

# Caracterizando Departamentos e Programas de Computação Utilizando Análise de Redes Sociais e Bibliometria

Luciano Digiampietri<sup>1</sup>, Ricardo Linden<sup>2</sup>, Lenin Barbosa<sup>1</sup>

<sup>1</sup>Escola de Artes, Ciências e Humanidades – Universidade de São Paulo (EACH-USP)

<sup>2</sup>CEPEL/Faculdade Salesiana Maria Auxiliadora (FSMA)

digiapietri@usp.br, rlinden@pobox.com, lenin.ferreira@gmail.com

**Abstract.** *In this paper we compare the graduate programs in Computer Science of levels 6 and 7 (which CAPES assume to be equivalent to the international best) and those programs best ranked in three international rankings. We used both bibliometric metrics and social networks analysis and found out that there is a great difference between the results achieved by the national and international programs. The results from a principal component analysis show that the Brazilian programs are a class of their own and do not share the characteristics of the best international graduate programs.*

**Resumo.** *Neste artigo apresentamos uma comparação entre os programas de pós-graduação em computação de nível 6 e 7 (os chamados de nível internacional pela CAPES<sup>1</sup>) e os programas melhores colocados em três rankings internacionais. Foram usados métricas bibliométricas e métricas de análise de redes sociais e pôde-se verificar uma grande diferença de resultados entre os programas nacionais e os internacionais. O resultado da análise de componentes principais mostra que os programas brasileiros residem em uma classe própria, não possuindo as características dos melhores programas internacionais.*

## 1. Introdução

A avaliação da qualidade de departamentos ou programas de pós-graduação (muitas vezes relacionada a análises bibliométricas) tem se tornado cada vez mais relevante, uma vez que os recursos de fomento são escassos em comparação com os pesquisadores que os desejam. Assim, necessita-se de uma análise mais abrangente e objetiva, tanto para melhor alocar os recursos quanto para elaborar políticas científicas eficazes.

Entender as características que levam um grupo a obter reconhecimento internacional pode servir de base para o desenvolvimento de outros grupos. Porém, avaliar programas ou departamentos é uma tarefa complexa, envolvendo a análise de diferentes métricas de características quantitativas e qualitativas [Digiampietri et al. 2014], sobre as quais não há consenso. Além da complexidade e possível subjetividade, os trabalhos da literatura usam índices diversos ou fontes de dados diferentes para a análise bibliométrica.

Na avaliação de grupos acadêmicos e, especialmente quando se foca em departamentos ou programas de pós-graduação, há algumas métricas mais utilizadas [Laender et al. 2008], muitas das quais usaremos neste trabalho. Além de análises bibliométricas, a

<sup>1</sup>[https://www.capes.gov.br/images/stories/download/avaliacao/COMPUTACAO\\_05mar10.pdf](https://www.capes.gov.br/images/stories/download/avaliacao/COMPUTACAO_05mar10.pdf)

análise de redes sociais vem sendo utilizada para tentar medir a influência ou importância de um grupo em um dado contexto, bem como identificar padrões de relacionamentos entre os membros de cada grupo. No Brasil, para a avaliação dos programas de pós-graduação a CAPES<sup>2</sup> (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) atribui aos programas de pós-graduação notas de três a sete, sendo que os programas com notas 6 ou 7 são considerados os programas de nível internacional.

Este trabalho visa a caracterizar 28 departamentos ou programas de pós-graduação em ciência da computação (oito brasileiros e vinte internacionais) considerando medidas bibliométricas e oriundas da análise de redes sociais, analisar a correlação entre estas medidas e o ranqueamento provido por três fontes internacionais e avaliar se os programas brasileiros apresentam padrões diferentes dos grupos internacionais avaliados.

O restante deste artigo está organizado da seguinte maneira. A Seção 2 contém a descrição de alguns conceitos básicos e trabalhos correlatos. A Seção 3 apresenta a metodologia utilizada. Na Seção 4 os resultados são apresentados e analisados. Por fim, a Seção 5 contém as conclusões e trabalhos futuros.

## 2. Conceitos Básicos e Trabalhos Correlatos

Os principais conceitos envolvidos no presente trabalho estão ligados a análises bibliométricas e de redes sociais e são descritos nesta seção. Existem diversos termos relacionados à análise quantitativa da produção científica, e dois dos mais usados são **Bibliometria** (*bibliometrics*) e a **Cientometria** (*scientometrics*), que estudam os aspectos quantitativos da produção e da ciência, respectivamente.

### 2.1. Análise de Redes Sociais

Rede social é uma estrutura composta por indivíduos (pessoas ou organizações) que são conectadas por um ou mais tipos de relações, como amizade, crença ou trabalho. O conceito de redes sociais surgiu há séculos nas áreas de humanas, porém, apenas nas últimas décadas, os estudos em redes sociais se intensificaram analisando tanto as características individuais de seus componentes, como também as características estruturais da rede [Lemieux and Ouimet 2008], baseando-se em dados que contemplam milhares ou milhões de indivíduos. Um dos métodos mais comuns para a representação computacional destas redes, também usado aqui, é utilizando grafos, nos quais cada indivíduo da rede social é um nó (ou vértice), e cada relação entre indivíduos é uma aresta, direcionada ou não.

### 2.2. Análise de grupos acadêmicos

Tipicamente, a análise de grupos acadêmicos é realizada em duas vertentes: análises bibliométricas e de redes sociais. Os grupos a serem analisados podem ser formados de diferentes maneiras, como por exemplo, de acordo com a formação/titulação dos indivíduos, sua localização geográfica, com suas áreas de atuação ou com o local/instituição em que trabalha (universidade, departamento ou programa de pós-graduação).

A colaboração científica é influenciada por diversos fatores como proximidade geográfica e os relacionamentos anteriores dos pesquisadores (como uma relação de orientação) [Menezes et al. 2009]. Para descobrir redes sociais acadêmicas e avaliar

<sup>2</sup><http://www.capes.gov.br/>

grupos de pesquisa precisa-se de informação tipicamente dispersa e de diferentes naturezas. Muitos desses tipos de informação são agrupados nos currículos dos pesquisadores, o que ajuda a enfrentar alguns dos desafios encontrados ao se utilizar apenas bases de citações bibliográficas, tais como: homonímia e polissemia, dispersão da produção bibliográfica em diferentes fontes, afiliação do pesquisador, relações de orientação, áreas de interesse, identificação das colaborações, etc.

Ao se pensar na avaliação de toda a rede social acadêmica de um país, o primeiro obstáculo (e muitas vezes um dos maiores) é a obtenção do grande volume de informação relacionada, normalmente distribuída em muitas fontes (como páginas Web, relatórios institucionais, bibliotecas digitais internacionais, etc). No entanto, esta atividade é simplificada no Brasil pela existência da Plataforma Lattes [Digiampietri et al. 2012]. Por outro lado, para a análise de grupos da área de ciência da computação utilizando dados internacionais, a principal fonte utilizada tem sido o projeto DBLP<sup>3</sup> [Laender et al. 2008, Freire and Figueiredo 2011].

### 2.3. Trabalhos correlatos

A relevância desta área pode ser vista no grande número de trabalhos correlatos existentes, entre os quais podemos destacar alguns, descritos a seguir. Por questões de limitação de espaço, não foi possível descrever todos os artigos relevantes da área.

Os estudos acadêmicos mais relacionados ao presente artigo são os trabalhos de caracterização das redes sociais acadêmicas por meio do cálculo de métricas utilizando, por exemplo, teoria dos grafos/teoria de redes [Ulrik and Erlebach 2005]. Estas pesquisas incluem a tentativa de oferecer uma visão geral sobre a produção científica brasileira [Leite et al. 2011] ou grupos de pesquisa específicos [Wainer and Vieira 2013].

Entre os trabalhos com enfoque bibliométrico, algumas das principais medidas bibliométricas utilizadas foram número de publicações, índice-h e número de citações [Hirsch 2005], sendo que alguns trabalhos ponderam a participação dos autores nas publicações de acordo com a posição do pesquisador na lista de autorias [Duffy et al. 2011]. Além da análise de grupos, alguns destes trabalhos também analisam se há influência/correlação entre gênero e as demais medidas.

Já Menezes et al. 2009 combinaram a análise bibliométrica com a de redes sociais para analisar a área de ciência da computação e suas subáreas em várias regiões do mundo (Estados Unidos, Europa e Brasil) no período de 1994 a 2006. Os autores destacaram os aspectos comuns e as diferenças observadas entre as redes de cada região.

O trabalho de Franceschet 2011 usou dados do DBLP a fim de analisar as redes de coautorias na área de ciência da computação e compará-las com dados de redes de outras áreas. Eles observaram que, em comparação com outras áreas, o nível de colaboração (em termos de números de coautores nas publicações) em ciência da computação é baixo.

Em uma análise da produtividade dos departamentos de computação brasileiros, Laender et al. 2008 usaram dados do DBLP e compararam indicadores de programas brasileiros considerados de nível internacional com alguns programas da América do Norte e Europa, concluindo que estes programas brasileiros atingiram a maturidade.

---

<sup>3</sup><http://dblp.uni-trier.de/db/>

Também utilizando dados do DBLP, Freire and Figueiredo 2011 analisaram a rede social acadêmica brasileira na área de ciência da computação. Os autores encontraram *super peers*, isto é, alguns indivíduos com grau muito acima dos demais da rede. Os autores também propuseram uma métrica para a avaliação de redes de colaboração que considera a importância de um nó na conexão de indivíduos de diferentes grupos.

### 3. Metodologia

A metodologia deste trabalho foi dividida em quatro atividades: seleção da amostra; obtenção dos dados; cálculo das métricas; e análise dos resultados. Ao se tratar de avaliação internacional, muitas vezes são considerados departamentos, unidades ou universidades enquanto que na avaliação nacional, tipicamente são considerados os programas de pós-graduação (que podem, inclusive, conter docentes de diferentes universidades). Ao longo deste texto o termo departamento será utilizado, por simplicidade, tanto para se referir a departamentos ou unidades quanto a programas de pós-graduação.

Para o presente trabalho optou-se por **selecionar** vinte departamentos internacionais com destaque na área de ciência da computação. Para isto foram utilizados três ranqueamentos internacionais de 2016, referentes a dados do ano de 2015: Shanghai<sup>4</sup> (considerando os 200 melhores departamentos do mundo), Top Universities<sup>5</sup> e The Guardian<sup>6</sup> (considerando os 100 melhores departamentos do mundo). Adicionalmente, foram inseridos os oito programas de pós-graduação em ciência da computação no Brasil considerados de nível internacional (isto é, com nota 6 ou 7 atribuída pela CAPES). Devido ao objetivo do presente trabalho que visa a caracterizar os principais departamentos da área de computação do mundo e os principais programas de pós-graduação em computação no Brasil, além dos departamentos melhor ranqueados também foram feitas algumas escolhas para manter a diversidade da amostra e para incluir departamentos com ranqueamentos mais parecidos com os programas brasileiros. A Tabela 3 apresenta os 28 departamentos ou programas analisados e suas posições nos três ranqueamentos utilizados. Estes departamentos estão ordenados pelo ranqueamento médio obtido.

Para cada um dos departamentos internacionais, a lista de seus pesquisadores foi obtida do site do respectivo departamento. Para os programas de pós-graduação brasileiros, a lista dos pesquisadores foi obtida do sistema Sucupira<sup>7</sup>. Ao todo, 1.232 nomes de pesquisadores foram identificados.

Neste trabalho foram **obtidos dados** de duas fontes principais. Os dados sobre as publicações dos pesquisadores selecionados foram obtidos do projeto DBLP (*Digital Bibliography & Library Project*). Já os dados sobre as citações das publicações foram obtidos do site *Microsoft Academic Search*<sup>8</sup>, conforme será apresentado a seguir.

Os mantenedores do projeto DBLP realizam um processo de desambiguação de nomes para evitar o cadastro repetido de um mesmo autor e a base conta com mais de um milhão e meio de registros de autores. Para realizar o casamento entre a lista de nomes de pesquisadores de cada departamento e os registros do DBLP adotou-se a estratégia para

<sup>4</sup>Academic Ranking of World Universities: <http://www.shanghairanking.com/>

<sup>5</sup>Worldwide university rankings: <http://www.topuniversities.com/>

<sup>6</sup>The Guardian: <http://www.theguardian.com/education/universityguide>

<sup>7</sup><https://sucupira.capes.gov.br/>

<sup>8</sup><http://academic.research.microsoft.com/Microsoft Academic Search>

**Tabela 1. Departamentos ou Programas de Computação Analisados**

Universidade	Abrev. usada	País	Shanghai	Top Universities	The Guardian
MIT - Electrical Eng. and Comp. Science Dep.	MIT EECS	Estados Unidos	2	1	1
Stanford University	Stanford	Estados Unidos	1	2	2
Harvard - School of Eng. and Applied Sciences	Harvard SEAS	Estados Unidos	4	4	4
Carnegie Mellon University	CMU	Estados Unidos	6	5	4
University of California - Berkeley	UCBerkeley	Estados Unidos	3	7	6
Princeton University	Princeton	Estados Unidos	5	10	9
Swiss Federal Institute of Technology in Zurich	ETH Zurich	Suíça	17	8	9
University of Cambridge	Cam	Reino Unido	23	6	8
University of Toronto	Toronto	Canada	11	11	16
University of Oxford	Oxford	Reino Unido	34	3	3
University of California - Los Angeles	UCLA	Estados Unidos	9	16	17
Hong Kong University of Science and Tech.	HKUST	China	39	14	8
The University of Texas - Austin	UTexas	Estados Unidos	7	34	22
Cornell University	Cornell	Estados Unidos	8	26	31
California Institute of Technology	Caltech	Estados Unidos	11	28	27
University of Illinois	Illinois	Estados Unidos	13	30	30
Chinese University of Hong Kong	CUHK	China	37	22	18
Ecole Polytechnique Federale de Lausanne	EPFL	Suíça	35	18	29
University of Washington	UW	Estados Unidos	51-75	28	20
Universidade Estadual de Campinas	UNICAMP	Brasil	101-150	51-100	51-100
Universidade de São Paulo - São Carlos	USP São Carlos	Brasil	151-200	51-100	51-100
Universidade de São Paulo - São Paulo	USP São Paulo	Brasil			
Boston University	BU	Estados Unidos	101-150	101-150	
Universidade Federal do Rio Grande do Sul	UFRGS	Brasil		151-200	
Universidade Federal do Rio de Janeiro	UFRJ	Brasil		151-200	
Universidade Federal de Minas Gerais	UFMG	Brasil		151-200	
Universidade Federal de Pernambuco	UFPE	Brasil		201-250	
Pontifícia Universidade Católica (RJ)	PUC-Rio	Brasil		201-250	

resolução de nomes apresentada por Mugnaini et al. (2012), que utiliza um sistema de pontuação para identificar o melhor casamento entre dois nomes considerando que eles podem estar no formato utilizado em citações bibliográficas (isto é, com a ausência de alguns “nomes do meio” e/ou com abreviações). Por meio deste processo, foram encontrados 1.170 registros de autores. Estes foram utilizados para identificar 73.232 registros de diferentes artigos (25.360 artigos em periódicos e 47.872 artigos em eventos).

Cada registro de publicação possui a lista de autores representados da maneira que estes autores são cadastrados na base do DBLP. Assim, a partir da lista de publicações é bastante simples criar as redes de coautoria dos pesquisadores envolvidos. Neste projeto foram calculadas métricas de 29 redes sociais de coautoria. A primeira rede possui como nó cada um dos departamentos e as arestas representam relações de coautorias entre pesquisadores de dois departamentos (Figura 1). A legenda de cores usada em todos os grafos apresentados neste artigo é apresentada na Figura 2.

A Figura 1 contém um grafo organizado automaticamente de acordo com as ligações. Neste grafo podemos ver que todos os programas brasileiros estão à esquerda do grafo e apresentam poucas ligações tanto entre si quanto com instituições estrangeiras, indicando um nível mais baixo de pesquisas conjuntas. A densidade de arestas em cada nó representando um departamento internacional é, na média, muito maior do que as nacionais, indicando que estes trabalham mais em cooperação com outras instituições, possivelmente aproveitando sinergias e “importando” ideias novas.

Adicionalmente foi gerada uma rede para cada departamento, na qual cada pesquisador é um nó e cada aresta representa relações de coautoria entre pesquisadores. Apenas

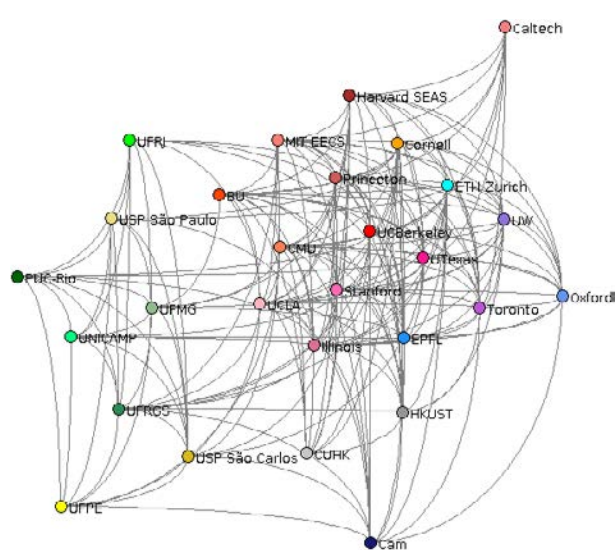


Figura 1. Rede de coautorias

- University of California - Berkeley
- Boston University
- California Institute of Technology
- Carnegie Mellon University
- Cornell University
- MIT - Electrical Engineering and Computer Science Department
- Harvard - School of Engineering and Applied Sciences
- Princeton University
- Stanford University
- The University of Texas - Austin
- University of California - Los Angeles
- University of Illinois
- University of Toronto
- University of Washington
- University of Oxford
- University of Cambridge
- Ecole Polytechnique Federale de Lausanne
- Swiss Federal Institute of Technology in Zurich
- Chinese University of Hong Kong
- Hong Kong University of Science and Technology
- Pontificia Universidade Católica do Rio de Janeiro
- Universidade Federal de Minas Gerais
- Universidade Federal do Rio Grande do Sul
- Universidade Federal do Rio de Janeiro
- Universidade Estadual de Campinas
- Universidade de São Paulo - São Paulo
- Universidade de São Paulo - São Carlos
- Universidade Federal de Pernambuco

Figura 2. Legenda de cores

para ilustrar estas relações entre os pesquisadores estudados, as Figuras 3 e 4 apresentam a rede de coautorias incluindo todos os pesquisadores estudados. Na Figura 3 são apresentadas as coautorias entre pesquisadores de diferentes departamentos (arestas em cinza) e entre pesquisadores do mesmo departamento (arestas coloridas). Já a Figura 4 contém apenas as relações de coautoria entre pesquisadores do mesmo departamento.

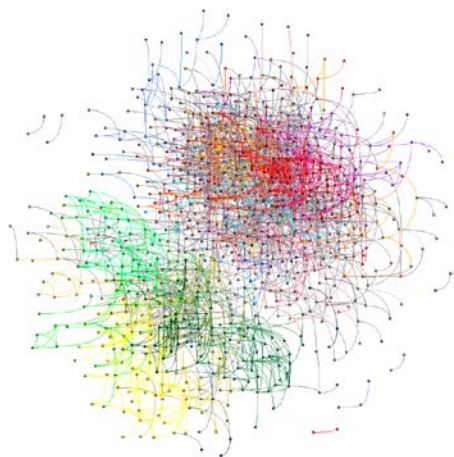


Figura 3. Rede de coautorias - todas as arestas

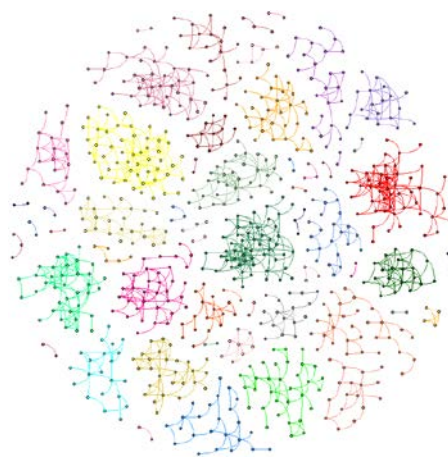


Figura 4. Rede de coautorias - arestas internas

O título de cada um dos artigos identificados foi buscado no site *Microsoft Academic Search*, e criamos um programa que verifica se cada artigo foi encontrado na busca (usando casamento aproximado de *strings* e permitindo uma distância de edição de no máximo 10% do tamanho do título do artigo) e, em caso afirmativo, o programa salva o número de citações recebido pelo artigo. Ao todos, os 73.232 artigos receberam 1.197.416 citações (658.721 para artigos em periódicos e 538.695 para artigos em eventos).

A escolha das métricas aplicadas ao estudo de cada rede social está diretamente relacionada com objetivos específicos do estudo a ser realizado [Scott 2009]. No caso

deste trabalho, com base nos dados obtidos, **dois tipos de métricas foram calculados**: bibliométricas e de análise de redes sociais. As medidas bibliométricas calculadas estão relacionadas ao número de artigos publicados e às citações recebidas. Oito das 16 medidas bibliométricas calculadas são apresentadas na Tabela 3, as demais oito medidas são as mesmas, só que calculadas em uma janela de tempo de cinco anos (de 2010 a 2014).

**Tabela 2. Medidas Bibliométricas Utilizadas**

Abrev. Utilizada	Descrição
Pubs	Quantidade total de publicações em periódicos ou anais de eventos dos pesquisadores de um dado departamento.
PubsCo	Quantidade total de publicações em periódicos ou anais de eventos em coautoria com pesquisadores de outro departamento do estudo.
Cits	Quantidade total das citações às publicações dos artigos do departamento.
CitsCo	Quantidade total das citações às publicações dos artigos do departamento envolvidas em coautoria com pesquisadores de outro departamento.
PubsPP	Taxa de publicações por pesquisador.
PubsCoPP	Taxa de publicações em coautoria por pesquisador.
CitsPP	Taxa de citações por pesquisador.
CitsCoPP	Taxa de citações por pesquisador envolvendo as publicações em coautoria com pesquisadores de outro departamento.

Para a análise de redes sociais, dois tipos de redes de coautoria foram construídas: uma em que cada nó corresponde a um departamento e uma aresta corresponde à relação de coautoria entre pesquisadores dos dois departamentos ligados pela aresta; e uma rede para cada departamento, na qual cada pesquisador corresponde a um nó e as arestas representam relações de coautoria entre pesquisadores de um departamento. Para o primeiro tipo de rede foram calculadas cinco medidas de centralidade individuais (ou locais) de cada nó (departamento), já para o segundo tipo foram calculadas nove métricas globais da rede de cada departamento. A Tabela 3 apresenta as métricas utilizadas.

**Tabela 3. Métricas da Análise de Redes Sociais Utilizadas**

Abrev. Utilizada	Descrição
Medidas Locais	
Betweenness <sub>l</sub>	<b>Centralidade de intermediação (betweeness)</b> - baseada na frequência em que um dado vértice aparece entre todos os caminhos mínimos da rede.
Closeness <sub>l</sub>	<b>Centralidade de proximidade (closeness)</b>
Degree <sub>l</sub>	<b>Centralidade de grau</b> - baseada no grau do vértice.
Eigen <sub>l</sub>	<b>Centralidade de autovalor (eigenvalue)</b> - baseada na importância dos vértices que estão ligados ao vértice em análise.
Page-Rank <sub>l</sub>	<b>Centralidade PageRak</b> - baseada na quantidade e qualidade das ligações existentes envolvendo o vértice em análise.
Medidas Globais	
Clique <sub>g</sub>	<b>Tamanho da clique máxima</b> - tamanho da maior clique na rede.
MCC <sub>g</sub>	<b>Média dos caminhos mínimos</b> - valor médio do caminho mínimo entre todos os pares de vértices da rede.
C.Clusterização <sub>g</sub>	<b>Coefficiente de clusterização</b> - medida da transitividade das relações da rede.
Degree <sub>g</sub>	<b>Centralização de grau</b> - medida do quão importante o vértice mais importante da rede é em relação aos demais (em relação ao grau).
Closeness <sub>g</sub>	<b>Centralização de proximidade</b> - medida do quão importante o vértice mais importante da rede é em relação aos demais (em relação à proximidade).
Eigen <sub>g</sub>	<b>Centralização de autovalor</b> - medida do quão importante o vértice mais importante da rede é em relação aos demais (em relação ao autovalor).
Assort.Grau	<b>Assortatividade de grau</b> - medida da tendência de nós de mesmo grau estarem ligados.
Diâmetro <sub>g</sub>	<b>Diâmetro</b> - diâmetro da rede (distância mínima entre os dois nós mais afastados na rede).
Densidade <sub>g</sub>	<b>Densidade da rede</b> - relação entre o número de arestas da rede e o número máximo de arestas possíveis na rede.

A **análise dos resultados** apresentada a seguir baseia-se na importância das métricas dos programas, considerando-se as análises de componentes principais e da correlação de postos entre características calculadas e ranqueamento dos departamentos.

### 4. Resultados

Os 28 departamentos (ou programas de pós-graduação) foram caracterizados por 34 métricas: o número de pesquisadores em cada departamento (Pesq), as 16 medidas bibliométricas apresentadas, as cinco medidas de centralidade de cada departamento na rede envolvendo os departamentos, nove medidas globais extraídas das redes de cada departamento, e as posições desses departamentos nos três ranqueamentos utilizados.

Para cada uma das medidas os departamentos foram ranqueados da 1ª posição (maior valor para a métrica) até a 28ª (menor valor para a métrica), conforme apresentado na Figura 5. Os fundos das células estão coloridos de acordo com o ranqueamento: quanto mais próximo de 1 a célula tem seu fundo com um verde mais intenso, por outro lado, células com valores próximos a 28 são coloridas com fundo vermelho. Destaca-se que o departamento com maior valor para uma dada métrica é ranqueado em 1º lugar, mas valores altos de uma métrica não necessariamente é uma característica positiva.

Departamento	Pesq	PubS	PubCo	Cits	CitsCo	PubCPP	PubCoPP	CitsPP	CitsCoPP	PubS_5anos	PubCo_5anos	Cits_5anos	CitsCo_5anos	PubCPP_5anos	PubCoPP_5anos	CitsPP_5anos	CitsCoPP_5anos	Betweenness_I	Closeness_I	Degree_I	PageRank_I	Eigen_I	Clque_g	MCC_g	C.Custeizacao_g	Degree_g	Closeness_g	Eigen_g	Assort_Gru	Diametro_g	Densidade_g	Shanghai	TopUniversities	TheGuardian	
MIT EECS	17	11	6	5	5	4	4	2	3	13	13	8	19	6	11	5	18	17	11	11	13	10	10	13	27	18	24	9	18	12	24	2	1	1	
Stanford	11	12	3	6	3	9	3	4	1	12	6	5	5	9	7	6	7	5	3	3	4	3	4	7	18	6	19	6	20	3	25	1	2	2	
Oxford	19	13	19	13	14	7	21	11	14	9	24	4	6	4	21	2	5	26	18	18	20	17	10	11	26	20	11	7	6	15	15	3	3		
Harvard SEAS	24	20	11	16	8	11	2	10	4	21	10	21	10	8	2	15	6	21	11	11	12	11	23	20	20	14	12	20	23	22	14	4	4	4	
CMU	2	5	2	4	4	13	7	9	7	6	2	2	18	13	8	8	19	3	3	3	3	4	10	5	26	16	23	4	14	3	28	6	5	4	
UCBerkeley	4	1	1	1	1	5	1	1	2	3	1	6	4	7	3	9	9	2	1	1	2	1	1	14	9	15	5	15	6	6	23	3	7	6	
Cam	27	26	27	18	26	15	26	17	22	27	27	20	8	17	24	10	2	13	18	18	18	18	23	26	23	2	14	14	24	26	7	14	6	7	
ETH Zurich	4	9	16	11	13	16	22	18	17	18	15	11	28	15	19	14	26	10	7	7	7	8	10	12	24	7	13	12	25	12	18	13	8	9	
Princeton	3	10	4	3	2	21	16	12	5	10	3	19	20	21	15	28	23	5	5	5	6	5	23	2	25	26	18	21	1	26	3	3	9	9	
HKUST	12	8	17	15	10	6	18	19	19	5	21	15	17	3	22	19	17	22	14	14	15	16	10	14	20	15	13	1	6	13	18	11	7	3	
Toronto	17	15	14	14	15	17	13	15	15	18	14	9	22	19	12	7	21	16	9	9	10	9	10	1	17	12	16	7	22	1	16	11	10	11	
UCLA	20	2	9	9	10	1	6	3	10	1	3	7	11	1	3	10	4	5	5	5	7	10	25	16	4	27	2	13	25	20	9	12	12	1	
CUHK	23	26	24	23	20	25	19	22	20	25	26	27	26	26	27	26	25	27	22	22	27	20	23	27	14	11	21	16	27	26	1	17	14	13	
Cornell	6	6	5	2	6	12	10	7	9	7	8	10	9	14	13	13	15	11	7	7	8	6	10	3	22	24	25	3	5	3	27	8	15	19	
UTexas	9	7	7	7	7	8	9	6	8	11	4	16	13	12	4	20	13	8	11	11	11	15	4	6	12	22	11	10	15	6	22	7	19	15	
FPTI	13	4	12	8	12	2	11	5	12	4	16	13	15	2	18	16	12	7	9	9	9	13	21	4	19	13	18	5	17	6	19	16	13	17	
UW	21	16	10	12	9	10	5	8	6	17	11	23	21	10	5	22	20	25	14	14	17	12	10	15	21	8	22	19	26	22	17	19	16	14	
Caltech	28	28	28	19	24	19	25	16	18	28	28	24	27	16	23	11	27	28	28	28	28	21	28	28	28	28	28	1	28	26	5	10	16	16	16
Illinois	8	3	8	10	11	3	12	14	13	2	12	14	12	5	17	21	16	1	1	1	1	2	10	8	13	23	17	18	9	6	21	12	18	18	
UNICAMP	10	18	15	22	19	26	15	24	21	16	8	25	14	22	9	24	14	23	26	22	23	26	1	21	2	9	3	25	3	19	8	20	20	20	
USP São Paulo	14	27	20	26	21	28	23	25	25	24	20	28	23	28	20	28	24	15	22	22	24	25	4	19	1	17	10	26	12	12	3	22	20	20	
USP São Carlos	21	24	21	24	22	25	23	24	20	23	24	22	17	1	3	18	16	1	3	14	21	21	24	4	23	3	10	6	27	19	22	20	20	20	
BU	25	23	22	17	17	18	17	11	11	26	23	18	24	23	14	12	22	19	14	14	16	14	10	22	10	5	9	23	4	19	6	20	23	23	
UFMG	15	21	13	20	16	23	14	21	16	19	7	3	1	21	6	4	1	12	20	20	19	22	3	16	5	3	2	21	11	12	11	24	24	23	
UFRRS	7	14	25	24	25	20	27	26	26	14	19	12	2	20	26	18	4	9	17	17	14	19	4	18	7	21	8	17	2	12	9	24	21	23	
UFRRJ	15	19	21	27	28	22	24	27	28	20	25	26	27	25	26	27	18	22	22	25	23	10	9	8	27	7	24	10	12	12	24	24	23	23	
UFPE	2	17	26	28	27	27	28	28	27	15	21	17	7	24	28	23	11	24	27	27	26	28	10	17	6	19	4	28	3	12	10	24	27	23	
PUC-Rio	25	22	18	21	22	14	8	20	23	23	17	22	16	11	10	17	0	20	22	22	22	22	27	4	24	4	1	1	22	16	22	2	24	27	21

Figura 5. Ranqueamento dos programas de acordo com as métricas utilizadas

A Figura 6 contém a correlação de postos de Spearman calculada para todos os pares de métricas utilizados no presente estudo. As três últimas linhas e colunas contêm as correlações entre cada medida e os três ranqueamentos internacionais utilizados. Destaca-se que os três ranqueamentos possuem uma correlação maior ou igual a 0,76 entre si. Ao se comparar os valores das métricas calculadas com os valores dos ranqueamentos, correlações positivas significam que quanto maior a medida melhor será a posição do departamento no ranqueamento (ou seja, o departamento será melhor ranqueado), de maneira simétrica, correlações negativas significam que quanto maior o valor da métrica pior será a posição do departamento no ranqueamento (ele será pior ranqueado).

As medidas com maiores correlações positivas com os ranqueamentos são todas bibliométricas, apresentadas conforme seus valores de correlação média com os três ranqueamentos: CitsCoPP, Cits, CitsPP e CitsCo. Observa-se que as citações recebidas por pesquisador em trabalhos em coautoria com pesquisadores de outro departamento foi a medida mais correlacionada com os ranqueamentos (seguida pelo total de citações recebidas pelos artigos publicados do departamento). Isto é compatível com o conceito que se



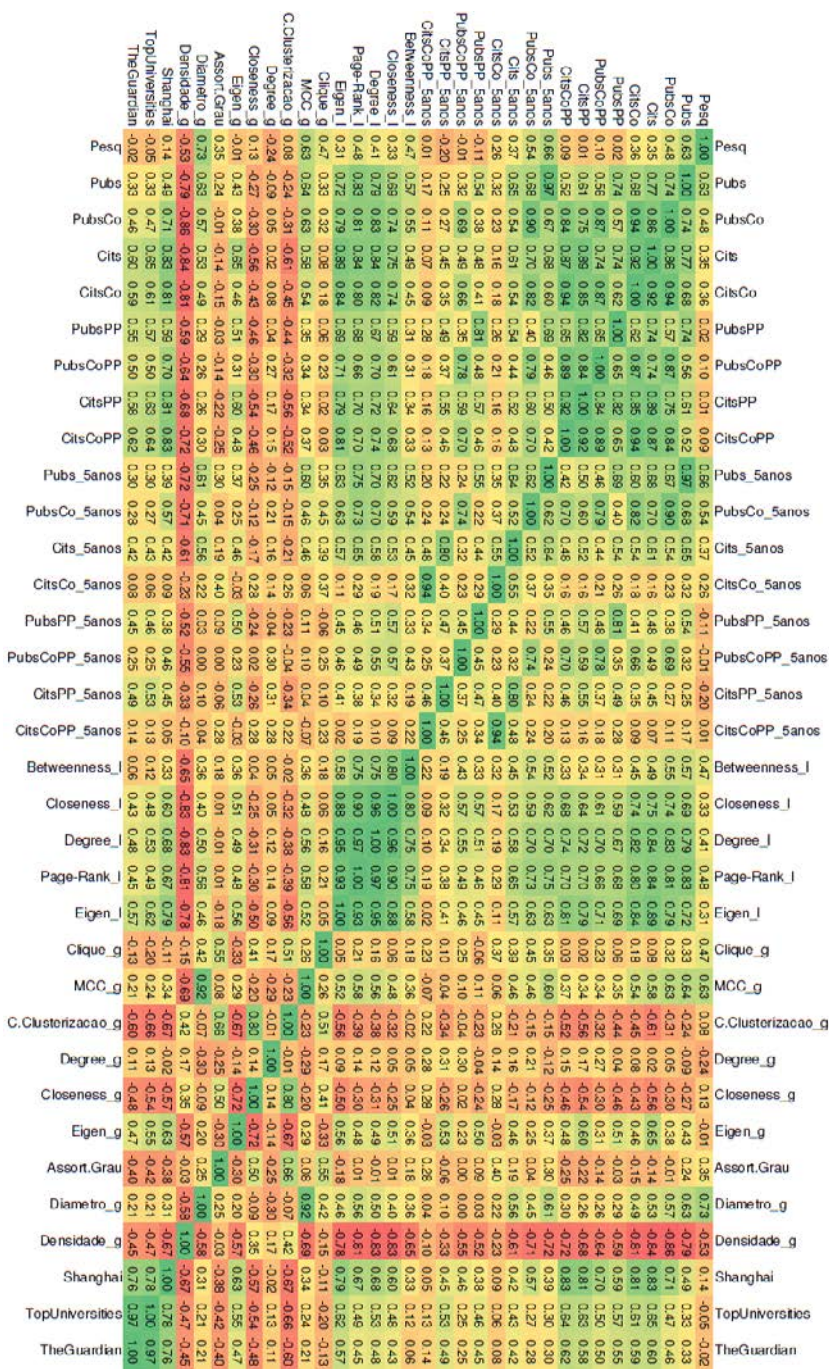


Figura 6. Correlações entre as métricas analisadas

tem da academia que bons departamentos (ou programas) publicam nos principais eventos ou revistas que são justamente aqueles cujos artigos publicados recebem mais citações.

Estas relações mostram outro motivo do posicionamento dos programas brasileiros. Fazendo a conta da média de citações por artigo com os números obtidos para este trabalho, temos o valor de 16,4. No caso dos melhores ranqueados, o valor sobe para cerca de 20 e, no caso dos brasileiros, cai para perto de 5. Isto permite que o ranqueamento vá ao encontro do conceito de que o número de citações é uma medida de qualidade da pesquisa e está de acordo com as correlações descritas no parágrafo anterior.

A língua não é uma limitação ao número de citações, posto que os dados das publicações foram obtidos do DBLP, que indexa principalmente artigos em revistas e conferências internacionais. Logo, este fator não afetou de forma relevante esta diferença significativa.

Das medidas de análise de redes sociais, a que apresentou a maior correlação positiva com os ranqueamentos foi a centralidade de autovalor ( $Eigen_l$ ). Todas as medidas de centralidade analisadas (considerando a rede em que cada departamento foi representado como um nó) possuem correlações positivas com o ranqueamento.

Ao se analisar as correlações negativas com os ranqueamentos destacam-se três medidas de redes (com correlações, na média, abaixo de -0,5). A primeira é o coeficiente de clusterização ( $C.Clusterização_g$ ) com correlações entre 0,6 e 0,67 com os ranqueamentos internacionais. Isto indica que quanto maior a transitividade entre as relações de um programa pior o seu ranqueamento.

A segunda medida de destaque é a correlação negativa entre os ranqueamentos e a centralização de proximidade ( $Closeness_g$ ), indicando que uma maior importância do pesquisador mais central de um dos departamentos em relação aos demais (considerando-se a proximidade como medida de importância) leva a uma tendência do departamento ser pior ranqueado. Isto é compatível com a ideia de que a diversidade de pensamento leva a resultados melhores da pesquisa e, por conseguinte, do ranqueamento.

Por fim, destaca-se a densidade da rede de cada um dos departamentos ( $Densidade_g$ ). A correlação indica que quanto mais densa a rede é, pior a posição do departamento, indicando que departamentos que tem muitas arestas internas tendem a ter um ranqueamento pior. Isto é compatível com o conceito de que a colaboração com pessoas “de fora” traz novas ideias e perspectivas, permitindo novas soluções e avanços.

Para estes três índices, a maioria dos programas brasileiros encontram na zona verde-verde amarelada, perto dos 10 primeiros lugares. Na métrica  $C.Clusterização_g$  os brasileiros estão nas 8 primeiras posições. Na  $Closeness_g$ , 10 primeiras. Para a medida  $Densidade_g$ , sete dos oito programas estão nas 12 primeiras posições. Estas medidas indicam um relacionamento mais intenso entre pesquisadores brasileiros de um mesmo programa e a dependência de cada programa de seu pesquisador mais central maior do que os melhores programas internacionais.

Foi utilizada a técnica de Análise de Componentes Principais para mapear as 31 métricas dos departamentos (todas as métricas menos as classificações nos ranqueamentos internacionais) para mapear os departamentos em um espaço bidimensional. A Figura 7 apresenta os 28 departamentos mapeados no espaço das duas componentes principais.

A primeira componente (eixo das abscissas na Figura 7) representa 51,4% da variância dos dados originais e está mais relacionada a medidas de importância dos programas (citações recebidas e as centralidades  $PageRank_l$ ,  $Degree_l$  e  $Closeness_l$ ). Já a segunda componente (eixo das ordenadas) representa cerca de 13% da variância dos dados e está mais relacionada ao modo de interação dos pesquisadores de um departamento. As três medidas mais intensamente representadas na segunda componente são Assortatividade de Grau, Coeficiente de Clusterização e Tamanho da Clique Máxima. A componente principal teve correlação maior ou igual a 0,59 com todos os ranqueamentos internacionais utilizados, destacando-se a correlação de 0,81 com o ranqueamento Shanghai.

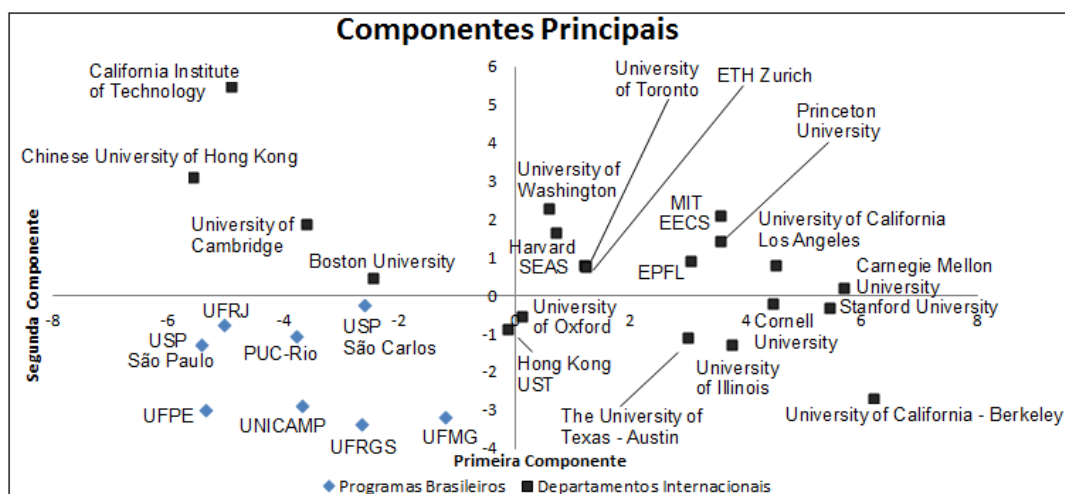


Figura 7. Plotagem das duas componentes principais

Destaca-se que nove dos dez melhor ranqueados encontram-se no lado direito da figura (valores positivos no eixo das abscissas). Por outro lado, dos dez piores ranqueados, nove se encontram do lado esquerdo da figura. Chama a atenção na figura que os oito programas de pós-graduação brasileiros analisados encontram-se na região esquerda inferior da figura (e são os únicos programas/departamentos nessa região), sugerindo que eles possuem características parecidas entre si e bastante diferente dos demais departamentos.

## 5. Conclusões

Neste trabalho, caracterizamos 20 departamentos internacionais da área de computação e oito programas de pós-graduação utilizando-se diferentes medidas bibliométricas e da análise de redes sociais. Foram calculadas as correlações entre todas as medidas e entre três ranqueamentos internacionais de universidades/departamentos.

A análise dos grafos gerados mostra que os programas brasileiros têm um nível de cooperação entre si e com departamentos internacionais abaixo do que é comum nestes últimos. Pode-se observar, por meio da análise de componentes principais, que os programas brasileiros não apresentam características similares àqueles que são considerados os melhores programas mundiais. Não é possível ainda estabelecer a razão desta diferença, nem se ela é positiva ou negativa, mas este trabalho indica uma desarmonia entre os critérios adotados pela CAPES e aqueles utilizados pelos ranqueamentos internacionais. Nesta análise também se vê que os oito programas de pós-graduação brasileiros analisados se diferenciam dos demais departamentos e se encontram muito próximos entre si. Assim, é possível inferir que existe uma “ciência à brasileira”, cujos critérios de qualidade parecem ser distintos daqueles adotados mundialmente.

Observou-se também que medidas bibliométricas estão fortemente correlacionadas com a posição dos programas nesses ranqueamentos, indicando a ausência de um “preconceito anti-periferia”. As posições dos programas são corretas dentro do modelo de valoração de pesquisa usados pelos mesmos (e que aproximadamente coincidem nos três ranqueamentos internacionais apresentados).

Como trabalho futuro pretende-se estender a presente pesquisa analisando uma quantidade maior de grupos e utilizando métricas adicionais.

## Agradecimentos

Este trabalho foi financiado pelo CNPq (processos 306046/2013-0 e 477246/2013-3).

## Referências

- Digiampietri, L. A., Mena-Chalco, J., Perez-Alcazar, J. J., Tuesta, E. F., Delgado, K., and Mugnaini, R. (2012). Minerando e caracterizando dados de currículos Lattes. In *BraSNAM 2012*, page 12, Curitiba, PR, Brasil.
- Digiampietri, L. A., Mena-Chalco, J. P., Vaz de Melo, P. O. S., Malheiro, A. P. R., 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*, 9(4):e94541.
- Duffy, R. D., Jadidian, A., Webster, G. D., and Sandell, K. J. (2011). The research productivity of academic psychologists: assessment, trends, and best practice recommendations. *Scientometrics*, 89(1):207–227.
- Franceschet, M. (2011). Collaboration in computer science: A network science approach. *Journal of the American Society for Information Science and Technology*, 62(10):1992–2012.
- Freire, V. and Figueiredo, D. (2011). Ranking in collaboration networks using a group based metric. *Journal of the Brazilian Computer Society*, pages 1–12.
- Hirsch, J. E. (2005). An index to quantify an individual’s scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46):16569–16572.
- Laender, A., de Lucena, C., Maldonado, J., de Souza e Silva, E., and Ziviani, N. (2008). Assessing the research and education quality of the top Brazilian Computer Science graduate programs. *ACM Special Interest Group on Computer Science Education*, 40(2):135–145.
- Leite, P., Mugnaini, R., and Leta, J. (2011). A new indicator for international visibility: exploring Brazilian scientific community. *Scientometrics*, 88:311–319.
- Lemieux, V. and Ouimet, M. (2008). *Análise Estrutural das Redes Sociais*. Instituto Piaget.
- Menezes, G. V., Ziviani, N., Laender, A. H. F., and Almeida, V. (2009). A geographical analysis of knowledge production. In *Computer Science In Proceedings of the 18th international conference on World Wide Web*, pages 1041–1050.
- Mugnaini, R., Digiampietri, L. A., de Oliveira, L. C., and Ferreira, S. M. S. P. (2012). Normalização de nomes de autores em fontes de informação institucionais: proposta de um método automático de verificação de erros. *Em Questão*, 18(3):263–279.
- Scott, J. (2009). *Social network analysis: a handbook*. SAGE, 2 edition.
- Ulrik, B. and Erlebach, T. (2005). *Network Analysis: Methodological Foundations*. Springer-Verlag.
- Wainer, J. and Vieira, P. (2013). Correlations between bibliometrics and peer evaluation for all disciplines: the evaluation of Brazilian scientists. *Scientometrics*, 96(2):395–410.