzida

	Nós	Arestas	Nós no Componente Gigante	Porcentagem de Nós no Componente Gigante	Densidade	Grau Médio	Coefficiente de Clusterização	Assortatividade de Grau	Centralidade de Grau	Centralidade de Proximidade	Diâmetro	Tamanho da Clique Máxima	Média dos Caminhos Mínimos
AC	317	110	24	7,57%	0,00220	0,694	0,342	0,164	0,071	0,003	8	4	3,19
AL	1109	850	379	34,17%	0,00138	1,533	0,214	-0,053	0,096	0,003	13	5	5,40
AM	1583	1595	706	44,60%	0,00127	2,015	0,195	-0,052	0,043	0,002	13	7	5,17
AP	161	55	40	24,84%	0,00427	0,683	0,057	-0,308	0,424	0,039	7	3	2,85
BA	5357	7291	2570	47,97%	0,00051	2,722	0,182	0,021	0,027	0,001	19	8	5,72
CE	3561	8288	2195	61,64%	0,00131	4,655	0,209	0,076	0,039	0,001	19	10	5,14
DF	5421	6791	2567	47,35%	0,00046	2,505	0,191	0,026	0,020	0,001	19	8	5,68
ES	1969	1832	843	42,81%	0,00095	1,861	0,251	0,033	0,040	0,002	19	7	6,91
GO	2987	4071	1458	48,81%	0,00091	2,726	0,232	0,039	0,029	0,001	20	11	5,69
MA	1154	807	334	28,94%	0,00121	1,399	0,323	0,208	0,043	0,003	17	6	5,83
MG	15234	45208	10137	66,54%	0,00039	5,935	0,158	0,067	0,021	0,000	21	11	5,31
MS	1828	2058	808	44,20%	0,00123	2,252	0,188	0,002	0,044	0,002	16	6	4,95
MT	1718	1182	585	34,05%	0,00080	1,376	0,217	0,062	0,044	0,002	17	7	6,05
PA	2576	3399	1405	54,54%	0,00102	2,639	0,194	-0,104	0,054	0,001	15	8	5,68
PB	3488	5163	1741	49,91%	0,00085	2,960	0,208	-0,031	0,042	0,001	17	9	5,32
PE	4842	10060	3106	64,15%	0,00086	4,155	0,178	0,036	0,028	0,001	18	8	5,13
PI	1026	774	345	33,63%	0,00147	1,509	0,212	-0,084	0,115	0,004	12	6	4,68
PR	10307	21472	6577	63,81%	0,00040	4,166	0,185	0,013	0,014	0,000	16	11	5,75
RJ	20639	50368	13638	66,08%	0,00024	4,881	0,146	0,081	0,012	0,000	18	18	5,62
RN	2627	3413	1372	52,23%	0,00099	2,598	0,213	-0,072	0,036	0,001	16	8	5,75
RO	411	203	92	22,38%	0,00241	0,988	0,320	-0,015	0,090	0,008	10	5	4,67
RR	279	62	15	5,38%	0,00160	0,444	0,287	-0,077	0,128	0,004	5	4	2,22
RS	13012	39254	9621	73,94%	0,00046	6,034	0,167	0,080	0,019	0,000	16	13	5,28
SC	5578	9751	3376	60,52%	0,00063	3,496	0,181	0,030	0,022	0,001	22	8	5,60
SE	1193	1165	485	40,65%	0,00164	1,953	0,255	0,117	0,058	0,003	15	8	5,45
SP	47314	179092	35836	75,74%	0,00016	7,570	0,130	0,057	0,007	0,000	18	14	5,24
TO	587	267	116	19,76%	0,00155	0,910	0,323	0,150	0,101	0,007	11	7	4,63
Brasil	156278	641825	118678	75,94%	0,00005	8,214	0,116	0,054	0,004	0,000	18	24	5,50

Grande do Sul e de Minas Gerais há mais de 70% dos nós no componente gigante. Já em Roraima e no Acre menos de 10% dos doutores estão no componente gigante. Estar no componente gigante costuma ser associado a estar mais próximo do principal fluxo de conhecimento/informação de sua rede, então valores elevados para esta métrica costumam ser considerados bons indicativos para as respectivas redes.

A métrica *densidade* representa a razão entre o número de arestas existentes em uma rede e o número máximo possível de arestas para a rede. A densidade da rede nacional é bastante baixa (cerca de 0,00005) e a rede estadual com maior densidade é do Amapá e a de menor é a de São Paulo. Na média, cada doutor da rede nacional se relaciona com outros 8,214 doutores (*grau médio*). Para as sub-redes estaduais destacam-se as redes de São Paulo, Rio Grande do Sul e Minas Gerais com maiores graus médios e Roraima, Amapá e Acre com os menores valores para esta métrica.

A métrica *coeficiente de clusterização* costuma ser associada a estabilidade (ou maturidade) de uma rede [Lemieux and Ouimet 2008] pois mede a quantidade de cliques de tamanho 3 (isto é, três nós ligados uns aos outros) em relação a quantidade de trios de nós onde haja ao menos duas arestas. Seguindo este raciocínio, uma rede totalmente estável teria o valor 1 para esta métrica e uma rede totalmente instável teria o valor 0 (zero). Na rede nacional o coeficiente de clusterização é de apenas 0,116 indicando que a rede de doutores é ainda instável ou imatura. Os três estados que se destacam positivamente de acordo com esta métrica são: Acre, Tocantins, Maranhão e Rondônia. Por outro lado, os menores valores para este coeficiente são do Amapá, São Paulo e Rio de Janeiro.

A métrica *assortatividade de grau* verifica se existe uma tendência de nós de mesmo grau se relacionarem (o valor -1 indica que não há nós de mesmo grau que se relacionam e o valor 1 indica que todas as arestas ocorrem entre nós de mesmo grau). Na rede nacional verifica-se que esta tendência é bastante baixa (0,054). Na rede do Maranhão observa-se o maior valor para esta tendência (0,208) e na rede do Amapá observa-se a maior tendência negativa (-0,308).

As medidas de *centralização de grau e de proximidade (closeness)* são baseadas em medidas de centralidade e servem para indicar o quão importante o nó mais central de cada rede é para a sua rede. Centralizações muito altas em redes sociais costumam não ser muito bem vistas pois indicam que um elemento exerce muita influência (ou centralização de conhecimento) naquela rede e a perda desse elemento pode trazer grandes prejuízos para a rede. O Amapá é o estado cuja rede possui maiores valores de centralização, já nas redes de São Paulo, Rio de Janeiro e Paraná estes valores são muito baixos.

A métrica *diâmetro* indica o maior caminho mínimo existente em uma rede (considerando-se apenas o componente gigante de cada rede). O diâmetro da rede nacional é de 18, significando que existe um par de doutores na rede nacional cujo caminho mínimo necessário para ligá-los passa por 18 doutores. Valores altos de diâmetro costumam indicar redes onde a transmissão de conhecimento é mais lenta. As redes com menor diâmetro são de Roraima (diâmetro igual a 5), Amapá (7) e Acre (8). Observa-se que são redes bastante pequenas. Por outro lado, os maiores diâmetros ocorrem nas redes de Santa Catarina (22), Minas Gerais (21) e Goiás (20). É interessante observar que o diâmetro da rede nacional é menor do que o diâmetro destas últimas redes, isto ocorre porque ligações entre estados “encurtam” o caminho entre alguns doutores de um mesmo estado. Outra medida que costuma ser associada à velocidade de transmissão de conhecimento/informação em redes sociais é a Média dos Caminhos Mínimos. Para esta medida, novamente as redes de Roraima, Amapá e Acre se destacam positivamente (com as menores médias). Por outro lado, as redes do Espírito Santo, Mato Grosso e Maranhão se destacam com os maiores valores para esta média.

Cliques indicam a quantidade de nós onde todos os nós estão ligados um ao outro, em redes sociais, uma clique indica um conjunto de pessoas no qual todos se relacionam. A maior clique da rede nacional é composta por 24 doutores. Já nas redes estaduais destacam-se as redes do Rio de Janeiro, São Paulo e Rio Grande do Sul com as maiores cliques e do Amapá, Roraima e Acre com as menores cliques.

A Tabela 3 apresenta a porcentagem de arestas existentes entre cada par de estados. Nesta tabela, cada linha representa um estado (e a soma de uma linha equivale a 100% das arestas daquele estado). Nas últimas colunas são apresentadas: a quantidade total de arestas do respectivo estado, a porcentagem de arestas que são intra-estado (entre doutores do mesmo estado) e inter-estados (entre doutores de diferentes estados). As células com fundo mais escuro possuem maiores valores. É fácil observar que as células da diagonal principal se destacam (indicando arestas entre doutores de um mesmo estado). Além disso, os estados de São Paulo, Minas Gerais e Rio de Janeiro são aqueles que, proporcionalmente, mais recebem arestas vindas de outros estados. Outros valores que chamam a atenção são: 18% das arestas dos doutores de Santa Catarina são com doutores do Rio Grande Sul, 14% das arestas do Amapá são com o Pará, 12% das arestas da Paraíba são com doutores de Pernambuco e 11% das arestas do Piauí são com o Ceará.

Tabela 4. Expressões Relativamente mais Frequentes em cada Estado

AC	acre brasil	branco acre	rio branco	acre brasil	chico mendes
AL	cidade maceio	maceio alagoas	mundau manguaba	alagoas brasil	alagoas brasil
AM	industrial manaus	cidade manaus	município manaus	manaus amazonas	amazonia central
AP	amapa brasil	amapa brasil	inflammatory analgesic	peso comprimento	osteichthyes characidae
BA	federal bahia	vitoria conquista	feira santana	santana bahia	reconcavo bahia
CE	federal ceara	cidade fortaleza	metropolitana fortaleza	município fortaleza	juazeiro norte
DF	universidade brasilia	distrito federal	brasilia brasil	federal district	spotted wilt
ES	hymenoptera bethylidae	cidade vitoria	sul espirito	vitoria espirito	município vitoria
GO	município jatai	município goiania	cidade goiania	federal goias	goiania goias
MA	cidade luis	ilha luis	luis maranhao	maranhao brasil	município luis
MG	clinicas ufmg	puc minas	município uberlandia	montes claros	federal minas
MS	kaiowa guarani	regiao dourados	sul mato	campo mato	ponta pora
MT	município cuiaba	tangara serra	cuiaba mato	chapada guimaraes	rio cuiaba
PA	metropolitana belem	sudeste paraense	nordeste paraense	cidade belem	município belem
PB	cidade campina	federal paraiba	sertao paraibano	joao pessoa	município campina
PE	cidade Recife	cabano santo	jaboatao guararapes	serra talhada	federal pernambuco
PI	teresina piaui	piaui brasil	serra capivara	piaui brasil	cutias dasyprocta
PR	originado projeto	litoral parana	francisco beltrao	campo mourao	federal parana
RJ	250 gev	cyprinodontiformes rivulidae	puc rio	noroeste fluminense	favelas rio
RN	cidade natal	natal brasil	município mossoro	município natal	natal brasil
RO	porto velho	velho rondonia	rondonia brasil	rio madeira	pastagens degradadas
RR	ferragem composicao	roraima brasil	pastagens degradadas	roraima brasil	andropogon gayanus
RS	universidade passo	clinicas porto	universitario santa	rio sinos	embrapa clima
SC	município florianopolis	florianopolis brasil	ilha santa	planalto catarinense	balneario camboriu
SE	federal sergipe	aracaju sergipe	sergipe brasil	sergipe brasil	hyptis pectinata
SP	clinico radiografica	reserva parque	sentry antimicrobial	clinicas fmpsp	liquido cefalorraqueano
TO	porto nacional	sul tocantins	tocantins brasil	vacas descarte	pastagem cultivada

4. Trabalhos Correlatos

Ao passo que nos últimos anos diversos trabalhos analisaram redes sociais acadêmicas, são poucas as iniciativas que analisaram redes que englobem todo um país e envolvam todas as áreas de conhecimento.

O trabalho de [Mena-Chalco et al. 2014] realizou uma análise focada na rede social de coautorias de um conjunto de mais de 1.100.000 de currículos extraídas do plataforma Lattes. Neste artigo foram analisadas diversas métricas de rede tanto da rede composta por todos os currículos analisados, bem como pelas sub-redes criadas a partir das grandes áreas de atuação declaradas nos currículos. A tese de [da Cunha e Melo 2011] visou a caracterizar a comunidade científica brasileira em relação à produtividade, internacionalização e visibilidade. Para isto os currículos Lattes de 51.080 doutores participantes de grupos de pesquisa foram analisados. Nesse trabalho, o agrupamento dos doutores utilizando-se informação geográfica foi feito por regiões e não por estados. Em [Medeiros and Mena-Chalco 2013] a rede brasileira de pessoas com currículos Lattes cuja grande-área de atuação são as Ciências Humanas é analisada. Neste trabalho são identificadas algumas características da dinâmica da grande área e as palavras mais frequentes nos títulos das publicações são analisadas utilizando-se nuvens de palavras.

Informações da DBLP⁴ foram utilizadas para estudar a rede brasileira de coautorias em ciência da computação [Freire and Figueiredo 2011]. Neste trabalho foram comparadas métricas de rede com a nota da avaliação CAPES atribuída aos programas de pós-graduação. Este tipo de comparação também foi realizado em [Digiampietri et al. 2014], porém neste trabalho foram utilizados dados extraídos da plataforma Lattes dos docentes de programas de pós-graduação em ciência da computação.

O trabalho apresentado neste artigo se diferencia dos demais por apresentar uma análise dos doutores atuando no Brasil mais ampla, envolvendo mais de 150.000 os dou-

de títulos nacionais.

⁴<http://dblp.uni-trier.de/db/>

tores cadastrados na plataforma Lattes. Além da análise da rede nacional, também foram realizadas e comparadas análises das redes estaduais e foram utilizadas técnicas de mineração de textos para identificar peculiaridades nas publicações estaduais.

5. Conclusões

Este artigo apresentou um análise geral das redes sociais dos doutores que atuam no Brasil. A análise realizada utilizou técnicas de análise de rede sociais e mineração de textos para entender características das redes as quais foram criadas a partir das informações extraídas dos currículos Lattes destes doutores.

Observou-se que as redes são heterogêneas, pouco densas, e que provavelmente ainda são instáveis (possuem pouca maturidade). Estas características podem ser justificadas pelo grande tamanho do país, bem como sua diversidade cultural e social. Apesar das redes não serem densas, seus componentes gigantes, de um modo geral, contém a grande maioria dos nós (tanto na rede nacional quanto nas redes estaduais), além disso, a média dos caminhos mínimos também não foi elevada em nenhuma das redes mostrando uma boa conectividade entre os doutores pertencentes aos componentes principais.

A partir da mineração de textos foi possível observar que se destacam expressões regionais (tanto na análise de frequência quanto na análise de frequência relativa). Apesar disto foi possível identificar algumas expressões (não regionais) que se destacaram nas publicações dos doutores de alguns estados. Como trabalhos futuros pretende-se aprofundar a análise realizada.

Agradecimentos

O trabalho realizado neste artigo foi parcialmente financiado pelo Programa de Educação Tutorial do MEC, FAPESP, CNPq, CAPES e Pró-reitoria de graduação da USP.

Referências

- da Cunha e Melo, P. L. (2011). *Produtividade, Internacionalização e Visibilidade da Comunidade Científica Brasileira na Virada do Milênio*. PhD thesis, Universidade Federal do Rio de Janeiro.
- Digiampietri, L. A., j. P. Mena-chalco, Melo, P. O. V., Malheiros, A. P., Meira, D. N. O., Franco, L. F., and Oliveira, L. B. (2014). BraX-Ray: An X-Ray of the Brazilian Computer Science Graduate Programs. *Plos-One*, (no prelo):20.
- Freire, V. and Figueiredo, D. (2011). Ranking in collaboration networks using a group based metric. *Journal of the Brazilian Computer Society*, pages 1–12.
- Lemieux, V. and Ouimet, M. (2008). *Análise Estrutural das Redes Sociais*. Instituto Piaget.
- Medeiros, C. B. and Mena-Chalco, J. P. (2013). The dynamics of multidisciplinary research networks - mining a public repository of scientists CVs. In Council, I. S. S., editor, *World Social Science Forum 2013*, pages 1–17, Montréal, Canada.
- 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*.