

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Genealogia acadêmica dos doutores atuantes em matemática: um mapeamento macro na ciência brasileira

Luciano Rossi¹, Igor Leite Freire², Rafael J. P. Damaceno³, Jesús P. Mena-Chalco⁴
Centro de Matemática, Computação e Cognição, UFABC

Resumo. Neste trabalho apresentamos um mapeamento macro dos doutores atuantes na área de matemática registrados na Plataforma Lattes. Foram computacionalmente prospectados todos os relacionamentos de orientação/coorientação acadêmica concluída para a criação do grafo de genealogia acadêmica, composto de 367.684 vértices e 388.603 arestas. Este mapeamento cientométrico permite: (i) evidenciar a influência da Matemática na comunidade acadêmica e (ii) identificar os matemáticos mais influentes.

Palavras-chave. Matemática, Genealogia acadêmica, Cientometria, Brasil.

1 Introdução

A análise da seleta comunidade dos doutores atuantes em matemática no Brasil, sob a ótica da genealogia acadêmica (GA) permite (i) evidenciar a influência exercida/recebida pela matemática em relação à outras áreas do conhecimento, (ii) identificar os matemáticos mais influentes segundo seus desempenhos na formação de recursos humanos.

A GA se define por meio do estudo das relações orientador/orientado e seus desdobramentos. Comumente, os dados genealógicos são estruturados em um grafo de genealogia, onde os vértices representam os acadêmicos e as arestas direcionadas seus relacionamentos de orientação. Análises de GA podem apresentar diversos objetivos, dentre os quais destacamos o reconhecimento de padrões que auxiliem na correta interpretação da formação da comunidade. O desenvolvimento de métricas topológicas facilitam as análises genealógicas.

Atualmente, diversas comunidades acadêmico-científicas despendem esforços no sentido de captar e armazenar conjuntos de dados genealógicos. Dentre diversas iniciativas, podemos destacar o *Mathematics Genealogy Project – MGP*⁵, uma iniciativa da *North Dakota State University* cujo objetivo é reunir dados genealógicos de todos os doutores em matemática do mundo. Outra plataforma *Web* de grande importância é *The Academic Family Tree*⁶ e fornece suporte para a estruturação de dados genealógicos para as mais diversas áreas do conhecimento (e.g., Neurociência, Ciência da Computação, Astronomia,

¹luciano.rossi@ufabc.edu.br

²igor.freire@ufabc.edu.br

³rafael.damaceno@ufabc.edu.br

⁴jesus.mena@ufabc.edu.br

⁵genealogy.math.ndsu.nodak.edu último acesso em 25 de Março de 2017.

⁶academictree.org, último acesso em 25 de Março de 2017.

Música, Epidemiologia, dentre outras). No Brasil, a Plataforma Lattes, é a mais completa interface de informação curricular que reúne mais de 4,9 milhões de currículos acadêmicos.

Neste trabalho apresentamos uma análise, baseada em GA, sobre a comunidade de doutores em matemática no Brasil, cujos dados utilizados foram prospectados da Plataforma Lattes. Utilizamos um conjunto de 6 métricas especialmente desenvolvidas para a caracterização do conjunto de dados. Adicionalmente, analisamos a influência existente entre as áreas do conhecimento segundo seus relacionamentos de orientação formal e concluída.

2 Trabalhos correlatos

O crescente interesse por parte da comunidade científica em GA, motivam trabalhos que têm por objetivo o mapeamento de pesquisadores em diferentes áreas do conhecimento. A comunidade dos matemáticos em todo o mundo é objeto de análise em diferentes trabalhos, pois conta com uma base de dados importante (MGP). O trabalho de Gargiulo et al. [4] utiliza, além do MGP, dados obtidos de fontes distintas. Os resultados refletem o processo de transferência de conhecimento entre uma rede de países e outra de disciplinas.

A comunidade dos neurocientistas também foi objeto de análises de GA. No estudo, conduzido por David e Hayden [1], destacamos a utilização de métricas, calculadas a partir do grafo de genealogia, que caracterizam os acadêmicos e, conseqüentemente, o grupo.

Identificar os acadêmicos que, historicamente, formaram a comunidade do Protozoologistas é o objetivo no trabalho de Elias et al. [3]. Este mapeamento descreve desde os fundadores até os mais recentes acadêmicos desta área do conhecimento.

Finalmente, um mapeamento amplo, realizado por Dores et al. [2], foi conduzido utilizando a base de dados de teses e dissertações (*Networked Digital Library of Theses and Dissertations – NDLTD*). Este trabalho apresenta um *framework* para extração de árvores de genealogia e descreve análises para a extração de conhecimento destas estruturas.

3 Materiais e métodos

3.1 Conjunto de dados

O conjunto de dados considerado neste trabalho é composto por cerca de 367.684 doutores, distribuídos por mais de 100 áreas do conhecimento, e 388.603 relacionamentos de orientação acadêmica. Deste total, concentramos as análises no subconjunto dos registros dos matemáticos que representa 4.906 matemáticos e 8.025 relacionamentos de orientação⁷ entre eles e as demais áreas (relacionamentos declarados nos currículos Lattes). Estes dados foram obtidos junto à Plataforma Lattes, em Janeiro de 2017, e possuem diversos atributos de interesse, dentre os quais consideramos: (i) o nome do acadêmico, (ii) a área do conhecimento a qual o acadêmico pertence, e (iii) todos os seus orientados.

Os currículos disponíveis na Plataforma Lattes são estruturados de acordo com as grandes áreas e suas áreas do conhecimento. Neste trabalho consideramos 101 áreas do

⁷Foram encontrados 1.800 relacionamentos onde o orientador não identificou sua área e 1.521 relacionamentos onde o orientado não identificou sua área.

conhecimento que são distribuídas em 9 grandes áreas. A área da Matemática possui 4.906 membros com o título de doutor (correspondente a 1,33% do total de acadêmicos).

Os dados genealógicos foram estruturados em um grafo de genealogia acadêmica, onde os acadêmicos são representados por vértices e os relacionamentos de orientação ou coorientação (formal e concluída), existentes entre eles, são representados por arestas direcionadas. O grafo de genealogia dos doutores possui: 4.192 nós isolados (1,14%), densidade é $5,75 \times 10^{-6}$, o grau médio dos vértices é 2,11 e o grau máximo encontrado é 137, há 29.376 componentes conexas e a maior componente conexa concentra 292.479 vértices (79,5%).

3.2 Métricas em grafos de genealogia

As métricas são importantes para a formulação de atributos quantitativos que representem alguma característica da topologia do grafo genealógico e possam ser considerados para: (i) a classificação do acadêmico (vértice), (ii) a caracterização de comunidades (*clusters*) e (iii) a identificação das principais linhagens acadêmicas (conjunto de vértices relacionados de forma linear). Consideramos 6 métricas para a caracterização dos acadêmicos no grafo de genealogia. Seja $\vec{G} = (V, E)$ um grafo de genealogia, fixamos que: o *território* do vértice (v) é definido por $T(v) = \{u \in V : \exists v \rightsquigarrow u - \text{caminho em } \vec{G}\}$, o conjunto dos vértices adjacentes à v é $D(v) = \{u \in V : (v, u) \in E\}$, o conjunto dos vértices que não possuem adjacentes é $NF(v) = \{u \in T(v) : |D(u)| = 0\}$, o conjunto dos caminhos de comprimento k é $C^{(k)}(v) = \{v \rightsquigarrow u - \text{caminho de comprimento } k \text{ em } \vec{G} : u \in V\}$, e a descendência k -fértil de v é $D^{(k)}(v) = \{u : u \in D(v) \text{ e } |D(u)| \geq k\}$ para $k \in \mathbb{N}$. A seguir formalizamos as métricas consideradas neste trabalho:

- A *fecundidade* representa o total de vértices descendentes que um dado vértice de interesse possui, e é definida por $f(v) = |T(v)|$;
- A *profundidade* é o número de arestas existentes entre o vértice de interesse e o vértice mais distante que possa ser alcançado, ou seja, é o tamanho do maior caminho existente no grafo, e é definida por $p(v) = \max\{k \in \mathbb{N} \cup \{0\} : v \rightsquigarrow u \in C^{(k)}(v) \text{ e } u \in NF(v)\}$;
- A *largura* representa o número de descendentes diretos a partir de um vértice de interesse, e é definida por $l(v) = |D(v)|$;
- A *distância média* consiste da razão entre a soma do número de caminhos de tamanho k e o número de caminhos. Considera a proximidade do vértice de interesse em relação a todos seus descendentes, e é dada por $dm(v) = (\sum_{k=1}^n |C^{(k)}(v)| \times k) / (\sum_{k=1}^n |C^{(k)}(v)|)$, em que n é o tamanho do maior caminho no grafo;
- O *número de folhas* é a quantidade de vértices não fecundos existentes em um grafo de genealogia, i.e., aqueles que não têm nenhum filho, e é dado por $nf(v) = |NF(v)|$;
- O *índice genealógico* [5] de um vértice é definido como o maior número de conexões existentes entre ele e seus vértices adjacentes que possuem, pelo menos, o mesmo número de conexões cada um. O objetivo desta métrica é considerar a quantidade e a qualidade genealógica (no sentido de perpetuidade) do grafo de genealogia e é dado por $g(v) = \max\{k \in \mathbb{N} \cup \{0\} : |D(v)| \geq k \text{ e } |D^{(k)}(v)| \geq k\}$.

Tabela 1: *As áreas do conhecimento e as relações de orientação com a área da Matemática.*

Área	Influência	Influenciada	Área	Influência	Influenciada
Administração	8	7	Engenharia Mecatrônica	0	3
Agronomia	3	36	Eng. Naval e Oceânica	0	1
Antropologia	0	1	Engenharia Nuclear	2	31
Artes	1	2	Engenharia Química	2	25
Astronomia	3	13	Engenharia Sanitária	2	9
Biofísica	2	4	Farmacologia	0	1
Biologia Geral	1	0	Filosofia	6	18
Bioquímica	0	2	Física	29	116
Biotecnologia	0	1	Fonoaudiologia	0	2
Botânica	2	0	Genética	2	4
Ciência da Computação	110	181	Geociências	16	62
Ciência da Informação	1	1	Geografia	0	6
Ciência e Tec. de Alimentos	1	4	História	5	23
Ciência Política	0	2	Letras	1	0
Ciências Ambientais	0	3	Linguística	2	2
Comunicação	4	2	Matemática	2274	2274
Demografia	0	2	Medicina	5	9
Desenho Industrial	1	0	Medicina Veterinária	2	0
Divulgação Científica	1	0	Microbiologia	0	1
Ecologia	3	3	Nutrição	1	1
Economia	31	20	Oceanografia	1	8
Educação	133	408	Odontologia	2	1
Educação Física	1	0	Plan. Urbano e Regional	0	1
Enfermagem	1	0	Probabilidade E Estatística	33	99
Engenharia Aeroespacial	1	30	Psicologia	0	13
Engenharia Agrícola	1	11	Química	2	15
Engenharia Biomédica	5	18	R. Florestais e E. Florestal	3	22
Engenharia Civil	12	60	R. Pesqueiros e E. de Pesca	0	1
Engenharia de Energia	1	0	Rob. Mecatrônica e Autom.	0	2
Eng. de Mat. e Metalúrgica	4	24	Saúde Coletiva	6	4
Engenharia de Minas	0	6	Sociologia	0	4
Engenharia de Produção	26	104	Teologia	1	0
Engenharia de Transportes	1	16	Turismo	1	0
Engenharia Elétrica	23	228	Zoologia	0	3
Engenharia Mecânica	38	210	Zootecnia	1	0

4 Resultados

4.1 Influência

A interdisciplinaridade, sob a ótica da GA, ocorre quando acadêmicos de uma determinada área do conhecimento orientam acadêmicos de áreas distintas. Analisamos as relações interdisciplinares entre os doutores da área da Matemática em relação aos doutores das 100 outras áreas do conhecimento com o objetivo de identificar a influência entre os grupos.

A Tabela 1 contém um total de 70 áreas do conhecimento. Para 30 áreas não encontramos relacionamentos de orientação envolvendo a Matemática. Cada uma das linhas apresenta o nome da área em questão, a segunda coluna indica o número de orientações que os acadêmicos desta área receberam de acadêmicos da Matemática. A coluna seguinte indica o número de orientações que os matemáticos receberam dos acadêmicos desta área.

Para a maior parte dos relacionamentos, observamos que a área da Matemática foi

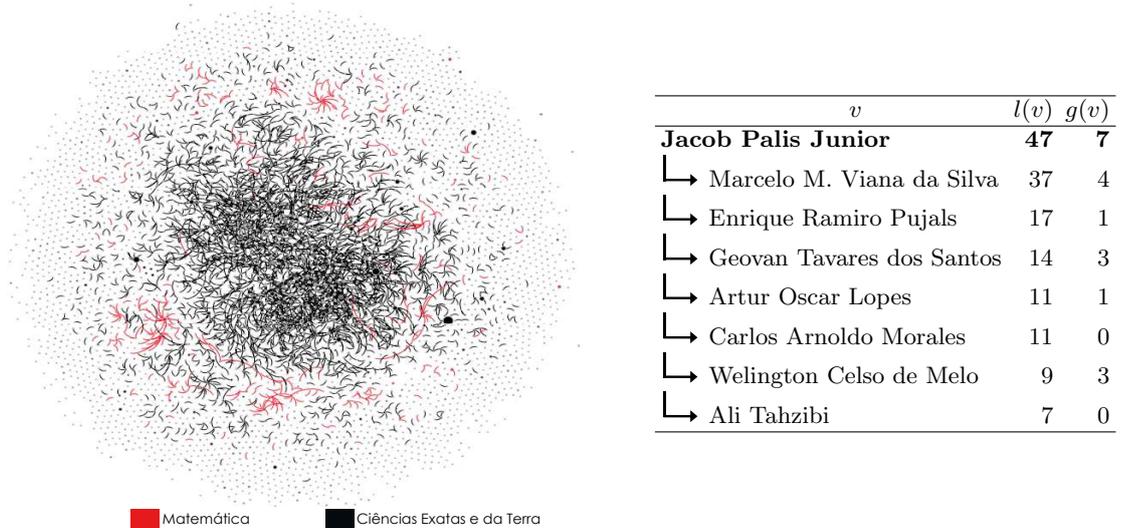


Figura 1: *Grafo de genealogia dos matemáticos e seus relacionamentos com a grande área de Ciências Exatas e da Terra (Figura a esquerda). Foram omitidos os acadêmicos sem descendentes para facilitar a visualização. Descendência direta parcial de Jacob Palis Junior e os valores de fecundidade e índice genealógico (Tabela a direita).*

mais orientada do que orientou. Este padrão parece contrário ao esperado, visto que a Matemática constitui uma disciplina presente em grande parte das outras áreas do conhecimento. A constatação pode ser parcialmente justificada observando-se um exemplo, um matemático fez seu doutorado sob a orientação de um acadêmico da área de Zoologia, visto que o seu tema de pesquisa tinha por objetivo desenvolver o modelo matemático no qual representa-se o padrão de desova de um determinada espécie de tartaruga marinha.

A Figura 1 apresenta o grafo de genealogia considerando a área da Matemática e as outras áreas com as quais se relacionou diretamente (descendência ou ascendência direta), bem como, a descendência direta do matemático Jacob Palis Junior, destaque na formação de recursos humanos. Note que os relacionamentos são representados por arestas coloridas e o tamanho dos vértices representa a quantidade de conexões que ele possui. Todas as áreas representadas fazem parte da grande área das Ciências Exatas e da Terra.

4.2 Caracterização dos matemáticos

As métricas consideradas foram calculadas para o conjuntos de doutores e resultaram em atributos quantitativos com características específicas. A Figura 2 mostra as distribuições de frequências, a média, a mediana e o maior resultado para cada métrica. Um matemático com destaque em orientação acadêmica é Jacob P. Junior.

De acordo com os dados da plataforma Lattes, Palis é um matemático brasileiro premiado pelos seus trabalhos com sistemas dinâmicos e orientou 47 alunos em nível de doutorado (largura). Sua descendência é composta por 272 matemáticos (fecundidade), deste total, 226 matemáticos não orientaram nenhum aluno (número de folhas). Estes

Tabela 2: *Ranking dos 10 matemáticos com melhores resultados para cada métrica.*

Métrica	Matemático	Valor	Métrica	Matemático	Valor
Fecund.	Jacob Palis Junior	272	Folhas	Jacob Palis Junior	226
	Manfredo Perdigao do Carmo	148		Julio Cesar Ruiz Claeysen	125
	Julio Cesar Ruiz Claeysen	141		Manfredo Perdigao do Carmo	120
	Luiz Adauto da Justa Medeiros	140		Luiz Adauto da Justa Medeiros	117
	Raul Antonino Feijóo	124		Agenor Cortarelli	106
	Agenor Cortarelli	119		Raul Antonino Feijóo	100
	Luiz Carlos Wrobel	97		Luiz Carlos Wrobel	88
	Paulo Roberto Oliveira	91		Manuel Antolino Milla Miranda	76
	Manuel Antolino Milla Miranda	88		Paulo Roberto Oliveira	71
	Clovis Caesar Gonzaga	85		Clovis Caesar Gonzaga	70
Profund.	Maurício Matos Peixoto	5	Largura	Jacob Palis Junior	47
	Agenor Cortarelli	4		Aloisio Pessoa de Araujo	37
	Alexandre A. M. Rodrigues	4		Luiz Adauto da Justa Medeiros	37
	Aloisio Pessoa de Araujo	4		Marcelo Miranda Viana da Silva	37
	Antonio Iván Ruíz Chaveco	4		Manuel Antolino Milla Miranda	36
	Carlos Tomei	4		Pavel Sobolevskii	35
	Clovis Caesar Gonzaga	4		Luiz Carlos Wrobel	33
	Fernando A. F. C. da Silva	4		Paulo Roberto Oliveira	33
	Jacob Palis Junior	4		Jose Mario Martinez Perez	32
Jorge Manuel Sotomayor Tello	4	Manfredo Perdigao do Carmo	31		
Distância Média	Agenor Cortarelli	3,23	Índice Genealógico	Jacob Palis Junior	7
	Antonio Iván Ruíz Chaveco	3,17		Raul Antonino Feijóo	6
	Maurício Matos Peixoto	2,27		Clovis Caesar Gonzaga	5
	Jacob Palis Junior	2,23		Manfredo Perdigao do Carmo	5
	Julio Cesar Ruiz Claeysen	2,20		Paulo Roberto Oliveira	5
	Angelo Barone Netto	2,19		Djairo Guedes de Figueiredo	4
	Luiz Adauto da Justa Medeiros	2,17		John Andrew Fossa	4
	Clovis Caesar Gonzaga	2,12		Jose Mario Martinez Perez	4
	Oswaldo Sangiorgi	2,11		Luiz Adauto da Justa Medeiros	4
Ramón O. Mendoza Ahumada	2,05	Luiz Carlos Wrobel	4		

descendentes são distribuídos ao longo de 4 gerações (profundidade) e o tamanho médio das linhagens a partir de Palis é de 2,23 (distância média). Em termos de qualidade de orientação, Palis destaca-se com o valor de índice genealógico igual a 7 (veja na Figura 1 os matemáticos considerados para este resultado). A Tabela 2 apresenta os 10 matemáticos com os melhores resultados de acordo com cada uma das métricas topológicas⁸.

Considerando a GA e os resultados das métricas topológicas podemos descrever o perfil do doutor em matemática brasileiro. Este matemático orientou 4,94 alunos, sua linhagem é composta por 7,95 descendentes, sendo que, deste total, 7,09 não orientaram alunos, e são distribuídos em 1,34 gerações. Além disto, os caminhos do grafo de genealogia a partir deste matemático possui tamanho médio de 0,83 e seu índice genealógico é 1,48.

⁸Para a Profundidade os seguintes matemáticos apresentaram resultado igual a 4: José Mario Martinez Perez, Julio Cesar Ruiz Claeysen, Luiz Adauto da Justa Medeiros, Manfredo Perdigao do Carmo, Oswaldo Sangiorgi, Paulo Roberto Oliveira e Raul Antonino Feijóo. Para Índice Genealógico o matemático Marcelo Miranda Viana da Silva apresentou resultado igual a 4.

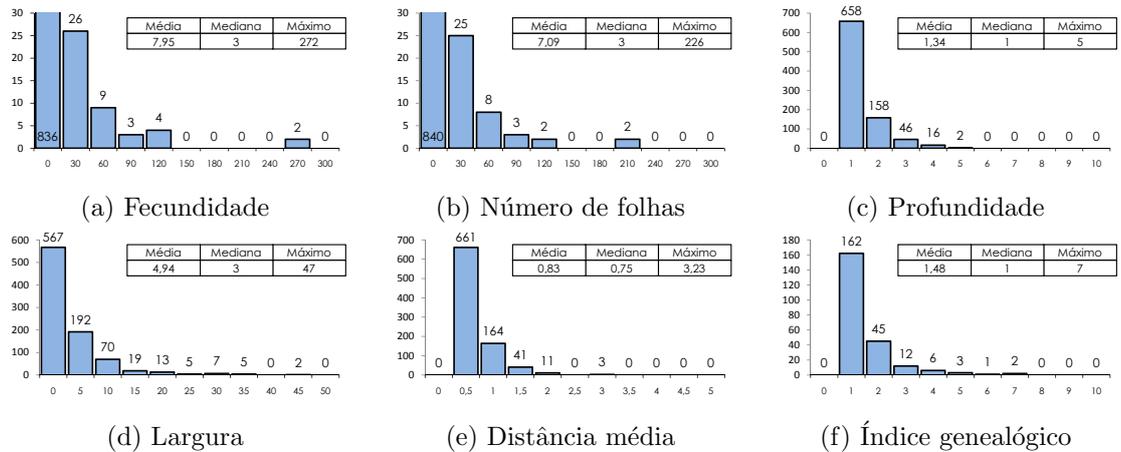


Figura 2: Distribuição de frequências para as métricas consideradas. Os gráficos apresentam, também, a média, a mediana e o maior valor observado para cada métrica.

5 Conclusões

A comunidade dos doutores atuantes em Matemática no Brasil apresenta-se como uma área multi-disciplinar quando nos referimos a GA. Apesar de influente na maior parte das áreas do conhecimento, a Matemática recebeu mais orientações acadêmicas do que as realizou. As métricas indicam que a Matemática é uma jovem área do conhecimento no Brasil, de acordo com os dados utilizados, quando comparada à outros países.

Utilizar GA para a realização de mapeamento de comunidades acadêmico-científicas, sob o ponto de vista da orientação acadêmica, surge com uma opção às tradicionais avaliações de desempenho de pesquisadores baseadas em análise de citações.

Referências

- [1] S. V. David and B. Y. Hayden. *Neurotree: A collaborative, graphical database of the academic genealogy of neuroscience*, *PloS ONE*, 7:e46608, 2012.
- [2] W. Dores, F. Benevenuto and A. H. Laender. *Extracting academic genealogy trees from the networked digital library of theses and dissertations*, In *Digital Libraries (JCDL), 2016 IEEE/ACM Joint Conference on*, pages 163–166, IEEE, 2016.
- [3] M. Elias, L. M. Floeter-Winter and J. P. Mena-Chalco. *The dynamics of brazilian protozoology over the past century*, *Mem I Oswaldo Cruz*, 111:67–74, 2016.
- [4] F. Gargiulo, A. Caen, R. Lambiotte and T. Carletti. *The classical origin of modern mathematics*, *EPJ Data Sci*, 5:26, 2016.
- [5] L. Rossi, I. L. Freire and J. P. Mena-Chalco. *Genealogical index: A metric to analyze advisor–advisee relationships*, *J Informetr*, 11:564–582, 2017.