



UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
Câmpus de Marília

Rafael Gutierrez Castanha

**Coeficiente de acoplamento bibliográfico genealógico para mensurar a
intensidade da preservação das correntes teórico-metodológicas nas
redes de genealogia acadêmica**

Marília
2023

Rafael Gutierrez Castanha

Coeficiente de acoplamento bibliográfico genealógico para mensurar a intensidade da preservação das correntes teórico-metodológicas nas redes de genealogia acadêmica

Tese apresentada ao Programa de Pós-Graduação em Ciência da Informação como parte das exigências para a obtenção do título Doutor em Ciência da Informação pela Faculdade de Filosofia e Ciências, Universidade Estadual Paulista (UNESP), Campus de Marília.

Área de Concentração: Informação, Tecnologia e Conhecimento

Orientadora: Prof. Dra. Maria Cláudia
Cabrini Grácio

Coorientadora: Prof. Dra. Leilah Santiago
Bufrem

Marília
2023

C346c Castanha, Rafael Gutierrez

Coeficiente de acoplamento bibliográfico genealógico para mensurar a intensidade da preservação das correntes teórico-metodológicas nas redes de genealogia acadêmica / Rafael Gutierrez Castanha. -- Marília, 2023

149 f. : tabs.

Tese (doutorado) - Universidade Estadual Paulista (Unesp), Faculdade de Filosofia e Ciências, Marília

Orientadora: Maria Cláudia Cabrini Grácio

Coorientadora: Leilah Santiago Bufrem

1. Acoplamento Bibliográfico. 2. Bolsistas de Produtividade em Pesquisa. 3. Bibliometria. 4. Genealogia Acadêmica. 5. Medidas de Similaridade. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Filosofia e Ciências, Marília. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

Rafael Gutierrez Castanha

Coeficiente de acoplamento bibliográfico genealógico para mensurar a intensidade da preservação das correntes teórico-metodológicas nas redes de genealogia acadêmica

Tese apresentada ao Programa de Pós-graduação em Ciência da Informação da Universidade Estadual Paulista “Júlio de Mesquita Filho” (Unesp), Faculdade de Filosofia e Ciências, Campus Marília, de como requisito parcial para a obtenção do título de Doutor em Ciência da Informação.

Área de concentração: Informação,
Tecnologia e Conhecimento
Linha de pesquisa: Produção e Organização
da Informação

Banca Examinadora

Orientadora: _____

Profa. Dra. Maria Cláudia Cabrini Grácio
Universidade Estadual Paulista (UNESP) - Campus Marília

1ª Examinadora: _____

Profa. Dra. Ely Francina Tannuri de Oliveira
Universidade Estadual Paulista (UNESP) - Campus Marília

2º Examinador: _____

Prof. Dr. Daniel Martínez-Ávila
Universidade Estadual Paulista (UNESP) - Campus Marília

3º Examinador: _____

Prof. Dr. Leandro Innocentini Lopes de Faria
Universidade Federal de São Carlos (UFSCAR)

4º Examinador: _____

Prof. Dr. Antonio Perianes-Rodriguez
Universad Carlos III de Madrid (UC3M) - Departamento de Biblioteconomia y Documentacion

Marília, 31 de janeiro de 2023

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

À orientadora Dra. Maria Cláudia Cabrini Grácio.

À co-orientadora Dra. Leilah Bufrem.

Aos membros da banca: Dra. Ely Oliveira, Dr. Antonio Perianes-Rodriguez, Dr. Daniel Martínez-Ávila, Dr. Leandro Innocentini.

À todos familiares.

RESUMO

Objetivo: Propõe um coeficiente de perspectiva relacional de citação e recitação, denominado Coeficiente de Acoplamento Bibliográfico Genealógico, para mensurar a intensidade das proximidades teórico-metodológicas entre pesquisadores no escopo das relações de genealogia acadêmica. Especificamente, fundamenta o Coeficiente de Acoplamento Bibliográfico Genealógico, comparando-o com outras formas de normalização do Acoplamento Bibliográfico de Autores, Índice de Jaccard e Cosseno de Salton, aplicado ao universo dos pesquisadores de Produtividade em Pesquisa seniors do CNPq da Matemática, caracteriza a floresta de genealogia acadêmica valorada pelo coeficiente proposto, e avalia a conexão entre os principais influentes da floresta de genealogia acadêmica. **Método:** Identifica e recupera toda produção científica indexada na base de dados Web Of Science dos pesquisadores Sênior da área de Matemática do Instituto de Matemática Pura e Aplicada e seus descendentes acadêmicos egressos de doutorado. Utiliza o coeficiente de correlação intraclasse e o método de Bland-Altman para análises de correlação e concordância entre o coeficiente proposto e o Cosseno de Salton e o Índice de Jaccard. Constrói a floresta de genealogia acadêmica valorada pelas perspectivas de acoplamento bibliográfico discutidas e composta por todos pesquisadores analisados. Identifica e ilustra a conexão entre os principais influentes teóricos responsáveis pelo acoplamento por meio das cocitações entre estes autores. **Resultados:** demonstra que o coeficiente proposto é correlacionado, mas não concordante com as demais medidas comparadas apresentando-se como um indicador alternativo e confiável para análises de acoplamento bibliográfico entre pesquisadores com relação de genealogia acadêmica. Constata que o coeficiente é capaz de mensurar a herança intelectual, representada pelo conjunto de autores recitados, perpassada de orientador para orientando, alinhando-se à ideia central dos estudos genealógicos da ciência. Constrói a floresta de genealogia acadêmica e verifica diferenças topológicas, em termos de conectividade, entre as três perspectivas de acoplamento analisadas, e, ratifica o novo indicador. Evidencia as conexões entre os principais influentes de toda floresta e testemunha fortes influências, do ponto de vista da genealogia científica, de evoluções relacionadas aos sistemas dinâmicos, equações diferenciais, folheações holomórficas e economia matemática. Sob a ótica da genealogia intelectual, revela uma escola de pensamento influenciada por Stephen Smale, Jacob Palis Junior, Vladimir Igorevich Arnold, César Camacho, Marco Bruenella e Robert Lucas. **Considerações Finais:** conclui que o Coeficiente de Acoplamento Bibliográfico Genealógico é funcional e robusto para análises de genealogia acadêmica ao estar correlacionado mas não concordar com os demais métodos de normalização amplamente difundidos na literatura por ter como principal mérito admitir acoplamento perfeito mesmo que as listas dos pesquisadores analisados sejam diferentes, aferido quando o orientando herda toda lista de autores recitados de seu orientador. Reconhece que os principais influentes da floresta de genealogia acadêmica analisada se alinham às temáticas de pesquisas dos pesquisadores analisados. Admite dificuldades em questões operacionais manuais.

Palavras-chaves: Acoplamento Bibliográfico; Bolsistas de Produtividade em Pesquisa; Bibliometria; Genealogia Acadêmica; Medidas de Similaridade.

ABSTRACT

Objective: It proposes a coefficient of relational perspective of citation and recitation, named Genealogical Bibliographic Coupling Coefficient, to measure the intensity of theoretical and methodological proximities between researchers in the scope of academic genealogy relations. Specifically, the Genealogical Bibliographic Coupling Coefficient is grounded by comparing it with other forms of normalization of Author Bibliographic Coupling, Jaccard Index, and Salton Cosine applied to the universe of senior CNPq Research Productivity researchers in Mathematics. Characterizes the forest of academic genealogy valued by the proposed coefficient, and evaluates the connection between the main influencers of the forest of academic genealogy. **Method:** Identifies and retrieves all scientific production indexed in the Web of Science database of senior researchers in the field of Mathematics from the Institute of Pure and Applied Mathematics and their academic descendants who graduated from a doctoral program. The intraclass correlation coefficient and Bland-Altman method are used for correlation and agreement analyses between the proposed coefficient and Salton Cosine and Jaccard Index. The forest of academic genealogy valued by the discussed bibliographic coupling perspectives and composed of all analyzed researchers is constructed. The connection between the main theoretical influencers responsible for coupling is identified and illustrated through cocitations among these authors. **Results:** Demonstrates that the proposed coefficient is correlated but not in agreement with the other compared measures, presenting itself as an alternative and reliable indicator for bibliographic coupling analyses between researchers with an academic genealogy relationship. Confirms that the coefficient can measure intellectual inheritance, represented by the set of recited authors, passed down from advisor to advisee, aligning with the central idea of genealogical studies of science. The forest of academic genealogy is constructed, and topological differences in terms of connectivity between the three analyzed coupling perspectives are verified, confirming the new indicator. The connections between the main influencers of the entire forest are evidenced, and strong influences, from the perspective of scientific genealogy, of developments related to dynamical systems, differential equations, holomorphic foliations, and mathematical economics are witnessed. From the perspective of intellectual genealogy, reveals a school of thought influenced by Stephen Smale, Jacob Palis Junior, Vladimir Igorevich Arnold, César Camacho, Marco Bruenella, and Robert Lucas. **Final Considerations:** Concludes that the Genealogical Bibliographic Coupling Coefficient is functional and robust for academic genealogy analyses since it is correlated but does not agree with other widely spread normalization methods in the literature, mainly because it admits perfect coupling even if the analyzed researchers' lists are different. This is measured when the advisee inherits the entire list of recited authors from their advisor. Recognizes that the main influencers of the analyzed forest of academic genealogy align with the research themes of the analyzed researchers. Admits difficulties in manual operational issues.

Keywords: Academic Genealogy; Bibliometrics; Bibliographic Coupling; Research Grants; Measure of similarity.

LISTA DE FIGURAS

Figura 1 - Relação entre os tipos de estudos genealógicos	30
Figura 2 - Perspectivas relacionais de citação (citação, cocitação e acoplamento bibliográfico).....	36
Figura 3 - Fluxograma de etapas de formulação do Coeficiente de Acoplamento Bibliográfico Genealógico	57
Figura 4 - Identificação e extração da lista de autores	70
Figura 5 - Identificação e exportação da lista de autores e das relações de citação.....	71
Figura 6 - Conjugação das listas de exportação de autores e relações de citação.....	72
Figura 7 - Organização completa das identificações dos autores e dos padrões de citação e recitação por documento.....	72
Figura 8 - Cálculos de acoplamento via The Genealogical Coupler.....	74
Figura 9 - Identificação das unidades de acoplamento via The Genealogical Coupler.....	76
Figura 10 - Fluxograma dos procedimentos metodológicos	81
Figura 11 - Gráfico comparativo entre as medidas Jaccard (recit), Salton (recit.) e CABG.....	86
Figura 12 - Gráfico de Bland-Altman entre o CABG e o Índice de Jaccard segundo os padrões de recitação de autores	92
Figura 13 - Gráfico de Bland-Altman entre o CABG e o Cosseno de Salton segundo os padrões de recitação de autores	94
Figura 14 - Gráfico comparativo entre Índice de Jaccard, Cosseno de Salton e CABG segundo padrões de citação e recitação.....	97
Figura 15 - Floresta de genealogia acadêmica dos pesquisadores bolsistas PQ-Sr. do IMPA e seus orientandos egressos de doutorado com arestas valoradas pelo CABG	100

Figura 16 - Floresta de genealogia acadêmica dos pesquisadores bolsistas PQ-Sr. do IMPA e seus orientandos egressos de doutorado com arestas valoradas pelo Índice de Jaccard (recit.).....	101
Figura 17 - Floresta de genealogia acadêmica dos pesquisadores bolsistas PQ-Sr. do IMPA e seus orientandos egressos de doutorado com arestas valoradas pelo Cosseno de Salton (recit.).....	102
Figura 18 - Rede de recitação da identidade científica da floresta de genealogia acadêmica	111
Figura 19 - Estrutura das influências intelectuais da floresta de genealogia acadêmica.....	113
Figura 20 - Rede de identidade científica da floresta de genealogia acadêmica de pesquisadores PQ-Sr do IMPA	120

LISTA DE TABELAS

Tabela 1 - Bolsistas PQ sênior da área de Matemática	66
Tabela 2 - Delimitação dos bolsistas PQ-Sr. e acoplamentos possíveis	67
Tabela 3 - Descrição da produção científica recuperada na WoS mediante os critérios estabelecidos	68
Tabela 4 - Cálculo dos acoplamentos segundo os padrões de citação e recitação de autores	83
Tabela 5 - Correlação intraclasse entre Índice de Jaccard, Cosseno de Salton e CABG.	89
Tabela 6 - Valores descritivos do gráfico de Bland-Altman.....	91
Tabela 7 - Força de conectividade média de cada rede de genealogia acadêmica	104

LISTA DE QUADROS

Quadro 1 - Acoplamento bibliográfico de autores segundo os valores mínimos de recitação em documentos distintos	41
Quadro 2 - Medidas de similaridade e possível adaptação ao ABA	50
Quadro 3 - Autores acopladores de toda floresta de genealogia acadêmica	107

SUMÁRIO

1 INTRODUÇÃO	12
2 ESTUDOS DE GENEALOGIA DA CIÊNCIA	24
3 ANÁLISES RELACIONAIS DE CITAÇÃO: COCITAÇÃO E ACOPLAMENTO BIBLIOGRÁFICO	32
4 ACOPLAMENTO BIBLIOGRÁFICO ENTRE AUTORES E ÍNDICES DE NORMALIZAÇÃO PARA INDICADORES RELACIONAIS DE CITAÇÃO	42
5 COEFICIENTE DE ACOPLAMENTO BIBLIOGRÁFICO GENEALÓGICO....	54
6 METODOLOGIA.....	63
6.1 UNIVERSO E CORPUS DA PESQUISA.....	64
6.2 PROCEDIMENTOS METODOLÓGICOS.....	77
7 RESULTADOS	82
8 CONSIDERAÇÕES FINAIS	122
REFERÊNCIAS.....	132
APÊNDICE	139
APÊNDICE 1 – Matriz de recitação.....	138
APÊNDICE 2 – Código em R do The Genealogical Coupler.....	144

1 INTRODUÇÃO

A ciência pode ser compreendida como um amplo sistema socialmente construído, cujas características estão relacionadas a disseminação e validação de novos conhecimentos e ao reconhecimento de trabalhos que contribuíram para o desenvolvimento de novas ideias nos diferentes campos e domínios científicos (PRADO; CASTANHA, 2020).

Compreender a ciência como atividade social perpassa diretamente por assimilar as relações estabelecidas pelas pessoas que compõem os diferentes meios científicos. A ciência evolui por meio das interações entre acadêmicos e pela produção e transferência de conhecimento decorrida destas interações que geram, por consequência, o conhecimento científico, uma das principais bases para o progresso socioeconômico e tecnológico de uma comunidade (DAMACENO; ROSSI; MENA-CHALCO, 2017).

Parte fundamental deste processo é a díade formada por orientador e orientando. Esta conexão constitui um relacionamento científico específico entre um pesquisador mais experiente (orientador) e outro em processo de formação (orientando). Especificamente em programas de pós-graduação (PPG), Ferreira et al. (2009) destacam que um discente de PPG é um futuro pesquisador a caminho de sua autonomia científica, porém, ainda dependente de um orientador.

Para melhor compreensão dessas relações, emergem os estudos de genealogia acadêmica (GA). Estudar a GA de determinado grupo de pesquisadores ou de um pesquisador individual visa compreender a herança intelectual perpassada de orientador para orientando¹, possibilitando compreender como se constitui a relação orientador-orientando e de que maneira ocorre a transmissão e difusão do conhecimento entre eles (SUGIMOTO, 2014; MOREIRA; DIAS; MOITA, 2017). Nesse cenário, os estudos empíricos contribuem para a visualização da dinâmica dessas relações e do seu papel para o avanço do conhecimento em um campo científico.

O relacionamento entre um pesquisador mais experiente e outro em formação, institucionalizado e formalizado por algum tipo de orientação,

¹ Se esclarece que nesta pesquisa o termo orientando se alinha ao presente em Miyahara (2011) sendo designado somente a orientações formalizadas e já concluídas.

supervisão e/ou mentoria acadêmica, visa fomentar a formação de novos recursos humanos e conseqüentemente a produção científica advinda desta vinculação.

Sugimoto (2014) descreve a genealogia acadêmica como estudos de cunho empírico da herança intelectual oriunda da orientação acadêmica, pelos quais é possível introduzir meios para análises de como as orientações estão sendo conduzidas.

A principal forma de representar as estruturas genealógicas é por meio dos grafos (rede social; sociograma) que é também, o principal recurso dos estudos genealógicos. Esta representação pode ser compreendida como uma representação matemática que descreve um conjunto de elementos com algum tipo de ligação entre si. Entende-se este conjunto como um composto de vértices (atores da rede) interligados por arestas.

Os grafos de GA, também chamados de árvore de GA, são caracterizados pela sua estrutura hierárquica na qual cada vértice representa um acadêmico e as arestas (de natureza binária - não valorada), o relacionamento de orientação entre eles. O enfoque das redes de GA pode ser descendente, quando se analisam os orientandos e as gerações subsequentes de um determinado pesquisador e/ou ascendente, que analisa as gerações anteriores de orientadores de um determinado pesquisador.

De acordo com Moura e Faria (2021), as árvores de GA favorecem a compreensão da dinâmica evolutiva de pesquisadores (ou grupos de pesquisadores). Podem, ainda, auxiliar agências de fomento e instituições acadêmicas a avaliar o impacto da formação dos novos recursos humanos, além de complementar a avaliação dos resultados das pesquisas geradas por estes pesquisadores. Os autores apontam, também, que a GA aplicada às métricas avaliativas no âmbito dos PPG, em conjunto à Bibliometria, tem o potencial de averiguar a formação de novos pesquisadores e a produção científica.

Nesse sentido, Gabriel Junior e Bufrem (2020) argumentam que a produção da literatura científica fornece elementos para a reflexão acerca da renovação das estruturas dos domínios científicos, a partir da formação de uma linha genealógica científica específica do campo analisado. Analisar a produção científica relacionada a um pesquisador ou a um grupo de pesquisadores no âmbito da GA pode contribuir para uma melhor compreensão do progresso

científico do domínio analisado, seja no âmbito das novas formações, das produções literárias e/ou da perpetuação de linhagens científicas/intelectuais.

Rossi et al. (2018) indicam que várias são as métricas passíveis de aplicabilidade no âmbito da GA, sobretudo com relação aos aspectos topológicos de uma rede (árvore) de GA. É importante ressaltar que a maioria das métricas contempla a evolução das redes de GA, isto é, dá enfoque à visualização das relações existentes e dos novos pesquisados (e gerações) que se inserem e se consolidam na rede analisada. Estas métricas apresentam bons resultados acerca da observação da perpetuação das correntes científicas, contribuindo para a visualização do impacto de um determinado pesquisador (ou grupo de pesquisadores) em seu campo de atuação com relação à formação de recursos humanos e até mesmo da produção científica advinda desta rede.

Entretanto, essas análises não contemplam os valores de conectividade das arestas da rede. Do ponto de vista matemático dos grafos, valorar as ligações (arestas) de uma rede implica em avaliar as intensidades das conexões entre dois atores, ou seja, quanto maior o valor atribuído a uma aresta, mais intensa é a ligação entre dois vértices (atores).

No âmbito da Ciência da Informação, a valoração das arestas significa a mensuração de uma relação, como a valoração da relação de coautoria pela frequência de trabalhos redigidos em coautoria entre dois pesquisadores (ou países, instituições, entre outros) ou a relação de coocorrência de palavras-chave pela frequência com que elas aparecem juntas. Ainda, uma aresta valorada em um grafo representando a relação de citação pode significar tanto a quantidade de citações em comum recebidas em uma literatura analisada (frequência de cocitação) pelos dois pesquisadores que formam a aresta, como a quantidade de referências citadas em comum por esses dois pesquisadores (intensidade de acoplamento bibliográfico), entre outras possibilidades de relação de citação entre dois autores.

Grácio (2020) observa que por meio das análises relacionais de citação, sustentadas na observação simultânea de unidades de análise (autores, documentos, periódicos, entre outros), é possível identificar e analisar as relações que formam a estrutura teórico-metodológica de um campo científico, permitindo visualizar as proximidades, a vizinhança, as associações e a interlocução estabelecida entre as publicações científicas, documento, autores,

periódicos, entre outros. Ainda de acordo com a autora, sob esta ótica, utilizam-se medidas de similaridade de citações entre estas unidades de análises (citação, cocitação e acoplamento bibliográfico).

Especificamente com relação ao acoplamento bibliográfico de autores (indicador relacional de citação) contribui para a visualização da semelhança entre os conjuntos de referenciais teórico-metodológicos adotados pelos dois autores analisados. Especificamente à similaridade no sentido de sobreposição e/ou intersecção entre conjuntos, tem-se a pioneira análise de acoplamento bibliográfico (AB). Formulado inicialmente por Kessler (1962; 1963), o AB prevê a intersecção entre as referências de dois documentos, considerando-os acoplados se possuírem ao menos uma referência em comum.

Do ponto de vista do acoplamento bibliográfico entre pesquisadores (ABA), a análise apresenta-se como um método de mapeamento de autores ativos em um campo de pesquisa para um retrato mais realista do estado atual de suas atividades de pesquisa (ZHAO; STROTMANN, 2008). Dessa forma, esta tese se alinha aos estudos de Shen et al. (2019) ao propor uma nova ótica do AB e compará-las com outras medidas de similaridade. Em seu estudo, Shen e colaboradores analisam além do Índice de Jaccard, o Cosseno de Salton, (também denominado ângulo de acoplamento), a função inversa de documento, e a força de acoplamento de Vladutz e Cook (1984). Apresentam uma nova formulação intitulada *força de acoplamento refinada*, por ser uma combinação entre estas diversas medidas de similaridade e o conceito de AB proposto por Kessler (1963).

Atualmente, várias são as possibilidades de aplicação do acoplamento bibliográfico. Destacam-se, entre os estudos, os esforços de Zhao e Strotmann (2008) que introduziram o acoplamento bibliográfico entre autores (ABA), pelo qual, em alternativa à comparação de documentos proposta por Kessler (1963), dois autores são considerados acoplados se apresentarem ao menos um item citado em comum.

Se esclarece que estudos via AB podem ser realizados entre diversas unidades de análises como: i) entre documentos, considerando as referências citadas em comum; ii) entre pesquisadores, considerando os documentos citados em comum; ii) entre documentos, computando a quantidade de autores

citados em comum. As referências citadas em comum são denominadas unidades de acoplamento por Kessler (1963).

Ao retomar a ideia de similaridade, é possível apontar o AB como a sobreposição de listas de referências entre duas unidades como apresenta Hjørland (2013). Em linhas gerais, independentemente da unidade de análise (entre documentos e/ou autores), o AB configura a intersecção entre dois conjuntos e a força (ou intensidade; frequência) de acoplamento, a quantificação (cardinalidade) dos elementos em comum entre eles.

Ao compreender o AB enquanto uma medida de similaridade entre conjuntos de referências (intersecção), é possível apontar caminhos que relativizam essa medida, visto que o fato de dois autores ou documentos estarem acoplados por determinado número de elementos em comum, não aponta o quanto isso representa em relação ao conjunto total de referências analisado. No âmbito da Ciência da Informação, trabalhos como Sen e Gan (1983), Vladutz e Cook (1984), Glänzel (2003), Zhao e Strotmann (2008), Grácio (2020), Grácio e Oliveira (2013), Oliveira e Alves (2018), Hamers et al. (1989), Leydesdorff (2007), Leydesdorff, Tekles, Bornmann (2021), Shen et al. (2019) aplicam e discutem índices de similaridade como o coeficiente de correlação de Pearson, o Índice de Jaccard e o Cosseno de Salton aplicados ao AB e às análises de cocitação.

Small (1973) já apontava a relevância do índice de similaridade de Jaccard para o contexto da cocitação de documentos. Nesse contexto, destaca-se assim a viabilidade desta medida para a relativização (ou normalização) também em análises relacionais entre outras unidades, como periódicos, coocorrência de palavras-chaves, autores, entre outros, relativizando a intensidade da conexão entre elas e possibilitando uma compreensão mais fidedigna das reais proximidades estabelecidas no universo analisado.

Já em Castanha, Bochi e Bufrem (2020) e Castanha e Grácio (2020) apontam o método de ABA como potencial indicador para as análises de genealogia acadêmica, bem como para identificação das correntes teóricas de um determinado domínio.

A fim de expandir os estudos relacionados ao acoplamento bibliográfico elencado à genealogia acadêmica, surge a necessidade da aplicação de novas métricas que sejam capazes de refletir de forma mais eficiente e representativa

as influências intelectuais e científicas geradas por meio das orientações acadêmicas, vista a importância deste tipo de relacionamento. Desta forma, explorar a intensidade de acoplamento bibliográfico entre orientador e orientando por meio de índices de similaridade, poderá apontar novos caminhos para a aplicação desta e das demais métricas baseadas em citações.

Por mais que seja clara a aplicabilidade do acoplamento bibliográfico, poucos estudos enfatizam o valor da intersecção gerada pela comparação entre os conjuntos de referências, isto é, o número de itens em comum. Geralmente, apresenta-se o parâmetro para validar se existe proximidade ou não entre os pesquisadores, não havendo uma discussão acerca do que, de fato, representa este valor em relação ao conjunto de referências e autores analisados.

Trabalhos como Zhao e Strotmann (2008) e Grácio (2020) indicam caminhos para a aplicação do acoplamento bibliográfico por meio da recitação de autores, sustentados nas ideias e conceitos propostos por White (2001), no intuito de analisar a sobreposição das identidades científicas de citação dos autores analisados.

Em seu trabalho, White (2001) discorre sobre identidade e padrões de citação, de modo que a identidade é representada pelo conjunto de autores recitados por um determinado autor. O autor aponta dois tipos de autores citados: aqueles citados somente uma vez e os autores citados duas vezes ou mais (recitados), sendo que as autocitações podem se encaixar em qualquer um dos casos.

White (2001) define que os padrões de recitação são parte fundamental da identidade de citação dos autores e destaca que a recitação pode ser subdividida em: i) sincrônica: quando o pesquisador cita determinado autor mais de uma vez em uma única obra; ii) recitação diacrônica: quando o pesquisador cita um mesmo autor em mais de um dos seus trabalhos.

Segundo Zhao e Strotmann (2021), autores representam escolas de pensamento, enquanto os artigos redigidos por eles representam evidências ou descobertas individuais sobre conceitos, teorias e/ou métodos. Nesse contexto, indivíduos tendem a desenvolver um conjunto de fontes de informação que preferem consultar quando contribuem para a produção do conhecimento e, assim, quanto mais informação dois indivíduos compartilham, mais intimamente seus interesses e crenças podem estar relacionados.

Do ponto de vista teórico-metodológico desta tese, vários são os estudos que abarcam os conceitos do acoplamento bibliográfico ou da genealogia acadêmica, porém, a junção entre as temáticas ainda é incipiente, tendo Castanha (2019) e Castanha e Grácio (2020) como seus precursores. Assim, associar o acoplamento bibliográfico a redes de genealogia acadêmica e aos padrões de recitação diacrônica de autores descritos por White (2001) constitui uma perspectiva original que enriquece a análise de proximidades teóricas entre os pesquisadores no escopo da relação orientador-orientando.

Alguns trabalhos que trazem como metodologia de análise o acoplamento bibliográfico aplicado nos mais diversos universos e domínios científicos podem ser ressaltados, como: Zhao e Strotmann (2008; 2014; 2021); Thijs, Zhang e Glanzel (2014), Lucas e Zorita (2014), Grácio (2016), Castanha e Grácio (2017; 2020), Franco (2018), Castanha (2019), Castanha, Bochi e Bufrem (2020).

Em busca², por meio do descritor “*bibliographic coupling*” (termo em língua inglesa) e “acoplamento bibliográfico” (língua portuguesa), nas bases de dados *Web of Science* (WoS), *Scopus* e na Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação (BRAPCI), foram encontrados 786, 942 e 15 documentos, respectivamente.

Já com relação à produção relacionada a genealogia acadêmica, enquanto relacionamento entre orientador e orientado, como metodologia para estudos que analisam as heranças intelectuais, registros históricos da formação de novos pesquisadores, propagação de determinada geração de orientações, importância de determinado pesquisado para o progresso do campo que está inserido, análises de multi/interdisciplinaridade, impactos de um determinado pesquisador, além da análise de florestas³ de genealogia acadêmica, é possível destacar os estudos de Sugimoto (2014), Rossi e Mena-Chalco (2014); Mena-Chalco (2016), Bufrem, Silva e Sobral (2017), Damaceno, Rossi e Mena-Chalco (2018), Oliveira, Moreira e Dias (2018), Rossi et. al (2018), Viana et al. (2019), Castanha e Grácio (2020), Gabriel Junior e Bufrem (2018; 2020) e Moura e Faria (2021).

² Pesquisa realizada em 30 de dezembro de 2022.

³“Uma árvore de genealogia acadêmica é uma estrutura em que cada vértice é um orientador acadêmico e as arestas (direcionadas) representam as relações de orientação. Um conjunto de árvores pode ser denominado floresta” (ROSSI; MENA-CHALCO, 2014, p.22)

A fim de se identificar a presença de pesquisas dedicadas à temática, realizaram-se buscas⁴ nas mesmas bases de dados supracitadas com o descritor “*academic genealogy*” (termo em língua inglesa para genealogia acadêmica) e “genealogia acadêmica”. Foram encontrados 55 documentos na WoS, 69 documentos na *Scopus* e 15 documentos na BRAPCI.

Destacam-se aqui também alguns projetos com foco afim aos estudos baseados na GA, como Rossi, Damaceno e Mena-Chalco (2018) mencionam: *Mathematics Genealogy Project*; *The Neurotree Project*; *The Academic Family Tree*; *RePEc Genealogy*; Plataforma Acácia. Especificamente com relação a *Mathematics Genealogy Project*, os autores o destacam como uma base relevante para análises relacionadas ao processo de transferência de conhecimentos de doutores da área da Matemática.

Já ao reunir os descritores⁵ (“*bibliographic coupling*” e “*academic genealogy*”; “acoplamento bibliográfico” e “genealogia acadêmica”) por meio da conjunção *AND* nas mesmas bases, foram recuperados somente um documento em cada uma das três bases. E mais, a fim de flexionar a busca⁶, optou-se pela conjugação entre “*bibliographic coupling*” e “*genealogy*”, obtendo cinco documentos na WoS, cinco na *Scopus* e um na BRAPCI. Sendo que os cinco artigos das bases WoS e *Scopus* são os mesmos.

Estes cinco documentos referem-se aos seguintes trabalhos: Hu et al. (2022), Galvagno e Pisano (2020), Sun (2011 et. al), Sun e Zhu (2012) e Castanha e Grácio (2020). No primeiro, os autores utilizam o conceito de genealogia na perspectiva da evolução para avaliar a literatura sobre *digital pathology* (patologia digital). Por meio de análises de cocitação e acoplamento bibliográfico, caracterizam a tendência evolutiva dos documentos encontrados em *upstream*, *midstream*, *downstream*. Já no segundo documento, os autores analisam a genealogia da literatura por análises de citação, cocitação e acoplamento dos documentos da área de *Internationalization of Family Business* (IFB). Nos outros dois artigos, os autores introduzem o conceito de genealogia genética das citações, definida a partir das relações de citação para ilustrar as

⁴ Pesquisa realizada em 30 de dezembro de 2022.

⁵ Pesquisa realizada em 30 de dezembro de 2022.

⁶ Pesquisa realizada em 30 de dezembro de 2022.

heranças da literatura científica (documentos) e operacionalizada (estas relações) a partir cocitação e do acoplamento bibliográfico.

O último artigo dessa busca refere-se ao único resultado oriundo da busca conjugada entre genealogia acadêmica e acoplamento bibliográfico. Os autores abordam a análise de genealogia acadêmica utilizando o acoplamento enquanto método para análise de proximidades entre pesquisadores. Trata-se do trabalho intitulado *Indicadores de Acoplamento Bibliográfico para a avaliação da proximidade teórico-metodológica em redes de genealogia acadêmica um estudo aplicado aos descendentes bolsistas PQ de Aldo Barreto de Castanha e Grácio (2020)*. Neste estudo, os autores analisam as proximidades entre o único pesquisador sênior do campo de Ciência da Informação no Brasil, Aldo Barreto, e seus orientandos que se tornaram bolsistas de produtividade em Pesquisa (CNPq) por meio do acoplamento bibliográfico de autores.

Nesse contexto, esta tese pretende responder ao seguinte problema: Qual o coeficiente de acoplamento bibliográfico próprio para a mensuração da intensidade da transmissão das correntes teórico-metodológicas no escopo das relações de genealogia acadêmica?

O estudo tem como hipótese: O Coeficiente de Acoplamento Bibliográfico Genealógico, derivado do coeficiente normalizado de Acoplamento Bibliográfico de Autores, mensura de forma representativa a intensidade de transmissão da corrente teórico-metodológica na relação de genealogia acadêmica ao normalizar o cálculo de acoplamento pela cardinalidade da lista de referências do pesquisador orientador.

Desse modo, como objetivo, esta tese propõe um coeficiente de perspectiva relacional de citação⁷ e recitação, denominado Coeficiente de Acoplamento Bibliográfico Genealógico (CABG), para mensurar a intensidade das proximidades teórico-metodológicas entre pesquisadores no escopo das relações de genealogia acadêmica.

De maneira específica: i) fundamentar o Coeficiente de Acoplamento Bibliográfico Genealógico; ii) comparar três formas de normalização do Acoplamento Bibliográfico de Autores: Índice de Jaccard, Cosseno de Salton e

⁷ análises relacionais de citação baseiam-se em pares de dados ao examinar as referências de uma literatura e têm por finalidade identificar e evidenciar a coocorrência de um par de referências na literatura científica (GRÁCIO, 2020, p. 89)

Coeficiente de Acoplamento Bibliográfico Genealógico; iii) para o universo dos pesquisadores Produtividade em Pesquisa seniores da Matemática vinculados ao Instituto de Matemática Pura e Aplicada (IMPA), caracterizar a floresta de genealogia acadêmica valorada pelo coeficiente proposto; iv) avaliar a conexão entre os principais influentes da floresta de genealogia acadêmica, gerando a estrutura intelectual dos pesquisadores que a compõem.

A escolha deste universo para a pesquisa empírica tem a intenção de realizar uma prova de conceito do novo indicador, e sob o ponto de vista matemático e social da ciência, analisar as proximidades teóricas entre orientadores e orientandos (egressos de doutorado) do Instituto de Matemática Pura e Aplicada (IMPA), visto que estas relações preveem uma proximidade não só de cunho acadêmico, mas também social, que auxiliam no desenvolvimento e consolidação de novos pesquisadores. Além disso, o campo em questão (Matemática), como supracitado, é alvo de diversos projetos no âmbito da GA.

Esta pesquisa, ao formular o Coeficiente de Acoplamento Bibliográfico Genealógico (CABG), baseado no acoplamento bibliográfico, poderá validar sob a ótica relacional dos padrões de citação de autores, a propagação de conhecimento dos orientadores em relação aos orientandos. Além do mais, possibilitará, a partir de um modo inovador e vantajoso perante outras medidas de similaridade, a detecção dos autores que historicamente se consolidaram como grandes influenciadores do pensamento e obra deste grupo de pesquisadores, sendo responsáveis pelas linhagens intelectuais construídas ao longo de suas carreiras.

Toda a análise é baseada na mensuração do número de autores citados em comum na relação orientador-orientando, configurando-se então, um estudo relacional de citação aplicado à GA. A elaboração de um indicador alternativo (e inédito) comparado a índices de similaridade consolidados – índice de Jaccard e Cosseno de Salton - contribui para o refinamento e avanço dos estudos referentes a proximidades teóricas entre pesquisadores, pelo ajuste aos âmbitos em que são aplicados.

As análises relacionais de citação, independentemente de sua configuração (citante-citado: citação; citante-citante: acoplamento bibliográfico; citado-citado: cocitação) captam ou não a presença de conexões entre unidades de análise e quantificam a similaridade entre elas. Aplicar métricas relacionais

em redes de GA poderá dar continuidade a estudos relacionados a questões topológicas no âmbito das análises de redes sociais (de genealogia acadêmica) ao valorar as arestas destas redes, e também aos estudos epistemológicos e históricos da perpetuação da produção científica de um determinado grupo de pesquisadores ao observar a herança intelectual perpassada dos orientadores para os orientandos.

Tanto em âmbito nacional quanto internacional, considera-se que uma análise da genealogia acadêmica de pesquisadores da área da Matemática, juntamente com a criação de um novo indicador e sua comparação a outras medidas, trará uma nova perspectiva a respeito do acoplamento bibliográfico e poderá contribuir com novos olhares sobre a dinâmica científica das relações de orientação de pesquisadores renomados da área de Matemática no Brasil.

Assim, esta pesquisa auxilia estudos bibliométricos e cientométricos voltados a compreensão, discussão e visualização do papel dos indicadores métricos (relacionais de citação) no processo de construção do conhecimento e análise da dinâmica da ciência de um determinado campo por meio dos estudos de genealogia acadêmica.

A fim de situar este estudo nos três grupos de pesquisas descritos em Glänzel (2003), retoma-se: G1) Bibliometria para profissionais da Bibliometria: que busca desenvolver e debater a Bibliometria como metodologia e está preocupada com o seu próprio desenvolvimento; G2) Bibliometria aplicada às disciplinas científicas: grupo que tem por objetivo fomentar a pesquisa bibliométrica aplicada, sendo o maior e mais diversificado grupo de interesse na Bibliometria; G3) Bibliometria para a política científica e gestão: esse é o domínio da avaliação da pesquisa com fins de orientar políticas científicas.

O estudo do acoplamento bibliográfico, enquanto indicador de similaridade entre pesquisadores, comparando o novo coeficiente proposto a outras medidas já reconhecidas na literatura, por meio da aplicação ao universo das orientações acadêmicas dos bolsistas PQ Sênior da área de Matemática, alinha-se ao grupo G1 de Glänzel, pois estas medidas são analisadas de forma empírica e geram insumos para o desenvolvimento metodológico da própria Bibliometria. O domínio da pesquisa bibliométrica considerada “de base”, em relação a outras pesquisas, adequa-se, assim, às necessidades de seu próprio desenvolvimento conceitual-teórico-metodológico. A oportunidade de aplicar

esta análise a pesquisadores de uma determinada disciplina científica (Matemática) revela uma aderência, também, ao grupo G2. Desta maneira, esta pesquisa promove a junção dos grupos G1 e G2.

Esta tese estrutura-se em oito seções. Na primeira, é introduzida a contextualização, objeto de estudo, problema e objetivos da pesquisa e a justificativa da sua contribuição para o desenvolvimento do campo científico da Ciência da Informação, especialmente dos Estudos Métricos da Informação (EMI) (especificamente à Bibliometria).

Na segunda, terceira e quarta seções, estão os aspectos teóricos que embasam toda a análise. A segunda seção traz o embasamento referente aos estudos genealógicos da ciência e apresenta questões relacionadas a genealogia acadêmica, científica e intelectual. A terceira seção discorre sobre as análises relacionais de citação, cocitação e acoplamento bibliográfico, trazendo aspectos históricos e definições conceituais voltadas a estes estudos. Esta seção se desdobra ainda em uma subseção que contempla a relação entre o acoplamento bibliográfico e medidas de similaridade.

A quinta seção refere-se à formulação do Coeficiente de Acoplamento Bibliográfico Genealógico e contextualiza a nova medida do ponto de vista matemático e genealógico. A sexta seção apresenta todos os aspectos metodológicos desta tese e subdivide-se em duas frentes: a primeira sobre o universo e o corpus da pesquisa e a segunda, sobre os procedimentos metodológicos. Aqui foram discutidos e apresentados todas ferramentas e meios de aplicação da análise.

A penúltima seção apresenta e discute todos os resultados obtidos. Inicia-se com os cálculos de acoplamentos, depois tem-se a validação do CABG e segue para a identificação e construção da estrutura intelectual analisada. Esta última composta por análises de citação e cocitação de autores. A última seção traz as considerações finais referentes aos resultados obtidos, às dificuldades encontradas durante o seu desenvolvimento e os próximos passos a serem desenvolvidos na continuidade da pesquisa.

Esta tese ainda conta com dois apêndices, contendo a matriz de recitação dos autores acopladores de toda floresta de GA analisada e o código em R do *The Genealogical Coupler*.

2 ESTUDOS DE GENEALOGIA DA CIÊNCIA

A ciência evoluiu ao longo dos séculos como um sistema que promove o progresso do método científico e dos processos de ensino, em que a atividade de tutoria acadêmica favorece o progresso científico e fomenta o surgimento de novos pesquisadores (DORES et al., 2017). Os autores também alertam que apesar de toda complexidade do sistema científico, geralmente os estudos buscam analisar a produtividade de pesquisa por meio de publicações produzidas, citações recebidas e colaborações estabelecidas, não levando em conta a formação de novos pesquisadores.

Tendo em vista as possibilidades de visualização e entendimento histórico dos processos de formação, orientação e conseqüentemente da difusão do conhecimento, realizar a modelagem e caracterização de árvores genealógicas acadêmicas surge como uma alternativa viável para a análise de como a ciência se propaga nas várias áreas de estudo (MOREIRA; DIAS; MOITA, 2017).

Os estudos de genealogia acadêmica, por meio de relação entre orientador e orientando, são de grande utilidade para o registro histórico de pesquisadores atuantes nos diversos campos do conhecimento. Este tipo de análise visa contribuir para a visualização diacrônica da difusão do conhecimento e, por meio dos estudos métricos da informação, pode oferecer subsídios para a sua caracterização e compreensão de sua evolução (ROSSI; MENA-CHALCO, 2014).

Do ponto de vista matemático, várias métricas topológicas podem ser aplicadas aos gráficos de GA de modo a caracterizar um indivíduo ou um grupo de pesquisadores, ou seja, é possível atribuir valores numéricos às relações presentes na rede que caracterizem aspectos que englobam a formação e capacidade de propagação de suas linhagens acadêmico-científicas.

Rossi et. al (2018) apresentam várias delas, a saber: i) a fecundidade (quantidade de orientandos; filhos acadêmicos); ii) a descendência (quantidade total de descendentes de um pesquisador; todas as gerações são consideradas: filhos, netos, bisnetos etc.); iii) gerações (quantidade de linhagens acadêmicas de uma rede; horizontalidades de uma rede).

Em complemento, segundo Damaceno, Rossi, Mena-Chalco (2017) e Damaceno et al. (2019), tem-se: Fertilidade (Número de descendentes que um

acadêmico orientou que orientaram cada um, no mínimo, um descendente direto. É o número de alunos orientados que já são orientadores); Grande área (Primeira grande área do conhecimento registrada pelo acadêmico na Plataforma Lattes); Índice genealógico "g" (índice genealógico g de um acadêmico é o maior número g de descendentes (diretos) que, por sua vez, possuem, no mínimo, g descendentes diretos cada); Primos (Número de acadêmicos que foram orientados por pais diferentes que, por sua vez, possuem o mesmo pai); Relações (Número de relações na descendência de um acadêmico. Reflete o número de arestas que existem ao longo da descendência do acadêmico analisado).

Já do ponto de vista histórico, é possível destacar os trabalhos pioneiros de Sachtleben (1951) e de Montoye e Washburn (1980) intitulados respectivamente de *Nobel prize winners descended from liebig: A table of academic genealogy* e *Research Quarterly Contributors: An Academic Genealogy*, os mais antigos recuperados na busca em busca feita na utilizando-se de “*academic genealogy*” (ou “genealogia acadêmica”) nas bases de dados Scopus e WoS (seção 1 – Introdução).

O primeiro apresenta a árvore de genealogia acadêmica dos vencedores do Prêmio Nobel entre os descendentes acadêmicos do pesquisador alemão Justus von Liebig. Justus foi um professor e químico renomado do século XIX e famoso pelo *modelo de Giessen*, nomeado em homenagem a cidade em que o pesquisador iniciou suas atividades de ensino e pesquisa. Além disso, Liebig é conhecido pelo seu famoso laboratório (Laboratório de Giessen) como aponta Maar (2006, p. 1132):

O laboratório de Liebig tinha ênfase maior na análise química (o laboratório de Stromeyer já apontava nesse sentido); não era seu objetivo manter um laboratório de demonstração, para o ensino de uma Química prática, ou quem sabe a redescoberta de princípios químicos a partir do experimento: o laboratório de Giessen ensinava a pesquisar em Química, e essa era a grande novidade. Embora cada aluno de Liebig pesquisasse seu próprio assunto, todos os assuntos eram discutidos em grupo, para proveito comum: surge pela primeira vez em um curso de química de uma universidade uma equipe de pesquisadores (nesse aspecto difere do laboratório de Stromeyer, no qual nunca se atuou em ‘equipe’).

Este viés justifica, no trabalho de Sachtleben (1951), a identificação de 61 descendentes acadêmicos de Leibig, difundidos em cinco gerações de descendentes, sendo 34 deles, vencedores do prêmio Nobel.

Já o segundo trabalho mais antigo recuperado, homenageia os 50 anos do periódico da área de Educação Física *Research Quarterly for exercise and Sport*. Os autores construíram a árvore de genealogia acadêmica dos principais autores que contribuíram com a revista durante seu primeiro meio século de existência. O intuito do estudo foi observar se estes autores influenciaram novos pesquisadores a publicar estudos científicos na área.

A análise identificou 135 autores, dando destaque a aqueles com cinco ou mais contribuições entre de 1930 até 1976, e as universidades de formadoras dos pesquisadores. Os autores conseguiram recuperar estas informações de 106 autores e construíram a rede de genealogia acadêmica ao qual se referiram como “gráfico genealógico” (*genealogical charts*), “árvore genealógica” (*Family tree*) e “arvore de genealogia acadêmica” (*academic Family tree*).

Montoye e Washburn (1980, p. 266) concluem a pesquisa buscando motivar atuais e futuros pesquisadores “a assumir a responsabilidade de treinar e inspirar os alunos a avançarem nas fronteiras do conhecimento por meio de uma vida de pesquisa. Só assim o futuro da Educação Física como disciplina estará assegurado”. Este estudo pode ser entendido como pioneiro na conjugação entre bibliometria e GA, visto que analisar autores que mais contribuem em uma revista científica é algo de interesse dos estudos bibliométricos.

É possível observar que ambos os trabalhos supracitados estão alinhados à Rossi e Mena-Chalco (2014) ao descobrir informações que auxiliem na compreensão a respeito de sua formação, capacidade de propagação e diferenciação e à Castanha e Grácio (2015) em que as autoras mencionam que árvores e florestas de GA auxiliam na visualização e caracterização de grupos e correntes científicas atuantes em um determinado domínio. Nestes casos, o primeiro domínio é representado por descendentes de um renomado autor e o segundo, um periódico em específico.

Nesse contexto, Lv e Chang (2021) sugerem que ao estudar a genealogia acadêmica de pesquisadores de maneira individual, é possível observar a relação disciplinar entre formação e atuação de acadêmicos, restaurar a

trajetória histórica da ciência e da tecnologia, evoluções de disciplinas, comparar os desenvolvimentos da ciência e da realidade tecnológica e prever progressos de disciplinas e direções inovadoras.

De maneira coletiva, a genealogia de grupos de cientistas pode revelar o espírito científico e o estilo acadêmico, identificar tradições científicas ou promover a formação de tradições científicas e incubar culturas científicas e inovadoras para explorar o crescimento do talento científico e tecnológico (LV; CHANG, 2021).

Nesse sentido, Mena-Chalco (2016) classifica os estudos de genealogia acadêmica em cinco tipos, em que cada uma delas dá ênfase a uma característica específica: i) genealogia honorífica: ênfase na descendência de um pesquisador para evidenciar o impacto nas gerações de acadêmicos; ii) genealogia egocentrista: ênfase na ascendência de um pesquisador para evidenciar os ancestrais mais importantes; iii) genealogia histórica: ênfase na contextualização histórica (análise histórica) do desenvolvimento de uma área. Geralmente considera os pioneiros/fundadores das áreas analisadas; iv) genealogia paradigmática: ênfase em como o conhecimento e práticas epistemológicas são transmitidas através das relações de orientação; v) genealogia analítica: Ênfase na aplicação de métodos matemáticos e computacionais para criar e estudar as relações de orientação.

Castanha e Grácio (2015) propõem o uso da GA enquanto uma nova abordagem da análise de domínio, teoria idealizada por Biger Hjørland.

A análise genealógica acadêmica, forma particular de rede social, firma-se em um paradigma social, uma vez que a atividade científica apresenta um modelo social hierárquico, a partir das relações entre os pesquisadores e da forma que se consolida a transmissão de conhecimentos e práticas epistêmicas por meio dessas relações. Especificamente, no vínculo orientador-orientando, observa-se uma relação hierárquica social de interdependência, em que há uma forma particular de disseminação e transferência de conhecimento, que se reflete na comunidade discursiva e carrega a subjetividade existente no domínio em que estão inseridos (CASTANHA; GRÁCIO, 2015, p. 111).

Um conjunto de relacionamentos de orientação enquanto domínio específico torna-se um meio útil para análises bibliométricas e cientométricas visto que aspectos científicos, históricos e epistêmicos podem ser descobertos

e, dessa maneira, apontar caminhos futuros para o desenvolvimento científico, intelectual e da formação de recursos humanos deste domínio.

Além do mais, é possível destacar outros estudos genealógicos da ciência para além da genealogia acadêmica, como a genealogia científica (GC) e a genealogia intelectual (GI). Sugimoto (2014) e Mena-Chalco (2016) alertam para a distinção da terminologia entre as genealogias em que a primeira se refere às descobertas, investigações e evoluções em nível de ciência, e, a segunda, a qualquer tipo de influência intelectual. Nesta última, se encontram pesquisadores que se intitulam, por exemplo, *freudianos*, *marxistas*, *foucaultianos* e etc.

Tanto a genealogia científica quanto a intelectual se diferem da GA, essencialmente, por não serem relações necessariamente institucionalizadas, diferentemente da GA, em que uma orientação (ou supervisão) é formalizada por algum tipo de vínculo e, como indica Miyahara (2011) é consolidada somente quando finalizada.

Em Andraos (2005), é estudada a genealogia científica de cientistas da área de Química que trabalham em laboratórios em universidades canadenses e no *National Research Council of Canada* que contribuíram com estudos relacionados a *organic reaction mechanisms*. O autor construiu a árvore de genealogia científica destacando orientações de doutorado e supervisões de pós-doutorado entre cientistas e as conexões entre as ideias científicas presentes na árvore. Nota-se que neste caso há uma associação entre as genealogias acadêmica (orientações e supervisões) e científica (ideias científicas).

De maneira semelhante, Gabriel Junior e Bufrem (2018; 2020) e Castanha e Grácio (2020) analisam a genealogia acadêmica e científica a partir de Aldo Barreto, expoente da Ciência da Informação brasileira. Os trabalhos identificaram a rede de orientandos de Aldo e por meio de análises relacionais apresentaram os principais influentes teóricos da rede de orientação.

Com relação à genealogia intelectual, este tipo de estudo visa analisar o fluxo de ideias, influências e correntes de pensamento perpassados ao longo do tempo, sem a obrigatoriedade de vínculos institucionalizados e formalizados entre pesquisadores. Esta perspectiva visa a identificação e compreensão das influências de autores predecessores sobre determinada temática e/ou domínio científico.

Safer e Tang (2009) indicam que na área de Psicologia a análise de citações tem sido usada para melhorar a compreensão da genealogia intelectual de ideias, o desenvolvimento de estudos interdisciplinares e colaborativos. Nesse contexto, em Bufrem, Silva e Sobral (2017), é realizada uma análise do fluxo genealógico intelectual, por meio da técnica de análise de citações, dos bolsistas PQ da área de Ciência da Informação, visando compreender as influências e bases intelectuais sobre este grupo de pesquisadores.

Os estudos genealógicos, independentemente de sua vertente acadêmica, científica e/ou intelectual, aplicados a um pesquisador ou a um grupo de pesquisadores, conjugando métodos e técnicas dos Estudos Métricos da Informação, especificamente a Bibliometria, aliados à produção e comunicação científica, fornecem subsídios para o avanço de estudos relacionados a origem e progresso de campos do conhecimento. Ao analisar a literatura científica de determinados campos, os EMI têm capacidade de identificar influências por meio dos estudos relacionais de citação “representando um tipo de “afinidade eletiva”, demonstrada durante a trajetória de um pesquisador ao selecionar e citar, por meio direto ou indireto, autores seminais ou fundantes para sua produção” (GABRIEL JUNIOR; BUFREM, 2020, p. 10).

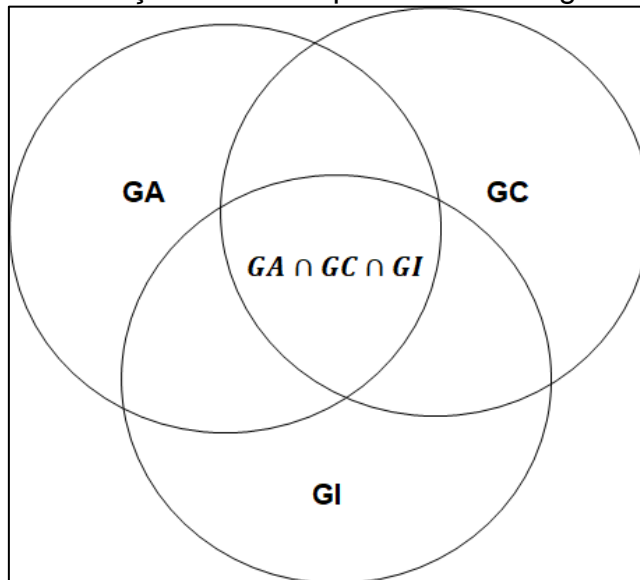
Esta afinidade e demonstração pode ser entendida como uma perspectiva relacional dos estudos de genealogia atrelados aos estudos bibliométricos voltados à análise da produção científica, uma vez que influências acadêmicas, científicas e/ou intelectuais podem ser expressas por meio da citação de autores ou obras que de alguma forma contribuíram para o desenvolvimento daquilo que foi pesquisado e conseqüentemente publicado.

A influência da escolha e das razões de citar determinado documento ou autor é discutida em Silveira, Caregnato e Bufrem (2017), ao debaterem as categorias de citação propostas por Ahmed et. al (2004) e Araújo (2009) no âmbito da Ciência da Informação. As autoras esclarecem que a análise de citação é um importante meio para melhor compreender a apropriação de conhecimentos científicos por pesquisadores, grupos e/ou instituições sendo utilizados para avaliar relacionamentos científico-sociais, tendências, enfoques e influências entre pesquisadores e/ou domínios científicos.

Gabriel Junior e Bufrem (2020, p 12) inferem, após realizar uma análise de genealogia acadêmica e científica, utilizando-se da análise de cocitação que

“foi possível também experimentar uma metodologia fundamentada nas relações entre os autores seminais, cujas teorias, conceitos e métodos serviram de base para os estudos da linhagem científica e dos princípios da genealogia científica”. A relação entre os tipos de estudos genealógicos é apresentada na Figura 1.

Figura 1 - Relação entre os tipos de estudos genealógicos



Fonte: Adaptado de Castanha (2019)

A relações presentes na Figura 1 não são mutuamente exclusivas, visto que é possível estabelecer relações par a par de combinações, como o relacionamento entre GA e GC, GA e GI, GC e CI e entre GA, GC e GI. A primeira combinação pode ser entendida como estudos que buscam identificar e quantificar orientações (ou supervisões) acadêmicas concluídas, além de identificar as evoluções e influências científicas advindas de orientador para orientando.

Já estudos que promovem a intersecção entre GA e GI poderão, a partir de relações entre orientador e orientando, formalizada, identificar os fluxos de ideias perpassados ao longo do tempo e ao longo das gerações de orientação acadêmica. A relação entre GI e GC possibilitará a análise de evoluções científicas juntamente com fluxos de ideias, sem a necessidade da estrutura de orientação acadêmica institucionalizada.

Como maior grau de sobreposição, a relação entre GA, GC e GI pode ser descrita como estudos capazes de analisar evoluções científicas e fluxos de ideias no âmbito das relações formalizadas entre orientador e orientando.

Exemplifica-se esta sobreposição: considere um pesquisador orientador de um programa de pós-graduação em Psicologia dito *freudiano* que orienta trabalhos relacionados a psicanálise e formou um orientando (egresso) que também se intitula *freudiano*. Neste caso, tem-se a genealogia intelectual representada pela escola de pensamento *freudiana*, a genealogia científica representada pelos avanços nos estudos psicanalíticos e a genealogia acadêmica por se tratar de um programa de pós-graduação que formaliza e institucionaliza orientações acadêmicas.

Em complemento, Mena-Chalco (2016) alerta que os estudos de genealogia acadêmica oferecem uma visão alternativa às abordagens tradicionais da Bibliometria e Cientometria, que raramente focam a formação e projeção de novos recursos humanos. Esta alternativa não deve ser entendida como algo menos expressivo, mas sim como um tipo de estudo tão robusto quanto os tradicionais. E mais, o autor indica o *genealogista acadêmico (cientista de dados de genealogia)* como profissional fundamental para a descoberta de conhecimento em grandes conjuntos de dados (MENA-CHALCO; 2016).

3 ANÁLISES RELACIONAIS DE CITAÇÃO: COCITAÇÃO E ACOPLAMENTO BIBLIOGRÁFICO

A produção científica oriunda dos diversos campos do conhecimento tem papel fundamental pois comunica à comunidade científica que novas pesquisas estão sendo fundamentadas. Sob esta ótica, ao tornarem público suas pesquisas, os pesquisadores submetem seus resultados à avaliação, a fim de apresentar novos conceitos, métodos e teorias. Assim, a atividade científica está atrelada à avaliação e à divulgação.

De acordo com Bufrem e Prates (2005), compreender essa atividade e sua importância indica que a ciência como um todo pode ser analisada sob a ótica da Sociologia da Ciência por meio de instrumentos metodológicos como a Bibliometria, Cientometria e Informetria. Historicamente, tais instrumentos, ou de maneira geral os EMI, originam-se da “necessidade de visualizar, analisar e avaliar a dinâmica e evolução da atividade científica e sua produção; e a gestão de livros e bibliotecas” (FREITAS et. al, 2017, p. 29). Desta maneira, as interações sociais no âmbito científico, resulta, por diversas vezes, em documentos científicos, unidade fundamental para analisar a evolução dos diversos campos do conhecimento.

Compreender que a produção científica evolui a partir de publicações anteriores é imprescindível para que investigações a respeito da dinâmica científica de determinado domínio seja realizada, emergindo as análises de citação.

As citações representam uma importante forma de reconhecimento e o ato de citar distingue as contribuições feitas em trabalhos anteriores no novo trabalho construído, seja como uma afirmação ou uma contestação da obra citada (AJIFERUKE; LU; WOLFRAM, 2010).

Tradicionalmente, atribui-se ao trabalho de Gross e Gross (1927) a primeira análise de citação, em que os autores contabilizaram referências do *The Journal of the American Chemical Society*. Posteriormente, entre as décadas de 1950 e 1960, as análises de citação ganham maior notoriedade a partir dos esforços de Eugene Garfield e, mais precisamente em 1963, com o surgimento do índice de citações *Science Citation Index* (SCI) atrelado ao *Institute of Scientific Information* (ISI), fundado por Garfield.

Dessa maneira,

A análise das referências de um autor ou de um conjunto de produções bibliográficas em uma literatura científica permite conhecer seus principais interesses e cerne científicos. Desse modo, a medida que a obra de um pesquisador cresce, o conjunto de autores referenciados em mais de uma das suas publicações reflete sua identidade científica, em especial, a partir daqueles autores cuja reincidência ocorra com maior frequência (GRÁCIO, 2020, p. 79).

É possível apontar que os Estudos Métricos da Informação (EMI) buscam, por meio de suas metodologias e técnicas, a construção de indicadores nos diversos âmbitos e domínios, inclusive relacionados à CT&I (PRADO; CASTANHA, 2020) baseados em diversos aspectos da produção científica, como as das citações recebidas e/ou do conjunto de referências de artigos ou pesquisadores.

Os indicadores buscam auxiliar o entendimento da dinâmica de Ciência, Tecnologia e Informação (CT&I), além de se apresentarem como instrumentos para o planejamento de políticas públicas e tomadas de decisões referentes à ciência e de ter a capacidade de equacionar problemas científicos e buscar soluções de como propor e validar tais indicadores relativos ao tema, definindo seus alcances e limitações (SANTOS, 2003).

Desta maneira, atividades de produção de indicadores quantitativos se fortalecem com a necessidade de dispor de ferramentas e meios para definição de diretrizes, investimentos, formulação de programas e avaliação de atividades relacionadas ao desenvolvimento científico e tecnológico no país (MUGAINI; JANUZZI; QUONIAM, 2004).

No âmbito dos EMI, os indicadores de citação e/ou produção científica tem ganhando cada vez mais espaço também porque parte dos esforços da ciência se concentra na elaboração e formulação de indicadores, de modo que estes instrumentos possam favorecer a avaliação da dinâmica de produção da ciência e auxiliar no monitoramento da produção científica, possibilitando uma estimativa de como os países contribuem para o desenvolvimento científico e tecnológico (CASTANHA, 2019).

Em um contexto de pesquisa acadêmica, os indicadores de citação podem ser usados como ferramentas para examinar relações universais entre variáveis dentro de uma estrutura de conhecimento. Já, no contexto político ou

de gestão, os indicadores de citação podem ser usados para se chegar a algum tipo de decisão política (MOED, 2017).

Dado o aumento global da produção científica nos diversos níveis, analisar a produção científica de instituições, pesquisadores, áreas do conhecimento, países, dentre outros, remete à utilização de indicadores específicos que podem ser subdivididos, segundo Prado e Castanha (2020) em: i) indicadores de produção; ii) indicadores de citação; iii) indicadores de ligação (ou de relação).

Em resumo, os indicadores de produção podem ser entendidos como a frequência (contagem; quantificação) de publicações por tipo de documento como livros, artigos, publicações em anais de eventos, relatórios e etc., nos diferentes níveis de agregação, seja individual (pesquisador) ou institucional, por área de conhecimento, por país, dentre outros (PRADO; CASTANHA, 2020).

Estes indicadores podem ser considerados os precursores no âmbito dos EMI, principalmente com relação a Bibliometria e suas leis clássicas como a Lei de Lotka (distribuição da produtividade de autores – a frequência de pesquisadores que publicam n documentos é aproximadamente $1/n^2$ dos que produzem somente um documento), a Lei de Bradford (distribuição da produtividade por periódico – uma pequena quantidade de periódico publicam a maioria dos artigos de determinado tema, formando um *core* da produção relacionada a essa temática) e a Lei de Zipf (contagem de palavras em documentos – um pequeno número de palavras é muito frequente e um alto número de palavras é pouco frequente). Além disso, cabe salientar e destacar a Lei de Elitismo de Price (maiores produtores de documentos – em um conjunto de n autores, a \sqrt{n} corresponde aos autores mais produtivos deste conjunto e são responsáveis por aproximadamente 50% de toda produção).

Já os indicadores de citação, remetem ao impacto e são obtidos a partir da quantificação de citações recebidas⁸ por determinada unidade de análise (livro, artigos, entre outros). Contudo, o discernimento entre a quantidade de citações recebidas e sua qualidade deve ser minuciosa, uma vez que citação e qualidade da produção não são necessariamente sinônimos.

⁸ Citação recebida pela unidade de análise se difere da quantificação e identificação das referências citadas pelo item analisado.

Em relação ao terceiro item supracitado, os indicadores de ligação (ou relação), é possível apontá-los como medidas (ou frequências) de ocorrência e/ou co-ocorrências “de autoria, de citações e de palavras, entre outras formas de relação observadas na produção científica. São empregados para identificar, visualizar e analisar as ligações ocorridas na geração do conhecimento” (PRADO; CASTANHA, 2020, p. 65), nos diversos níveis de agregação. Nesta classificação, encontram-se os estudos sobre coautorias, coocorrências de termos e os estudos relacionais de citação, cocitação e acoplamento bibliográfico (foco desta pesquisa), análises comumente realizadas no campo da Ciência da Informação.

Os indicadores relacionais de citação têm ganhado cada vez mais espaço no âmbito da Bibliometria e da Cientometria pois se apresentam como meios de avaliação e análise do comportamento de diferentes domínios científicos.

A lista de referências de documentos científicos pode, nesse contexto, ser analisada como reflexo de uma comunidade científica discursiva, visto que apresenta a interlocução não só entre os autores citantes e citados, mas também o reconhecimento da proximidade teórica e/ou metodológica entre os citados por esta comunidade, apontando assim, possíveis características do domínio científico analisado (GRÁCIO, 2016).

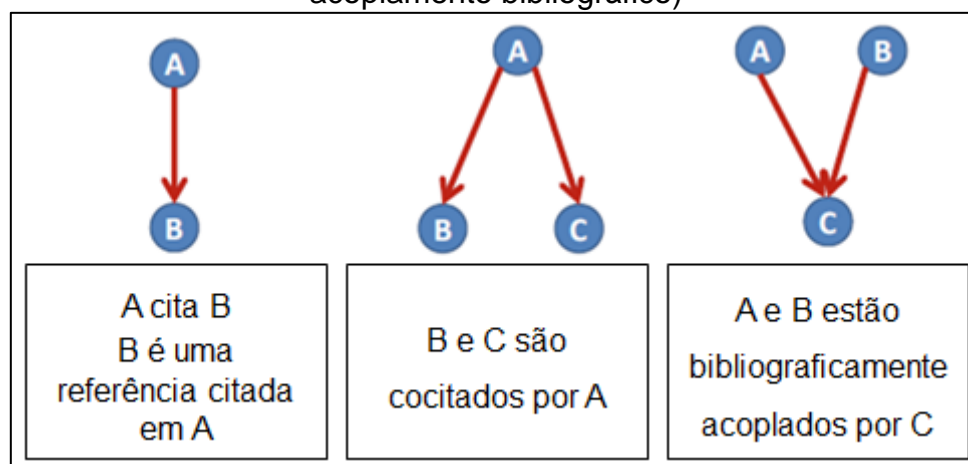
Ainda, segundo Grácio (2016), para se avaliar a interlocução entre pesquisadores, os estudos relacionais de citação ganham maior importância dentro deste tipo de análise por contribuírem para a visualização do processo comunicativo e interativo entre eles, bem como da estrutura do específico domínio em estudo.

A perspectiva relacional da análise das referências aponta um caminho para evidenciar a proximidade entre documentos, autores, instituições, periódicos e/ou países. As análises de citação têm se mostrado uma alternativa para compreender a disseminação e a apropriação de novos conhecimentos pelos pesquisadores e dessa maneira, a interação das citações entre as diversas unidades de análise é perceptível no domínio ao qual está inserido.

É possível subdividir as análises de citação em dois grupos, segundo Grácio (2020): i) uni-variados: cada sujeito do universo de estudo é investigado individualmente, segundo uma característica; ii) relacionais: buscam identificar e visualizar a semelhança ou dessemelhança entre os atores. Exemplificam-se as

análises uni-variadas com os indicadores de impacto (indicadores de citação), como o índice H e o Fator de Impacto, aplicados a pesquisadores e a periódicos, respectivamente. Já os relacionais de citação têm-se as análises de citação, cocitação e acoplamento bibliográfico, como apresenta a Figura 2.

Figura 2 - Perspectivas relacionais de citação (citação, cocitação e acoplamento bibliográfico)



Fonte: Adaptado de Moed (2017- tradução própria)

As análises relacionais de citação apresentam-se como método mais comumente usados para identificar a estrutura de conhecimento e a evolução dinâmica de uma especialidade (HOU; YANG; CHEN, 2018). A citação entre duas unidades de análise (documentos, autores, periódicos, instituições etc.) explicita a relação entre citante (aquele que promove a citação) e citado (aquele que recebe a citação; aquele que é referenciado na publicação do citante). O ato de citar estabelece um vínculo inicialmente assimétrico entre citante-citado, pois esta relação não necessariamente constitui um vínculo bidirecional entre as partes.

Matematicamente, as análises relacionais de citação podem ser descritas como operações de matrizes. Assim, uma matriz de citação (C) de dimensão $n \times m$, em que n corresponde à quantidade de citantes e m à quantidade de citados, é denotada por $C_{n \times m}$. Quando a matriz é composta por números naturais, os números presentes nas caselas da matriz representam as intensidades das relações de citação existentes entre os n elementos citantes e os m elementos citados, ou seja, para $1 < i < n$ e $1 < j < m$, cada valor f_{ij} presente na matriz representa a frequência (quantidade) de citação de um item

citante i para outro item citado j . Estas matrizes são denominadas matrizes valoradas. Caso a matriz possua somente valores unitários (não valorada ou matriz booleana), f_{ij} assume valores iguais a 0 ou 1. Matrizes binárias (somente com valores 0 ou 1) são denominadas booleanas (ou não valoradas) e matrizes com valores variando nos números naturais são denominadas matrizes valoradas, de dimensão $n \times m$.

Segundo Glänzel (2003), a relação entre unidades citantes e referências citadas de um conjunto de publicações a ser analisada pode ser representada por uma matriz booleana de dimensão $n \times m$ (citantes-citados). Esta representação é útil para descrever somente a presença ou ausência de ligação entre dois elementos, isto é, se $f_{ij} = 0$, o elemento i não cita o elemento j , e se $f_{ij} = 1$, i cita j . Desta forma, em uma análise de citação, a matriz de citação explicita a relação, não necessariamente simétrica, entre citantes e citados.

Do ponto de vista matemático, a partir da multiplicação da matriz de citação booleana ($C_{n \times m}$) e sua transposta C^T ($C_{m \times n}$), concebem-se as matrizes de acoplamento bibliográfico e de cocitação, que descrevem a intensidade de itens em comum entre elementos do mesmo tipo. A multiplicação $C \times C^T$ gera a matriz quadrada e simétrica $C_{n \times n}$ que descreve a quantidade de autores citados em comum da relação citante-citante em que é baseado o acoplamento bibliográfico (AB). A multiplicação $C^T \times C$ gera a matriz quadrada e simétrica $C_{m \times m}$ que descreve a quantidade de autores que citam conjuntamente os citados componentes da relação citado-citado na qual é baseada a cocitação (CC). Ambas as multiplicações geram matrizes quadradas e simétricas à diagonal, isto é, os elementos da parte triangular superior têm os mesmos valores dos elementos da parte triangular inferior.

Assim, a partir da matriz de citação booleana, na matriz $C_{n \times n}$ correspondente às relações de acoplamento bibliográfico, f_{ij} denota a quantidade de itens (autores citados) que acoplam dois indivíduos. Por outro lado, na matriz $C_{m \times m}$, correspondente às relações de cocitação, f_{ij} , denota em quantos documentos citantes (ou listas de referências de autores) dois indivíduos são citados concomitantemente. Se a matriz $C_{n \times n}$ é oriunda da multiplicação de uma matriz de citação valorada com sua transposta, cada valor f_{ij} presente na matriz (correspondendo ao citante i e citado j) descreve a

intensidade de acoplamento⁹ entre dois elementos. Raciocínio análogo ocorre com a matriz de cocitação gerada a partir de uma matriz valorada de citação.

Destaca-se que os valores presentes na diagonal principal dessas matrizes podem ser considerados uma representação da ligação reflexiva de um elemento com ele mesmo. Em análises de acoplamento, a diagonal pode ser preenchida com a quantidade de citações recebidas por cada documento ou autor citante ou a média de citações recebidas pelos componentes do campo analisado ou com a indicação de valores ausentes (ou perdidos ou *missing*). No caso da matriz de cocitação, obtida a partir do produto entre $C^T \times C$, com C unitária, a diagonal principal pode indicar em quantos documentos (ou listas de referências) em que o autor (ou documento) em questão está presente ou ser adotado o “valor” *missing*.

A análise de cocitação foi apresentada inicialmente por Henry Small (1973) como alternativa ao acoplamento bibliográfico proposto por Kessler (1962; 1963). A análise de cocitação pode ser entendida como a frequência com que dois documentos são citados concomitantemente, ou seja, citados de maneira conjunta em um mesmo rol de referências citadas. Ao contabilizar e identificar a frequência de cocitação entre documentos, é possível descrever a estrutura de um determinado domínio científico analisado por meio de suas principais obras e como as mesmas se conectam entre si a partir da perspectiva de reconhecimento da comunidade ativa nesse domínio.

A análise de cocitação foi estendida à ligação entre autores citados como alternativa à cocitação entre documentos a partir dos estudos de White (1981) e White e Griffith (1981). Analisar a cocitação entre autores explicita a ligação entre os principais influentes do domínio observado e com que frequência estes são citados de maneira conjunta. A cocitação frequente a um grupo de pesquisadores é um indicativo da sua influência teórico-metodológica no domínio analisado.

Fundamentado por Kessler (1962; 1963), o acoplamento bibliográfico ganhou notoriedade no campo da Ciência da Informação no âmbito da Bibliometria pelo seu ineditismo entre os estudos da época. Kessler propôs a comparação entre documentos por meio de suas listas de referências a fim de

⁹ Esta operação é similar a *full counting* em Perianes-Rodriguez, Waltman e Van Eck (2016).

identificar os documentos citados em comum entre eles. A quantidade de documentos citados em comum é descrita como frequência (intensidade; força) de acoplamento bibliográfico. Dessa maneira, quanto maior o número de documentos citados em comum por dois documentos, maior a proximidade entre eles.

O método de AB visa agrupar documentos por meio de referências citadas em comum por estes documentos (unidades de acoplamento) de modo que, quanto mais unidades de acoplamento em comum entre dois documentos, maior a proximidade (intensidade de acoplamento) entre eles. Kessler (1963) sugere dois critérios de agrupamento G_A e G_B , em que o primeiro analisa quantos documentos se acoplam a um documento específico (P_0) e o segundo prevê identificar um grupo de documentos em que todos se acoplam entre si.

Dado a simplicidade de sua formulação, torna-se fácil sua aplicabilidade. Do ponto de vista matemático, a comparação entre dois conjuntos de referências significa estabelecer (e contabilizar) a intersecção entre estes dois conjuntos.

Pouco mais de quatro décadas após o estudo precursor de Kessler (1963), Zhao e Strotmann (2008) fundamentam uma nova ótica do AB, intitulado acoplamento bibliográfico entre autores (ABA), em seu trabalho *Evolution of research activities and intellectual influences in information science 1996–2005: introducing author bibliographic-coupling analysis*.

Zhao e Strotmann (2008) fundamentam o ABA de modo que a unidade de análise são os autores, ou seja, apresentam uma alternativa às ideias de Kessler pelas quais as unidades a serem acopladas são documentos. Contudo, Zhao e Strotmann (2008) apontam dois caminhos para que se estabeleça o acoplamento entre autores: i) pela quantidade de documentos citados em comum por dois pesquisadores; ii) considera-se toda obra do autor citado como única, isto é, verifica-se quais autores e não documentos foram citados, e a quantidade de autores citados em comum por dois pesquisadores, é dita frequência de acoplamento bibliográfico.

Por mais que a adaptação do AB para autores pareça simples, uma das questões resolvidas pela nova formulação foi a seguinte: a frequência de acoplamento entre dois documentos, como proposto por Kessler, é fixa e perpetuada, uma vez que publicados, suas referências são imutáveis, logo esta frequência é a mesma independentemente do momento em que foi calculada.

Nesse contexto, o ABA resolve este problema, pois a frequência de acoplamento entre dois autores continua a evoluir se pelo menos um deles continuar a publicar.

A força do ABA está relacionada ao número de documentos ou autores citados em comum por dois autores. Analogamente às ideias de Lucas e Garcia-Zorita (2014) aplicadas ao AB entre documentos, é possível indicar que o tempo é determinante e faz com que esta força possa variar sensivelmente com o passar dos anos, visto que novos autores podem adentrar no rol de obras citadas pelos autores analisados e assim, a quantidade de unidades de análise citada em comum entre dois autores pode estar em constante aumento configurando uma medida dinâmica e não necessariamente estática. Alinhado a isto, Thijs, Zhang e Glänzel (2015) apontam que a ligação entre os documentos, uma vez estabelecida, permanecerá constante ao longo do tempo.

Tanto o acoplamento bibliográfico entre documentos proposto inicialmente por Kesller, quanto o ABA, estabelecem uma relação entre as unidades citantes a fim de explicitar as referências em comum entre eles, seja por meio de documentos citados em comum, seja ao considerar a toda obra do autor citada como única, e obedecem aos preceitos matemáticos das operações matriciais das análises relacionais de citação descritos anteriormente.

Zhao e Strotmann (2008) apontam a análise de ABA como um método para mapear autores ativos em um campo de pesquisa para que seja obtida uma imagem mais realista do estado atual de suas atividades de pesquisa. Além disso, os autores apontam a análise de cocitações de autores (ACA) para mapear os pesquisadores que influenciaram esses autores ativos, a fim de mostrar a estrutura das influências intelectuais externas na pesquisa. Afirmam ainda que a análise de cocitação e o acoplamento não são mutuamente exclusivos, e mais, ao combiná-las é possível obter uma visão completa da estrutura intelectual e ter uma visão da trajetória evolutiva do domínio analisado.

Ao discorrerem sobre o cálculo da frequência do ABA, os idealizadores indicam que: i) se a obra do *autor A* contiver X artigos, o conjunto de referência ponderada do *autor A* incluirá todos os diferentes itens citados nas listas de referência desses X artigos; ii) se um item aparecer em N , das X listas de referência do autor A , ele aparecerá no conjunto de referência ponderada do *autor A* com peso N ; iii) se um mesmo item (autor citado em comum) aparecer

no conjunto de referência do *autor B* com um peso M , então este item tem peso de ponderação entre o valor mínimo entre N e M , como presente no Quadro 1.

Quadro 1 - Acoplamento bibliográfico de autores segundo os valores mínimos de recitação em documentos distintos.

Pesquisadores Acoplados	Autor em comum (1)	Autor em comum (2)	Autor em Comum (3)	...	Autor em comum (n)
Pesquisador citante A	M_1	M_2	M_3	...	M_n
Pesquisador citante B	N_1	N_2	N_3	...	N_n
Valor Mínimo de recitações	$Min (M_1, N_1)$	$Min (M_2, N_2)$	$Min (M_3, N_3)$...	$Min (M_n, N_n)$
Força de Acoplamento: $\sum_{n=1}^n \min\{M_n; N_n\}$	$= Min(M_1, N_1) + Min(M_2, N_2) + Min(M_3, N_3) + \dots + Min(M_n, N_n)$				

Fonte: Adaptado de Castanha (2019).

Esta perspectiva não utiliza os padrões de recitação diacrônica como critério para selecionar as unidades de acoplamento, pois permite que uma mesma unidade de acoplamento responsável por acoplar dois autores esteja presente somente em um documento. Exemplifica-se: seja Y , um determinado autor citado 5 vezes em 3 documentos na obra do pesquisador A e este mesmo autor (Y) citado 5 vezes em somente 1 documento da obra do pesquisador B, o peso de recitação atribuído a esta unidade de acoplamento (autor acoplador) será o valor mínimo entre 3 e 1 ($\min\{3; 1\} = 1$). O autor acoplador Y está presente em 3 obras do pesquisador A e em apenas uma obra do pesquisador B.

Ainda com relação ao cálculo de acoplamento, Grácio (2020, p. 135) elucida que: “a frequência de AB entre dois pesquisadores A e B é definida como a soma dos pesos (número de artigos em que foi citado) mínimos dos autores com recitação, simultaneamente, nas obras dos pesquisadores A e B”.

Em complemento, Zhao e Strotmann apontam as ideias de White (2001) a respeito das análises das referências aplicadas ao cálculo das frequências de ABA, de modo que os padrões de recitação (o fato de um autor citar recorrentemente os mesmos autores ou documentos ao longo dos anos), podem ser úteis para definição da identidade de citação de um autor analisado.

O conceito de recitação diacrônica de White (2001), enquanto reincidência de autores citados em diferentes obras de um determinado autor, unido ao ABA,

torna-se terreno fértil para a construção de redes de conhecimento e influências intelectuais em domínios científicos.

Essa junção de conceitos permite gerar redes de acoplamento bibliográfico, em que os atores são os pesquisadores analisados, as arestas determinam quais autores estão acoplados, com a espessura das arestas proporcional à frequência de ABA segundo os padrões de recitação nas obras dos autores analisados.

Assim, é possível descrever tal situação como análise de acoplamento bibliográfico entre autores segundo padrões de recitação diacrônica de autores, ao considerar toda obra do autor citado como única.

Ao compreender o ABA como uma análise de proximidade teórico-metodológica entre dois autores e que, quanto maior o número de referências em comum, maior esta proximidade e conexão entre eles (ZHAO; STROTMANN, 2008), pouco se discute a respeito do valor gerado por esta análise a partir da frequência de acoplamento. Nesse sentido, o ABA pode representar uma intensidade maior ou menor do que de fato ela é ao observar apenas o quantitativo de referências em comum, sem relativizá-la e sem avaliar se estas referências são de fato representativas da identidade científica dos autores analisados.

Thijs, Zhang e Glänzel (2015) destacam que a vantagem do acoplamento bibliográfico é que não há demora para o cálculo do vínculo entre unidades de análises, pois todos os dados necessários estão presentes no momento da publicação ou indexação na base de dados. Com isso, desvendar formas de relativizar os cálculos do ABA perante toda situação analisada propicia meios de comparar diferentes unidades de análise inseridas em diferentes contextos.

4 ACOPLAMENTO BIBLIOGRÁFICO ENTRE AUTORES E ÍNDICES DE NORMALIZAÇÃO PARA INDICADORES RELACIONAIS DE CITAÇÃO

A fim de propor um novo indicador relacional, baseado no acoplamento bibliográfico, deve-se primeiramente defini-lo do ponto de vista matemático e sua funcionalidade enquanto indicador relacional que explicita uma conexão entre referências de um dado conjunto de pesquisadores a ser analisado. São descritos os conceitos matemáticos relacionados ao AB, ao ABA e suas

normalizações (relativizações) por meio do ao Índice de Jaccard e do Cosseno de Salton.

O acoplamento bibliográfico proposto por Kessler (1962; 1963), que visa medir a ligação entre dois artigos com base no número de referências em comum citadas por dois documentos (medida estática) e o ABA (medida dinâmica) que avalia a ligação entre autores citantes, podem fornecer uma visão da estrutura, da característica e do desenvolvimento das redes de conhecimento científico do domínio analisado (GRÁCIO, 2016). Por conseguinte, matematicamente, o Acoplamento Bibliográfico entre dois documentos pode ser descrito na Equação 1 como:

$$AB(D1, D2) = ref(D1) \cap ref(D2) \quad [1]$$

Na equação, $ref(D1)$ e $ref(D2)$ são as respectivas listas de referências dos documentos $D1$ e $D2$ a serem acoplados. A intersecção entre $ref(D1) \cap ref(D2)$ é definida como frequência de acoplamento bibliográfico e este valor representa a cardinalidade da intersecção entre os conjuntos de referências dos documentos $D1$ e $D2$.

Já com relação ao ABA, é possível descrever esta medida, antes da ponderação das citações, como apresentado no Quadro 1, de duas maneiras: i) o ABA segundo a frequência de documentos citados em comum por dois autores A e B; ii) Se considera toda obra do autor citado como única e calcula-se a frequência de autores citados em comum por dois pesquisadores A e B.

A primeira perspectiva do ABA pode ser equacionada na Equação 2:

$$ABA(A, B) = doc(A) \cap doc(B) \quad [2]$$

Na equação, $doc(A)$ e $doc(B)$ são os conjuntos documentos citados pelos autores A e B, respectivamente. A frequência de ABA é dada pela cardinalidade desta intersecção, isto é, pela quantidade de documentos citados em comum por A e B.

A segunda perspectiva do ABA, ao considerar a obra do autor citado como única é apresentada na Equação 3:

$$ABA(A, B) = Aut(A) \cap Aut(B) \quad [3]$$

Na equação $Aut(A)$ e $Aut(B)$ são os conjuntos dos autores citados, respectivamente, por A e B. Neste caso, a frequência de ABA é dada pelo número de autores citados em comum por A e B. Em ambos os casos do ABA, a frequência de acoplamento bibliográfico é passível de evolução, dependendo do momento em que for computado, caracterizando-se como medida diacrônica, diferentemente da frequência sincrônica inicialmente proposta por Kessler (1963, 1965).

Visando trazer novas perspectivas para o ABA que fundamentem e justifiquem a formulação de um novo indicador, apresenta-se o Cosseno de Salton e o Índice de Jaccard ajustados ao ABA. O primeiro utilizado para análises de cocitação por Vanz e Stumpf (2010), Grácio e Oliveira, (2015), Oliveira e Alves (2017), entre outros, e o segundo muito utilizado em estudos referentes da Ecologia e em estudos de cocitação como Leydesdorff (2008), Egghe e Leydesdorff (2009), Oliveira e Alves (2017), entre outros, e de coautoria como Grácio, Oliveira, Wolfram (2019). Especificamente com relação ao ABA, tem-se os estudos de Castanha e Grácio (2020) e de Grácio (2020).

Tanto o Índice de Jaccard quanto o Cosseno de Salton são entendidos como índices de similaridade relativa entre conjuntos, ou seja, estas medidas buscam mensurar o quanto dois conjuntos apresentam proximidades (similaridades) ou distanciamentos proporcionalmente ao universo analisado, o que faz com que assumam valores no intervalo entre zero e um: zero significando que os dois conjuntos são mutuamente excludentes e o valor um, que os dois conjuntos são idênticos.

Nesta pesquisa, propõe-se usá-los para medir proximidades teórico-metodológicas a partir das referências (autores citados) de dois pesquisadores com relação de orientação acadêmica. Além disso, ambas são medidas de cardinalidade, isto é, quantificam elementos em comum de dois conjuntos.

O Índice de Jaccard, originalmente aplicado (e utilizado até nos dias de hoje) no campo da Ecologia e da Biologia para análises de biodiversidade foi fundamentado por Jaccard em 1908 e muito utilizado para conservação e classificação de espécies. É considerado um dos índices de similaridade mais utilizados na literatura dentre tantos existentes (REAL; VARGAS, 1996).

A formulação original do Índice de Jaccard entre dois elementos A e B - $IJ(A, B)$ - pode ser descrita matematicamente na Equação 4 como:

$$IJ(A, B) = \frac{A \cap B}{A \cup B} = \frac{A \cap B}{(A) + (B) - (A \cap B)} \quad [4]$$

Na equação, A e B são conjuntos numéricos a serem comparados. Tão próximo quanto da formulação inicial do Índice de Jaccard, é possível estendê-lo ao acoplamento bibliográfico, visto que esta medida de similaridade se utiliza da intersecção entre dois conjuntos (princípio do acoplamento bibliográfico) em relação ao total de elementos dos conjuntos (união entre os conjuntos).

A força de acoplamento bibliográfico entre dois documentos dividida pelo número de itens da união de duas listas de referências foi denominada por Sen e Gan (1983) de força de acoplamento bibliográfico relativa (*relative bibliographic coupling strength*). Em sua essência, é a aplicação do próprio Índice de Jaccard no contexto da Bibliometria, mais especificamente do acoplamento bibliográfico, como apresenta a Equação 5:

$$IJ_{AB} = \frac{ref(A) \cap ref(B)}{ref(A) \cup ref(B)} = \frac{AB(A, B)}{ref(A) + ref(B) - ref(A \cap B)} \quad [5]$$

Segundo a Equação 5, IJ_{AB} é a força de acoplamento bibliográfico relativa. Destaca-se que, em paralelo ao acoplamento bibliográfico, sob o olhar da cocitação, Leydesdorff (2008), Grácio e Oliveira (2013) apresentam o Índice de Jaccard da seguinte forma (Equação 6):

$$IJ = \frac{Cocit(X, Y)}{Cit(X) + Cit(Y) - Cocit(X, Y)} \quad [6]$$

$Cocit(X, Y)$ representa a frequência em que os autores X e Y são cocitados e $Cit(X)$ e $Cit(Y)$ representam o total de citações recebidas pelos autores X e Y, respectivamente.

Ao retomar ao conceito da formulação do acoplamento bibliográfico, apresenta-se na Equação 7, a reformulação do Índice de Jaccard segundo as premissas do acoplamento bibliográfico entre autores, semelhante à Equação 5.

$$IJ_{ABA} = \frac{Aut(A) \cap Aut(B)}{Aut(A) \cup Aut(B)} = \frac{ABA(A, B)}{Aut(A) + Aut(B) - Aut(A \cap B)} \quad [7]$$

Na equação, $Aut(A)$ e $Aut(B)$ representam os conjuntos de autores citados pelos pesquisadores A e B, respectivamente. Salienta-se que a intersecção entre dois conjuntos de referências representa a ideia central do acoplamento bibliográfico: sobreposição de dois conjuntos de referências a fim de observar quantas delas são comuns, ou seja, a intersecção entre os conjuntos. Na Equação 7, considera-se a obra dos autores citados como única para realização do cálculo.

Esta ideia pode ser estendida também ao ABA segundo a quantidade de documentos citados em comum por dois autores A e B. Esta formulação é dada pela Equação 8:

$$IJ_{ABA} = \frac{doc(A) \cap doc(B)}{doc(A) \cup doc(B)} = \frac{ABA(A, B)}{doc(A) + doc(B) - doc(A \cap B)} \quad [8]$$

Com relação à medida de Cosseno de Salton entre dois conjuntos A e B – $CS(A, B)$ - Glänzel (2003) discorre a respeito da medida aplicada ao acoplamento bibliográfico sob o entendimento de que a literatura científica (citada) a ser analisada gera um espaço vetorial booleano que é, particularmente, definido pela relação entre n elementos citantes e m citados (artigos; autores). Sob esta definição, o autor aponta que cada linha da matriz booleana (de dimensão $n \times m$) pode ser compreendido como um vetor booleano de citação e as colunas representam os vetores de referências citadas. Assim, o acoplamento entre dois itens citantes pode ser como o produto entre dois vetores: $A \cdot B$.

Contudo, esse cálculo não consiste em uma medida relativa e, nesse sentido, Sen e Gan (1983) propuseram o ângulo de acoplamento (*coupling angle* – CA) como forma normalizada do acoplamento bibliográfico (Equação 9). Esta medida corresponde a uma adequação da formulação do Cosseno de Salton. A medida de Salton pode ser descrita como a razão entre o número de referências conjuntas e a média geométrica do número de referências dos dois artigos (GLÄNZEL, 2003) a serem acoplados. A generalização do Cosseno de Salton, segundo ambas as definições de Sen e Gan (1983) e Glänzel (2003), é apresentada na Equação 9:

$$CA(A, B) = \frac{A \cdot B}{||A|| \cdot ||B||} \sim CS(A, B) = \frac{A \cap B}{\sqrt{A \times B}} \quad [9]$$

Na Equação 9, A e B são conjuntos a serem comparados e a intensidade da sua sobreposição relativizada pelo denominador (raiz quadrada do produto de ambos). Após o conceito de força de acoplamento bibliográfico relativa (Equação 5 - IJ_{AB}) de Sen e Gen (1983), é possível, a partir da Equação 9, apresentar método de acoplamento bibliográfico de dois documentos A e B, a partir do Cosseno de Salton, como apresenta a Equação 10:

$$CS(A, B) = \frac{ref(A) \cap ref(B)}{\sqrt{ref(A) \times ref(B)}} \quad [10]$$

Na equação, $ref(A)$ e $ref(B)$ são as referências dos documentos A e B. Assim como o Índice de Jaccard, é possível descrever que o Cosseno de Salton, segundo Grácio e Oliveira (2015), aplicado à cocitação de autores é dado pela razão entre a frequência de cocitação de dois autores (A e B) e a raiz quadrada da multiplicação dos números de citações recebidas individualmente pelos autores A e B. Matematicamente, ele é expresso na Equação 11 por:

$$CS_{cocit} = \frac{Cocit(A, B)}{\sqrt{Cit(A) \times Cit(B)}} \quad [11]$$

Na equação, $Cocit(A, B)$, $Cit(A)$ e $Cit(B)$ são os mesmos valores da Equação 6. Ainda, segundo as autoras, o Cosseno de Salton tem a capacidade de mensurar a proximidade de dois autores cocitados, independentemente do tamanho da amostra e referências coletadas.

Entretanto, como o enfoque desta proposta de pesquisa é a análise dos autores citantes, pode-se reescrever esta fórmula sob a perspectiva ABA, conforme apresentado na Equação 12.

$$CS_{ABA} = \frac{ABA(A, B)}{\sqrt{Aut(A) \times Aut(B)}} = \frac{Aut(A) \cap Aut(B)}{\sqrt{Aut(A) \times Aut(B)}} \quad [12]$$

Na fórmula, $ABA(A, B)$ representa a intersecção entre os autores citados em comum entre os autores analisados, A e B, representados por $Aut(A)$ e

$Aut(B)$, respectivamente. Esta reformulação do Cosseno de Salton relativiza as referências em comum entre os autores citantes e pode ressaltar as proximidades entre eles.

Ao justificar o uso do Cosseno de Salton aplicado ao ABA, Grácio (2020, p. 135) aponta em seu trabalho que no processo de relativização dos valores dos conjuntos comparados:

Embora CS e IJ apresentem comportamentos similares, em casos de acoplamento envolvendo dois pesquisadores, um com identidade de citação composta por muitos autores com alto peso e o outro com identidade de citação formada por poucos autores com baixa recitação, considera-se que CS representa mais efetivamente o elo de intensidade de acoplamento entre eles.

Assim, a autora formula o ABA segundo os pesos das recitações de autores, como proposto por Zhao e Strotman (2008), em sua análise para identificação de domínios científicos, conforme apresenta-se na Equação 13.

$$CS_{ABA} = \frac{f_{ABA}(A, B)}{\sqrt{Pesq(A) \times Pesq(B)}} \quad [13]$$

Na fórmula, $f_{ABA}(A, B)$ representa frequência de acoplamento bibliográfico entre os pesquisadores A e B e $Pesq(A)$ e $Pesq(B)$ representam a soma total dos pesos de recitação dos autores citados pelos pesquisadores A e B, respectivamente. Nesta perspectiva, encontram-se os padrões de recitação dos valores mínimos apresentados anteriormente no Quadro 1.

Deve-se destacar que já em 1984, Vladutz e Cook (1984) propuseram uma reformulação da força de acoplamento entre dois documentos acrescentando a variável R_i como apresenta a Equação 14.

$$AB(A, B) = \frac{4}{L(A) + L(B)} \sum_i \frac{1}{freq(R_i)} \quad [14]$$

Na equação, $L(A)$ e $L(B)$ representam o tamanho das listas de referências de dois artigos A e B a serem acoplados, e, R_i denota a i -ésima referência que os artigos A e B têm em comum e $freq(R_i)$ é o número de vezes que R_i ocorre nas listas de referência do conjunto completo de documentos em estudo. Esta

medida, bem como as apresentadas em Sen e Gan (1983) influenciaram Shen et al. (2019) a proporem um novo método de cálculo do acoplamento bibliográfico entre documentos intitulado força de acoplamento refinada (*refined bibliographic coupling formula*), baseado na função inversa de documento função inversa (*frequency-inverse document frequency: TF-IDF*). Ela é apresentada na Equação 15.

$$CS(A, B) = \frac{1}{L(A) \cdot L(B)} \sum_i \log \left(\frac{N}{freq(R_i)} \right) \quad [15]$$

em que N é o número total de documentos do conjunto de dados a ser analisados e $CS(A, B)$ é a *Coupling Strength* (força de acoplamento) entre A e B.

Outras duas medidas vistas em cálculos de acoplamento bibliográfico que ganharam notoriedade são a *Full Couling* e a *Fractional Couling*, debatidas em Perianes-Rodriguez, Waltman e Van Eck (2016) e utilizadas no *VosViewer*, *software* de análises e construção de redes bibliométricas. A *Full Couling* (contagem total) remete à multiplicação entre a matriz de citação ponderada ($C_{n \times m}$) e sua transposta ($AB = C \times C^T$) em que cada elemento interno (f_{ij}) da matriz AB representa a *full couling bibliographic coupling* entre dois elementos i e j (autores e/ou documentos) citantes, podendo ser entendida como força de acoplamento. Assim, a *Full Couling* corresponde à soma dos produtos entre cada elemento das linhas (ou vetores) de citação C_i e C_j composta por elementos c_{ik} e c_{jk} de uma matriz de citação com n citantes (linhas) e m citados (colunas): $f_{ij} = \sum(c_{ik} \cdot c_{jk})$, tal que $k \in [1; m]$, conjunto das m publicações citadas.

Por outro lado, a *Fractional Couling* relativiza a citação de cada item pela quantidade de citações recebidas. Assim, para a matriz de citação $C_{n \times m}$, e n_k o total de citações recebidas (por todos os autores) pela publicação k , a *fractional couling bibliographic coupling* (fc_{ij}) entre os pesquisadores i e j é a soma dos produtos entre as citações dos pesquisadores ponderados por $n_k - 1$, como apresenta a Equação 16:

$$C = \begin{matrix} C_1 \\ \vdots \\ C_n \end{matrix} \begin{bmatrix} c_{11} & \cdots & c_{1k} & \cdots & c_{1m} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ c_{n1} & \cdots & c_{nk} & \cdots & c_{nm} \end{bmatrix}_{n \times m} \quad \rightarrow \quad fc_{ij} = \sum_{k=1}^m \frac{(c_{ik} \cdot c_{jk})}{n_k - 1} \quad [16]$$

Na equação, c_{ik} e c_{jk} representam elementos das linhas da matriz de citação (elementos citados por i e j) referentes às unidades (autores; documentos) i e j a serem acopladas, tal que $i, j \in [1; n]$. Ou seja, fc_{ij} calcula a *fractional couting bibliographic coupling* entre duas das n unidades citantes.

Embora não seja o foco deste trabalho, apresentam-se outras medidas passíveis de adaptação ao contexto do AB ou ABA: Czekanowski (1913), Kulczynski (1927), Braun-Blanquet (1932), Simpson (1943), o índice Soeresen-Dice (1945;1948), Ochiai (1957), Sorgenfrei (1959), Mountford (1962), McConaughey (1964), Johnson (1967), Van der Maarel (1969), Anderberg (1973), Coeficiente de Matching, Distância Euclidiana de Clifford & Stephenson (1975) e Consonni & Todeschini (2012) - todas elencadas por Brusco, Cradit e Steinley (2021) - além das propostas específicas ao ABA de Vladutz and Cook (1984) e Shen et al. (2019). Estas medidas são apresentadas e adaptadas ao ABA (tendo autores citados como unidade de acoplamento) no Quadro 2. Além disso, adiciona-se a proposta de Vladutz and Cook (1984), a força de acoplamento refinada de Shen et. al (2019) e as contagens *full couting bibliographic coupling* e *fractional couting bibliographic coupling* presentes em Perianes-Rodriguez, Waltman e Van Eck (2016).

Quadro 2 – Medidas de similaridade e possível adaptação ao ABA

Medidas	Formulação	Adaptação ao ABA
Czekanowski	$\frac{2c}{2c + a + b}$	$\frac{2.ABA}{2.ABA + (A) + (B)}$
Kulczynski	$\frac{c}{2} \left(\frac{1}{a} + \frac{1}{b} \right)$	$\frac{ABA}{2} \left(\frac{1}{(A)} + \frac{1}{(B)} \right)$
Braun-Blanquet	$\frac{c}{\text{máx.} \{ (c + a); (c + b) \}}$	$\frac{ABA}{\text{máx.} [ABA + (A); ABA + (B)]}$
Simpson	$\frac{c}{\text{mín.} (c + a; c + b)}$	$\frac{ABA}{\text{mín.} [ABA + (A); ABA + (B)]}$
Sorensen-Dice	$\frac{2c}{a + b}$	$\frac{2.ABA}{(A) + (B)}$

Ochiai	$\frac{c}{\sqrt{(c+a)(c+b)}}$	$\frac{ABA}{\sqrt{(ABA+(A))(ABA+(C))}}$
Sorgenfrei	$\frac{c^2}{(a+c)(b+c)}$	$\frac{ABA^2}{[(A+ABA)][(B+ABA)]}$
Mountford	$\frac{2c}{2ab+ac+bc}$	$\frac{2.ABA}{2(A)(B)+(A).ABA+ABA.(B)}$
McConaughey	$\frac{c^2-ab}{(a+c)(b+c)}$	$\frac{ABA^2-(A).(B)}{[(A+ABA)][(B+ABA)]}$
Johnson	$\frac{c}{c+a} + \frac{c}{c+b}$	$\frac{ABA}{ABA+(A)} + \frac{ABA}{ABA+(B)}$
Van der Maarel	$\frac{2c-a-b}{2c+a+b}$	$\frac{2ABA-(A)-(B)}{2ABA+(A)+(B)}$
Anderberg	$\frac{c}{(c+2a+2b)}$	$\frac{ABA}{[ABA+2(A)+2(B)]}$
Coeficiente de Matching	$\frac{c}{\text{máx.}(a,b)}$	$\frac{ABA}{\text{máx.}[(A),(B)]}$
Consonni & Todeschini	$\frac{\ln(1+c)}{\ln(1+a+b+c)}$	$\frac{\ln(1+ABA)}{\ln[1+(A)+(B)+ABA]}$
Distância Euclidiana	$D = \sqrt{\sum (x_{i,A} - x_{i,B})^2}$	$D = \sqrt{\sum (Cit_{i,A} - Cit_{i,B})^2}$
Vladutz & Cook	$\frac{4}{L(A)+L(B)} \sum_i \frac{1}{freq(R_i)}$	$\frac{4}{(A)+(B)} \sum_i \frac{1}{freq(R_i)}$
Shen et. al	$\frac{1}{L(A).L(B)} \sum_i \log\left(\frac{N}{freq(R_i)}\right)$	$\frac{1}{(A).(B)} \sum_i \log\left(\frac{N}{freq(R_i)}\right)$
Full Counting	$\sum_{k=1}^m (c_{ik} \cdot c_{jk})$	$\sum_{k=1}^m (c_{Ak} \cdot c_{Bk})$
Fractional Counting	$\sum_{k=1}^m \frac{(c_{ik} \cdot c_{jk})}{n_k - 1}$	$\sum_{k=1}^m \frac{(c_{Ak} \cdot c_{Bk})}{n_k - 1}$

Fonte: Elaboração do Autor

Em que:

- a : quantidade de elementos do conjunto A.
- b : quantidade de elementos do conjunto B.
- c : quantidade de elementos em comum entre A e B: $A \cap B$.
- ABA : frequência de acoplamento entre dois autores A e B.
- (A) : quantidade de autores citados pelo autor A.
- (B) : quantidade de autores citados pelo autor B.
- $Cit_{i,A}$: citações recebidas pelo autor i na lista do pesquisador analisado A.
- $Cit_{i,B}$: citações recebidas pelo i na lista do pesquisador analisado B.
- R_i : i -ésima autores que os autores A e B têm em comum.
- $freq(R_i)$: frequência em que os autores R_i ocorre nas listas analisadas.
- N : número total de autores do conjunto analisado.
- c_{Ak} : elementos da linha/vetor de autores citados pelo autor A na matriz de citação $C_{n \times m}$.
- c_{Bk} : elementos linha/vetor de autores citados pelo autor B na matriz de citação $C_{n \times m}$.
- n_k : total de citações recebidas (de todos os autores) pela publicação k .

A maioria das medidas apresentadas, assim com o Índice de Jaccard e o Cosseno de Salton, são oriundas de estudos relacionados às Ciências Biológicas e, também, às Ciências da Computação. A intenção do Quadro 2 é fomentar novas discussões a respeito de medidas de similaridades aplicada ao acoplamento bibliográfico. Similares ao Índice de Jaccard e ao cosseno de Salton, medidas que não foram cunhadas especificamente para o fim bibliométrico e ganharam adaptações, todas estas medidas são passíveis de adaptações, como exposto anteriormente.

Deve-se destacar no Quadro 2 que a distância euclidiana, amplamente difundida em processamentos de análises textuais baseada na frequência de palavras entre dois documentos, é adaptado ao ABA por meio da frequência de citações dos autores citados pelos pesquisadores A e B a serem acoplados. Pelo fato de ser uma medida de distância, elaborada inicialmente para distâncias entre

dois pontos no plano cartesiano, quanto mais próxima a zero esta medida, maior a proximidade entre dois elementos.

Para a aplicação dessa medida ao ABA e a obtenção de $Cit_{i,A}$ e de $Cit_{i,B}$, adotam-se os autores citados por A e B como vetores, em que cada elemento do vetor representa a frequência de citação (citações recebidas) por determinado elemento (artigo ou autor) das listas de referências analisadas. Caso haja necessidade de relativizar a distância euclidiana entre 0 e 1, é possível dividir a medida pelo tamanho destes vetores.

Ainda, no trabalho de Brusco, Cradit e Steinley (2021), os autores elencam e correlacionam 71 coeficientes de similaridade diferentes, a fim de avalia-los para aplicações no campo da psicologia. Além disso, apontam que há mais de 100 anos desenvolve-se estudos relacionados a estes coeficientes.

Estas medidas podem ser entendidas como indicadores de proximidades que levam em conta as variáveis a, b, c , descritas acima e acrescidas de uma variável d , em que d representa os elementos ausentes tanto em A quanto em B. Este elemento não se aplica ao ABA, porém também pode ser adaptável em casos em que se possa esperar a presença de autores, ou artigos, nas listas de referências de A e B.

Os coeficientes podem ser divididos em medidas de: i) coocorrência: geralmente varia de 0 a 1; ii) associação: geralmente varia de -1 a 1. Para além disto, outras adaptações podem ser conjuradas como por exemplo a reformulação, baseada no acoplamento bibliográfico, de Leydesdorff, Tekles e Bornmann (2021) do *disruption index*, inicialmente proposta para análise de patentes por Funk e Owen Smith (2017) e mais tarde adaptado a citações por Wu et al. (2019).

5 COEFICIENTE DE ACOPLAMENTO BIBLIOGRÁFICO GENEALÓGICO

Apresentados os conceitos teóricos, formulações e reformulações do ABA, Cosseno de Salton e Índice de Jaccard, é proposto o Coeficiente de Acoplamento Bibliográfico Genealógico (CABG) que, por meio de concepções do ABA aplicada a GA, configura uma nova perspectiva de acoplamento bibliográfico. Assim como as demais medidas, destina-se a avaliar a similaridade entre conjuntos, neste caso, representada pelos conjuntos de autores recitados por pesquisadores com relação de orientação acadêmica (orientador-orientando).

Para isso, inspira-se na concepção de Moed (2017, p. 98):

Uma tarefa fundamental dos *informetricians* é examinar a validade do indicador, definida como o grau em que um indicador mede o aspecto que afirma medir. Por exemplo, as contagens de citações são um indicador válido da qualidade da pesquisa? A coautoria é um indicador válido de colaboração científica? Essas são questões científico-metodológicas a serem investigadas no domínio dos estudos quantitativos-empíricos de C&T e decididas com base em critérios científicos-acadêmicos.

Formular uma métrica com potencial para avaliar a proximidade teórica-metodológica entre pesquisadores perpassa por delimitar e formular, de maneira clara e lógica, um indicador que represente as relações entre eles e consiga apontar caminhos para análise do contexto ao qual estes pesquisadores estão inseridos.

Dessa maneira, baseado no conceito de recitação diacrônica descritos por White (2021) e do ABA de Zhao e Strotman (2008), ao considerar toda obra dos autores citados como única, o índice CABG propõe relativizar o acoplamento das identidades científicas (autores recitados) em função da identidade científica do pesquisador orientador, uma vez que o conceito de genealogia acadêmica remete à herança intelectual perpassada do orientador para orientando. O CABG busca avaliar a presença da identidade científica do orientador na identidade do orientando, ou seja, o coeficiente busca mensurar o quanto da identidade científica do orientador está contida na identidade do orientando.

Em suma, o CABG centra-se no conceito do acoplamento bibliográfico entre autores, porém, o cálculo da frequência de acoplamento bibliográfico adota como unidade de acoplamento somente os autores recitados pelo pesquisador

(sua identidade científica) a fim de as proximidades teórico-metodológicas serem decorrentes somente dos autores reincidentes nas obras dos dois pesquisadores com relação de orientação acadêmica. Ao considerar a obra dos autores citados como única, acoplar dois pesquisadores orientador e orientando, por meio dos autores recitados (aqueles citados em 2 documentos ou mais), pode explicitar as influências dos autores que compõem a identidade científica do orientador perante o orientando, resgatando a ideia de herança intelectual.

Nesse contexto, o CABG é formulado segundo as etapas:

1) identifica-se a produção científica dos pesquisadores (orientador e orientando) a serem analisados¹⁰;

2) são recuperadas e extraídas as listas de referências do orientador e do orientando segundo o critério de que: nas publicações do orientador, este deve constar como primeiro autor e seu orientando não deve ser coautor da publicação. De maneira análoga, nas publicações do orientando, o mesmo critério de ser seguido, ou seja, utiliza-se apenas as publicações nas quais o orientando é primeiro autor e seu orientador não seja coautor da publicação;

3) nas listas extraídas, tanto do orientador, quanto do orientando (já egresso, isto é, com trabalho já defendido; orientação finalizada), considera-se a obra dos autores citados como única;

4) constroem-se as listas dos autores recitados (citados em pelo menos 2 documentos) de ambos os pesquisadores: orientador ($L1$) e orientando ($L2$);

5) calcula-se a cardinalidade da intersecção (L) por meio dos autores recitados em comum nas listas do orientador ($L1$) e orientando ($L2$) segundo a Equação 17:

$$L = L1 \cap L2 \quad [17]$$

6) relativiza-se a cardinalidade da intersecção pela cardinalidade da lista de referência do orientador ($L1$), obtendo o índice CABG (Equação 18):

$$CABG = \frac{[L1 \cap L2]}{L1} \quad [18]$$

¹⁰ A base de dados escolhida para extração é critério do pesquisador segundo sua metodologia.

Como pode ser observado na Equação 18, o CABG é obtido pela cardinalidade da intersecção entre $L1$ e $L2$ (quantidade de autores recitados em comum), dividida pelo número de autores recitados pelo orientado ($L1$), fazendo com que o coeficiente varie entre 0 e 1, em que: 0 significa que não há qualquer proximidade entre orientador e orientando com relação aos autores recitados em comum, e 1 significa que a lista de autores recitados pelo orientador está completamente contida (ou é igual) na lista $L2$ do orientando, implicando em uma relação de inclusão entre as listas: $L1 \subseteq L2$. Assim, um acoplamento bibliográfico perfeito (igual a 1) entre pesquisadores com relação de orientação acadêmica segundo o CABG é definido como (Equação 19):

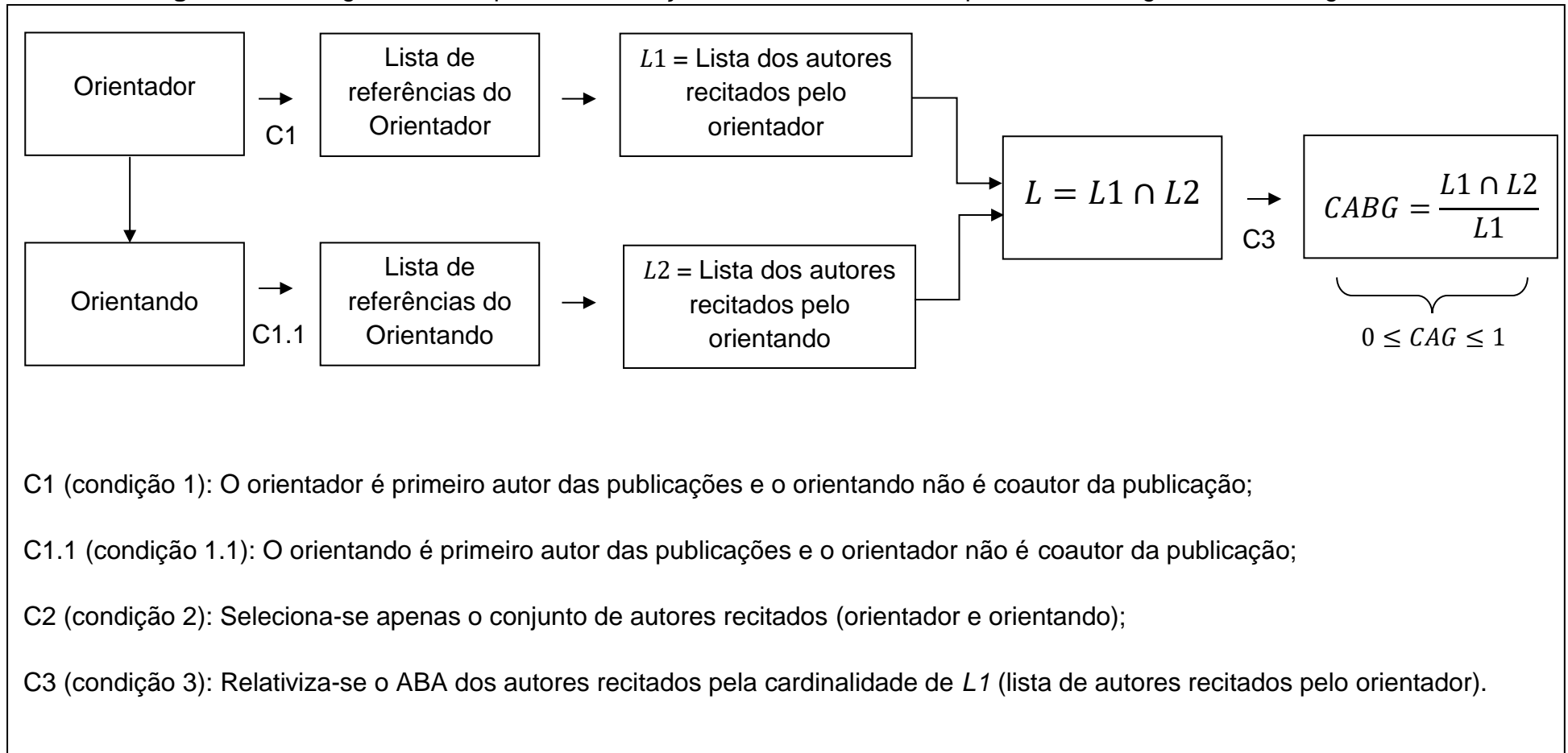
$$L1 \subseteq L2 \leftrightarrow (\forall Aut_{recit} \text{ tal que } Aut_{recit} \in L1 \rightarrow Aut_{recit} \in L2) \quad [19]$$

em que $L1$ é subconjunto de $L2$ se, e somente se, para todo autor recitado (Aut_{recit}), se um autor recitado pertence ao conjunto $L2$, então este mesmo autor recitado pertence ao conjunto $L1$. Dito de outra forma: orientador e orientando estão totalmente acoplados, se, e somente se, todo autor recitado em $L1$ está presente em $L2$. Caso isso ocorra, é possível afirmar que a identidade científica do orientador está contida na identidade científica do orientando, isto é, o orientando herdou integralmente a identidade científica de seu orientador.

Em complemento, é possível aplicar o CABG em listas de referências de autores citados somente uma vez. Todavia, nesse contexto, o conceito de identidade científica, representado pela recitação diacrónica de autores, não é necessariamente preservado, uma vez que as listas de referências a serem acopladas incluem autores citados e recitados (sincronicamente e diacronicamente). Contudo, a ideia de herança de referentes teóricos do orientador para o orientando se mantém.

Para melhor compreensão e visualização, a Figura 3 aponta as etapas para aplicação do CABG.

Figura 3 – Fluxograma de etapas de formulação do Coeficiente de Acoplamento Bibliográfico Genealógico



Fonte: Elaboração do autor

O CABG propõe $L1$ em seu denominador (relativização) como valor pré-fixado no cálculo de acoplamento. Neste ponto, o coeficiente se difere do Índice de Jaccard e do Cosseno de Salton, visto que ambas medidas também variam entre 0 e 1, mas atingem o valor máximo $IJ = 1$ ou $CS = 1$, somente se os conjuntos A e B comparados forem iguais, como apresenta a Equação 20 e 21, respectivamente:

$$IJ = \frac{A \cap B}{A \cup B} = \frac{A \cap B}{(A) + (B) - A \cap B} = 1 \leftrightarrow A \cap B = A \cup B \leftrightarrow A = B \quad [20]$$

$$CS = \frac{A \cap B}{\sqrt{A \times B}} = 1 \rightarrow A \cap B = \sqrt{A \times B} = \sqrt{A} \cdot \sqrt{B} \leftrightarrow A = B \quad [21]$$

Ao fixar o denominador $L1$, o índice CABG atinge o valor máximo (igual a 1) sem necessariamente $L1$ ser igual a $L2$, visto que (Equação 22):

$$\text{Se } L2 > L1 \text{ tal que } L1 \subseteq L2 \rightarrow CABG = 1 \quad [22]$$

Neste caso, o pesquisador orientando acopla toda lista de autores recitados de seu orientador. Assim, este acoplamento indica que toda identidade científica de seu orientador está presente em sua lista de autores recitados. Este fato justifica a relativização da intersecção entre as listas por $L1$ pela circunstância da ideia basilar da genealogia acadêmica estar intimamente relacionada à propagação de conhecimento de orientador para orientando. Em situação contrária, destaca-se que (Equação 23):

$$\text{Se } L1 > L2 \rightarrow L1 \not\subseteq L2 \rightarrow CABG < 1 \quad [23]$$

Caso a lista de autores recitados pelo orientador seja maior que de seu orientando, o valor de CABG é necessariamente menor que 1 e, então, não permitirá a existência de um acoplamento perfeito. É importante ressaltar que nem sempre a lista do orientador é maior que a lista de seu orientando, visto que pesquisadores podem parar suas pesquisas por motivos diversos como aposentadoria ou óbito, e seus orientados continuarem a produzir. Ao prosseguir suas carreiras, é provável que novos autores incorporados na fundamentação das suas pesquisas, sejam citados, participando, conseqüentemente, das suas listas de referências.

Desta forma, o CABG pode ser entendido como uma medida diacrônica e mais dinâmica que o próprio ABA de Zhao e Strotmann (2008) de modo que, caso ambos os pesquisadores continuem a produzir trabalhos, esta medida pode aumentar ou diminuir ao longo dos anos, e não somente aumentar como os autores apresentaram em resposta ao acoplamento entre documentos formulado por Kesller.

A Equação 24 evidencia que o CABG não é uma formulação igual às medidas que relativizam a intersecção entre conjuntos pelos valores máximos ou mínimos das cardinalidades dos conjuntos comparados, justamente pelo valor pré-fixado de $L1$, pois, a depender dos conjuntos a serem analisados, o valor $L1$ pode ser maior ou menor que $L2$ tornando-se valor máximo ou mínimo.

$$\frac{[L1 \cap L2]}{L1} = \frac{[L1 \cap L2]}{\text{máx.}\{L1; L2\}} \leftrightarrow L1 \geq L2 \quad \text{ou} \quad \frac{[L1 \cap L2]}{L1} = \frac{[L1 \cap L2]}{\text{mín.}\{L1; L2\}} \leftrightarrow L1 \leq L2 \quad [24]$$

Ainda, o CABG pode se igualar ao Cosseno de Salton quando as cardinalidades $\#L1$ e $\#L2$ são iguais, visto que (Equação 25).

$$\text{Se } \#L1 = \#L2 \rightarrow CS = \frac{A \cap B}{\sqrt{A \times B}} = \frac{L1 \cap L2}{\sqrt{L1 \times L2}} = \frac{L1 \cap L2}{\sqrt{L1^2}} = \frac{L1 \cap L2}{L1} = CABG \quad [25]$$

Do ponto de vista do Índice de Jaccard, o CABG iguala-se ao índice se o valor da cardinalidade da intersecção entre $L1$ e $L2$ se igualar a $\#L2$, isto é, se $\#(L1 \cap L2) = \#L2$, como demonstrado na Equação 26.

$$\text{Seja } \#(L1 \cap L2) = \#L2, \quad IJ = \frac{A \cap B}{A \cup B} = \frac{L1 \cap L2}{L1 + L2 - (L1 \cap L2)} = \frac{L1 \cap L2}{L1 + L2 - L2} = \frac{L1 \cap L2}{L1}$$

$$\text{Logo, } IJ = \frac{L1 \cap L2}{L1} = CABG; \quad \therefore \#(L1 \cap L2) = \#L2 \rightarrow IJ = CABG \quad [26]$$

É importante destacar que o CABG pode ser entendido com uma medida essencialmente maior ou igual ao Índice de Jaccard ($CABG \geq IJ$), mas nunca menor, pois, para que o índice seja maior que o CABG, seu denominador deverá ser menor que o denominador de CABG, como se demonstra na Equação 27.

$$\text{Se, } CABG < IJ \rightarrow \frac{L1 \cap L2}{L1} < \frac{L1 \cap L2}{L1 + L2 - L1 \cap L2} \leftrightarrow L1 > L1 + L2 - (L1 \cap L2)$$

$$\text{Logo, } L1 > L1 + L2 - (L1 \cap L2) \rightarrow (L1 \cap L2) > L2$$

$$\text{Contudo, } (L1 \cap L2) \leq L2, \text{ tal que, } \nexists (L1 \cap L2) > L2, \therefore CABG \geq IJ \quad [27]$$

A situação em que $L1 \cap L2 > L2$ não existe, pois, $L1 \cap L2$ é no máximo igual a $L2$, mas jamais maior. A intersecção entre dois conjuntos nunca é maior que um dos dois conjuntos comparados, sempre será menor ou igual.

Em relação ao Cosseno de Salton, o CABG atinge valores maiores que esta medida quando a cardinalidade da lista de autores recitados pelo orientando ($L2$) é maior que a cardinalidade da lista de seu orientador ($L1$) como apresenta a Equação 28:

$$CABG > CS \rightarrow \frac{L1 \cap L2}{L1} > \frac{L1 \cap L2}{\sqrt{L1 \times L2}} \leftrightarrow \sqrt{L1 \times L2} > L1 \rightarrow L1 \times L2 > L1^2 \rightarrow L2 > L1 \quad [28]$$

Explica-se, na Equação 28, que para listas não nulas, o CABG será maior que o Cosseno de Salton se, e somente se, o denominador do Cosseno de Salton for maior que o denominador do CABG. Caso contrário, tem-se $L1 > L2$.

Além do mais, alerta-se que a mesma cardinalidade não é sinônimo de mesmos elementos, visto que se $L1 = L2$ e $\#L1 = \#L2$ (mesmos elementos e mesma cardinalidade), as três medidas se igualam: $CS = IJ = CABG$.

Os valores $L1$ e $L2$ representam as listas de autores recitados por dois pesquisadores, ou seja, duas distribuições de frequência (quantidade de vezes em que um autor é recitado ao longo dos documentos recuperados) e o valor da intersecção $L = L1 \cap L2$, a frequência de acoplamento entre dois pesquisadores. A comparação entre os valores de recitação de cada autor acoplador, entre dois pesquisadores A e B a serem acoplados, pode ser representada por dois vetores (colunas) (Pesquisadores A e B) e L linhas ($L1 \cap L2$ – autores recitados em comum), com as respectivas frequências de recitação de autores. Esta lista, contendo em suas linhas os valores das respectivas frequências de recitação de autores pelos dois pesquisadores A e B, é passível de aplicação do teste de Wilconxon para distribuições pareadas em caso de necessidade de vincular um teste de hipóteses (com nível de significância α) à situação. O teste indicará se as distribuições de frequências das recitações são ranqueadas de maneira similar, podendo ser estatisticamente iguais, ou aferindo diferenças significativas entre os valores das recitações de cada autor.

É importante apontar que o critério de seleção das obras a serem analisadas é replicado de Castanha e Grácio (2020) e alinha-se a uma releitura de Freitas e Hilário (2020, p. 88) em que as autoras apontam, ao discutir a coautoria de documentos e formas de colaboração científica:

consideramos que os referenciais bibliográficos podem se constituir como recurso para análise de interações e colaborações entre pesquisadores. Acreditamos que, nas Ciências Humanas, esse tipo de análise pode ser relevante para medir as colaborações, em especial, oriundas da formação de pesquisadores e de suas orientações científicas, já que há uma tendência de adoção dos mesmos referenciais teóricos compartilhados nessas situações. [...] Evidenciamos que a indicação de autor correspondente é geralmente destinada aos principais autores (primeiro autor) ou autores mais experientes e estabelecidos na ciência, e que, nas coautorias heterogêneas que podem decorrer de relações de genealogia acadêmica (orientador - orientando), os últimos autores listados tendem a ser os autores mais experientes. A função de último autor listado é retratada como função de orientador e supervisor, tecendo contribuições qualitativas e validação do conteúdo da pesquisa proporcionada por sua expertise e experiência.

Justamente pelo fato de pesquisadores com relação de genealogia acadêmica poderem adotar referências semelhantes e orientadores muitas vezes supervisionarem trabalhos e fornecerem contribuições substanciais, documentos redigidos sem a presença do orientador ou do orientando podem ser compreendidos como trabalhos em que o pesquisador (orientador ou orientando) em questão não tem influência direta da outra parte no manuscrito.

Desse modo, os documentos redigidos em que orientador ou orientando são primeiros autores e a outra parte não consta na lista de coautores, apontam um caminho em que a atividade científica em comum não está presente na obra aumentando a representatividade do acoplamento a ser realizado, pois se há recitação de autores em comum, mesmo com os pesquisadores não atuando de maneira conjunta na redação dos documentos, é possível inferir que há proximidade de identidades de citação entre pesquisadores com relação de genealogia acadêmica. Ainda, é sabido que documentos em coautoria preveem que os seus coautores estão obrigatoriamente acoplados pelas referências citadas nestes documentos.

Retomando a inspiração de Moed (2017) supracitada, o índice CABG visa avaliar o acoplamento das identidades científicas entre dois pesquisadores com relação de orientação acadêmica, a fim de contribuir para a visualização da

estrutura intelectual, das proximidades teórico-metodológicas e da herança intelectual entre eles, evidenciadas pelos padrões de citação diacrônica descritos por White (2001) e consolidado em Zhao e Strotmann (2008) e Grácio (2020). Deste modo, busca-se validar o CABG enquanto indicador viável para mensuração de heranças de identidades científicas perpassadas do orientador para o orientando, podendo ser replicado para diversos campos ou domínios científicos.

6 METODOLOGIA

Esta tese é caracterizada pela sua fundamentação bibliométrica, cientométrica e da análise de domínio. Nota-se que a Cientometria se preocupa com a dinâmica da ciência, como atividade social, tendo como objetos de análise a produção, a circulação e o consumo da produção científica (SANTOS; KOBASHI, 2009), além de possuir fundamentos bibliométricos em seu meio para que esta dinâmica possa ser analisada de forma consistente a partir do seu principal produto: publicações científicas.

Ainda, retoma-se a classificação de Glänzel (2003) para reiterar que esta tese está inserida nos grupos G1 e G2 propostos pelo autor, e aqui se desenvolve esta inserção: em G1 pela formulação e avaliação do CABG perante outros indicadores de similaridade baseados no padrões de citação e recitação de autores, e em G2 ao aplicar estes índices de similaridade ao domínio científico dos bolsistas PQ Sênior do campo da Matemática no Brasil vinculados ao IMPA e seus respectivos orientandos egressos de doutorado.

É possível afirmar que um domínio científico é descrito como comunidades discursivas que se assemelham em diversos aspectos, como linguagem, teoria, conhecimento (HJØRLAND; ALBRECHTSEN, 1995) ou ainda uma comunidade cujos membros que possuem afinidades intelectuais, sociais e acadêmicas. Destaca-se ainda a genealogia acadêmica (enquanto relação entre orientador e orientando) como uma configuração de domínio específica como apresentado por Castanha e Grácio (2015).

Dessa maneira, analisar um determinado domínio com proximidades sociais e acadêmicas, representado por orientadores e orientandos do campo da Matemática, pode apontar formas para compreender as comunidades discursivas por meio de aspectos sócio-contextuais, ou seja, os dados a serem estudados em cada domínio devem ser observados sob um olhar crítico do contexto em que estão inseridos.

Os Estudos Métricos da Informação apresentam-se como método viável para análise deste domínio, e mais especificamente por meio das medidas baseadas no ABA. Hjørland (2002) aponta os estudos bibliométricos como um dos 11 aportes da análise de domínio. Ainda segundo o autor, também os estudos de cunho histórico e epistemológico configuram aporte à análise de domínio. Estes dois itens estão presentes nesta pesquisa visto que, o aspecto temporal é essencial ao analisar as heranças intelectuais, representadas pelas identidades de citação, perpassada ao

longo dos anos do orientador para seus orientandos. No que diz respeito ao aspecto epistemológico, a identificação dos principais influentes da floresta de genealogia acadêmica a ser analisada pode contribuir para caracterizar as correntes científicas a partir das influências intelectuais dos autores recitados sobre o domínio analisado. Evidenciar os principais influentes de um domínio é primordial para compreender possíveis desdobramentos e pesquisas futuras a serem desenvolvidas, bem como para investigar como pesquisas antecessoras se perpetuaram ao longo dos anos.

As redes ou florestas de genealogia acadêmica podem revelar origens, evoluções, proximidades e disseminação do conhecimento entre pesquisadores, testemunhando o desenvolvimento e consolidação de correntes e tendências científicas do domínio analisado. Em suma, a GA enquanto abordagem para a análise de domínio contribui para a compreensão de um domínio e de sua estrutura, quando associada aos estudos bibliométricos, históricos e epistemológicos (CASTANHA; GRÁCIO, 2015).

Os métodos bibliométricos e cientométricos tornaram-se cada vez mais essenciais, visto que se apresentam como valiosa ferramenta, tanto para o estudo das questões relacionadas a disciplinas, quanto para a produção científica de um determinado país, tema específico (BUFREM; PRATES, 2005) ou domínio em particular.

Dessa maneira, considera-se que a formulação de um indicador de perspectiva relacional (CABG) baseado no ABA e ajustado à análise da proximidade entre autores presentes em uma rede (floresta) de genealogia acadêmica, contribui não só para o refinamento e avanço metodológico da Bibliometria aos diversos âmbitos da sua aplicação e dos estudos métricos como um todo, mas também para a caracterização do impacto das relações de orientação acadêmica na dinâmica científica de um campo do conhecimento, aqui circunscrito à Matemática e representada pelos pesquisadores sênior do IMPA.

6.1 UNIVERSO E CORPUS DA PESQUISA

A tese adota como universo de pesquisa, para prova de conceito, os bolsistas PQ Sênior da área de Matemática vinculados ao Instituto de Matemática Pura e Aplicada (IMPA) que têm sua bolsa em vigência na data da coleta de dados¹¹ e seus

¹¹ Pesquisa realizada em 28 de abril de 2021.

respectivos orientandos egressos de doutorado com produções científicas que satisfaçam os critérios de autoria C.1 e C.1.1 apresentados na Figura 3: o orientador é primeiro autor de suas publicações e o orientando não é coautor da mesma publicação; o orientando é primeiro autor de suas publicações e o orientador não consta como coautor da publicação.

O CNPq divide os bolsistas de produtividade em pesquisa em 3 níveis: nível 2 (PQ-2), nível 1 (PQ-1) e nível sênior (PQ-Sr.), em que o nível 2 representa o menor nível de excelência e o nível sênior representa o maior. Destaca-se ainda que o nível 1 é subdividido em outros quatro níveis: 1-D, 1-C, 1-B e 1-A (sendo o nível 1-A o maior destes).

As bolsas são distribuídas aos pesquisadores que constituem a elite da produção científica no país, visto que devem cumprir diversos requisitos estipulados pelo CNPq, além de se dedicarem ao cumprimento dos propósitos de seus projetos de pesquisa, mantendo atividades acadêmico-científicas oficialmente vinculadas a instituições de pesquisa e ensino (BUFREM, MASCARENHAS, SOBRAL, 2017).

Esse grupo de pesquisadores atua massivamente em pesquisas relacionadas às instituições às quais estão vinculados e são responsáveis pela formação de recursos humanos ao orientarem diversas pesquisas em programas de pós-graduação no Brasil (DIAS; MOREIRA; DIAS, 2018).

Guimarães, Grácio e Matos (2014) apontam que, devido ao aumento de ciência produzida por tais pesquisadores, é de grande importância o desenvolvimento de novas metodologias que analisem suas produções científicas e gerem indicadores que auxiliem não só em suas tomadas de decisões, mas também no planejamento de suas respectivas pesquisas.

Especificamente com relação aos bolsistas PQ sêniores, esta modalidade é “destinada ao pesquisador que se destaque entre seus pares como líder e paradigma na sua área de atuação, valorizando sua produção científica e tecnológica, segundo requisitos e critérios normativos” (CAPES, 2021).

Na área de Matemática, atualmente o CNPq mantém 302 pesquisadores com bolsa em vigência, subdivididas em: 92 bolsas PQ-2; 109 bolsas PQ-1D; 28 bolsas PQ-1C; 37 bolsas PQ-1B; 20 bolsas PQ-1A; 16 bolsistas PQ-Sr. A Tabela 1 apresenta os vínculos institucionais dos 16 pesquisadores bolsistas PQ-Sr.

Tabela 1 – Bolsistas PQ sênior da área de Matemática

Bolsista	Instituição
Alcides Lins Neto*	IMPA
Aloisio Pessoa de Araújo*	IMPA
Jacob Palis Junior*	IMPA
Luiz Carlos Pacheco Rodrigues Velho*	IMPA
Paulo Roberto Grossi Sad*	IMPA
Cesar Leopoldo Camacho Manco**	FGV
Renato de Azevedo Tribuzy	UFAM
Israel Vainsencher	UFMG
Marcio Gomes Soares	UFMG
Fernando Antonio Figueiredo Cardoso da Silva	UFPE
Artur Oscar Lopes	UFRGS
Clovis Caesar Gonzaga	UFSC
Djairo Guedes de Figueiredo	UNICAMP
Flavio Ulhoa Coelho	USP
Hildebrando Munhoz Rodrigues	USP
Jorge Manuel Sotomayor Tello	USP

Fonte: Conselho Nacional de Desenvolvimento Científico e Tecnológico (2022);
*Pesquisadores do IMPA. **Tornou-se nível sênior em 2022.

Dentre os 16 pesquisadores (apresentados na Tabela 1) que compõem o universo dos pesquisadores bolsistas PQ-Sr da área de Matemática, foram analisados os 5 pesquisadores do Instituto de Matemática Pura e Aplicada (IMPA). Fundada em 1952 e localizado no Rio de Janeiro – RJ, a instituição foi a primeira unidade de pesquisa do CNPq e tem reconhecido papel desenvolvimento da Matemática brasileira, latino-americana e mundial.

Para esta tese, por meio da Plataforma Lattes de Currículos, foram identificadas todas orientações de doutorado dos 5 pesquisadores do IMPA, bem como as publicações (somente artigos) de todas as díades orientador-orientando. Com relação à produção, optou-se pela base de dados *Web of Science* (WoS) vinculada a empresa *Clarivate*. As citações do banco de dados do *Clarivate Analytics Web of Science* (WoS) são as medidas mais comumente usadas para auxiliar a reavaliação da pesquisa em

termos de impacto acadêmico em vários níveis de agregação, como instituições, periódicos ou indivíduos (WANG, GLANZEL, CHEN, 2020).

A seguir é apresentada, na Tabela 2, a descrição desta amostra. Foram identificados, via currículos Lattes, a quantidade de orientandos com teses de doutorado já concluídas (orientação finalizada), o número de documentos indexados na WoS em que os orientadores são primeiro autor, e a quantidade de acoplamentos possíveis, segundo os critérios C.1 e C.1.1 da formulação do CABG.

Tabela 2 – Delimitação dos bolsistas PQ-Sr. e acoplamentos possíveis.

PQ-Sr	Orientandos	#Docs. WoS	Primeiro Autor	Acoplamentos Possíveis*
Alcides Lins Neto	12	36	14	8
Aloisio P. Araújo	43	48	46	15
Jacob Palis Junior	42	34	24	36
Luiz C. P. R. Velho	17	127	23	2
Paulo R. G. Sad	2	14	5	1
Total	116	259	112	62

*Número de orientandos que possuem 2 ou mais documentos indexados na woS e o orientador não é coautor. **Fonte:** Elaboração do Autor

A Tabela 2 delimita a quantidade de acoplamentos a serem realizados: 62 análises de acoplamentos. Esse total de análises envolve a comparação das listas de autores citados de 62 orientandos com as listas de autores citados pelos respectivos orientadores. Ilustrativamente, o bolsista PQ-Sr Aloisio P. Araújo teve sua lista de autores recitados comparada com a lista de autores recitados por 15 dos seus orientandos egressos de doutorado, isto é, para este pesquisador PQ-Sr foram calculados os valores de CABG, Índice de Jaccard e Cosseno de Salton 15 vezes segundo cada uma das medidas.

Assim, cada uma das 62 comparações envolveu 8 cálculos de similaridade: 1) ABA utilizando-se todos os autores citados (Equação 3); 2) ABA utilizando-se apenas autores recitados em dois ou mais documentos ($L1 \cap L2$); 3) ABA adaptado ao Índice de Jaccard (Equação 7); 4) ABA adaptado ao Índice de Jaccard somente com os autores recitados em dois documentos ou mais; 5) ABA adaptado ao Cosseno de Salton (Equação 12); 6) ABA adaptado ao Cosseno de Salton somente com os autores

recitados em dois documentos ou mais; 7) Coeficiente de Acoplamento Bibliográfico Genealógico (CABG); 8) CABG sem padrões de recitação.

De maneira global, a amostra tratada na presente pesquisa analisa 5 dos 16 pesquisadores PQ-Sr da área de Matemática sendo calculados 62 acoplamentos (de um total de 116 orientandos) entre os bolsistas analisados e seus orientandos egressos de doutorado. Em complemento a amostra, tem-se a comparação entre os autores recitados nos 112 dos 259 artigos publicados pelo conjunto de 5 bolsistas PQ-Sr em que os mesmos se figuram como primeiro autor e seus respectivos orientandos a serem comparados não são coautores, e os autores recitados nos artigos produzidos pelos orientandos que atendem o mesmo critério de autoria. O conjunto dos autores citados e recitados é detalhado no Tabela 3.

Tabela 3 – Descrição da produção científica recuperada na WoS mediante os critérios estabelecidos

Pesquisadores	#Docs. Critério (Orientador)	#Docs. WoS (Orientando)	#Docs. Critério (Orientando)	Autores Citados	Autores Recitados
PQ-Sr - Alcides Lins Neto**	Total = 14	-	-	89	28
Wanderson Costa e Silva	14	4	3	59	9
Liliana Puchuri Medina	14	2	2	24	4
Arturo U. Fernandez Perez	14	17	10	77	28
Leonardo M. Camara	14	5	5	51	13
Hossein Movasati	14	38	26	218	76
Sergio Mariano Licanic	14	2	2	26	6
Luis Gustavo Mendes	14	9	7	57	13
Jose Omegar Calvo Andrade	14	11	11	122	25
PQ-Sr. Luiz CPR Velho**	Total = 23	-	-	226	65
Thales Miranda Vieira	23	15	4	131	3
Esdras Medeiros	23	8	3	85	4
PQ-Sr. Paulo Grossi Sad**	Total = 5	-	-	27	7
Gabriel Calsamiglia	5	6	6	73	14
PQ-Sr. Jacob Palis Junior**	Total = 24	-	-	134	71
Ali Tahzibi	24	28	6	49	14
Artur Oscar Lopes**	24	72	44	360	129
Aubin Arroyo	24	13	9	100	16
Bernando A. S. M. Rebolledo	24	19	4	33	11
Bladismir Leal	24	3	2	26	2
Carlos Arnoldo Morales Rojas	24	90	42	284	86
Carlos F. Borges Palmeira	24	7	3	5	2
Carlos G. T. Araújo Moreira	24	66	15	68	13
Carlos Vasquez	24	28	4	75	30
Eduardo Colli	24	11	6	76	10

Eleonora D. C. Garcia	24	23	22	216	76
Enrique Ramiro Pujals	24	65	17	121	49
Fernando J. Sanchez-Salas	24	5	3	30	9
Frederico Rodriguez-Hertz	24	45	21	180	68
Iaci Pererira Malta	24	2	2	7	1
Jaques Gheiner	24	3	3	26	16
Jorge Beloqui	24	9	6	32	11
Jorge M. Martins da Rocha	24	29	4	45	17
Leonardo Enrique Mora	24	9	5	35	8
Julio L. J. Diaz Casado	24	37	21	155	77
Luiz F. C. da Rocha	24	8	2	37	0
Maria I. T. Camacho	24	14	8	28	12
Martin Sambarino	24	22	3	55	14
Paulo Roberto Grossi Sad**	24	14	5	27	7
Pedro Mendes	24	14	6	28	5
Marcelo M. da Silva Viana	23	9	8	128/116*	70/33*
Pedro M. N. R. Dias Duarte	24	35	24	185	71
Rafael Labarca	24	27	15	247	34
Raul Mario Ures	24	29	5	66	16
Ricadro Mañe	24	43	41	116	35
Roberto Markarian	24	25	12	87	27
Roger Metzger	24	17	8	53	15
Sergio Eugeniu Plaza	24	10	8	62	21
Stefano Luzzatto	24	39	19	389	80
Vilton Pinheiro	24	17	2	38	8
Wellington Celso de Melo	24	13	5	42	19
PQ-Sr. Aloisio P. Araújo**	Total = 46	-	-	479	160
Alvaro José Riascos Villegas	46	19	6	84	3
Elvio Accinelli Gamba	46	23	23	192	43
Gyorgy Varga	46	14	8	41	4
Marco Shinobu Matsumura	46	4	4	79	14
Rodrigo Jardim Raad	46	4	4	60	21
Alvaro Sandroni	45	50	21	477/305*	153/79*
José V. Machado Vicente	45	22	4	479/81*	160/12*
Marcos Hiroyuki Tsuchida	45	15	4	479/70*	153/9*
Bruno Funchal	44	11	3	479/92*	149/5*
Luciano Irineu de Castro Filho	44	28	19	479/362*	152/88*
Wilfredo Leiva Maldonado	44	25	11	444/139*	158/27*
Susan Schommer	43	24	5	468/56*	155/4*
José Heleno Faro	42	12	4	420/134*	128/17*
Humberto Luiz Ataide Moreira	40	42	6	479/54*	128/1*
Paulo Klinger Monteiro	41	55	23	459/126*	150/46*

Fonte: Elaboração do Autor. *(Autores citados pelo orientador/autores citados pelo orientando);

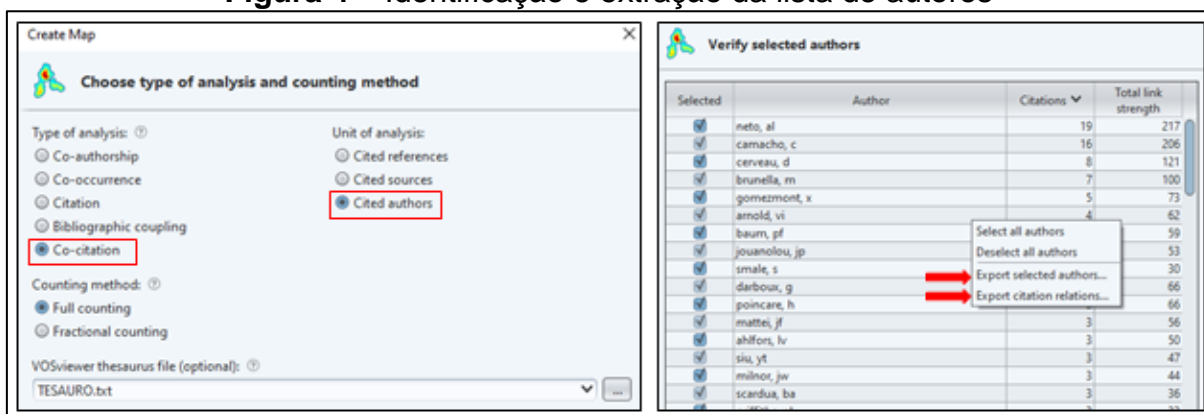
**Bolsista PQ-Sr.

A segunda coluna do Tabela 3 apresenta, para cada orientador bolsista PQ-Sr., o total de documentos indexados na WoS que atendem ao critério de autoria e a quantidade de documentos a serem acoplados com cada orientando. A terceira e quarta coluna referem-se à quantidade total de documentos identificados de cada orientando e a quantidade de documentos presentes na base de dados que atendem ao critério de autoria, respectivamente. Ou seja, serão acoplados os autores recitados nos documentos da segunda (orientador) e quarta (orientando) coluna. As duas colunas subsequentes (quinta e sexta) representam a quantidade de diferentes autores citados e recitados, respectivamente. Os valores destas colunas (quinta e sexta) podem ser entendidos como o tamanho dos conjuntos a serem comparados pelas medidas de similaridade enunciadas anteriormente.

Após identificação de toda produção a ser analisada, exportou-se¹² todo registro das publicações e extraiu-se, por meio do software *VosViewer*, utilizando a função *export selected authors*, todos autores citados por cada pesquisador. O software é uma ferramenta para construção e visualização de redes bibliométricas de citação, cocitação, acoplamento bibliográfico, coautorias e co-ocorrência baseado em diversas unidades de análise como autores, periódicos, documentos, organizações e países.

A ferramenta permite acessar o registro de referências por meio da opção de análise de cocitações de autores e esta função possibilita a identificação dos autores citados (considerando toda obra do autor citado como única) e a extração do relacionamento de citações, como apresenta a Figura 4.

Figura 4 – Identificação e extração da lista de autores



Fonte: Elaboração do autor baseado no software *VosViewer*

¹² Na WoS: Exportação→Outros formatos de arquivo→Registro completo e referências citadas→separado por tabulações (*win, utf-8*)

Ao seleccionar a exportação de autores (*export selected authors*) e as relações de citação (*export citation relations*), tem-se a identificação dos autores segundo o *Author ID* (número de identificação do autor) com seus respectivos nomes, e, a relação de citações apresenta o *document ID* (número de identificação do documento) e a frequência de citação dos autores por documento (*frequency*), como exemplificado na Figura 5.

Figura 5 – Identificação e exportação da lista de autores e das relações de citação

The figure consists of two side-by-side screenshots of text files. The left screenshot, titled '*Author ID - Blo...', shows a table with three columns: 'id', 'author', and 'citations'. The right screenshot, titled 'Citation Relations - Bloco de Notas', shows a table with three columns: 'document id', 'author id', and 'frequency'.

id	author	citations
1	"abraham, r"	1
2	"ahlfors, lv"	3
3	"arnold, vi"	4
4	"baldassarri, f"	1
5	"barth, w"	1
6	"baum, pf"	4
7	"belliart, m"	1
8	"birkhoff, gd"	1
9	"brunella, m"	7
10	"camacho, c"	16

document id	author id	frequency
1	5	1
1	9	1
1	16	1
1	22	1
1	31	1
1	64	2
1	77	1
1	81	1
2	2	2
2	10	2

Fonte: Documentos gerados a partir da exportação via software *VosViewer*

Os documentos exportados a partir da seleção de autores gera os documentos de texto presentes na Figura 5 em que, à esquerda, Figura 5-a, estão presentes a identificação dos autores e o total de citações que cada um recebeu. Exemplifica-se na segunda linha: o autor "*ahlfors, lv*" identificado como *id 2* foi citado 3 vezes no rol de documentos extraídos. À direita, na Figura 5-b, está presente a relação de citações em que se interpreta, na 9ª linha: o autor de *id 2* ("*ahlfors, lv*") foi citado no documento de *id 2*, duas vezes.

Para cada pesquisador (orientador ou orientando) geraram se estes dois documentos similares ao da Figura 5. Com intenção de relacionar as listas de identificação de autor com as relações de citação por documento, foi realizado, por meio do software Excel, o cruzamento entre estes arquivos de modo que, para cada atributo *Author ID* fosse relacionado o nome do autor citado em questão. Para isso, as listas oriundas da "exportação de autores" (*export selected authors*) e das "relações

de citação” (*export citation relations*) foram conjugadas com a função *procv* do software. A função retornou toda nomenclatura dos autores segundo o *Author ID* de cada um deles. Após este procedimento, foi aplicada a função *cont.se* para contagem das recitações diacrônicas por documentos, ou seja, em quantos documentos cada autor foi citado. Apresenta-se essa conjugação de listas na Figura 6 para exemplificação.

Figura 6 – Conjugação das listas de exportação de autores e relações de citação.

	A	B	C	D	E	F	G	H	I	J
1	Author id	author	citations		document id	author id	frequency		Autores	Recitações/Documento
2	1	afraimovich, vs	1		1	2	1		=PROCV(F2;A:B;2;0)	=CONT.SE(I:I/1/2)
3	2	aoki, n	1		1	4	1			
4	3	araujo, v	1		1	11	1			
5	4	aubin j-p	1		1	13	2			
6	5	baladi, v	2		1	14	1			
7	6	bautista, s	2		1	16	1			
8	7	beltran, j	1		1	26	1			
9	8	benedicks, m	2		1	38	2			
10	9	bonatti, pc	5		1	39	1			
11	10	bouwer, ej	2		1	46	1			
12	11	bowen, r	3		1	48	2			
13	12	cao, yl	1		1	50	1			
14	13	carrasco-olivera, d	2		2	1	1			
15	14	cordeiro, w	1		2	3	1			
16	15	drinker, p	1		2	6	1			
17	16	eisenberg, murray	1		2	7	1			
18	17	fowler, ac	1		2	9	4			

Fonte: Elaboração do autor via software Excel

Após o cruzamento das listas exportadas, a identificação dos *Author ID* de cada autor e a contagem das recitações por documento, foram construídas quatro novas colunas para as frequências de citações e recitações por documentos excluindo os valores duplicados como exemplifica-se na Figura 7.

Figura 7 – Organização completa das identificações dos autores e dos padrões de citação e recitação por documento.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Author id	author	citations		document id	author id	frequency		Autores	Recitações/Doc.		Autores	Cits/Doc.		Autores	Recitações/Doc.
2	1	afraimovich, vs	1		1	2	1		aoki, n	1		bowen, r	3		bowen, r	3
3	2	aoki, n	1		1	4	1		aubin j-p	1		guckenheimer, j	3		guckenheimer, j	3
4	3	araujo, v	1		1	11	1		bowen, r	3		katok, ab	3		katok, ab	3
5	4	aubin j-p	1		1	13	2		carrasco-olivera, d	1		lorenz, en	3		lorenz, en	3
6	5	baladi, v	2		1	14	1		cordeiro, w	1		metzger, rj	3		metzger, rj	3
7	6	bautista, s	2		1	16	1		eisenberg, murray	1		rovella, a	3		rovella, a	3
8	7	beltran, j	1		1	26	1		kelly, jp	1		bautista, s	2		bautista, s	2
9	8	benedicks, m	2		1	38	2		pilyugin, sy	1		bonatti, pc	2		bonatti, pc	2
10	9	bonatti, pc	5		1	39	1		raines, be	1		young, ls	2		young, ls	2
11	10	bouwer, ej	2		1	46	1		utz, wr	1		baladi, v	2		baladi, v	2
12	11	bowen, r	3		1	48	2		walters, p	1		kifer, y	2		kifer, y	2
13	12	cao, yl	1		1	50	1		williams, rk	1		pallis, j	2		pallis, j	2
14	13	carrasco-olivera, d	2		2	1	1		afraimovich, vs	1		ruelle, d	2		ruelle, d	2
15	14	cordeiro, w	1		2	3	1		araujo, v	1		sinai, yag	2		sinai, yag	2
16	15	drinker, p	1		2	6	1		bautista, s	2		viana, m	2		viana, m	2
17	16	eisenberg, murray	1		2	7	1		beltran, j	1		aoki, n	1			
18	17	fowler, ac	1		2	9	4		bonatti, pc	2		aubin j-p	1			
19	18	guckenheimer, j	3		2	17	1		fowler, ac	1		carrasco-olivera, d	1			
20	19	guimond, rj	1		2	20	1		hayashi, s	1		cordeiro, w	1			
21	20	hayashi, s	1		2	21	1		hirsch, mw	1		eisenberg, murray	1			
22	21	hirsch, mw	1		2	29	1		li, m	1		kelly, jp	1			

Fonte: Elaboração do autor via software Excel

As colunas com cabeçalho azul, *Cits./Doc.* e *Recitações/Doc.*, representam as citações por documento (citado em pelo menos um documento) e as recitações por documento (citado em dois documentos ou mais). Após esta organização foi possível comparar estas duas listas com as listas de seu orientador ou orientando segundo as medidas de similaridade descritas anteriormente (procedimento 3 – seção 6.2). Para cada orientando, foi construída uma planilha com todas as colunas presentes na Figura 7 acrescida de uma coluna auxiliar com as listas de autores citados e recitados por seu orientador.

Comumente em análises bibliométricas de acoplamento bibliográfico constrói-se uma matriz de acoplamento bibliográfico para identificação das conexões entre autores e para o cálculo da frequência de acoplamento. Contudo, como esta pesquisa é baseada no acoplamento entre orientador e orientando, a comparação par a par é, como descrita anteriormente, mais aconselhada, visto que uma matriz com todos os 66 pesquisadores analisados se torna algo inviável, pois nem todos os atores estão submetidos a acoplamento, somente os pesquisadores com relação de genealogia acadêmica entre si. Ou seja, busca-se o acoplamento somente entre orientador e orientando e não entre orientando e orientando.

Tendo disponível as listas de citações e recitações por documento de todos os 66 pesquisadores a serem analisados, calcularam-se todas as 8 perspectivas de acoplamento bibliográfico. Para o cálculo de acoplamento, implementou-se, em linguagem de programação R, um código para o cálculo de acoplamento entre orientador e orientando que deu origem ao web-aplicativo (*web-app*) intitulado *The Genealogical Coupler*¹³ (CASTANHA, 2022a). Este web-app é um aplicativo derivado do código fonte do também *web-app The Coupler*¹⁴ formulado pelo autor desta tese em Castanha (2022b). O *The Genealogical Coupler* (código no Apêndice 2), bem como o *The Coupler*, calculam de maneira pareada, o acoplamento bibliográfico entre unidades a serem acopladas em valores absolutos (quantidade de autores citados ou recitados em comum) e em valores normalizados, via Cosseno de Salton e Índice de Jaccard. A principal diferença entre os dois aplicativos é que o *The Genealogical Coupler* realiza comparações entre uma unidade fixa (orientador) e as demais (orientandos) e, conseqüentemente, forcene os valores de CABG para o conjunto

¹³ Disponível em: <https://rafaelcastanha.shinyapps.io/thegenealogicalcoupler>

¹⁴ Disponível em: <https://rafaelcastanha.shinyapps.io/thecoupler>

analisado. A Figura 8 ilustra os cálculos de acoplamento entre orientador e orientando realizados via *The Genealogic Coupler*.

Figura 8 – Cálculos de acoplamento via *The Genealogic Coupler*

The screenshot shows the 'The Genealogic Coupler' web application. The interface includes a sidebar on the left with options for file selection, separator, and normalization. The main area displays a table with columns for 'Orientador', 'Orientandos', 'L1', 'L2', 'Coupling', 'Saltons_Cosine', 'Jaccard_Index', and 'CABG'. The table contains 8 rows of data for the 'LINS NETO, A' network. Below the table, there are input fields for each column and a 'Download Data' button.

Orientador	Orientandos	L1	L2	Coupling	Saltons_Cosine	Jaccard_Index	CABG
LINS NETO, A	ANDRADE, J. O. C	28	25	8	0.3023716	0.17777778	0.28571429
LINS NETO, A	CAMARA, L. M	28	13	6	0.3144855	0.17142857	0.21428571
LINS NETO, A	COSTA SILVA, W	28	9	5	0.3149704	0.15625000	0.17857143
LINS NETO, A	LICANIC, S. M	28	6	5	0.3857584	0.17241379	0.17857143
LINS NETO, A	MEDINA, L. P	28	4	2	0.1889822	0.06666667	0.07142857
LINS NETO, A	MENDES, L. G	28	13	5	0.2620712	0.13888889	0.17857143
LINS NETO, A	MOVASATI, H	28	76	6	0.1300665	0.06122449	0.21428571
LINS NETO, A	PEREZ, A. U. F	28	28	8	0.2857143	0.16666667	0.28571429

Fonte: Elaborado pelo autor via *The Genealogic Coupler*

Na Figura 8, é possível identificar o orientador e os orientandos a serem acoplados nas duas primeiras colunas e nas colunas subsequentes: *L1*, *L2*, *Coupling*, *Saltons_Cosine*, *Jaccard_Index* e *CABG*, em que *L1* e *L2* são os tamanhos das listas de referências de orientador e orientando, respectivamente. *Coupling*, *Saltons_Cosine*, *Jaccard_Index* e *CABG* representam os valores do cálculo dos acoplamentos não normalizados (ABA ou $L1 \cap L2$), normalizados via Cosseno de Salton, Índice de Jaccard e *CABG*, respectivamente. Todos os cálculos de acoplamento baseados nas diferentes medidas de similaridade calculadas estão presentes na Tabela 4.

Assim, a segunda coluna da Tabela 4 - *ABA* - refere-se à quantidade de autores citados (ao menos uma vez) em comum pelo orientando e seu orientador. Este valor representa a intersecção entre todos os autores citados nas obras recuperadas do orientando analisado com relação ao seu orientador. A terceira coluna da Tabela 4 - $L1 \cap L2$ - representa a quantidade de autores recitados (citados em dois documentos ou mais) em comum pelo orientando (*L2*) e seu orientador (*L1*). Esta coluna representa o acoplamento bibliográfico entre autores enquanto intersecção entre as listas de autores recitados por orientador e orientando. Ambas as colunas não são relativizadas (normalizadas) por qualquer valor.

Adiante, a quarta e quinta coluna da Tabela 4 expressam os cálculos do Índice de Jaccard adaptado ao ABA segundo os padrões de citações totais – *Jaccard (total)* - e de recitações – *Jaccard (recit.)* – respectivamente. A medida *Jaccard (total)* alinha-se a Equação 7 em que o cálculo da união entre os conjuntos leva em conta a quantidade de autores citados (5ª coluna do Tabela 3) por orientador e orientando, e, a intersecção pelo valor *ABA* da Tabela 4.

Com relação ao *Jaccard (recit.)*, a medida também se assemelha a Equação 7, mas admite apenas os autores recitados para realização de seu cálculo (6ª coluna da Tabela 3) e utiliza-se da intersecção $L1 \cap L2$. Exemplifica-se o cálculo desta medida, utilizando o primeiro acoplamento da Tabela 4, entre o orientador bolsista PQ-Sr. Alcides Lins Neto e seu orientando Wanderson Costa e Silva, na Equação 29.

$$IJ_{ABA} = \frac{ABA(A, B)}{Aut(A) + Aut(B) - Aut(A \cap B)} \left\{ \begin{array}{l} Jaccard(total) = \frac{12}{89 + 59 - 12} = 0,0882 \\ Jaccard(recit.) = \frac{5}{28 + 9 - 5} = 0,1563 \end{array} \right. \quad [29]$$

Em sequência, a sexta e sétima coluna descrevem o Cosseno de Salton adaptado ao ABA de acordo com os padrões de citações totais – *Salton (total)* – e de recitação – *Salton (recit.)* – respectivamente. O cálculo de *Salton (total)* remete a Equação 12 em que os valores $Aut(A)$ e $Aut(B)$ representam o tamanho da lista de autores citados pelos Autores A e B a serem acoplados representados na quinta coluna do Tabela 3 enquanto a intersecção dos conjuntos é dada por *ABA*.

De maneira similar, a medida *Salton (recit.)* utiliza-se a sexta coluna do Tabela 3 para descrever os valores de $Aut(A)$ e $Aut(B)$ e a intersecção entre estes valores será $L1 \cap L2$. Para exemplificar estes cálculos apresenta-se, na Equação 30, o primeiro acoplamento presente na Tabela 4 (entre Alcides Lins Neto e Wanderson Costa e Silva).

$$CS_{ABA} = \frac{ABA(A, B)}{\sqrt{Aut(A) \times Aut(B)}} \left\{ \begin{array}{l} Salton(total) = \frac{12}{\sqrt{89 \times 59}} = 0,1656 \\ Salton(recit.) = \frac{5}{\sqrt{28 \times 9}} = 0,3150 \end{array} \right. \quad [30]$$

Na oitava coluna da Tabela 4 está presente o cálculo do CABG. O novo coeficiente é descrito na Equação 18 em que os valores $L1$ e $L2$, descritos como a

cardinalidade das listas de autores recitados por orientador e orientando, respectivamente, estão presentes na sexta coluna do Tabela 3 e a intersecção entre estes itens é $L1 \cap L2$. Analogamente ao apresentado nas Equações 29 e 30, exemplifica-se o cálculo do CABG entre Alcides Lins Neto e Wanderson Costa e Silva na Equação 31.

$$CABG = \frac{(L1 \cap L2)}{L1} = \frac{5}{28} = 0,1786 \quad [31]$$

A última coluna da Tabela 4 apresenta uma adaptação do CABG as citações totais dos autores, isto é, não leva em conta a recitação dos autores. Esta perspectiva não é foco desta pesquisa, porém, convém apresentar para que possíveis comparações possam ser realizadas.

Ademais, ambas aplicações (*The Coupler* e *The Genealogical Coupler*) fornecem quais unidades (autores citados ou recitados) foram responsáveis por acoplar (unidades de acoplamento; autores acopladores; autores recitados em comum) as unidades analisadas. Nesta pesquisa, correspondem aos responsáveis por acoplar orientador e orientandos. Este procedimento é ilustrado na Figura 9.

Figura 9 – Identificação das unidades de acoplamento via *The Genealogical Coupler*

Unidades de Análise	Unidades de Acoplamento
LINS NETO, A...COSTA SILVA, W	neto, al; camacho, c; cerveau, d; jouanolou, jp; gomezmont, x
LINS NETO, A...MEDINA, L. P	neto, al; cerveau, d
LINS NETO, A...PEREZ, A. U. F	neto, al; camacho, c; brunella, m; cerveau, d; arnold, vi; baum, pf; gomezmont, x; mattei, jf
LINS NETO, A...CAMARA, L. M	camacho, c; brunella, m; jouanolou, jp; mattei, jf; gunning, rc; seidenberg, a
LINS NETO, A...MOVASATI, H	neto, al; camacho, c; cerveau, d; arnold, vi; gomezmont, x; mattei, jf
LINS NETO, A...LICANIC, S. M	neto, al; camacho, c; brunella, m; gomezmont, x; griffiths, ph
LINS NETO, A...MENDES, L. G	neto, al; camacho, c; brunella, m; gomezmont, x; seidenberg, a
LINS NETO, A...ANDRADE, J. O. C	neto, al; camacho, c; brunella, m; cerveau, d; baum, pf; gomezmont, x; scardua, ba; griffiths, ph

Fonte: Elaborado pelo autor via *The Genealogical Coupler*

Isto posto, o procedimento presente na Figura 9 explicita todos os autores que compõem as identidades científicas dos pesquisadores a serem acoplados. Desta maneira, estes autores constituem a estrutura intelectual e possível linhagem científica

perpassada de orientador para orientando. Esclarece-se que para os cálculos em que o orientador possui documentos redigidos em coautoria com diversos orientandos, como por exemplo, a relação entre Aloísio Araújo e seus orientandos, utilizou-se a ferramenta par a par, similar ao cálculo manual, diferente da relação entre Lins Neto (exemplificado na Figura 9) e seus orientandos, em que o orientador não possui documentos redigidos em coautoria com os orientandos como primeiro autor.

6.2 PROCEDIMENTOS METODOLÓGICOS

Após as identificações dos bolsistas PQ-Sr da área da Matemática vinculados ao IMPA juntamente com seus respectivos orientandos (egressos de doutorado) e de toda a produção científica (apenas artigos) de cada bolsista e de seus orientandos na base de dados *Web Of Science* apresentadas na seção anterior, esta pesquisa se desenvolveu nas seguintes etapas ilustradas na Figura 10:

1) Recuperação da produção científica dos orientadores e orientandos: Identificados os artigos produzidos pelos orientadores e orientandos que compõem a amostra, foram recuperados todos os artigos em que o orientador é primeiro autor e seu orientando a ser acoplado não consta como coautor, e todos artigos em que o orientando é primeiro autor e seu orientador não é coautor. O processo de aplicação deste critério e de seleção do rol de artigos utilizados foi realizado de maneira manual na base de dados WoS e, após isso, exportou-se o registro completo de cada artigo (metadados e referências citadas). Como os padrões de recitação são parte fundamental da pesquisa, como critério, para compor a amostra, cada orientando teve que ter pelo menos dois artigos que atendessem ao critério de autoria anteriormente mencionado (ser primeiro autor e o orientador não ser coautor) indexados na WoS.

2) Extração dos referentes de cada documento de cada pesquisador analisado: Para cada um dos 5 orientadores e dos 62 orientandos, foram registrados em planilhas eletrônicas (Excel) todos os autores por cada um dos 66 pesquisadores. A extração da lista de referências dos artigos recuperados foi realizada utilizando o software *VosViewer*. O software permite identificar e exportar os autores citados por um determinado conjunto de documentos, além de exportar as relações de citação explicitando a distribuição de autores citados por artigo por das funções *export select authors* e *export citation relations*. Estas funções permitem detectar em quais artigos cada autor é citado e quantas vezes cada um foi citado, como apresentado nas

Figuras 5, 6 e 7. Ressalta-se que este procedimento considera toda obra do autor citado como única e as autocitações não foram excluídas, pois considera-se que este fenômeno pode ser significativo para análise, ainda mais quando se trata de relações de genealogia acadêmica. A extração foi submetida a tesauro a fim de evitar duplicidade dos nomes dos autores citados.

3) Organização e aplicação das medidas similaridades baseadas no ABA para cada par de orientação: Após a extração em planilha eletrônica dos autores citados pelos bolsistas e seus orientandos, organizaram-se as listas a serem comparadas de maneira que fosse possível identificar os padrões de citação e recitação no conjunto de artigos de cada pesquisador (orientando e orientador) participante da relação de genealogia acadêmica em análise. Assim, dado um orientador e seus orientandos, compararam-se as listas de autores citados e recitados resultante do conjunto de artigos recuperados de cada um deles e calculou-se a proximidade entre estas duas listas por meio das 8 perspectivas descritas na seção anterior. Esta etapa fornece a quantidade de autores diferentes citados e recitados em comum segundo as normalizações via Índice de Jaccard, Cosseno de Salton e CABG. Todos cálculos foram realizados via *The Genealogical Coupler* e conferidos via Excel e manualmente. A conferência se dá por redundância para constatar validade do código.

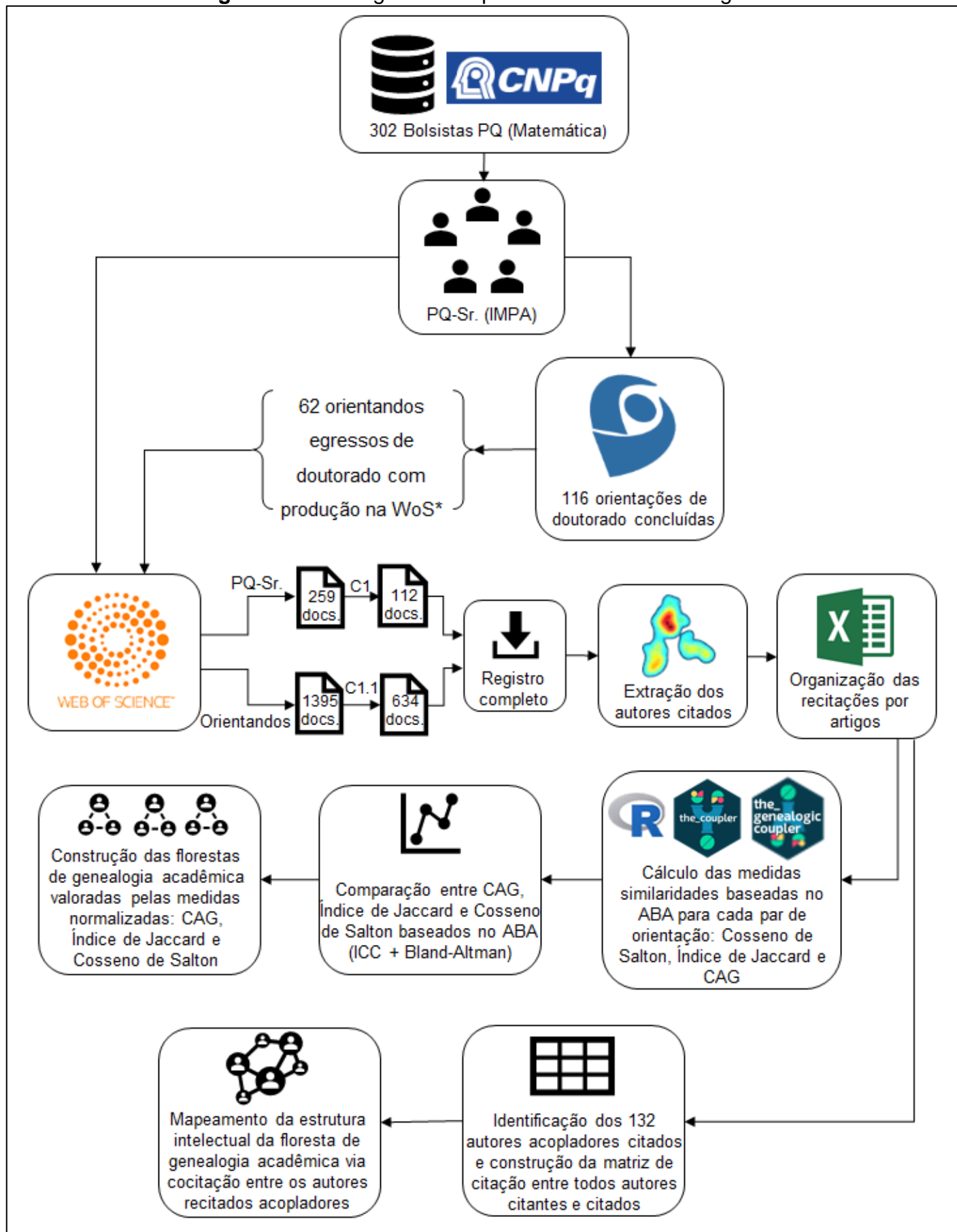
4) Comparação entre o Coeficiente de Acoplamento Bibliográfico Genealógico, Índice de Jaccard e Cosseno de Salton baseados ABA: a comparação do CABG à medidas já consagradas na literatura, como o Índice de Jaccard e o Cosseno de Salton, segundo os padrões de recitação diacrônica de autores, permite avaliar a pertinência da propositura desta tese – reivindicação do uso do CABG como um índice representativo da proximidade entre autores envolvidos em uma relação de genealogia acadêmica. Para esta comparação, calculou-se o coeficiente de correlação intraclasse (ICC) entre as três medidas, a fim de observar quão “calibradas” entre si estão as medidas, principalmente em relação ao CABG (Tabela 5). O uso da correlação intraclasse é inspirado na comparação entre métodos de avaliação, visto que o ICC é comumente utilizado para analisar o quão alinhados estão dois ou mais métodos. Esta medida remete à confiabilidade entre métodos ou avaliadores submetidos a comparação e tem sua interpretação semelhante ao coeficiente de correlação de Pearson ou Spearman, variando no intervalo $[-1; 1]$, em que quanto maior a proximidade a 1, maior a proximidade entre os métodos avaliativos. No âmbito da Ciência da Informação, Denisczwick e Kern (2013) aplicam esta medida para

analisar a confiabilidade de pareceres de mestrado no campo de Ciência da Informação. Desta maneira, uma correlação intraclasse entre o CABG, o Cosseno de Salton e o Índice de Jaccard indica o quão ajustados estão as medidas perante o novo indicador. Tal fato pode apontar os primeiros indícios para que o CABG seja uma medida tão viável quanto as medidas já consagradas na literatura. Em seguida, o CABG é submetido à análise de concordância em relação ao Cosseno de Salton e Índice de Jaccard por meio do método de Bland-Altman. Esta metodologia consiste na aplicação do Teste T de *Student* pareado (ou entre as diferenças entre as medidas a serem analisadas e o valor de referência zero), ou seja, quanto mais próxima a diferença entre os métodos de zero, maior a concordância; ou ainda, o teste T compara as médias das diferenças entre as medidas com o valor zero. Destaca-se que uma forte correlação não implica, necessariamente, em concordância significativa. Hirakata e Camey (2009) indicam que, em geral, a correlação mede a associação entre as variáveis comparadas e não propriamente a concordância entre elas. Nesse sentido, o procedimento de cálculo de correlação seguido de teste estatístico para analisar diferenças significativas entre medidas também está presente no campo de Ciência da Informação e se assemelha a Larivière e Gingras (2011) em que os autores comparam as medidas de citação normalizada por campo de conhecimento *AoR* (*average of ratios* – média das razões) e *RoA* (*ratio of averages* – razões das médias) e concluem que mesmo com forte correlação, as medidas são estatisticamente diferentes. A seguir, o método de Bland-Altman gera um gráfico em que é possível visualizar o quanto as diferenças calculadas entre os métodos se distanciam de zero, além de outliers e possíveis tendências (Figuras 12 e 13). As aplicações do ICC e de Bland-Altman foram feitas por meio dos softwares *SPSS* e *Excel*.

5) Mapeamento da estrutura intelectual da floresta de genealogia acadêmica composta pelos bolsistas PQ-Sr e seus orientandos: foram construídas redes de genealogia acadêmica a fim de representar o domínio analisado. Para cada bolsista, foi construída uma árvore de genealogia acadêmica e posteriormente todas foram reunidas em uma só representação, constituindo uma floresta de genealogia acadêmica. Entretanto, deve-se notar que para cada perspectiva de acoplamento (CABG, Cosseno de Salton e Índice de Jaccard) foi construída uma floresta específica para que as comparações entre estas perspectivas sejam mais bem visualizadas, uma vez que, em cada rede, as espessuras das arestas (ligações) foram valoradas por

cada uma das três perspectivas (Figuras 15, 16 e 17). Para a construção dessas florestas, foi utilizado o software Pajek. Após a construção das florestas de GA, avaliou-se a força de conectividade média segundo as três perspectivas de normalização do ABA analisadas (Tabela 7). Avaliar a força de conectividade em grafo remete diretamente a analisar topologicamente a intensidade das conexões entre os atores que compõe o grafo. Neste caso, como orientador e orientando estão conectados segundo o Índice de Jaccard, Cosseno de Salton e CABG. Além disso, foram apresentados os referentes recitados que compõem toda identidade científica do domínio analisado. Para este processo, foram construídas as redes de citação (Figura 18) e cocitação (Figuras 19 e 20) utilizado o software *VosViewer* em que a parte citante é representada pelos orientandos e os citados pelos autores acopladores recitados (Quadro 3).

Figura 10 – Fluxograma dos procedimentos metodológicos



Fonte: Elaboração do autor. *Dados oriundos da Tabela 2.

7 RESULTADOS

A fim de identificar, recuperar e promover a intersecção entre os referentes teóricos da produção científica de artigos de cada bolsista PQ-Sr e de seus orientandos, presentes na base de dados *Web Of Science*, foram recuperados todos os documentos mediante os critérios C.1 e C.1.1 em que o orientador seja primeiro autor e seu orientando a ser acoplado não é coautor da publicação e todos artigos em que o orientando seja primeiro autor e seu orientador a ser acoplado não é coautor do documento como descrito nos procedimentos metodológicos.

O cálculo do acoplamento, utilizando-se como unidade de acoplamento todos os autores citados (citados ao menos uma vez), será dado pela intersecção entre os elementos presentes na quinta coluna da Tabela 3, enquanto o cálculo segundo a recitação de autores (citados em dois documentos ou mais), pela intersecção entre os elementos da sexta coluna (Tabela 3), para cada par de orientação.

A identificação das obras segundo o critério de autoria estabelecido foi feita de maneira manual e foi constatado que 2 dos 5 pesquisadores do IMPA (Alcides Neto e Luiz Velho) não publicaram artigos em coautoria com seus orientandos indexados na WoS. Para estes três pesquisadores, a lista de autores citados e recitados é oriunda de toda a publicação na qual estes pesquisadores são primeiros autores. Analogamente, este critério se aplica aos orientandos.

Do ponto de vista topológico, a floresta de genealogia acadêmica analisada é composta por 5 orientadores em que o nó (pesquisador) mais fértil é Jacob Palis com 36 orientandos (egressos de doutorado) autores de publicações que atendem ao critério de autoria estabelecido. Ainda, Jacob Palis orientou o bolsista PQ-Sr. Paulo G. Sad, também presente no rol de orientadores analisados.

Dessa forma, esta floresta tem duas gerações de orientações a partir de Jacob Palis Junior. O pesquisador também orientou o bolsista PQ-Sr Artur Lopes, além de Ali Tahzibi, Carlos A. M. Rojas, Carlos G. T. A. Moreira, Marcelo M. Silva Viana, Vilton Pinheiro e Luciano Castro Filho, também bolsistas PQ.

Em relação aos demais orientadores, Alcides Lins Neto orientou o também bolsista PQ Hossein Movasati. Ainda, o pesquisador Aloisio Araújo se apresenta como o segundo orientador mais fértil da rede analisada com 15 orientações que respondem ao critério de autoria. Já Paulo Sad e Luiz Velho podem ser considerados os menos férteis nesse contexto com uma e duas orientações, respectivamente.

A seguir são apresentados, na Tabela 4, todos cálculos de acoplamento bibliográfico baseados nas medidas de similaridade.

Tabela 4 – Cálculo dos acoplamentos segundo os padrões de citação e recitação de autores

Pesquisador	ABA	$L1 \cap L2$	Jaccard (total)	Jaccard (recit)	Salton (total)	Salton (recit)	CABG	CABG (total)
PQ-Sr - Alcides Lins Neto	-	-	-	-	-	-	-	-
Wanderson Costa e Silva	12	5	0,0882	0,1563	0,1656	0,3150	0,1786	0,13483
Liliana Puchuri Medina	11	2	0,1078	0,0667	0,2380	0,1890	0,0714	0,12360
Arturo U. Fernandez Perez	22	8	0,1528	0,1667	0,2658	0,2857	0,2857	0,24719
Leonardo M. Camara	14	6	0,1111	0,1714	0,2078	0,3145	0,2143	0,15730
Hossein Movasati	26	6	0,0925	0,0612	0,1867	0,1300	0,2143	0,29213
Serigio Mariano Licanic	10	5	0,0952	0,1724	0,2079	0,3858	0,1786	0,11236
Luis Gustavo Mendes	14	5	0,1061	0,1389	0,1966	0,2621	0,1786	0,15730
Jose Omegar Calvo Andrade	20	8	0,1047	0,1780	0,1919	0,3024	0,2857	0,22472
PQ-Sr. Luiz C. P. R. Velho	-	-	-	-	-	-	-	-
Thales Miranda Vieira	3	0	0,0085	0,0000	0,0174	0,0000	0,0000	0,01327
Esdras Medeiros	14	0	0,0471	0,0000	0,1010	0,0000	0,0000	0,06195
PQ-Sr. Paulo Grossi Sad	-	-	-	-	-	-	-	-
Gabriel Calsamiglia	9	2	0,0989	0,1053	0,2027	0,2020	0,2857	0,33333
PQ-Sr. Jacob Palis Junior	-	-	-	-	-	-	-	-
Ali Tahzibi	17	6	0,1024	0,0759	0,2098	0,1903	0,0845	0,1269
Artur Oscar Lopes	31	15	0,0670	0,0811	0,1411	0,1567	0,2113	0,2313
Aubin Arroyo	25	9	0,1196	0,1154	0,2160	0,2670	0,1268	0,1866
Bernando A. S. M. Rebolledo	18	6	0,1208	0,0789	0,2707	0,2147	0,0845	0,1343
Bladismir Leal	15	2	0,1034	0,0282	0,2541	0,1678	0,0282	0,1119
Carlos A. M. Rojas	42	22	0,1117	0,1630	0,2153	0,2815	0,3099	0,3134
Carlos F. Borges Palmeira	0	0	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Carlos G. T. A. Moreira	22	6	0,1222	0,0769	0,2305	0,1975	0,0845	0,1642
Carlos Vasquez	16	4	0,0829	0,0412	0,1596	0,0867	0,0563	0,1194
Eduardo Colli	27	8	0,1475	0,1096	0,2675	0,3002	0,1127	0,2015
Eleonora D. C. Garcia	48	23	0,1589	0,1855	0,2821	0,3131	0,3239	0,3582
Enrique Ramiro Pujals	48	26	0,2319	0,2766	0,3770	0,4408	0,3662	0,3582
Fernando J. Sanchez-Salas	9	1	0,0581	0,0127	0,1419	0,0396	0,0141	0,0672
Frederico Rodriguez-Hertz	28	12	0,0979	0,0945	0,1803	0,1727	0,1690	0,2090
Iaci Pererira Malta	0	0	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Jaques Gheiner	10	5	0,0667	0,0610	0,1694	0,1483	0,0704	0,0746
Jorge Beloqui	10	5	0,0641	0,0649	0,1527	0,1789	0,0704	0,0746
Jorge M. Martins da Rocha	15	4	0,0915	0,0476	0,1932	0,1151	0,0563	0,1119
Leonardo Enrique Mora	16	4	0,1046	0,0533	0,2336	0,1678	0,0563	0,1194
Julio L. J. Diaz Casado	46	23	0,1893	0,1840	0,3192	0,3111	0,3239	0,3433
Luiz F. C. da Rocha	4	0	0,0240	0,0000	0,0568	0,0000	0,0000	0,0299
Maria I. T. Camacho	6	3	0,0385	0,0375	0,0980	0,1028	0,0423	0,0448

Martin Sambarino	27	9	0,1667	0,1184	0,3145	0,2855	0,1268	0,2015
Paulo Roberto Grossi Sad	8	2	0,0523	0,0263	0,1330	0,0897	0,0282	0,0597
Pedro Mendes	12	4	0,0800	0,0556	0,1959	0,2123	0,0563	0,0896
Marcelo M. da Silva Viana	40	13	0,1961	0,1444	0,3283	0,2705	0,1857	0,3125
Pedro M. N. R. Dias Duarte	22	13	0,0741	0,1008	0,1397	0,1831	0,1831	0,1642
Rafael Labarca	26	12	0,0732	0,1290	0,1429	0,2442	0,1690	0,1940
Raul Mario Ures	24	6	0,1364	0,0741	0,2552	0,1780	0,0845	0,1791
Ricadro Mañe	26	10	0,1161	0,1042	0,2085	0,2006	0,1408	0,1940
Roberto Markarian	17	7	0,0833	0,0769	0,1574	0,1599	0,0986	0,1269
Roger Metzger	15	8	0,0872	0,1026	0,1780	0,2451	0,1127	0,1119
Sergio Eugeniu Plaza	1	1	0,0051	0,0110	0,0110	0,0259	0,0141	0,0075
Stefano Luzzatto	42	21	0,0873	0,1615	0,1840	0,2786	0,2958	0,3134
Vilton Pinheiro	9	3	0,0552	0,0395	0,1261	0,1259	0,0423	0,0672
Welington Celso de Melo	15	8	0,0932	0,0976	0,1999	0,2178	0,1127	0,1119
PQ-Sr. Aloisio P. Araújo	-	-	-	-	-	-	-	-
Alvaro J. Riascos Villegas	0	0	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Elvio Accinelli Gamba	31	10	0,0484	0,0518	0,1022	0,1206	0,0625	0,0647
Gyorgy Varga	0	0	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Marco Shinobu Matsumura	4	1	0,0072	0,0058	0,0206	0,0211	0,0063	0,0084
Rodrigo Jardim Raad	20	8	0,0385	0,0462	0,1180	0,1380	0,0500	0,0418
Alvaro Sandroni	50	7	0,0683	0,0311	0,1311	0,0637	0,0458	0,1048
José V. Machado Vicente	7	1	0,0127	0,0058	0,0355	0,0228	0,0063	0,0146
Marcos Hiroyuki Tsuchida	0	0	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
Bruno Funchal	5	2	0,0088	0,0132	0,0238	0,0733	0,0134	0,0104
Luciano I. Castro Filho	67	22	0,0866	0,1009	0,1609	0,1908	0,1447	0,1399
Wilfredo Leiva Maldonado	22	5	0,0392	0,0278	0,0886	0,0766	0,0316	0,0495
Susan Schommer	11	2	0,0214	0,0127	0,0679	0,0803	0,0129	0,0235
José Heleno Faro	25	7	0,0473	0,0507	0,1054	0,1501	0,0547	0,0595
Humberto L. A. Moreira	3	0	0,0057	0,0000	0,0187	0,0000	0,0000	0,0063
Paulo Klinger Monteiro	46	22	0,0853	0,1264	0,1913	0,2648	0,1467	0,1002

Fonte: Elaboração do Autor

Expostos todos os cálculos das 8 medidas baseadas no ABA, aponta-se a presença de cinco acoplamentos (coluna ABA – Tabela 4) iguais a zero, isto é, quatro díades orientador-orientando não se acoplaram pelos documentos, que atendem o critério de autoria, indexados na WoS. A saber: Jacob Palis e Carlos F. Borges Palmeira; Jacob Palis e Iaci Pererira Malta; Aloisio Araújo e Álvaro J. Riascos Villegas; Aloisio Araújo e Gyorgy Varga; Aloisio Araújo e Marcos Hiroyuki Tsuchida.

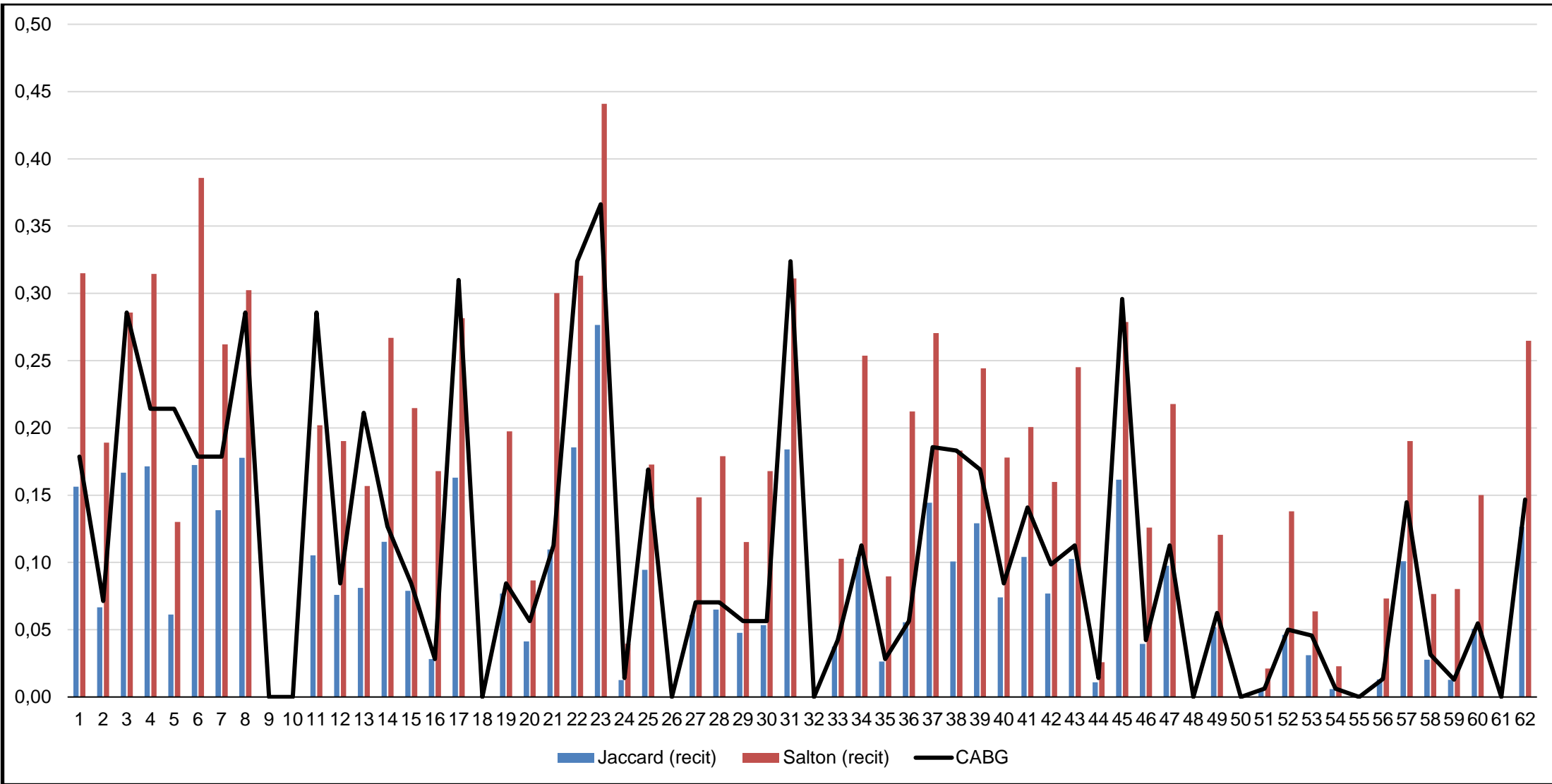
Ao observar o acoplamento segundo a recitação de autores ($L1 \cap L2$), o orientador bolsista PQ-Sr Luiz Velho não se acoplou a seus orientandos Thales Miranda Vieira e Esdras Medeiros via autores recitados. Embora haja acoplamento

pelos autores citados (citados pelo menos uma vez em toda obra dos pesquisadores analisados), este fato aponta um não acoplamento entre as identidades científicas de orientador e orientandos. Neste caso, o valor do CABG, bem como de Jaccard (*recit*) e Salton (*recit.*), são zero, apontando que não há similaridade entre os conjuntos de autores recitados.

Os acoplamentos entre Jacob Palis Junior e Luiz F. C. da Rocha, e, Aloisio Araújo e Humberto L. A. Moreira também resultaram em zero. Com isso, dos 62 acoplamentos calculados ($L1 \cap L2$), nove são iguais a zero. Mesmo que zero remeta a nulidade, estes valores não devem ser excluídos da análise, pois indicam uma não conexão teórico-metodológica entre orientador e orientando. Esclarece-se que um valor de ABA (coluna 2 – Tabela 4) igual zero implica em $L1 \cap L2 = 0$, entretanto, a recíproca não é válida, pois, pode haver $ABA \neq 0$ e $L1 \cap L2 = 0$. Este caso é representado pelas outras quatro orientações supracitadas.

Adiante, adentra-se na comparação entre o CABG, Índice de Jaccard e Cosseno de Salton baseados no acoplamento bibliográfico entre autores (procedimento 4 – seção 6.2). Primeiramente, são apresentadas as comparações entre as medidas mediante a recitação de autores. A partir da construção da Tabela 4, foi possível ilustrar, graficamente, o comportamento de *Jaccard (recit)*, *Salton (recit)* e do CABG, na Figura 11.

Figura 11 – Gráfico Comparativo entre as medidas Jaccard (recit), Salton (recit.) e CABG



Fonte: Elaboração do Autor

O gráfico da Figura 11 é composto por barras verticais (em azul e vermelho), que representam o Índice de Jaccard e o Cosseno de Salton conforme os padrões de recitação de autores, e pela linha (em preto) que indica os valores do CABG. É sabido que gráficos de linha em si são aconselhados para explicitações de séries temporais, o que não é o caso da Figura 11, porém, a linha do CABG ajusta-se as barras e tem função de enfatizar o comportamento do coeficiente perante das demais medidas.

Ao visualizar os topos das barras verticais (quanto maior a barra, maior o valor calculado) é possível observar que, em geral, o Cosseno de Salton destoa das outras duas medidas por atingir valores calculados maiores. De maneira contrária, o Índice de Jaccard revela, em sua maioria, menores valores calculados perante o Cosseno de Salton e o CABG.

A linha que representa o CABG no gráfico da Figura 11 apresenta valores, na maior parte dos casos, maiores que o Índice de Jaccard e menores que o Cosseno de Salton. O CABG é maior que o Índice de Jaccard em 52 dos 62 casos (83,87% do total) e igual em 10 outros casos (16,13%), em que 9 deles o acoplamento é igual a zero e apenas um único caso é diferente de zero e possuem valores iguais. Este cálculo ocorre entre o bolsista PQ-Sr. Jacob Palis e seu orientando Bladismir Leal. Tal fato alinha-se à Equação 26, em que o tamanho (cardinalidade) da lista de autores recitados pelo orientando - $\#L2$ - é igual ao valor da cardinalidade da intersecção entre orientador e orientando $\#(L1 \cap L2)$.

Este acoplamento aponta a situação em que a lista $L2$ do orientando está completamente contida na lista $L1$ de seu orientador. Ressalta-se que esta situação não configura um acoplamento perfeito (igual a 1), visto que, apesar dos autores que compõe a identidade científica do orientando estar presente na lista de seu orientador, a lista do orientador é maior que a de seu orientando e, conseqüentemente, não está contida na lista de seu orientando. Com isso, retoma-se a Equação 27 e reafirma-se que $CABG \geq IJ$, e, em caso de igualdade, o cenário será semelhante ao anteriormente exposto entre Jacob Palis e Bladismir Leal.

Em relação ao Cosseno de Salton, o CABG atingiu valores maiores em 7 cálculos realizados (11,29% do total). Este fenômeno ocorre caso $L2 > L1$ (Equação 28). O Cosseno de Salton, aplicado a recitação de autores atingiu valores maiores que o CABG em 44 cálculos (70,97% do total). Esta situação pode estar relacionada a maior longevidade de carreira dos orientadores bolsistas PQ-Sr, visto que, uma das premissas da bolsa concedida pelo CNPq é a constante produção científica de seus

contemplados. No entanto, os 7 casos em que $L2 > L1$, aponta caminhos para afirmar uma clara influência em termos de volume de produtividade científica perpassada de orientador para orientandos.

Destes sete orientandos que superaram o tamanho da lista de autores recitados em relação a seu orientador, tem-se: i) Hossein Movasati: orientando de Alcides Neto e bolsista PQ-1C também do IMPA, ii) Gabriel Calsamiglia: orientando de Paulo Sad e professor da Universidade Federal do Fluminense (UFF); iii) Artur Oscar Lopes: orientando de Jacob Palis e bolsista PQ-Sr. vinculado a Universidade Federal do Rio Grande do Sul (UFRGS); iv) Carlos Arnoldo Morales Rojas, orientando de Jacob Palis Junior e bolsista PQ-1B vinculado a Universidade Federal do Rio de Janeiro (UFRJ); v) Eleonora Dolores Catsigeras Garcia: orientanda de Jacob Palis Junior e pesquisadora da *Universidad de la República Uruguay*; vi) Júlio Lorenzo Justino Diaz Casado: orientando de Jacob Palis Junior, pesquisador e professor da Pontifícia Universidade Católica (PUC) do Rio de Janeiro e associado da academia brasileira de ciências; vii) Stefano Luzzatto; orientando de Jacob Palis Junior e pesquisador do *International Centre for Theoretical Physics (ICTP)*.

Tais fenômenos descritos anteriormente podem ocorrer segundo as características do domínio genealógico a ser analisado. Esta possibilidade torna o CABG uma medida bastante dinâmica, uma vez que está condicionada à continuidade da produção científica do orientador e/ou do orientando, e não somente de uma das partes. Isto é, pode, ao longo do tempo, aumentar, caso $L2$ aumente em maior proporção que $L1$, ou diminuir caso $L1$ aumente em maior proporção que $L2$.

Além disso, o CABG e o Cosseno de Salton têm os mesmos valores se as cardinalidades de $\#L1$ e $\#L2$ forem iguais, como descrito na Equação 25 ou se a intersecção $L1 \cap L2$ for igual a zero. No rol de pesquisadores analisados, foi constatado 11 cálculos em que estas medidas se equiparam, em que 9 deles são iguais a zero, e em 2 cálculos tem-se as cardinalidades $\#L1 = \#L2$; a saber: Arturo Ulises Fernandez Perez, orientando de Alcides Lins Neto, e Pedro Miguel Duarte, orientando de Jacob Palis.

Descritas as características gerais de comportamento das três medidas baseadas no ABA conforme a recitação de autores, as medidas foram submetidas à análise de correlação intraclasse. Para isso, calcularam-se os coeficientes de correlação intraclasse (ICC) entre as três medidas, como apresenta Tabela 5.

Tabela 5 – Correlação intraclasse entre Índice de Jaccard, Cosseno de Salton e CABG.

Medidas	Jaccard (recit.)	Salton (recit.)	CABG
Jaccard (recit.)	1	0,715*	0,873*
Salton (recit.)		1	0,840*
CABG			1

Fonte: Elaboração do Autor. *significativo a nível de 5%.

O ICC apresentado na Tabela 5 aponta uma forte correlação entre as três medidas comparadas, com destaque para as correlações entre CABG e Jaccard (recit) e CABG e Salton (recit). O ICC pode ser definido com o um descritor de confiabilidade entre dois meios de avaliação e até que ponto itens em um determinado teste, ou domínio permanecem alinhados (CICCHETTI, 1994). O coeficiente é amplamente utilizado nas Ciências da Saúde para analisar a correlação entre dois examinadores, mas é possível observar trabalhos no âmbito da Ciência da Informação de Deniszczwicz e Kern (2013) e Bornmann (2011), que utilizam o ICC para analisar a confiabilidade entre pareceristas.

Similar à classificação das análises de correlação (Pearson, Spearman, Kappa, entre outros), é possível classificar coeficiente de correlação intraclasse em: i) pobre para ICC menor que 0,40; ii) razoável para ICC entre 0,40 e 0,59; iii) bom para ICC entre 0,60 e 0,74; iv) excelente para valores de ICC 0,75 e 1,00 (CICCHETTI, 1994).

A correlação entre o Índice de Jaccard e o CABG, e entre o Cosseno de Salton e o CABG são consideradas excelentes (fortes) com valores iguais a 0,873 e 0,840, respectivamente. Estes valores representam uma forte correlação entre as medidas e aponta um forte alinhamento entre os cálculos de similaridade baseados nas recitações de autores por dois pesquisadores com relação de genealogia acadêmica.

Contudo, apesar da intensa, as correlações não são perfeitas. Caso fossem, bastaria conhecer o valor de Cosseno de Salton ou Índice de Jaccard para obter o CABG e vice-versa, o que não é o caso, como explicitado anteriormente. Ainda, mesmo que tal fato ocorra, é possível que este fenômeno seja uma especificidade do domínio, dadas as características que diferenciam as medidas (apresentadas pelas Equações de 20 a 28).

Com intuito de submeter o CABG à comparação com outras medidas amplamente difundidas no âmbito dos Estudos Métricos da Informação para possível validação enquanto indicador viável para análises de acoplamento bibliográfico, foi

constatado que o Cosseno de Salton, Índice de Jaccard e CABG variaram de maneira conjunta entre si. Desta maneira, dado o alinhamento do CABG a estas medidas, tem-se o primeiro indício da possibilidade de uso da medida enquanto indicador de intensidade de acoplamento bibliográfico entre autores com relação de orientação acadêmica. E mais, o ICC revela que o alinhamento entre o CABG e o Índice de Jaccard, e, entre o CABG e o Cosseno de Salton são maiores que a correlação entre o Cosseno de Salton e o Índice de Jaccard (0,715 – classificado como bom).

Entretanto, este resultado não indica concordância entre as medidas. Sabendo que as medidas de Jaccard e Salton são reconhecidas e bem estabelecidas na literatura especializada no campo de Ciência da Informação, o CABG estar correlacionado a estas medidas, indica que está "calibrado" com elas. Em suma, o CABG possui a capacidade de "captar" as variabilidades tão bem quanto Jaccard e Salton, mas guarda distinções em seus resultados, sendo sua principal característica admitir acoplamento perfeito entre pesquisadores mesmo que possuam lista de referências de tamanhos diferentes.

Bland e Altman (1983) indicam que medidas de correlação¹⁵ não necessariamente apontam uma concordância entre métodos. Para isso, o CABG foi submetido a análise de concordância por meio do método de Bland-Altman em relação ao Índice de Jaccard e ao Cosseno de Salton para justamente avaliar as distinções de valores entre as medidas.

O método proposto por Bland e Altman, oriundo da Ciências da Saúde, e amplamente aplicado em estudos bioestatísticos, matemáticos e epidemiológicos, visa avaliar a concordância entre duas variáveis, utilizando uma visualização gráfica a partir da dispersão da diferença entre estas duas variáveis e a média entre elas (HIRAKATA; CAMEY, 2009). O gráfico de Bland-Altman ilustra o viés (o quanto as diferenças se afastam do valor zero), o erro (a dispersão dos pontos das diferenças em torno da média), os limites de concordância ($média \pm 1,96 \times Desvio\ Padrão$)¹⁶ além dos possíveis outliers e tendências.

As autoras supracitadas apontam que este teste para concordância busca descrever se as diferenças entre as medidas dependem ou não de seus valores. Desta maneira, a hipótese de o viés ser ou não igual a zero pode ser testada pelo teste T de

¹⁵ Apesar da escolha pelo ICC, o coeficiente de correlação de Pearson e Spearman também indicaram forte correlação entre as medidas.

¹⁶O valor 1,96 refere-se a significância $\alpha=0,05$

Student para amostras emparelhadas (teste T aplicado a uma amostra entre as diferenças das medidas e valor de referência igual a zero).

O teste conclui pela concordância entre as variáveis se o *p-valor* for maior ou igual ao nível de significância estabelecido ($p - valor \geq \alpha$) e pela não concordância entre as variáveis se $p - valor < \alpha$. Assim, se a diferença entre os métodos é estatisticamente igual a zero, há concordância, caso contrário, não. Em geral o teste remete a seguinte ideia: dado dois métodos, se a diferença numérica entre os valores observados é estatisticamente igual a zero, eles são estatisticamente iguais, logo, concordam.

Para a análise de concordância entre o CABG e o Índice de Jaccard e Cosseno de Salton, foram construídos dois gráficos de Bland-Altman, além de calculados os testes T pareados para ambas as situações.

O gráfico descreve a dispersão entre a diferença dos métodos e a média das diferenças entre os métodos e apresenta, na Figuras 12 e 13, três elementos que guiam a análise: os limites de concordância (superior e inferior) em vermelho; o valor do viés da análise (média das diferenças entre os valores calculados) em verde pontilhado; e o próprio valor de zero em preto. Estes valores estão descritos, primeiramente, na Tabela 6.

Tabela 6 – Valores descritivos do gráfico de Bland-Altman

	Jaccard x CABG	Salton x CABG
Média das diferenças (viés)	-0,032	0,055
Desvio Padrão das diferenças	0,049	0,069
Limite de Concordância Superior	0,064	0,191
Limite de Concordância Inferior	-0,128	-0,080

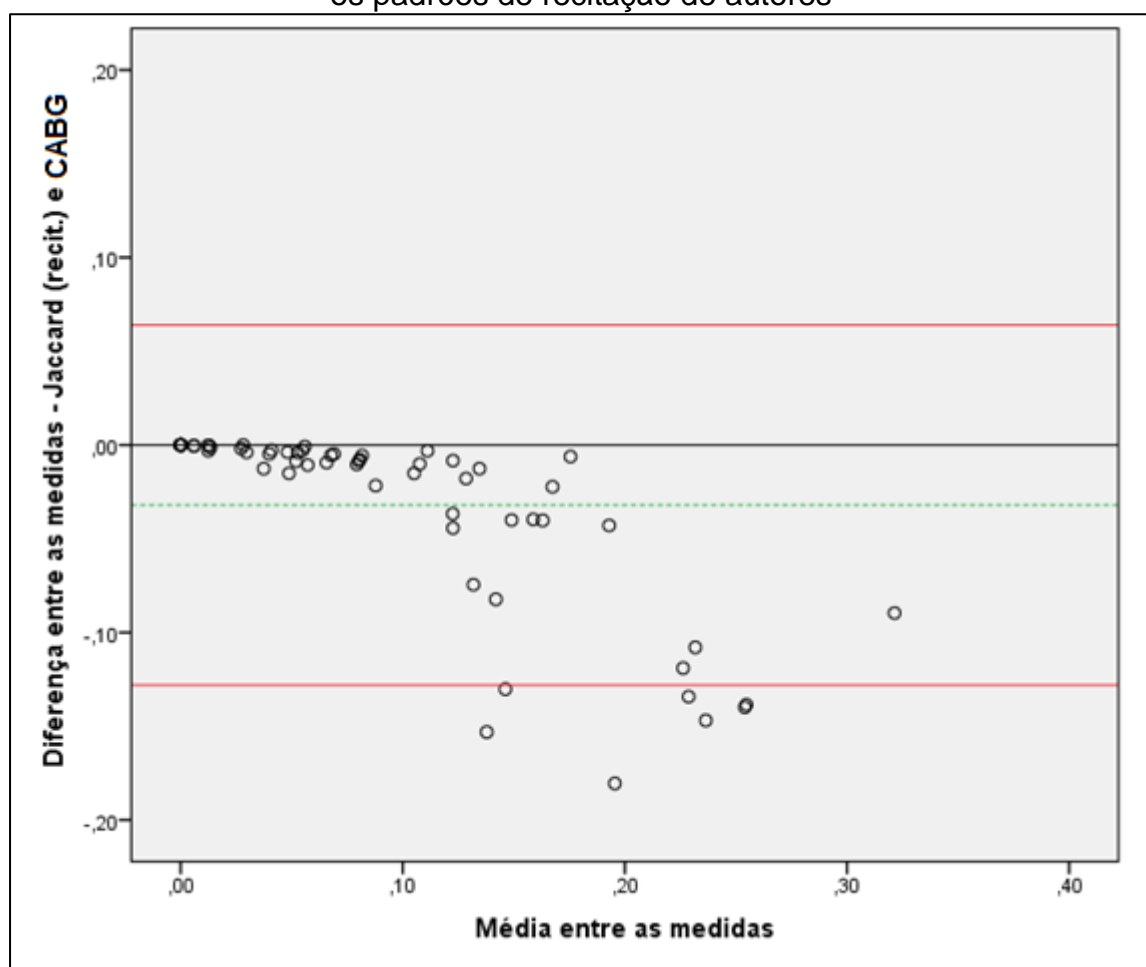
Fonte: Elaboração do Autor

A Tabela 6 aponta que a média das diferenças (viés) da análise entre as medidas de similaridade indica que, em média, o CABG e o Índice de Jaccard diferenciam-se em 0,032. Sabendo que $CABG \geq IJ$, o coeficiente é em média 0,032 maior que Índice de Jaccard.

Com relação à comparação entre o CABG e o Cosseno de Salton, o viés da análise é de 0,055. Assim, os métodos diferenciam-se, em média, em 0,055. Dado que nesta amostra os valores do Cosseno de Salton foram maiores que o CABG na

maioria dos casos, então este viés pode ser entendido como o Cosseno de Salton sendo, em média, 0,055 maior que o CABG. Estes resultados são testados por meio do teste T de *Student* para comparação entre médias e indicam se estas diferenças são estatisticamente diferentes de zero, concluindo se há concordância entre os métodos ou não. A primeira representação, presente na Figura 12, retrata o gráfico de Bland-Altman entre o CABG e o Índice de Jaccard com os valores descritos na Tabela 6.

Figura 12 – Gráfico de Bland-Altman entre o CABG e o Índice de Jaccard segundo os padrões de recitação de autores



Fonte: Elaboração via software SPSS

A análise gráfica da Figura 12 permite observar uma tendência de valores concentrados próximos ou abaixo do valor zero de referência. Este resultado é explicado pelo fato de o CABG ser essencialmente maior que o Índice de Jaccard, visto o denominador $L1$ do CABG (lista de referências de autores recitados do orientador) ser essencialmente menor que o denominador do Índice de Jaccard,

representado pelo valor da união entre $L1 \cup L2$. Os valores das diferenças entre os métodos são iguais a zero quando o CABG possui mesmo valor que o Índice de Jaccard. Como exposto anteriormente, este fenômeno ocorreu em 10 casos analisados e estão representados no gráfico de Bland-Altman pelos pontos sobre a linha horizontal (de cor preta; valor zero).

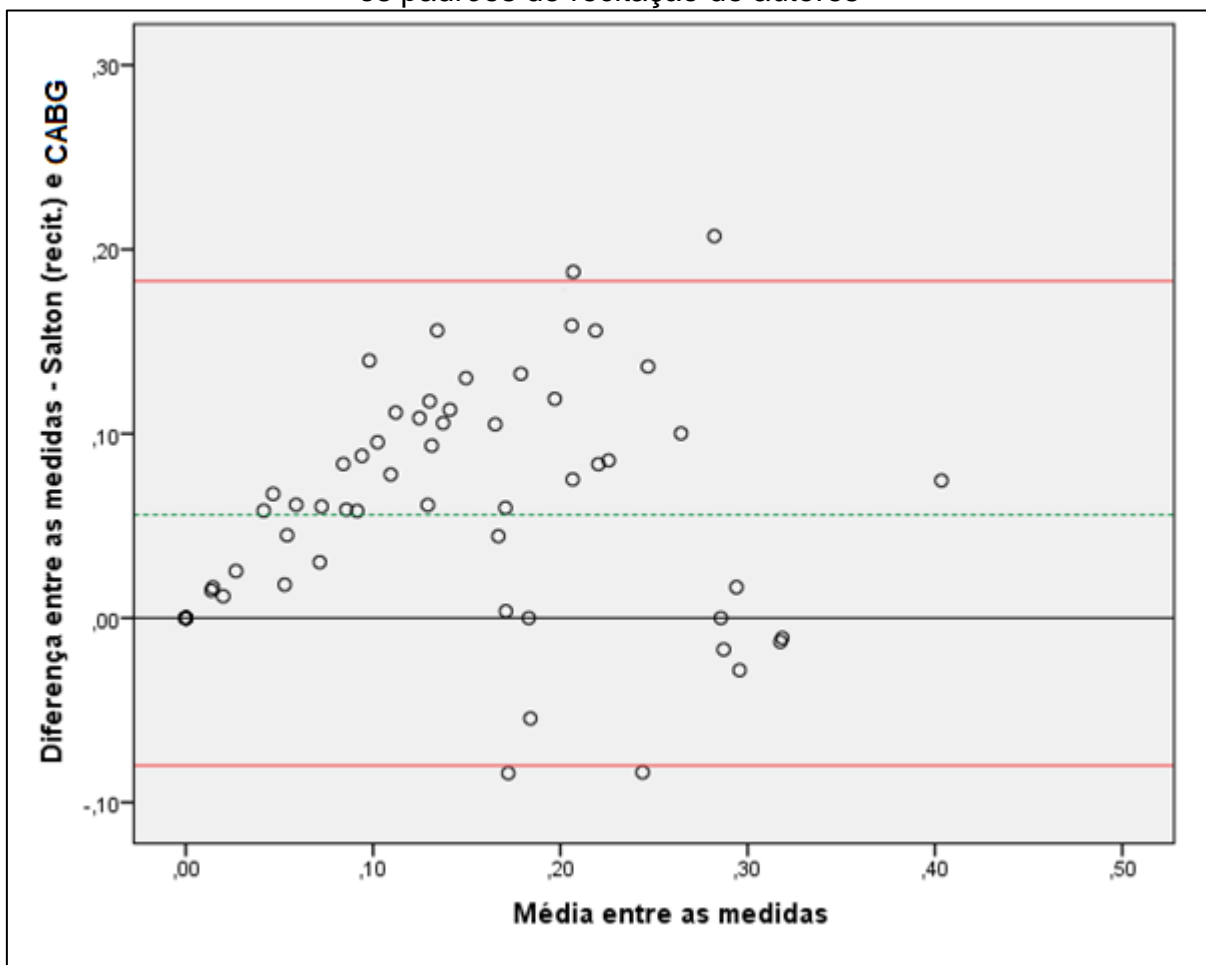
Ao aplicar o teste T pareado entre as medidas, o teste apontou diferenças significativas entre os métodos e zero ($p\text{-valor} < 0,05$). Assim, a média das diferenças entre CABG e Jaccard são significativamente diferentes de zero, e, portanto, tendem a não concordarem entre si. Desta maneira, os métodos estão correlacionados, mas não concordam entre si¹⁷ (as diferenças entre os valores observados dos métodos com o valor de referência zero são diferentes a nível de 5% de significância).

Este resultado não invalida o CABG, pelo contrário, aponta caminhos para que o coeficiente seja uma medida alternativa ao Índice de Jaccard, pois, há correlação, mas não há concordância justamente pela sua fundamentação. Enquanto para o Índice de Jaccard dois conjuntos terão similaridade perfeita (igual a 1) somente se forem idênticos (mesmos elementos e mesma cardinalidade), o CABG permite tal fenômeno mesmo com conjuntos de tamanhos diferentes, para isso, basta $L1$ estar contido em $L2$. Essa premissa alinha-se a ideia de herança intelectual, conceito central dos estudos de genealogia acadêmica.

Na sequência, construiu-se o gráfico de Bland-Altman entre o CABG e o Cosseno de Salton, apresentado na Figura 13.

¹⁷O teste não paramétrico de Wilcoxon também apontou que existem diferenças significativas entre os métodos.

Figura 13 – Gráfico de Bland-Altman entre o CABG e o Cosseno de Salton segundo os padrões de recitação de autores



Fonte: Elaboração via software SPSS

Ao ilustrar o relacionamento entre as diferenças do CABG e do Cosseno de Salton e a média das diferenças das medidas, foi possível observar a ausência de viés de proporção, ou seja, os valores distribuem-se acima e abaixo média das diferenças (linha verde). Isso ocorre devido à formulação do CABG, visto que o coeficiente pode atingir valor maior que o Cosseno de Salton, caso a lista de autores recitados do orientando seja maior que a de seu orientador ($L2 > L1$) (sete casos analisados). Ainda, as medidas concordam totalmente, caso não haja intersecção entre as listas de autores recitados por orientador e orientando ($L1 \cap L2 = 0$) ou caso as cardinalidades de $\#L1$ e $\#L2$ sejam iguais. Esta igualdade (CABG = Cosseno de Salton) ocorreu em 11 casos, como apontado anteriormente.

Após a aplicação do teste T de Student pareado entre o CABG e o Cosseno de Salton foi constatado diferenças significativas ($p\text{-valor} < 0,05$) entre as diferenças

das medidas e zero (as diferenças entre os valores observados dos métodos com o valor de referência zero são diferentes a nível de 5% de significância).

Desta maneira, o teste conclui que não há concordância entre os métodos. De maneira similar à análise entre o CABG e o Índice de Jaccard, este fato não invalida o coeficiente, mas sim, o fortalece apresentando-o como potencial medida de acoplamento bibliográfico entre orientador e orientando.

A análise de concordância, por meio do método de Bland-Altman indicou a não concordância entre os três métodos comparados. Apesar da forte correlação entre as medidas, a diferença entre os valores obtidos por meio dos métodos é significativamente diferente de zero. Para ambas comparações tal fato infere que: sabendo que, tanto o Índice de Jaccard, quanto o Cosseno de Salton, atingem seu valor máximo (100% ou 1) somente se os conjuntos analisados forem exatamente do mesmo tamanho e todos seus elementos iguais, o CABG, ao admitir a possibilidade de alcançar um acoplamento perfeito (valor máximo igual a 100% ou 1) mesmo comparando conjuntos de tamanhos diferentes (basta a lista de autores recitados pelo orientador estar contida na lista de seu orientando), se apresenta como medida de similaridade ajustada para o Acoplamento Bibliográfico entre Autores com relação de genealogia acadêmica, pois, um acoplamento perfeito nesta perspectiva, aponta que o orientando herdou toda lista de autores recitados de seu orientador (o orientando recita os mesmos autores que seu orientador). Um acoplamento perfeito, segundo o CABG, indica que o orientando herdou, por inteiro, a identidade científica de seu orientador. Assim, a identidade científica do orientando contém a identidade científica de seu orientador (a identidade científica do orientador está contida na identidade científica de seu orientando).

Dado o fato de que se duas medidas são estatisticamente concordantes entre si, não faz diferença entre usar uma ou outra. Assim, em complemento a formulação do CABG, constatou-se que o Índice de Jaccard e o Cosseno de Salton também não concordam entre si e mesmo assim são amplamente difundidas no âmbito dos Estudos Métricos da Informação. É neste sentido que a formulação do CABG se anuncia: uma medida de similaridade inédita fundamentada e concebida a partir dos conceitos teóricos das análises de genealogia acadêmica, objetiva, confiável e alternativa às demais presentes na literatura relacionada ao acoplamento bibliográfico.

A metodologia de Bland-Altman aponta que a concordância, o viés calculado, e os limites calculados devem levar em conta aspectos clínicos. Aqui, estes aspectos

podem ser representados como as atividades científicas de relacionamento entre pesquisadores orientadores e orientandos. Levando em conta este aspecto social da ciência, um coeficiente que admite possibilidade de total similaridade entre dois conjuntos distintos de autores, pode ser ideal para analisar, de forma precisa, o legado teórico perpassado de orientador para orientando ao longo do tempo, visto que a medida não é estática.

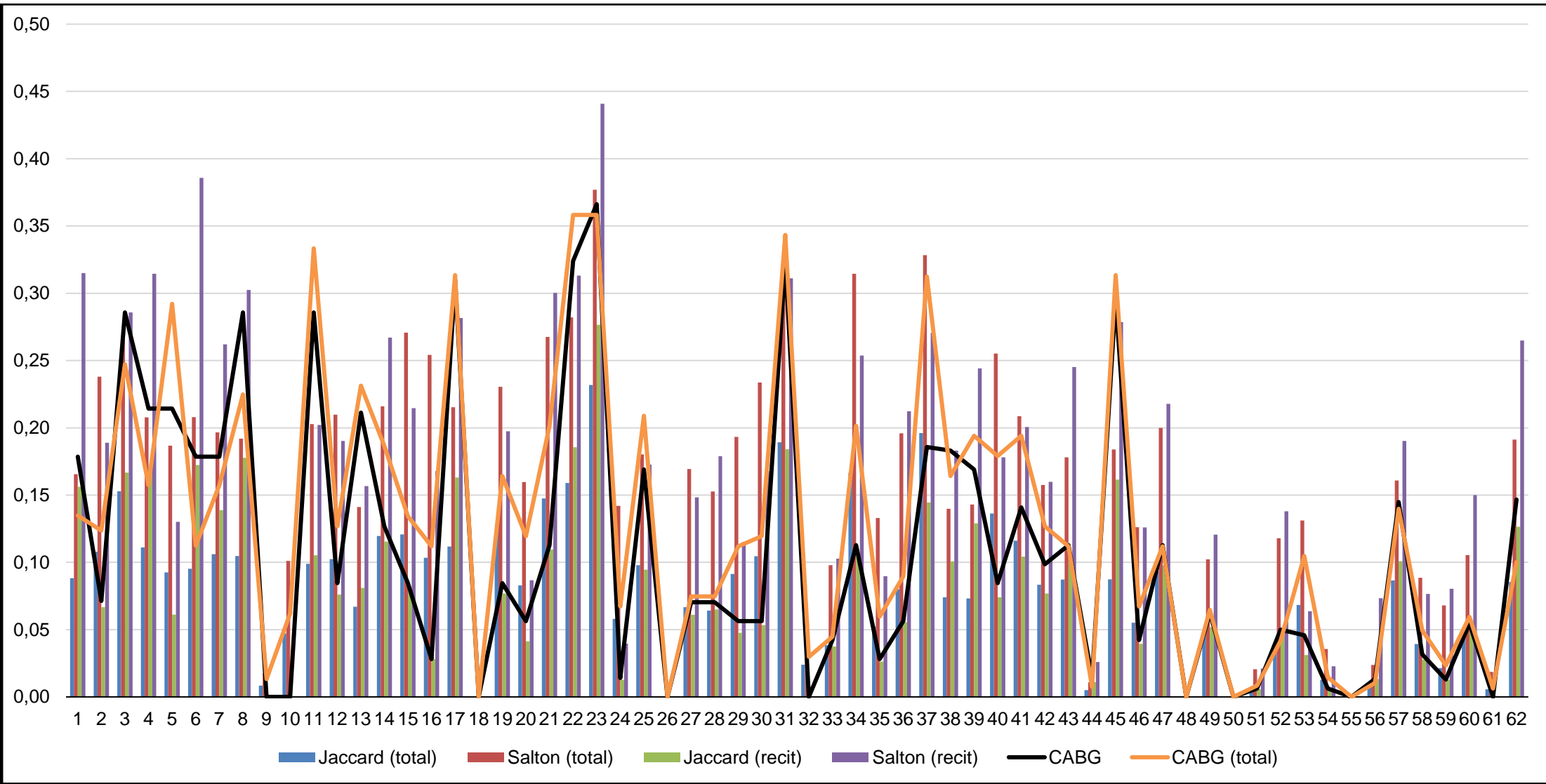
Por mais que o Índice de Jaccard e o Cosseno de Salton permitam alteração de valores ao longo tempo, caso o orientando tenha lista maior que a de seu orientador (supondo aposentadoria, morte ou aumento da atividade de publicação do orientando perante seu orientador), estas medidas jamais atingirão valor igual à 1 (100%), pois, só admitem esta situação se os dois conjuntos comparados forem idênticos.

Diante de toda fundamentação do CABG, é possível enquadrar o coeficiente como uma alternativa metodológica mais representativa para o fenômeno analisado (avaliação da intensidade da manutenção da herança teórico-metodológica do orientador entre seus orientandos), que une os preceitos teóricos da GA, para análises de acoplamento bibliográfico entre autores com relação de orientação acadêmica.

Este resultado é similar, ao apresentar forte correlação e diferenças estatisticamente significativas entre medidas (não concordância), a Larivière e Gingras (2011) em que os autores defendem o uso o *AoR* invés da *RoA* justamente pelas medidas não concordarem com mesmo estando fortemente correlacionada.

A fim de apresentar o comportamento de todas as medidas de maneira conjunta, foi elaborado o gráfico presente na Figura 14.

Figura 14 – Gráfico Comparativo entre Índice de Jaccard, Cosseno de Salton e CABG segundo padrões de citação e recitação



Fonte: elaboração do Autor

O gráfico da Figura 14 apresenta os valores de todas as medidas descritas na Tabela 4. Essa ilustração tem serventia para observar que, em geral, o Cosseno de Salton, independente de se utilizar das recitações ou não, é maior que as demais medidas. Tal fato pode ser explicado pelo denominador de sua formulação (raiz quadrada do produto do tamanho dos conjuntos).

Os valores do CABG e de *CABG (total)* apresentam-se como medidas essencialmente maiores que o Índice de Jaccard e menores que o Cosseno de Salton. Desta forma, o coeficiente, pela sua essência de relativizar a intersecção entre as listas de autores citados por orientador e orientando pela cardinalidade da lista do orientador, possui uma formulação, do ponto de vista sociológico da ciência, mais robusto que as outras medidas pelo fato do CABG levar em conta a própria definição de genealogia acadêmica. Ou seja, é uma medida concebida dentro deste contexto.

A ideia de herança intelectual relativizada pela lista de autores citados pelo orientador alinha-se à relação assimétrica entre orientador e orientando, corriqueiramente representada em redes de genealogia acadêmica por grafos direcionados. Desse modo, o registro de um pesquisador orientador PQ_0 que orienta um pesquisador em desenvolvimento (orientando) P_1 , é representado em uma rede de GA por: $PQ_0 \rightarrow PQ_1$, em que a relação contrária não é válida¹⁸. Este fenômeno é assimétrico, assim como a formulação do CABG que possui como item principal a cardinalidade da lista de citados somente do orientador. A representação assimétrica de orientação ilustra, nas redes de genealogia acadêmica, uma orientação concluída e auxilia na construção das redes e na representação de novas gerações de orientação acadêmica.

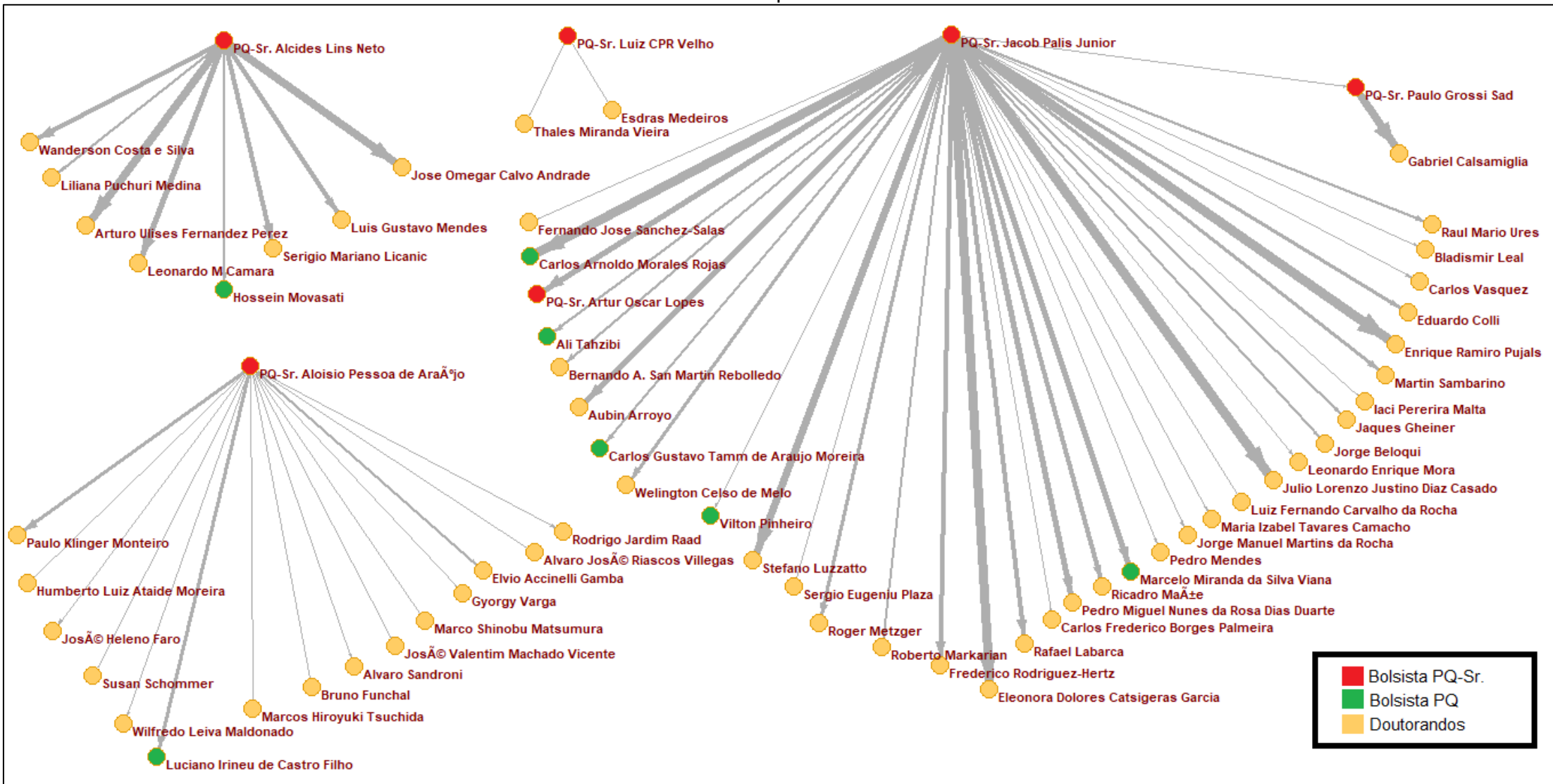
Assim, após a fundamentação do CABG e a comparação com outras medidas, foi construída a floresta de genealogia acadêmica com todos os orientadores do IMPA e seus orientandos (procedimento 5 – seção 6.2).

Na construção da floresta de genealogia acadêmica todas as arestas da rede foram valoradas pelo CABG, Jaccard (recit.) e Salton (recit.) a fim de evidenciar a proximidade teórico-metodológica entre orientador e orientando, medidas pela similaridade das listas de referências de autores recitados em comum. Este tipo de

¹⁸ Ocorrerá $PQ_1 \rightarrow PQ_0$ em caso, por exemplo, de supervisão de pós-doutorado, porém, somente após a finalização de $PQ_0 \rightarrow PQ_1$.

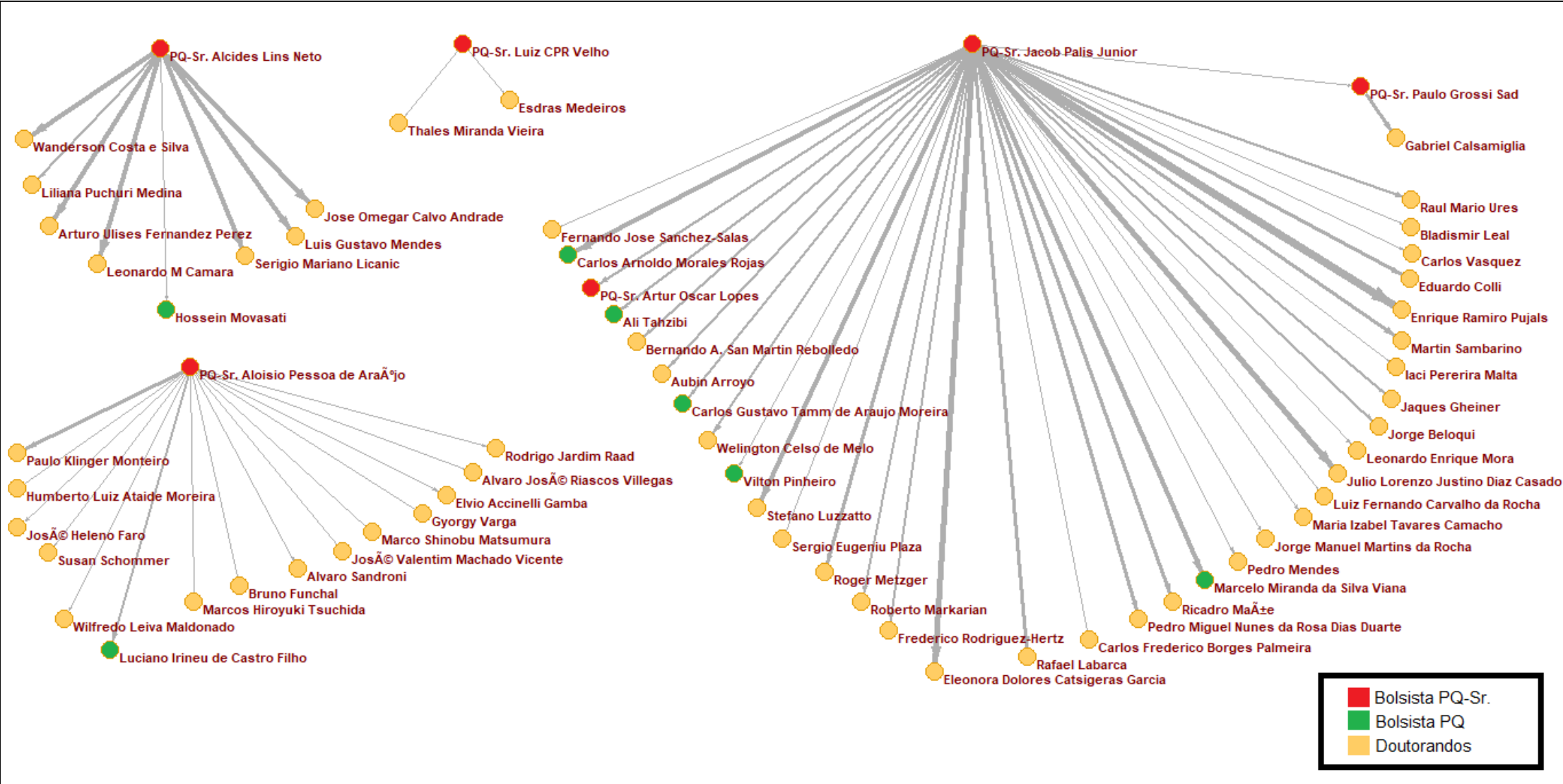
grafo auxilia na visualização destas proximidades e no melhor entendimento da difusão das orientações acadêmicas, como apresentam as Figuras 15, 16 e 17.

Figura 15 – Floresta de genealogia acadêmica dos pesquisadores bolsistas PQ-Sr. do IMPA e seus orientandos egressos de doutorado com arestas valoradas pelo CABG



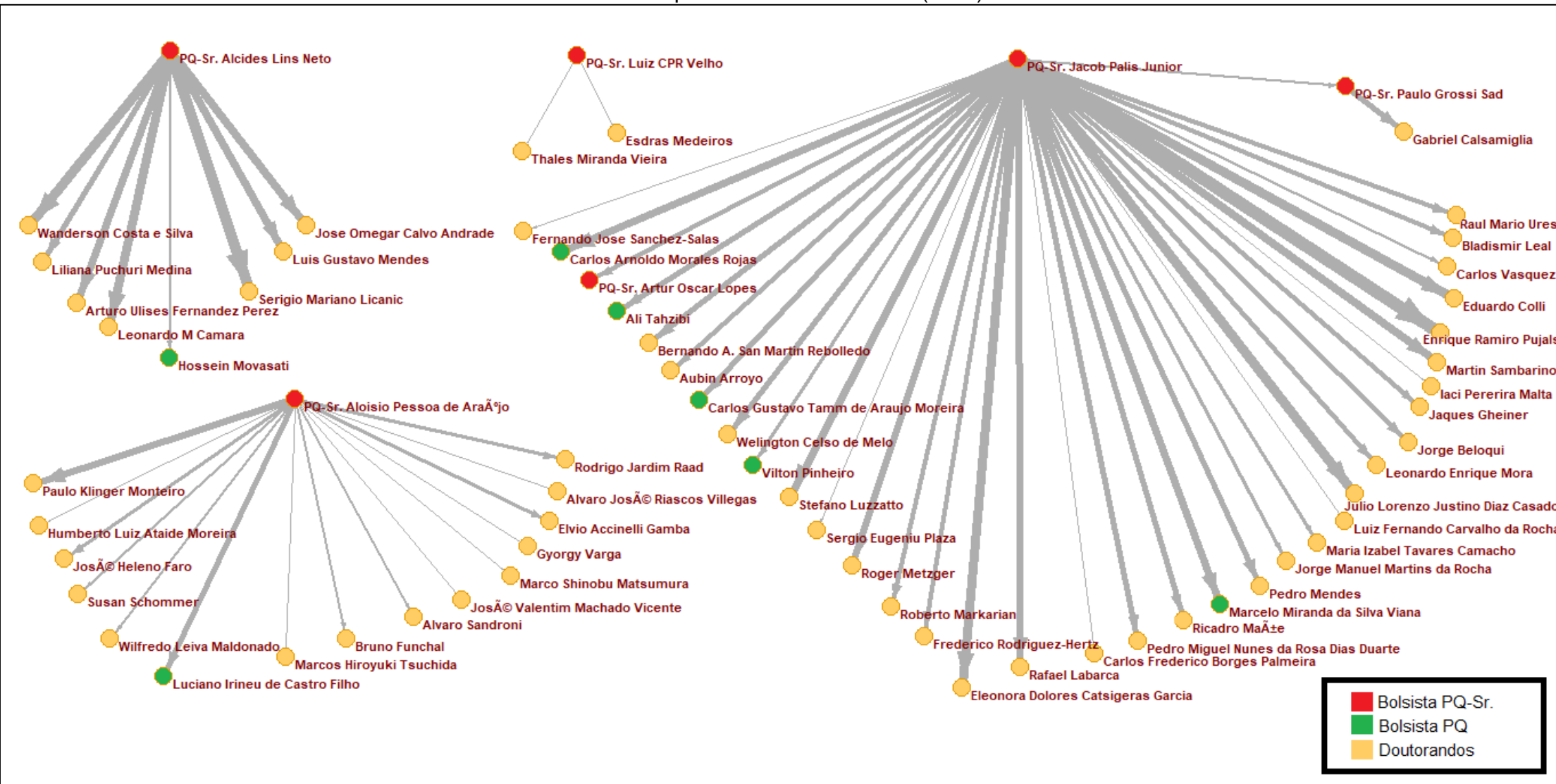
Fonte: Elaboração via software Pajek

Figura 16 – Floresta de genealogia acadêmica dos pesquisadores bolsistas PQ-Sr. do IMPA e seus orientandos egressos de doutorado com arestas valoradas pelo Índice de Jaccard (recit.)



Fonte: Elaboração via software Pajek

Figura 17 – Floresta de genealogia acadêmica dos pesquisadores bolsistas PQ-Sr. do IMPA e seus orientandos egressos de doutorado com arestas valoradas pelo Cosseno de Salton (recit.)



Fonte: Elaboração via software Pajek

É possível observar nas Figuras 15, 16 e 17 diferenças topológicas entre as 5 redes genealogia acadêmica que compõem a floresta da GA dos pesquisadores sêniores do IMPA. O orientador Jacob Palis configura como o mais fértil de toda a floresta. O pesquisador é responsável por 36 orientações, sendo 7 descendentes bolsistas PQ. Destaca-se ainda, seu orientando, também bolsista PQ-Sr, Paulo G. Sad, que deu continuidade à linhagem de Jacob. Desta forma, Jacob é avô acadêmico do orientando (egresso) Gabriel Calsamiglia (orientando de Paulo Sad).

O bolsista sênior Aloísio Araújo é o segundo mais fértil dentre os pesquisadores do IMPA analisados, sendo responsável por 15 orientações de doutorado concluídas (com produções que atendem ao critério de autoria). Dentre estas orientações, destaca-se o pesquisador Luciano Irineu Castro Filho, atualmente bolsista PQ-2 da área de Matemática.

Já o pesquisador Alcides Lins Neto orientou 8 pesquisadores com produções indexadas na WoS e que atendem ao critério de autoria, como destaque ao orientando Hossein Movasati, também bolsista PQ. Os demais pesquisadores do IMPA, Luiz Velho e Paulo Sad são responsáveis por duas e uma, orientações respectivamente.

Ainda do ponto de vista topológico da floresta de genealogia acadêmica, é possível observar diferentes espessuras das arestas que conectam os nós da rede, isto é, as linhas unidirecionadas que ligam o orientador ao seu orientando. Quanto mais espessa a linha, maior a proximidade entre os nós. Neste caso, por se tratar de uma representação de acoplamento bibliográfico entre pesquisadores baseados em autores recitados em comum, quanto maior a espessura da aresta, maior o número de autores recitados em comum entre orientador e orientando segundo as três perspectivas de normalização.

Como a conectividade entre orientador e orientando é parte fundamental desta pesquisa, calculou-se para cada rede de genealogia acadêmica da floresta, a força de conectividade média entre os pesquisadores analisados, apresentada na Tabela 7. “A força de conectividade média da rede, é definida como a média aritmética das forças de conectividade sobre todos os vértices” (PINTO, 2018, p. 10).

Salienta-se que, em cada uma das Figuras 15, 16 e 17, existem 5 redes de genealogia acadêmica. Cada rede se inicia a partir de seu orientador e as cinco redes juntas compõem a florestas de genealogia acadêmica. Evidentemente, que o bolsista PQ-Sr. Paulo Sad está inserido na rede de Jacob Palis, mas para efeito dos cálculos de conectividade, será tratada de maneira separada.

Tabela 7 – Força de conectividade média de cada rede de genealogia acadêmica

Bolsista PQ-Sr.	Força de conectividade média (citados)*	Força de conectividade média (recitados)**	Coefficiente Acoplamento Bibliográfico Genealógico médio**	Índice de Jaccard (recit.) médio	Cosseno de Salton (recit.) médio
Jacob P. Junior	20,47	8,38	0,1180	0,0840	0,1830
Aloísio P. Araújo	19,4	5,8	0,0383	0,0320	0,0800
Alcides L. Neto	16,13	5,63	0,2010	0,1400	0,2730
Paulo G. Sad	9	2	0,286	0,1050	0,2030
Luiz C.P.R. Velho	8,5	0	0	0	0
Média de toda floresta¹⁹	19,08	7,03	0,108	0,0760	0,164

Fonte: Elaboração do Autor. *citados em pelo menos um documento. **citados em dois documentos ou mais.

A Tabela 7 evidencia diferenças topológicas entre os pesquisadores presentes na floresta de genealogia acadêmica. Reforça-se o destaque à Jacob Palis com 20,47 autores citados, em média, em comum com seus orientandos. Ao observar os padrões de recitação, Jacob possui, em média, 8,38 autores recitados em comum com seus orientandos. Ao relativizar esta situação por meio do CABG, o coeficiente explicita um coeficiente médio de acoplamento genealógico (CABGM)²⁰ de 0,118. Isto é, a proximidade média entre as identidades científicas, medida pelo CABG, entre o orientador e seus descendentes é de 11,8%. Ainda, cada descendente herdou, em média, 11,8% da identidade científica do orientador. O Índice de Jaccard e o Cosseno de Salton apresentaram proximidades (similaridades) médias de 0,084 (8,4%) e 0,183 (18,3%), respectivamente.

O pesquisador Aloísio Araújo é o segundo pesquisador com o maior valor médio de autores citados e recitados em comum junto aos seus orientandos (19,4 e 5,8, respectivamente). Contudo, é o pesquisador com menores valores de CMAG, Índice de Jaccard e Cosseno de Salton médios. Tal fato é explicado pela alta quantidade de autores citados (479) e recitados (160) perante seus orientandos, isto é, pela extensa lista de referentes citados e recitados, diminuindo, conseqüentemente, o valor de cada CABG calculado bem como das outras medidas. Este fenômeno será observado sempre que a lista do orientador L_1 crescer em maior proporção perante

¹⁹ Cálculo considerando todos os 62 acoplamentos (e não média das médias dos cinco valores).

²⁰ $CABGM = \frac{1}{N} \sum_{i=1}^N \left(\frac{L_1 \cap L_2 i}{L_1} \right)$; N = número de orientações; i = orientando i .

o a intersecção $L1 \cap L2$, pois, quanto maior o denominador $L1$, menor o valor de CABG.

O valor do CMAG de Aloísio Araújo perante seus orientandos é de 0,0383, ou seja, a proximidade média entre Aloísio e seus descendentes analisados é de 3,83%: cada descendente herdou, em média 3,83% da identidade científica do orientador. Ao verificar as demais medidas, tem-se similaridades médias iguais a 0,032 (3,2%) e 0,080 (8%), respectivamente.

Em contraponto, o bolsista PQ-Sr. Alcides Neto citou e recitou respectivamente, em média, 16,3 e 5,63 autores em comum com seus orientandos. Os valores do CMAG, Índice de Jaccard e Cosseno de Salton indicam uma proximidade média de identidades científicas de 0,201 (20,1%), 0,14 (14%) e 0,273 (27,3%) entre o pesquisador e seus filhos acadêmicos em que cada descendente herdou, em média 20,1% da identidade científica do orientador. Visualmente é possível observar, nas Figuras 15, 16 e 17 que as arestas, desta rede, são mais espessas que das demais redes. Alcides se acoplou com todos seus descendentes segundo os padrões de citação e recitação, isto significa que possui autores citados e recitados em comum com todos seus orientandos. Este fato aponta fortes indícios de continuidade da corrente científica herdada de Alcides.

Com relação aos outros dois pesquisadores, menos expressivos em termos de quantidade de orientações, Paulo Sad, responsável por somente uma orientação acadêmica (segundo os critérios de autoria), citou e recitou, respectivamente, 9 e 2 autores e possui CABG, Índice de Jaccard e Cosseno de Salton iguais a 0,286, 0,105 e 0,203 em relação ao seu orientando. Apresentou, assim, uma proximidade de 28,6%, 10,5% e 20,3% respectivamente entre eles. Neste caso, as medidas CABG, Índice de Jaccard e Cosseno de Salton são iguais aos seus respectivos valores médios por se tratar somente de uma orientação.

Luiz Velho, bolsista PQ-Sr orientador de dois egressos de doutorado, se acoplou com seus orientandos em termos de autores citados em comum, mas não se acoplou por meio de autores recitados em comum. Este não acoplamento entre os pesquisadores via autores recitados indica, por enquanto, uma descontinuidade da identidade científica de Luiz Velho perante seus dois descendentes. Na floresta das Figuras 15, 16 e 17, as arestas advindas do orientador tem serventia apenas para explicitar a relação entre orientador e orientandos.

De modo geral, dentro da floresta analisada, é possível destacar duas orientações: i) entre o bolsistas PQ-Sr. Jacob Palis e Enrique Ramiro Pujals; ii) entre o PQ-Sr. Aloísio Araújo e Luciano Irineu Castro Filho.

Com relação à primeira orientação, a díade se destaca por representarem as maiores similaridades segundo as medidas do CABG, Índice de Jaccard e Cosseno de Salton. Os valores calculados foram de 36,62%, 27,66% e 44,08% respectivamente. Os dois estão acoplados por 26 autores recitados em comum (item 23 do eixo horizontal das Figuras 11 e 14).

Enrique Pujals é graduado pela *Universidad de Buenos Aires* (1992) e tornou-se doutor pelo IMPA, sob orientação de Jacob em 1996. Desde 2007, Enrique é docente do IMPA e dá continuidade a pesquisas relacionadas a sistemas dinâmicos, mesma área de atuação de seu orientador. O pesquisador está inserido também no programa de pós-graduação da instituição e já orientou 15 teses de doutorado além de já ter supervisionado 7 estágios de pós-doutorado.

Já a segunda dupla se destaca pela quantidade de autores citados em comum (citados em pelo menos um documento). Aloísio e Luciano citaram 67 autores em comum dentre os quais 22 foram recitados em ambas as listas de referências analisadas. Luciano, assim como seu orientador, possui formação multidisciplinar entre Ciências Econômicas e Matemática, atuando em pesquisas relacionadas a desenhos de mercado e decisão sob incerteza, temas de interesse também de Aloísio. Atualmente, ambos são professores e pesquisadores do IMPA e estão inseridos no programa de pós-graduação do instituto.

Tanto com relação ao primeiro, quanto a segundo par de orientação, é observada uma continuidade científica entre orientador e orientando, representada não só pela herança de referencial teórico (caracterizada pelo acoplamento), mas também em termos de acadêmicos, visto que ambos os orientandos se tornaram pesquisadores do IMPA, possuem formação acadêmica similar à de seus orientadores e realizam pesquisas com temáticas semelhantes a seus mentores.

É importante destacar que, apesar do foco desta pesquisa estar relacionado a proximidade teórica entre pesquisadores medidas a partir da recitação de autores em comum, que a orientação acadêmica é um processo de formação de recursos humanos que dura anos e envolve diversas nuances em que as heranças perpassadas de orientador para orientando podem ter reflexos não só em referentes citados em comum, mas também na formação como um todo. Exemplifica-se: aptidão

para desenvolver pesquisas semelhantes à do orientador, envolvimento em atividades de pesquisa, ensino e/ou extensão que o orientador tenha apresentado ao orientando, autonomia científica, processos de inovação, entre outros.

Ao retomar a análise de autores recitados em comum e conseqüentemente a identidade científica dos pesquisadores analisados, apresenta-se o Quadro 3 com todos os autores acopladores (unidades de acoplamento), isto é, todos autores recitados em comum entre orientadores e orientandos. A partir da exibição do Quadro 3, inicia-se a construção da estrutura intelectual da floresta de GA analisada.

Quadro 3 – Autores acopladores de toda floresta de genealogia acadêmica

Autor	#ABA	Autor	#ABA	Autor	#ABA	Autor	#ABA
palis, j*	23	bamon, r	4	mascolell, a	2	hansen, lp	1
newhouse, se	17	hayashi, s	4	peixoto, mm	2	hart, od	1
smale, s	17	labarca, r*	4	schmeidler, d	2	hartman, p	1
mane, r*	14	moreira, cgtd*	4	seidenberg, a	2	he, w	1
bowen, r	13	robinson, c	4	tsujii, m	2	kehoe, tj	1
shub, m	11	rovella, a	4	tucker, w	2	klibanoff, p	1
ruelle, d	10	aliprantis, cd	3	wilson, rb	2	kozlovski, os	1
viana, m*	10	arroyo, a	3	yorke, já	2	kreps, dm	1
de melo, w*	9	bewley, tf	3	abraham, r	1	kubler, f	1
guckenheimer, j	9	feigenbaum, mj	3	armstrong, m	1	laffont, jj	1
camacho, c	8	henon, m	3	azariadis, c	1	luenberger, dg	1
pugh, cc	8	kifer, y	3	benartzi, s	1	magill, m	1
mora, l	7	lyubich, my	3	birkhoff, gd	1	marstrand, jm	1
neto, al*	7	mattei, jf	3	blanchard, oj	1	maskin, e	1
benedicks, m	6	sotomayor, j	3	brown, dj	1	milgrom, pr	1
brunella, m	6	ures, r*	3	chateaufeuf, a	1	monteiro, pk*	1
diaz, lj*	6	wen, l	3	colli, e	1	mukerji, s	1
gomezmont, x	6	williams, rf	3	condie, s	1	quiggin, j	1
morales, ca*	6	hirsch, m	3	dana, ra	1	reny, pj	1
pujals, er*	6	arrow, kj	2	de castro, li*	1	rigotti, l	1
takens, f	6	aumann, rj	2	dubey, p	1	riley, jg	1
alves, jf	5	baum, pf	2	dunford, n	1	rockafellar, rt	1
anosov, dv	5	debreu, g	2	ellsberg, d	1	rychlik, mr	1
avila, a	5	geanakoplos, j	2	epstein, lg	1	santos, ms	1
bonatti, c	5	gilboa, i	2	falconer, kj	1	scardua, ba	1
cerveau, d	5	graczyk, j	2	friedman, m	1	schaeffer, hh	1
duffie, d	5	grandmont, jm	2	fudenberg, d	1	simon, cp	1
herman, mr	5	griffiths, ph	2	ghirardato, p	1	simon, lk	1
lorenz, em	5	holmstrom, b	2	gilles, c	1	sternberg, s	1
lucas, re	5	jakobson, mv	2	gorodetski, a	1	tresser, c	1

sullivan, d	5	jouanolou, jp	2	guesnerie, r	1	werner, j	1
araujo, ap*	4	kopell, n	2	gunning, rc	1	woodford, m	1
arnold, vi	4	maccheroni, f	2	hammond, pj	1	yannelis, nc	1

Fonte: Elaboração do Autor. #ABA: vezes em que o autor acoplou dois pesquisadores.
*pesquisadores presentes nas Tabelas 3 e 4 (citantes e citados)

O Quadro 3 apresenta todos os autores que conectaram orientadores e orientandos. Estes autores (acopladores) quantificam 132 autores recitados em dois documentos ou mais pelos pesquisadores analisados. Desta forma, é possível inferir que este rol compõe a identidade científica da floresta de genealogia acadêmica constituída pelos pesquisadores bolsista PQ-Sr do IMPA e seus orientandos egressos de doutorado.

Desses 132 autores recitados, destaca-se que 14 são citantes e citados, isto é, estão entre os 66 pesquisadores analisados. Dentre estes 14, três são orientadores bolsistas PQ-Sr: Jacob Palis (presente em 23 acoplamentos), Alcides Neto (presente em 7 acoplamentos) e Aloísio Araújo (presente em 4 acoplamentos). Dos 11 outros pesquisadores, *mane, r., viana, m., ures, r., de melo, w., morales, ca., pujals, er., labarca, r., diaz, jl. e, moreira, cgtd.* são orientandos de Jacob Palis. Já *de castro, li, e, monteiro, pk.,* são orientandos de Aloísio Araújo.

Estes 11 pesquisadores orientandos apontam a presença de simetria de citação perante o orientador. Desse modo, sabendo que a análise de genealogia acadêmica aqui proposta prevê o relacionamento de citação somente entre orientado e orientando, estes orientandos estão presentes na identidade científica de seus orientadores e herdaram seus orientadores em suas próprias identidades. E mais, pelo fato de as autocitações não terem sido excluídas, é perceptível que estes orientandos se consideram influentes dentro das próprias obras (em termos de recitações) e seus orientadores alinham-se a esta percepção, pois também se autocitam. Quando estas autocitações ocorrem de maneira concomitante, há um acoplamento cruzado²¹ entre orientador e orientando.

É notável que Jacob Palis (*palis, j*) se destaca sobre os demais por estar presente em 23 cálculos de acoplamento. Contudo, estes 23 cálculos são entre Jacob e 23 (dos 36 totais) orientandos. Assim, o pesquisador acoplou-se somente com elementos de sua rede de genealogia acadêmica. De maneira similar, os demais

²¹Dado dois pesquisadores A e B: A cita B e se autocita, B cita A e se autocita. Caso os dois estejam acoplados somente por si mesmos, a força de acoplamento absoluta neste caso é igual a 2.

orientadores, Alcides Neto e Aloísio Araújo, também estão presentes no cálculo de acoplamento com seus orientandos (7 e 4 cálculos respectivamente).

Nesse sentido, dos 62 acoplamentos calculados sob as diferentes perspectivas, em 34 deles (54,84% do total), o orientando acoplou-se ao seu orientador pelo próprio orientador. Como se trata de uma análise de recitação, é possível afirmar que o orientador compõe a identidade científica de 34 orientandos.

Por outro lado, destacam-se três autores recitados presentes em mais de uma rede de genealogia acadêmica que compõe a floresta analisada: i) *arnold, vi.*: presente no cálculo de dois acoplamentos na rede de Jacob Palis Junior e também em dois acoplamentos na rede de Alcides Neto; ii) *brunella, m.*: Presente no cálculo de cinco acoplamentos na rede de Alcides Neto e no único cálculo de acoplamento de Paulo Sad com seu orientando; iii) *camacho, c.*: acoplou sete orientandos à Alcides Neto, e, Paulo Sad a seu orientando.

Vladimir Igorevich Arnold (1937 – 2010) foi um matemático soviético-russo e membro da academia de ciência da Rússia e atuou, principalmente, nas áreas de teoria de sistemas dinâmicos, equações diferenciais e topologia, sendo orientando de Andrey Kolmogorov (famoso pelo teste de normalidade de Kolgomorov-Smirnov). A atuação de Arnold alinha-se diretamente às temáticas de pesquisa de Jacob Palis e Alcides.

Já Marco Brunella (1964-2012) foi um pesquisador italiano que atuou na universidade de *Dijon* e ficou mundialmente famoso pelos estudos da teoria geométrica das folheações holomórficas, temática amplamente presente nas pesquisas de Alcides e Paulo Sad. Lamentavelmente, Brunella faleceu dias após sua chegada ao Rio de Janeiro para um período de pesquisas no IMPA²². Em reconhecimento, o instituto lançou em 2017 o *pós-doutorado de excelência Marco Brunella*.

O terceiro pesquisador recitado em mais de uma rede de genealogia acadêmica, Cesar Camacho (1943), é um famoso matemático nascido no Peru que constituiu carreira no IMPA se tornando diretor do instituto. Atualmente, é professor e pesquisador da Fundação Getúlio Vargas (FGV) e atual bolsista PQ nível Sr. Camacho atuou em diversas áreas da Matemática com enfoque em equações diferenciais e, também, em teoria geométrica das folheações holomórficas. Camacho,

²² <https://www.mat.uniroma2.it/~fbracci/download/marco-per%20UMI.pdf>

Alcides e Paulo Sad alinham-se em temáticas teóricas sendo, inclusive, coautores de trabalhos científicos e livros.

Os três autores apontam uma tendência científica da floresta analisada. Esta tendência está relacionada aos estudos das equações diferenciais e dos sistemas dinâmicos. Por mais que sejam temáticas altamente abrangentes, é possível notar um direcionamento do instituto às pesquisas relacionadas a estes tópicos. Tal fato alinha-se às áreas de pesquisas dos cursos de pós-graduação do IMPA, sendo elas: Álgebra, Análise e Equações Diferenciais Parciais, Computação Gráfica, Dinâmica dos Fluidos, Economia Matemática, Geometria Complexa e Folheações Holomorfas, Geometria Diferencial, Geometria Simplética, Otimização, Probabilidade, Sistemas Dinâmicos e Teoria Ergódica.

Com intuito de identificar toda estrutura de influência teórica sobre a floresta analisada, construiu-se a rede de citação e cocitação tendo como base os autores acopladores de toda a floresta de genealogia acadêmica. Estas redes explicitam a relação entre citante-recitado e recitado-recitado, respectivamente.

Zhao e Strotmann (2008) indicam o uso da análise de cocitação entre autores para a análise das influências intelectuais de pesquisadores sobre determinado campo do conhecimento. Com isso, é possível compreender o desenvolvimento destas influências e suas estruturas ao longo dos anos analisados.

Para isso, foi elaborada a matriz de recitação de autores com todos os 132 autores recitados presentes no Quadro 3 e os orientandos citantes. Dessa maneira, a matriz de recitação (Apêndice 1), possui dimensão 132×53 (*citados x citantes*). É importante lembrar que alguns orientandos não se acoplaram a seus orientadores, logo, não estão presentes nesta matriz.

A partir da construção da matriz, foi elaborada, via software *VosViewer*, a rede de citação entre os orientandos e seus influentes (Figura 18). Como se trata de uma rede em que todos os elementos citados foram recitados em dois ou mais artigos, refere-se à rede de recitação da identidade científica da floresta de genealogia acadêmica. Na Figura 18, têm-se em letras minúsculas os autores recitados e, com as iniciais em letras maiúsculas, os orientandos citantes. A rede apresenta o padrão de recitação entre os orientandos analisados e os 132 autores que compõe a identidade científica da floresta de genealogia acadêmica.

Esta rede dá foco para a relação entre citante e recitado e, desta forma, é possível observar que os orientandos citantes ganham destaque por estabelecerem diversas conexões por meio dos autores recitados em comum com seus orientadores. Aqueles que mais se acoplaram, por meio de autores recitados, em valores absolutos, destacam-se nesta rede devido a tais conexões. Destacam-se os orientandos de Jacob Palis: Carlos Morales (22 autores recitados em comum), Eleonora Castigeras (23 autores recitados em comum), Enrique Pujals (26 autores recitados em comum), Julio Lorenzo (23 autores recitados em comum), e, Stefano Luzzatto (21 autores recitados em comum).

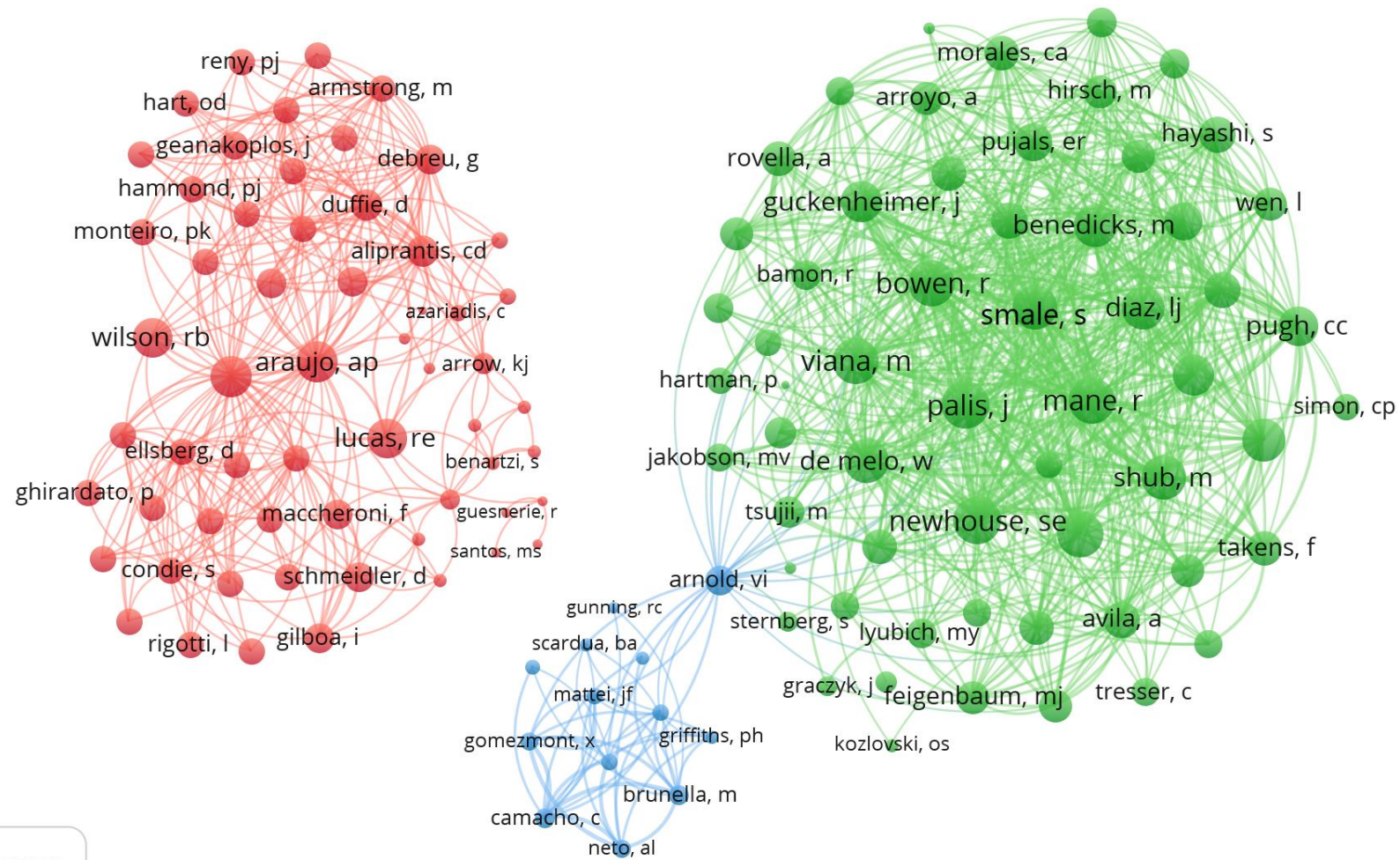
Estes cinco orientandos são os descendentes de Jacob Palis que mais se conectaram ao seu orientando, ganhando notório destaque na rede da Figura 18. Fato importante a se destacar é que estes cinco orientandos recitaram o orientador e se autocitaram em suas respectivas listas de referentes.

É possível observar a visibilidade do orientador na rede, visto que Jacob está presente na lista de autores recitados de 23 orientandos. O orientador também recita 10 orientandos, sendo eles: Aubin Arroyo, Wellington Celso de Melo, Julio Lorenzo, Rafael Labarca, Ricardo Mañe, Carlos Moreira, Enrique Pujals, Ricardo Ures, Marcelo Viana, e, Eduardo Colli, em que, os nove primeiros estabelecem acoplamento cruzado junto ao orientador.

Na Figura 18, pode-se destacar também os orientandos de Aloísio Araújo: Luciano Irineu e Paulo Monteiro. O primeiro, supracitado por ser bolsista PQ e por ser o orientando com maior número de autores citados (em pelo menos um documento) em comum como seu orientador (67 autores), além dos 22 autores recitados em comum, o que gera destaque a Luciano na rede de recitação. Analogamente, Paulo Monteiro também recitou 22 autores em comum com Aloísio e, por este fato, ganha notoriedade na rede de recitação.

É necessário destacar que, assim como na rede dos descendentes de Jacob Palis, tanto Luciano Irineu quanto Paulo Monteiro recitaram o orientador Aloísio Araújo em suas respectivas listas de referentes. Além disso, Aloísio Araújo também recita ambos em sua lista, configurando um acoplamento cruzado, como descrito anteriormente. Destacada a rede de recitação, apresenta-se, na Figura 19, a rede de cocitação dos 132 autores acopladores de modo a avaliar a conexão entre os principais influentes da floresta de genealogia acadêmica, gerando a estrutura intelectual dos pesquisadores que a compõe.

Figura 19 – Estrutura das influências intelectuais da floresta de genealogia acadêmica



VOSviewer

Fonte: Elaboração do autor via software VosViewer

A análise de cocitação entre autores (ACA) avalia um campo de pesquisa e examina um conjunto de autores altamente citados nesse campo do qual a estrutura que é revelada pela ACA representa a base de conhecimento (ZHAO; STROTMANN, 2014).

Zhao e Strotmann (2021) indicam que analisar autores, enquanto unidades de análise alternativa à análise de documentos em si, implica a representação e a construção de escolas de pensamento, visto que, ao considerar toda obra do autor citado como única, toda influência deste autor é carregada consigo no âmbito das análises relacionais de citação.

A Figura 19, ao ilustrar a cocitação entre autores acopladores e recitados em dois ou mais documentos pelos 66 pesquisadores analisados, constitui a base das influências destes autores. O fato da rede de cocitação entre autores ser composta exclusivamente por autores recitados configura-se então como uma análise de cocitação entre a identidade científica dos autores.

Do ponto de vista topológico, a rede possui densidade igual a 0,303, ou seja, está 30,3% conectada em relação a todas as conexões possíveis. A rede apresenta três subredes compostas pelos nós em verde e azul e em vermelho, em que as duas primeiras se refere aos autores recitados mais influentes das redes de orientação acadêmica de Jacob Palis, Alcides Neto e Paulo Sad, enquanto a terceira, à rede de Aloisio Araújo. Já o grau de conectividade (cocitação) médio entre os autores recitados é de 23,57. Isto significa que cada autor é cocitado, em média, com 23,57 outros autores.

A ACA, ao explicitar os autores recitados que mais se conectam uns aos outros, possibilita evidenciar as conexões mais fortes entre os autores recitados, isto é, as maiores frequências em que dois autores são citados de maneira conjunta. Destacam-se então as ligações de *bowen, r.* à *palis, j.*, *mane, r.* e *ruelle, d.* cocitados respectivamente em 12, 12 e 8 listas de referências, de *newhouse, se.* à *mane, r.* e *bowen, r.* cocitados (ambos) em 12 listas de referências, e de *palis j.* à *newhouse, se.*, *shub, m.*, *smale, s.*, *viana, m.*, cocitados em 16, 9, 14 e 9 listas de referências, respectivamente.

A partir destas constatações, é possível conceber os primeiros traços da estrutura intelectual da floresta de genealogia acadêmica analisada. Contudo, os autores supracitados estão inseridos somente na rede de orientação de Jacob Palis. Desta maneira, Bowen, Mañe, Newhouse, Shub, Smale, Ruelle, Viana e Palis Junior

compõem conjunto substancial das influências da floresta de genealogia acadêmica analisada, sobretudo, da rede oriunda de Jacob Palis. Estes oito autores também foram cocitados de maneira conjunta nas listas de referências de Enrique Pujals e Julio Lorenzo. É necessário ressaltar que os recitados Ricardo Mañe e Marcelo Viana, são também orientandos da rede de orientação de Jacob.

Robert Edward “Rufus” Bowen (1947 – 1978) foi um matemático conhecido pelos seus estudos relacionados aos sistemas dinâmicos e professor pesquisador da Universidade da Califórnia. Robert ficou conhecido pela fórmula de Bowen relacionada a pressão topológica e a dimensão de Hausdorff (IOMMI, 2008).

Já Michael Shub (1943) foi também pesquisador da Universidade da Califórnia, da Universidade de Brandeis, do *Queens College da City University of New York*, da *Universidad de Buenos Aires* e da *Graduate Center da City University of New York*. Michael direciona seus estudos aos sistemas dinâmicos e a teoria dos números complexos. O pesquisador é, também, membro da *American Mathematical Association*, da *American Association for the Advancement of Science*, da *New York Academy of Science* e do *Fields Institute*²³.

Com relação a Sheldon Earle Newhouse (1942), o pesquisador matemático estadunidense é membro da acadêmica brasileira de ciências e, também, possui foco de seus estudos aos sistemas dinâmicos. Sheldon atualmente é professor emérito da Universidade de Michigan.

O outro pesquisador recitado, David Ruelle²⁴ (1935), é um físico-matemático nascido na Bélgica e naturalizado francês. Atualmente, é professor emérito do *Institut des Hautes Études Scientifiques Université Paris-Saclay*. David possui estreita afinidade com estudos relacionados a física-matemática e principalmente a mecânica estatística, sendo amplamente premiado por suas contribuições ao campo, como o Prêmio Dannie Heineman de Física Matemática (1985), a Medalha Boltzmann (1986), o Prêmio Holweck (1993), a Medalha Matteucci (2004), o Prêmio Henri Poincaré (2006) e a Medalha Max Planck (2014).

Os orientandos de Jacob Palis, Ricardo Mañe (1948 – 1995) e Marcelo Viana (1962), reconhecidamente por suas pesquisas também relacionados aos sistemas dinâmicos e pela continuidade da temática de Jacob ganham destaque entre as influências da rede analisada visto que foram pesquisadores do IMPA. Ricardo foi um

²³ Disponível em: <https://www.cuny.cuny.edu/profiles/michael-shub>

²⁴ CV disponível em: <https://www.ihes.fr/~ruelle/index.html>

dos primeiros orientandos de Jacob e após sua morte prematura aos 47 anos, Marcelo Viana deu continuidade a seus estudos²⁵ fomentando a corrente científica de Jacob.

Último dos oito nomes a ser destacado, Stephen Smale (1930) é um matemático mundialmente conhecido pela prova da *conjectura de Poincaré* para dimensões maiores ou iguais a cinco, pelas análises relacionadas à teoria matemática do caos e pela formulação da *ferradura de Smale*. Além disso, Stephen Smale, em reconhecimento aos seus notórios trabalhos e contribuições para Matemática, recebeu, em 1966, a Medalha Fields, maior honraria da área da Matemática.

Contudo, a presença de Stephen Smale na rede analisada não está relacionada somente a sua importante contribuição à área. O pesquisador foi orientador²⁶ de Robert Bowen, Sheldon Newhouse, Michael Shub, e Jacob Palis e, conseqüentemente avô acadêmico de Ricardo Mañe e Marcelo Viana. Assim, a origem e a influência dos estudos relacionados a sistemas dinâmicos estão em pleno desenvolvimento graças a influências passadas do ascendente direto da rede de Jacob. Os pesquisadores Robert Bowen, Sheldon Newhouse e Michael Shub podem ser entendidos como irmãos acadêmicos do pesquisador bolsista PQ-Sr.

Ao analisar as outras redes a fim de encontrar os maiores influentes destacam-se também as conexões estabelecidas entre os autores recitados pelos descendentes do bolsista Alcides Lins Neto. Foi observada as conexões entre *camacho, c.* e *brunella, m.*, entre *camacho, c.* e *cerveau, d.*, e *camacho, c.* e *gomezmont, x.* cocitados em 5, 4 e 6 listas de referências, respectivamente. Ainda, tem-se as ligações entre *gomezmont, x.* e *brunella, m.*, e *gomezmont, x.* e *cerveaud, d.* cocitados em 4 listas de referências em ambas as conexões. E mais, foi observada que o próprio orientador (*neto, al.*) é cocitado à *brunella, m.*, *camacho, c.* e *cerveau, d.* em 4, 6 e 5 listas de referências respectivamente.

Estes cinco autores recitados sobressalentes (Cerveau, Gomez-Mont, Brunella, Camacho e Lins Neto) representam os autores acopladores da cadeia de orientação de Alcides Lins Neto e podem ser considerados como parte principal da identidade científica da rede advinda do bolsista PQ-Sr. Assim como na rede de Jacob, o orientador Alcides também compõe a identidade científica da rede.

Dentre estes pesquisadores, Dominique Cerveau é professor da Universidade de Rennes e membro do *Institut universitaire de France*. O pesquisador é especialista

²⁵ Disponível em: <https://piaui.folha.uol.com.br/materia/medalha-de-origem-controlada/>

²⁶ <https://www.mathgenealogy.org/>

em equações diferenciais no campo complexo e na teoria das folheações holomórficas. Estes estudos também são temáticas de interesse de pesquisa de Alcides Neto, corroborando a presença do autor enquanto parte da identidade científica da rede. Em complemento, Cerveau é coautor de diversos trabalhos do bolsista PQ.

Já Xavier Gomez-Mont (1951) é um matemático mexicano e professor da Faculdade de Ciências da Universidade Nacional Autônoma do México (UNAM). Xavier, assim como Alcides Neto e Cerveau, também se interessa pelas folheações holomórficas e dos sistemas dinâmicos. Similar ao relacionamento com Cerveau, Alcides Neto também possui documentos produzidos em coautoria com Gomez-Mont.

Os já destacados, Marco Brunella e César Camacho, por estarem presente em mais de uma rede de orientação, também fazem parte dos principais influentes da rede de orientações de Alcides. Além de ambos possuírem afinidade às folheações, Camacho foi orientador de Alcides, confirmando uma herança intelectual de proximidades teóricas entre orientador e orientando.

Dos oito orientandos analisados de Alcides, os cinco autores destacados (Cerveau, Gomez-Mont, Alcides, Brunella e Camacho) estão presentes, respectivamente, em cinco, seis, sete, cinco e sete listas de referências. De maneira conjunta, estes autores estão nas listas de referências de autores recitados dos orientandos Arturo Perez e José Omegar-Calvo. Assim, em ambos os casos os dois orientandos recitaram estes cinco autores. Ainda, Alcides Lins Neto só não é recitado por um orientando: Leonardo Camara.

A principal conexão, por meio dos autores recitados, entre as redes de influentes Alcides e de Jacob (redes em verde e azul) é estabelecida por Vladimir Igorevich Arnold. Fato a ser ressaltado é que Cesar Camacho, que orientou Alcides Lins Neto, também foi orientando de Stephen Smale, mesmo orientador de Jacob. Desse modo, Cesar e Jacob são irmãos acadêmicos e em algum momento de suas carreiras seguiram vertentes de pesquisas diferentes, mas fazem parte de uma mesma linhagem de estudos relacionados aos sistemas dinâmicos, influenciada por Stephen Smale. Desta maneira, Jacob Palis é dito tio acadêmico de Alcides Lins Neto (Alcides Lins Neto é sobrinho acadêmico de Jacob Palis).

Ao observar as influências do bolsista PQ-Sr. Paulo Sad sobre seu orientando Gabriel Calsamiglia, pode-se apontar que os autores recitados que os acoplaram foram Brunella e Camacho, evidenciando uma ramificação a partir de Jacob em

direção as folheações holomórficas, similar a pesquisas advindas de Alcides. Tal fato é corroborado ao constatar que Alcides Neto, Paulo Sad e César Camacho são coautores de alguns estudos, como destacado anteriormente.

Do ponto de vista genealógico, Paulo Sad é filho acadêmico de Jacob Palis, e Gabriel Calsamiglia é neto acadêmico de Jacob Palis. E mais, Paulo Sad é neto acadêmico de Stephen Smale e sobrinho acadêmico de César Camacho.

Ao adentrar na análise das influências dos autores recitados sobre a rede de orientação de Aloísio Araújo, três pesquisadores sobressaem-se: *duffie, d.*, *lucas, re.*, *araujo, ap*, este último, o próprio orientador. Araujo é cocitado com *duffie, d.* e *lucas, re.* em duas listas de referências cada. De maneira similar, *duffie, d.* e *lucas, re.* também são cocitados em duas de referências. Araújo, *duffie, d.* e *lucas, re.* estão presentes em 4, 5 e 5 listas, dos 11 orientandos que se acoplaram com Aloísio (além de estar presente na própria lista do orientador). De maneira conjunta, dentre os orientandos de Aloísio, os três autores somente aparecem em uma lista de referências de autores recitados, de Elvio Gamba.

Aloísio Araújo é um famoso pesquisador, que além de bolsista PQ-Sr da área de matemática, apresenta vínculo com o IMPA e com a Fundação Getúlio Vargas. Graduado em Economia e Estatística, Araújo é conhecido por dedicar seus estudos a economia matemática e por ser um dos precursores da temática no Brasil e no IMPA. Em meados da década de 1980, o pesquisador inaugurou a primeira turma relacionada a temática lecionou para alunos de renome como Sérgio Werlang, Carlos Ivan Simonsen Leal e Ricardo Paes de Barros, grandes expoentes do cenário econômico brasileira (IMPA, 2020).

A relação de Aloísio com a Matemática está diretamente relacionada com o IMPA, local onde o pesquisador se tornou mestre, sob orientação de Djairo Figueiredo. Foi durante este período (1966 – 1969) que o pesquisador percebeu que poderia contribuir muito a ciência brasileira. Em relato ao portal eletrônico do IMPA (2020), Aloísio aponta: “Quando comecei a cursar o IMPA no mestrado, em 1966, reconheci esta causa como parte desse projeto científico brasileiro”.

Com relação aos outros dois pesquisadores mais influentes nesta rede, James Darrell Duffie (1954) é economista e professor da *Stanford Graduate School of Business* concentrando seus estudos a teoria dinâmica dos preços de ativos, equilíbrios de mercado, modelagem matemática e econometria. Duffie é graduado em Engenharia Civil, mestre em Economia Estatística e Doutor em Engenharia

Econômica²⁷. Tanto com relação a formação quanto a atuação de pesquisa o autor recitado se alinha a Aloísio Araújo.

Já Robert Emerson Lucas Junior²⁸ (1937) é professor da Universidade de Chicago e tornou-se mundialmente reconhecido por ser ganhador do prêmio Nobel de Economia em 1995 por ter desenvolvido e aplicado a *hipótese de expectativas racionais*. Robert possui afinidade com a econometria e a macroeconomia e é considerado por Aloísio um clássico da área da economia matemática que em suas palavras ao IMPA (2020) aponta: “Fui convidado para dar cursos de economia matemática na Universidade de Chicago, onde aprendi economia mais aplicada com os novos clássicos, como Robert Lucas Jr [...]”.

Dessa maneira, é possível constatar que a linhagem científica da Aloísio para com seus descendentes está mais relacionada a economia matemática do que a teoria dos sistemas dinâmicos ou mais especificamente as folheações holomórficas, como os demais bolsistas PQ-Sr analisados, justificando a subrede, em vermelho, na Figura 19, não se conectar as demais redes.

Contudo, Aloísio aponta influências diretas de Jacob Palis em sua carreira ao dizer que “Tive contato com pesquisadores como Leopoldo Nachbin, Elon Lages Lima, Jacob Palis e Maurício Peixoto que me inspiraram profundamente” (IMPA, 2020).

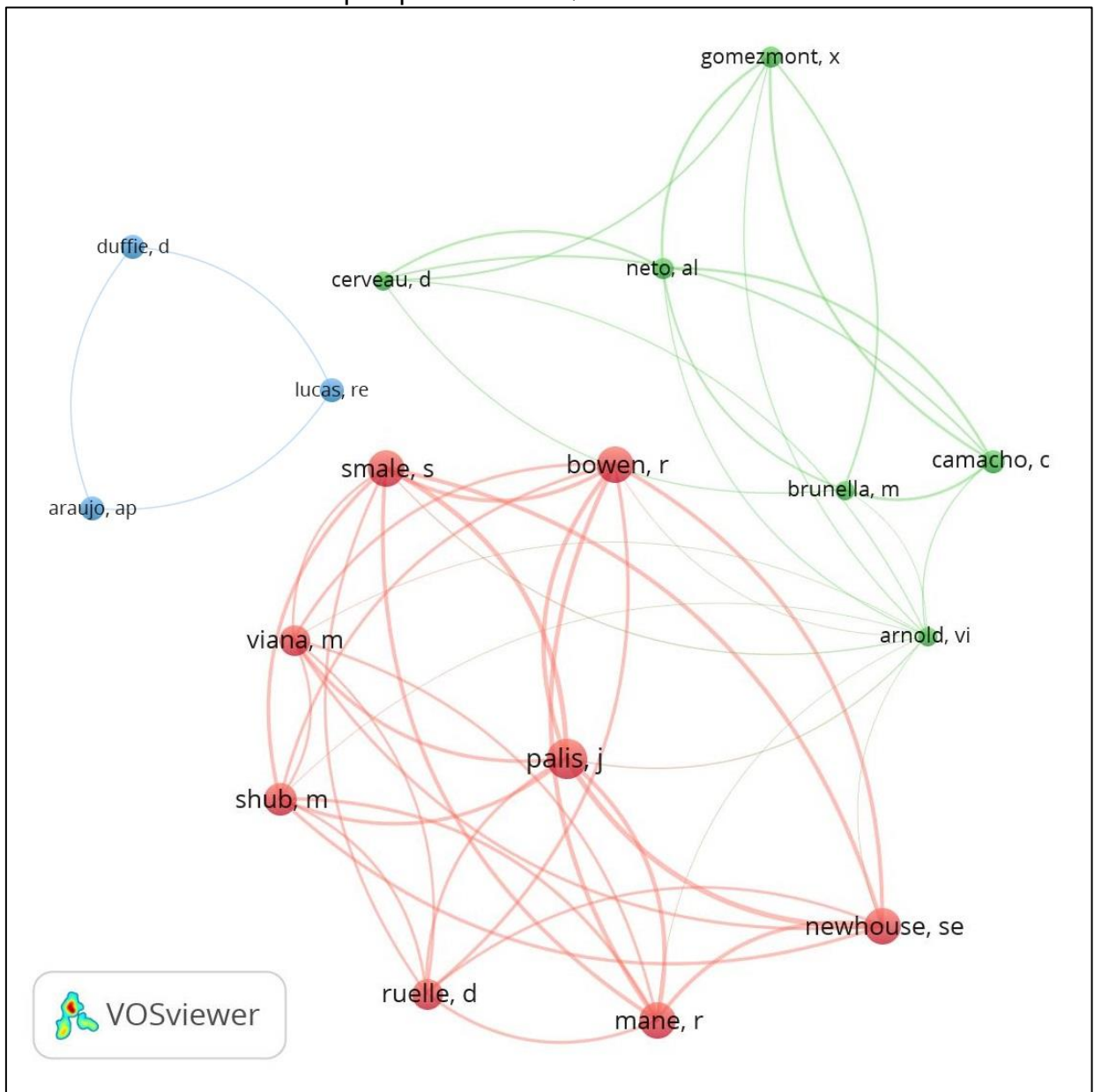
Já o bolsista PQ-Sr Luiz Velho, não mencionado, não se acoplou a seus orientandos que atenderam ao critério de autoria para recuperação de documentos na WoS. Entretanto, em suas temáticas de estudos, Luiz Velho destoa da formação e das temáticas dos demais bolsistas, visto que o pesquisador possui formação voltada a Informática, a Computação Gráfica e a Ciências da Computação, sendo expoente em pesquisas relacionadas a processamento digital de imagens, visualização e modelagem geométrica. Ele é o primeiro bolsista PQ-Sr da área de *Visual Computing* do IMPA e seu contato com o instituto está relacionado ao seu estágio de pós-doutorado após período de doutoramento em 1994.

Elencados os principais influentes da floresta analisada, construiu-se uma nova rede, somente com estes atores, compondo, na Figura 20, a rede da identidade científica segundo os padrões de recitação diacrônica de autores acopladores dos pesquisadores analisados.

²⁷ CV disponível em: <https://web.stanford.edu/~duffie/resume.pdf>

²⁸ Disponível em: <https://www.britannica.com/biography/Robert-E-Lucas-Jr>

Figura 20 – Rede de identidade científica da floresta de genealogia acadêmica de pesquisadores PQ-Sr do IMPA



Fonte: Elaboração do autor via software *VosViewer*

Dos pesquisadores orientadores bolsistas PQ-Sr do IMPA, Jacob Palis, Alcides Neto e Aloísio Araújo compõem a identidade exposta na Figura 20. É importante ressaltar novamente que Paulo Sad se acoplou, enquanto autor recitado, a seu orientador, e é filho acadêmico de Jacob. Paulo e seu orientando foram acoplados por Brunella e Camacho, autores influentes sobre a rede.

Além disso, pelo não acoplamento, por meio de autores recitados em comum, entre Luiz Velho e seus orientandos, não é possível inferir se Luiz é ou não influência sobre seus descendentes, como os demais orientadores destacados.

Nesse contexto, pode-se notar que a floresta de genealogia acadêmica analisada apresenta estreita relação com os sistemas dinâmicos, as folheações holomórficas e a economia matemática. A maior afinidade relacionada aos sistemas dinâmicos é advinda de Jacob Palis, às folheações holomórficas de Alcides Neto e Paulo Sad, à economia matemática de Aloísio Araújo, e, a *Visual Computing* de Luiz Velho.

Com relação aos orientandos Marcelo Viana e Ricardo Mañe, ambos ganham destaque por serem parte da identidade científica da floresta analisada, tanto dos orientadores, quanto dos orientandos em geral, principalmente ao observar a rede proveniente de Jacob.

Ainda, há grande influência de Vladimir Igorevich Arnold e Marco Brunella, que exercem um intermédio de conexão entre as redes, e dos ascendentes genealógicos César Camacho e Stephen Smale. A influência permanente de Jacob sobre o IMPA e a história do desenvolvimento e consolidação internacional do Instituto se entrelaçam e justificam a importância do pesquisador sobre toda floresta genealógica analisada.

Em resumo, a análise das influências dos pesquisadores PQ-Sr do IMPA perante seus descendentes configura-se, do ponto de vista da genealogia científica, evoluções (científicas) relacionadas aos sistemas dinâmicos, equações diferenciais, folheações holomórficas e economia matemática, e, sob a ótica da genealogia intelectual, representa uma escola de pensamento influenciada por Stephen Smale, Jacob Palis, Vladimir Igorevich Arnold, César Camacho, Marco Brunella e Robert Lucas. Poderia ser dizer que o IMPA conta com pesquisadores, sobretudo, *smaleanos*, *jacobianos* e *robertianos*. Dado o fato que esta pesquisa parte das relações de orientações acadêmicas, tem-se, portanto, uma sobreposição entre GA, GC e GI.

8 CONSIDERAÇÕES FINAIS

Esta pesquisa apresentou uma análise do desempenho da nova métrica denominada Coeficiente de Acoplamento Bibliográfico Genealógico (CABG)²⁹ na floresta de genealogia acadêmica dos bolsistas PQ-Sr da área da Matemática vinculados ao IMPA com intuito de mostrar sua pertinência como indicador representativo da conservação da corrente teórico-metodológica na linha sucessória da relação de orientação acadêmica. Baseado no conceito de acoplamento bibliográfico entre autores, o coeficiente busca, assim, avaliar a intensidade da transmissão da herança teórica entre orientador e orientando.

Para prova de conceito desta tese, foram selecionados os pesquisadores sêniores do IMPA e seus orientandos com produção científica que atendesse o critério de autoria em que o orientador deveria ser primeiro autor e seu orientando a ser acoplado não deveria ser coautor, e, reciprocamente, o orientando fosse o primeiro autor e seu orientador não fosse coautor da produção. A análise contou com 5 orientadores bolsistas PQ-Sr e 62 orientandos.

Esta pesquisa desdobrou-se visando os objetivos propostos e todo o estudo foi pautado em referencial teórico baseado nos estudos de genealogia da ciência e nas análises relacionais de citação, mais especificamente, no acoplamento bibliográfico. Os estudos genealógicos foram apresentados a fim de embasar não só o reconhecimento do contexto e a temática da pesquisa como um todo, mas também para subsidiar a formulação do CABG, visto que, foi possível identificar aspectos relacionais (de citação) associados aos estudos genealógicos da ciência, independentemente de sua classificação: acadêmica, científica ou intelectual.

A fundamentação teórica em torno do acoplamento bibliográfico, tanto do ponto de vista científico-social de proximidades entre autores, quanto do ponto de vista matemático de intersecção (e comparação) de conjuntos, foi percorrida com a finalidade de resgatar todos os conceitos envolvidos em sua formulação e seu potencial adaptativo a novas reformulações.

O acoplamento bibliográfico enquanto genuína medida de similaridade entre dois conjuntos foi apresentado como métrica aplicada à diferentes unidades de análise

²⁹ A proposta desta tese foi, previamente, submetida ao fórum doutoral do 8º EBBC (Encontro Brasileiro de Bibliometria e Cientometria) e ao VI FDECIN (Fórum de discentes e egressos do Programa de Pós-Graduação em Ciência da Informação da UNESP (campus Marília/SP).

como autores e/ou documentos e foram sugeridas 19 diferentes formas de adaptação, sobre medidas de similaridade, ao ABA (Quadro 2).

Toda formulação do ABA, bem como todas as sugestões, servem como importante registro e aporte para futuras pesquisas que busquem a elaboração e/ou reformulação de métricas baseado no acoplamento bibliográfico. Esta conceitualização forneceu subsídios científicos e matemáticos para elaboração do Coeficiente de Acoplamento Bibliográfico Genealógico (CABG).

O percurso metodológico originou-se dos objetivos propostos e, de maneira geral buscou expor, validar e propor um coeficiente de perspectiva relacional de citação e recitação (CABG), para mensurar a intensidade das proximidades teórico-metodológicas entre pesquisadores no escopo das relações de genealogia acadêmica no campo da Matemática no Brasil.

Para isso, fundamentou-se o Coeficiente de Acoplamento Bibliográfico Genealógico (objetivo específico - i) tanto do ponto de vista matemático quanto do ponto de vista teórico, embasando-o na principal ideia da GA: a herança intelectual de orientador para orientando. O novo coeficiente foi elaborado com base na teoria advinda dos estudos genealógicos com aporte da teoria dos conjuntos oriundos da Matemática aplicada às análises relacionais de citação.

Entende-se que a formulação do CABG enquanto medida de similaridade é confiável, válida e alternativa as medidas comparadas (Índice de Jaccard e Cosseno de Salton) visto que o coeficiente é uma medida dinâmica, capaz de apontar uma similaridade perfeita (igual a 1 ou 100%) entre dois conjuntos (pesquisadores) mesmo que esses conjuntos não sejam idênticos (diferentes cardinalidades das listas de referentes). Tal fato une a conceitualização matemática de pré-fixar o valor de relativização da intersecção entre os conjuntos pelo tamanho da lista do orientador junto à ideia de herança intelectual oriunda da genealogia acadêmica. Caso o orientando analisado herde toda a lista de referencial do orientador, mesmo que eles tenham listas de tamanhos diferentes, terão similaridade perfeita igual 1 ou 100%. As outras medidas comparadas não admitem tal situação, pois, só atingem valor máximo se os elementos que compõem os conjuntos forem idênticos (mesmos elementos e mesma cardinalidade).

O CABG emerge a partir da inspiração dos aspectos relacionais de citação provenientes das análises genealógicas em dois sentidos: i) o acoplamento bibliográfico por si só estabelece relação entre duas partes citantes (orientador e

orientando); ii) o CABG utiliza-se dos padrões de recitação diacrônica de autores. O primeiro item diz respeito à proximidade entre orientador e orientando e seu valor representa essencialmente a herança teórica herdada pelo orientando. Já o segundo item torna o coeficiente uma medida dinâmica em que seu valor poderá aumentar ou diminuir ao longo do tempo, a depender da continuidade de produção do orientador e do orientando e da recitação de novos autores.

Concebido o CABG, esta tese comparou o coeficiente à outras duas medidas citadas (objetivo específico – ii), visto que estas medidas são amplamente conhecidas, estabelecidas e utilizadas no campo de Ciência da Informação. Para a comparação, foram correlacionadas as três medidas, por meio do Coeficiente de Correlação Intraclasse (ICC) e posteriormente, aplicou-se a metodologia de Bland-Altman.

Previamente a estas comparações, foram apresentados todos os valores de acoplamento bibliográfico segundo os padrões de citação (citados ao menos em um documento) e recitação (citados em dois ou mais documentos) para as medidas comparadas: CABG, Índice de Jaccard e Cosseno de Salton.

O ICC apontou uma intensa correlação entre as medidas. Pelo fato de estas medidas possuírem relevância na literatura especializada, a correlação entre o CABG e elas confirmaram que o coeficiente está calibrado com as medidas e apresenta-se como medida objetiva, confiável e consistente para análises genealógicas. Contudo, isso não revelou concordância entre as medidas. Para análise de concordância, foi utilizado a método de Bland-Altman.

O método aplica o teste T de *Student* pareado entre as medidas a serem comparadas e ilustra a diferença entre os métodos por meio do gráfico de Bland-Altman (Figuras 12 e 13). Ao comparar o CABG ao Índice de Jaccard e ao Cosseno de Salton, o teste apontou diferenças significativas entre as medidas, ou seja, a diferença entre as medidas é estatisticamente diferente de zero e, portanto, não concordam entre si. A forte correlação seguida de não concordância aponta caminhos para o CABG se tornar uma medida alternativa às demais, dado que a formulação do coeficiente, pautada nos estudos genealógicos e nos padrões relacionais de recitação de autores embarca conceitos que as demais métricas não englobam.

A não concordância aliada a uma correlação entre medidas revela uma alternativa consistente à medida testada, justamente porque se duas medidas são estatisticamente concordantes entre si, não faz diferença entre usar uma ou outra. Nesse sentido, é necessário apontar que o Índice de Jaccard e o Cosseno de Salton

também não concordaram entre si e são amplamente difundidas no âmbito dos EMI. Reafirma-se que o principal mérito do CABG é permitir que a similaridade entre dois pesquisadores, segundo autores recitados em comum, seja perfeita, mesmo que a quantidade de autores recitados em comum seja diferente. Para isso, basta o orientando herdar a lista de autores recitados de seu orientando que terá 100% de similaridade com ele. Com isso, embora alguns domínios apresentem características diferentes em termos de correlação e concordância aos presentes nesta tese, o CABG ainda apresenta especificidades e é diferente, por definição, tanto do Índice de Jaccard quanto do Cosseno de Salton, como demonstrado nas Equações 20 a 28.

Testado o CABG, esta pesquisa adentrou na construção da floresta de genealogia acadêmica (objetivo específico – iii) e da avaliação da conexão entre os atores da floresta de GA e dos principais influentes desta floresta (objetivo específico – iv).

A construção da floresta de genealogia acadêmica apresentada contou com três versões, em que para cada versão, as arestas das redes foram valoradas cada uma com uma perspectiva de similaridade diferente baseados nos padrões de recitação (CABG, Índice de Jaccard e Cosseno de Salton). Foram ainda identificados nas florestas cada bolsista PQ-Sr, bolsistas PQ em geral e os egressos de doutorado.

As três construções serviram para a análise topológica da rede baseada na força de conectividade entre orientadores e orientandos. Foi constatado uma maior proximidade, segundo as medidas de similaridade, entre Alcides Neto e seus orientandos. Com relação à média, em termos de valores absolutos de autores citados e recitados em comum, Jacob Palis foi o orientador com força de conectividade média mais expressiva. O cálculo da força de conectividade sobre estas florestas de genealogia acadêmica é importante pois trata-se também, de redes de acoplamento bibliográfico em que cada arestas corresponde a um acoplamento calculado.

O computo da força de conectividade evidenciou, sob a ótica topológica da rede, que os valores de conectividade do CABG, Índice de Jaccard e Cosseno de Salton na floresta de GA analisada são substancialmente diferentes entre si (10,8%, 7,6% e 16,4%, respectivamente), em que o CABG se apresenta como “valor intermediário” entre as medidas. Dada sua formulação caracterizada pelo valor pré-fixado em seu denominador ($L1$ – cardinalidade da lista de autores recitados pelo orientador analisado), o CABG é possivelmente mais representativo do que as demais, pois busca quantificar quanto da identidade científica de seu orientador

(autores recitados em comum) um determinado orientando herdou. Neste caso, o valor médio desta herança é de 10,8%, isto é, cada orientando analisado herdou, em média, 10,8% das listas de autores recitados de seu orientador.

Por se tratar, fundamentalmente, de uma rede de acoplamento bibliográfico, os cálculos de acoplamento revelaram os autores acopladores (unidades de acoplamento) responsáveis pelas intersecções calculadas e a partir destes autores foi possível avaliar a conexão entre os principais influentes da floresta de genealogia acadêmica, gerando a estrutura intelectual dos pesquisadores que a compõe (objetivo específico – iv).

Após a construção das florestas de GA, foram identificados (Quadro 3) 132 autores acopladores recitados nas obras dos 66 pesquisadores analisados. Em suma, estes 132 autores representam a identidade científica de todo domínio analisado. Contudo, para refinar esta análise, foram destacados os autores que mais se destacaram em cada rede de genealogia específica e aqueles que permearam mais de uma rede, como os pesquisadores recitados César Camacho, Marco Brunella e Igorevich Arnold.

Após esta constatação, construiu-se a rede de citação de autores entre todos os orientandos analisados e os 132 autores acopladores. Esta rede explicitou a relação entre orientandos citantes e autores recitados. A rede de recitação apontou destaques aos orientandos de Jacob Palis por mais se conectarem, em valores absolutos, ao orientador: Carlos Morales (22 autores recitados em comum), Eleonora Castigeras (23 autores recitados em comum), Enrique Pujals (26 autores recitados em comum), Julio Lorenzo (23 autores recitados em comum), e, Stefano Luzzatto (21 autores recitados em comum). No mesmo sentido, tem-se os orientandos de Aloísio Araújo, Luciano Irineu e Paulo Monteiro, ambos recitando 22 autores em comum com o orientador.

Já com relação à análise das principais influências da floresta de genealogia acadêmica, foi construída a rede de cocitação entre todos os 132 autores acopladores recitados a fim de elaborar a estrutura de influências intelectuais de toda floresta de genealogia acadêmica analisada. Esta rede apresentou autores de grande influência sobre cada uma das cinco redes de genealogia acadêmica que compõe a floresta divididos em três grupos: as influências de Jacob Palis e Paulo Sad, as influências de Alcides Neto e as influências de Aloisio Araújo.

Foram identificados como os principais influentes sobre a rede dos bolsistas PQ-Sr: i) Jacob Palis Junior: Bowen, Mañe, Newhouse, Shub, Smale, Ruelle, Viana e Jacob Palis Junior; ii) Paulo Sad: Camacho e Brunella; iii) Alcides Lins Neto: Camacho, Brunella, Gomez-Mont, Cerveau, Alcides Neto; iv) Aloísio Pessoa Araújo: Duffie, Lucas e Araújo. Dado o fato de Luiz Velho não se acoplar, segundo autores recitados, a seus orientandos, não foi possível identificar as principais influências do orientador perante seus dois orientandos analisados. Ressalta-se que Luiz Velho se acoplou com seus orientandos caso sejam considerados os autores citados somente em um documento e não os recitados

Estes pesquisadores, juntamente com Vladimir Igorevich Arnold, compuseram uma nova rede de cocitação, apenas com os autores destacados anteriormente. Estes 17 atores compõem a essência da identidade científica dos pesquisadores bolsistas PQ-Sr. do IMPA perante seus orientandos.

Ao remeter ao problema proposto nesta tese, é possível afirmar e aceitar a hipótese proposta de que há um coeficiente de acoplamento derivado do coeficiente normalizado de Acoplamento Bibliográfico de Autores que mensura de forma representativa da intensidade de transmissão da corrente teórico-metodológica e a manutenção das correntes científica/intelectuais em grupos de pesquisadores com relação de genealogia acadêmica do campo da Matemática. A aceitação da hipótese se reforça pela dinamicidade do coeficiente proposto e pela capacidade de estimar o quanto da identidade científica de um orientando pode herdar de seu orientador.

Diferente das demais medidas comparadas (Índice de Jaccard e Cosseno de Salton), em que ambas irão servir para mensurar o quão similar dois conjuntos de referências são, e irão alcançar seu valor máximo somente se ambas listas forem exatamente iguais, excluindo a possibilidade do elemento central dos estudos de GA: a herança intelectual perpassada de orientador para orientando.

Neste ponto, O CABG, ao permitir que duas listas de referências de tamanhos diferentes atinjam valor máximo (acoplamento perfeito; igual a 1 ou 100%), alinha-se à ideia basilar das análises de genealogia acadêmica. Para isso, basta o orientando herdar a lista de autores recitados de seu orientando, e com isso, orientador e orientando estarão 100% acoplados pela identidade científica (lista de autores recitados) do orientador, ou seja, o orientando herdou 100% a identidade de seu orientador.

Assim, conjectura-se que o CABG, pela sua formulação matemática aliada ao contexto da genealogia acadêmica, é uma métrica aplicável a qualquer domínio científico compostos por pesquisadores com relação de orientação acadêmica.

Com isso, é possível resgatar as cinco classificações dos tipos de análises genealógicas de Mena-Chalco (2016), presentes na seção 2, e inferir que esta tese realiza um estudo, no âmbito da genealogia acadêmica, classificada como genealogia analítica, ao dar ênfase na aplicação de métodos matemáticos (formulação do CABG e comparação à outras medidas) e computacionais (aparato computacional e implementação do *The Genealogical Coupler*) para criar e estudar as relações de orientação acadêmica.

Ademais, o CABG configura-se como uma forma de normalização assimétrica de acoplamento bibliográfico, adequável a diferentes situações e que pode ser extrapolada aos diferentes contextos, independente de relações genealógicas ou não, em que se almeja analisar a herança de itens citados por determinada unidade de análise por outra. Por exemplo, nos estudos pioneiros de Kessler (1963), ao propor a forma de agrupamento em que são analisados quantos documentos se acoplam a um documento específico (P_0), o CABG pode ser aplicado para averiguar qual herança de referências de P_0 aos documentos acoplados a ele, com a lista de referências de P_0 tratada como $L1$ no contexto do CABG. Outra aplicação do CABG seria em análises de coautoria, em que desejasse mensurar a quantidade de documentos em coautoria que um grupo de pesquisadores compartilham com um pesquisador fixo (focal). Neste caso, $L1$ representaria o número total de artigos deste autor focal e $L2$ a quantidade de artigos produzidos pelo pesquisador a ser comparado.

É possível, ainda, tratar a lista de referentes citados por um determinado pesquisador, instituição, grupo de pesquisa ou unidade de análise qualquer, como um item focal, assumir essa lista como $L1$ e acoplá-la a diversos outros itens, sob a perspectiva assimétrica de heranças perspassadas de uma unidade para outra. Neste caso, esta lista poderia ser tratada com uma espécie de “grupo controle” de referências citadas. Estas sugestões preservam a ideia básica de genealogia (enquanto heranças) em suas aplicações.

É importante destacar, também, que esta pesquisa difere daquelas realizadas em Hu (2022), Galvagno e Pisano (2020), Sun (2011) e Sun e Zhu (2012), encontradas na WoS e Scopus e citadas na seção 1 (Introdução), visto que nestes três trabalhos se dá foco à identificação de genealogias de documentos.

Em Hu (2022), o acoplamento bibliográfico é utilizado enquanto método para a análise da literatura (artigos) analisada a respeito de *digital pathology* e o conceito de genealogia é adotado para a classificação de toda obra recuperada em artigos *upstream*, *midstream*, *downstream*, em que os artigos *upstream* são aqueles de base teórica sendo os primeiros trabalhos no campo de pesquisa, *midstream* seriam os artigos mais importantes (de destaque) da área de pesquisa e *downstream* os artigos publicados recentemente na área de pesquisa com tópicos de pesquisa mais recentes.

No segundo documento, os autores analisaram quais artigos são vistos como mais influentes sobre a literatura relacionada a *Family Business Internationalization (IFB)* e utilizam o conceito de genealogia para identificar os documentos antepassados e descendentes sobre IFB. Neste estudo, o acoplamento bibliográfico foi aplicado a documentos (perspectiva documental de Kesller) com intuito de analisar a proximidade entre os documentos descendentes identificados.

Já os outros dois documentos de Sun (2011) e Sun e Zhu (2012), os autores introduzem o conceito de genealogia genética de citações bem como os termos gene e genoma documental. Os autores definem sobre a perspectiva genética seis relações de citação, a saber: i) referência: documento pai; ii) citação: documento filho; iii) referência da referência: documento avô; iv) citação da citação: documento neto; v) documentos bibliograficamente acoplados: documentos irmãos; vi) documentos cocitados: documento cônjuge.

Em resumo, estes documentos utilizam o acoplamento bibliográfico entre documentos enquanto método bibliométrico e se apossam do conceito de genealogia para descrever origens e gerações de artigos.

Como dificuldades e limitações encontradas durante esta pesquisa, têm-se as operacionalidades, muitas vezes realizadas de maneira manual, como a construção das listas de autores recitados (Figuras 5, 6 e 7), a extração da produção científica indexada na base de dados WoS autor a autor segundo os critérios de autoria, e a construção das redes via Pajek (organização dos arquivos a serem processados). Especificamente como relação a este último, a construção de redes genealógicas valoradas não é tão corriqueira quanto outras representações em grafos, tornando os recursos para isso não muito difundidos e pouco acessíveis. Foram encontradas alternativas em algumas bibliotecas da linguagem de programação R, como

visNetwork e *networkD3*. Todavia, o ajuste manual para representações de redes de GA via Pajek pode ser considerado o mais simples.

Já como trabalhos futuros, pretende-se analisar outros para que o CABG seja colocado a prova em outros contextos além de ser comparado com outras medidas como aquelas apresentadas no Quadro 2. Outro anseio é a implementação computacional da extração referências oriundas de bases como Web Of Science, Scopus, Dimensions e OpenAlex, de maneira automática similar ao que ocorre em softwares como *VosViewer* ou *Bibliometrix*, porém, organizadas em colunas tabuladas.

Mesmo com a implementação das ferramentas *The Coupler* e *The Genealogical Coupler*, a integração total entre exportação de registros, extração de referências, visualização de redes juntamente com a valoração de arestas, atribuiria maior agilidade a processos envolvidos em análises bibliométricas e cientométricas. Ainda, como relação ao *The Genealogical Coupler*, pretende-se manter a ferramenta sempre atualizada e livre de eventuais problemas como falhas no código ou problemas de hospedagem. Anseia-se implementar a análise de redes de GA ponderadas por métricas diversas (presentes no Quadro 2) contendo várias gerações de orientação acadêmica. A construção de novos insumos computacionais é de extrema utilidade para os Estudos Métricos da Informação, uma vez que favorece novas análises em contextos diversos.

Assim, esta pesquisa contribui, não só com a formulação de uma nova métrica baseada no ABA respaldada nos estudos genealógicos, mas também com uma ferramenta para este tipo de análise, que leva em conta a valoração das arestas das redes de GA e não somente a difusão das redes, como geralmente é apresentado.

Além disso, realizou um resgate histórico das influências intelectuais, científicas e acadêmicas perpassadas ao longo dos anos no IMPA representado pela estrutura intelectual identificada, com destaque aos autores acopladores. A utilização da análise cocitação entre os autores acopladores auxiliou no entendimento da desta estrutura intelectual.

Por fim, espera-se que esta pesquisa contribua para o progresso do campo de Ciência da Informação, mais especificamente dos EMI, e em conjunto com a Matemática e Estatística, componham forte campo multi/interdisciplinar de atuação de bibliometristas e cientometristas no desenvolvimento de análises de cunho relacional de citação aplicadas a redes de genealogia acadêmica (e sociais como um todo), à

construção de novas métricas e indicadores científicos, além de consistentes formulações e debates bibliométricos que aliem o contexto sociológico da ciência à rigidez matemática.

REFERÊNCIAS

AJIFERUKE, I.; LU, K.; WOLFRAM, D. A comparison of citer and citation-based measure outcomes for multiple disciplines. **Journal of the American Society for Information Science and technology**, v. 61, n. 10, p. 2086-2096, 2010.

ALTMAN, D. G.; BLAND, J. M. Measurement in medicine: the analysis of method comparison studies. **Journal of the Royal Statistical Society: Series D (The Statistician)**, v. 32, n. 3, p. 307-317, 1983.

ANDRAOS, J. Scientific genealogies of physical and mechanistic organic chemists. **Canadian journal of chemistry**, v. 83, n. 9, p. 1400-1414, 2005.

BORNMANN, L. Scientific peer review. **Annual review of information science and technology**, v. 45, n. 1, p. 197-245, 2011.

BRUSCO, M; CRADIT, J. D.; STEINLEY, D. A comparison of 71 binary similarity coefficients: The effect of base rates. **Plos one**, v. 16, n. 4, p. e0247751, 2021.

BUFREM, L. S.; MASCARENHAS, F. S.; SOBRAL, N. V. Análise das influências intelectuais na produção científica da área de Ciência da Informação: um estudo sobre os bolsistas de produtividade em pesquisa (PQ-CNPq). **Em Questão**, v. 23, 2017.

BUFREM, L. S.; PRATES, Y. O saber científico registrado e as práticas de mensuração da informação. **Ciência da Informação**, v. 34, n. 2, 2005.

CAPES. Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. **Conselho Nacional de Desenvolvimento Científico e Tecnológico – Bolsas nos país e no exterior**. 2021. Disponível em: https://www.gov.br/cnpq/pt-br/aceso-a-informacao/bolsas-e-auxilios/copy_of_modalidades

CASTANHA, R. G. **THE GENEALOGICAL COUPLER**. [S. l.]: Zenodo. 2022a. Disponível em: DOI: <https://doi.org/10.5281/zenodo.7618526>

CASTANHA, R. G. The Coupler: uma nova ferramenta bibliométrica para análises relacionais de citação, acoplamento bibliográfico e cocitação. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, v. 20, p. e022029, 2022b. DOI: <https://doi.org/10.20396/rdbci.v20i00.8671208>

CASTANHA, R. C. G.; GRÁCIO, M. C. C. Estudos de genealogia acadêmica como abordagem para análise de domínio. **Organização do conhecimento e diversidade cultural. Sociedade Brasileira de Organização do Conhecimento (ISKO-Brasil). Chap**, p. 108-116, 2015.

CASTANHA, R. G. Acoplamento bibliográfico como proposta metodológica para a mensuração da intensidade das ligações em genealogia acadêmica: a influência teórica de Aldo Barreto na ciência da informação. **Dissertação de Mestrado (Ciência da Informação)** - Universidade Estadual Paulista Campus Marília. 2019.

CASTANHA, R. G.; BUFREM, L. S.; BOCHI, F. ESTUDOS RELACIONAIS DE CITAÇÃO: COCITAÇÃO, ACOPLAMENTO BIBLIOGRÁFICO E GENEALOGIA CIENTÍFICA. In: **Tópicos da Bibliometria para bibliotecas universitárias**, p. 134-162, 2020.

CASTANHA, R. G.; GRÁCIO, M. C. C. Bibliographic Coupling Indicators for the evaluation of theoretical-methodological proximity in academic genealogy networks: a study applied to the PQ's scholarships descendants of Aldo Barreto. **RDBCI: Digital Journal of Library and Information Science**; Vol. 18 (2020): Continuous Publishing; e020039, v. 24, n. 2, 2020.

CICCHETTI, D. V. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. **Psychological assessment**, v. 6, n. 4, p. 284, 1994.

Conselho Nacional de Desenvolvimento Científico e Tecnológico. CNPQ. **Bolsas em curso**. 2022. Disponível em:
http://plsql1.cnpq.br/divulg/RESULTADO_PQ_102003.curso

DAMACENO, R.; ROSSI, L. MENA-CHALCO, JP. Identificação do grafo de genealogia acadêmica de pesquisadores: Uma abordagem baseada na Plataforma Lattes. In: **Proceedings of the 32nd Brazilian Symposium on Databases**. p. 76-87, 2017.

DENISCZWICZ, M.; KERN, V. M. Indicadores de gestão na revisão por pares: confiabilidade da revisão recíproca anônima de propostas de mestrado. **Liinc em Revista**, v. 9, n. 1, 2013.

DIAS, T. M. R.; MOREIRA, T. H. J.; DIAS, P. M. Caracterização e análise das redes de colaboração científica dos bolsistas de produtividade em pesquisa do CNPq. In: **Anais do VII Brazilian Workshop on Social Network Analysis and Mining**. SBC, 2018.

DORES, W. et al. Building the Brazilian academic genealogy tree. In: **International Conference on Theory and Practice of Digital Libraries**. Springer, Cham, 2017. p. 537-543.

EGGHE, L.; LEYDESDORFF, L. The relation between Pearson's correlation coefficient r and Salton's cosine measure. **Journal of the American Society for information Science and Technology**, v. 60, n. 5, p. 1027-1036, 2009.

FERREIRA, L. M. et al. Relação orientador-orientando: o conhecimento multiplicador. **Acta Cir. Brasil.**, São Paulo, v. 24, n. 3, 2009.

FRANCO, N. M. G. **Análise relacional interinstitucional**: redes de coautoria e acoplamento de autores. Dissertação de Mestrado. UFSCAR. 2018.

FREITAS, J. L. et al. El interdominio de los estudios métricos de la información en Iberoamérica y Sudáfrica: análisis en la base SciELO en el período 1978-2013. **Revista Cubana de Información en Ciencias de la Salud**, v. 28, n. 1, 2017.

FUNK, Russell J.; OWEN-SMITH, Jason. A dynamic network measure of technological change. **Management science**, v. 63, n. 3, p. 791-817, 2017.

GABRIEL JUNIOR, R. F.; BUFREM, L. S. Influências teóricas de um pesquisador sênior da ciência da informação: genealogia acadêmica do professor Aldo Barreto. **Encontro Brasileiro de Bibliometria e Cientometria**, v. 6, n. 6, p. 6º. 2018.

GABRIEL JUNIOR, R. F.; BUFREM, L. S. Influências teóricas do Professor Aldo Barreto na Ciência da Informação: uma análise da genealogia acadêmica. **Ciência da Informação em Revista**. v. 7, n. 3, p. 7-20, 2020.

GALVAGNO, M.; PISANO, V. Building the genealogy of family business internationalization: a bibliometric mixed-method approach. **Scientometrics**, v. 126, n. 1, p. 757-783, 2021.

GLÄNZEL, W. **Bibliometrics as a research field a course on theory and application of bibliometric indicators**. 2003.

GRÁCIO, M. C. C. **Análises relacionais de citação para a identificação de domínios científicos: uma aplicação no campo dos Estudos Métricos da Informação no Brasil**. Editora Oficina Universitária, 2020.

GRÁCIO, M. C. C., Acoplamento bibliográfico e análise de cocitação: revisão teórico-conceitual. In: **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 21, n. 47, p. 82-99, 2016.

GRÁCIO, M. C. C.; DE OLIVEIRA, E. F. T. INDICADORES DE PROXIMIDADES EM ANÁLISE DE COCITAÇÃO DE AUTORES: um estudo comparativo entre coeficiente de Correlação de Pearson e Cosseno de Salton. **Informação & Sociedade: Estudos**, v. 25, n. 2, p. 116, 2015.

GRÁCIO, M. C. C.; OLIVEIRA, E. F. T.; WOLFRAM, D. Production and impact of Brazilian papers in Library and Information Science journals (1986-2015). **Ciência da Informação**, v. 48, n. 3, 2019.

GRÁCIO, M. C. C.; OLIVEIRA, E. F. T. Análise de cocitação de autores: um estudo teórico-metodológico dos indicadores de proximidade, aplicados ao GT7 da ANCIB. **LIINC em Revista**, p. 196-213, 2013.

GUIMARÃES, J. A. C.; GRACIO, M. C. C.; MATOS, D. F. O. Produção científica de bolsistas pesquisa em Ciência da Informação do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq): um estudo com artigos de periódicos. **Revista de Informação**, v. 15, n. 2, 2014.

HAMERS, L. et al. Similarity measures in scientometric research: The Jaccard index versus Salton's cosine formula. **Information Processing and Management**, v. 25, n. 3, p. 315-18, 1989.

HIRAKATA, V. N.; CAMEY, S. A. Análise de concordância entre métodos de Bland-Altman. **Clinical & Biomedical Research**, v. 29, n. 3, 2009.

HJØRLAND, B. Citation analysis: a social and dynamic approach to knowledge organization. **Information Processing and Management**, v. 49, n. 6, p. 1313–1325, 2013.

HJØRLAND, B. Domain analysis in information science: eleven approaches—traditional as well as innovative. **Journal of documentation**, v. 58, n. 4, p. 422-462, 2002.

HOU, J.; YANG, X.; CHEN, C. Emerging trends and new developments in information science: A document co-citation analysis (2009–2016). **Scientometrics**, v. 115, n. 2, p. 869-892, 2018.

HU, D. et al. Investigating the genealogy of the literature on digital pathology: a two-dimensional bibliometric approach. **Scientometrics**, p. 1-17, 2022.

IMPA. Instituto de Matemática Pura e Aplicada. - **O fascínio de Aloisio Araujo pela economia matemática**. 2020. Disponível em: <https://impa.br/noticias/o-fascinio-de-aloisio-araujo-pela-economia-matematica/>

IOMMI, G. The Bowen formula: Dimension theory and thermodynamic formalism. 2008.

KESSLER, M. M. Bibliographic coupling between scientific papers. **American documentation**, v. 14, n. 1, p. 10-25, 1963.

KESSLER, M. M. An experimental study of bibliographic coupling between technical papers. **MASSACHUSETTS INST OF TECH LEXINGTON LINCOLN LAB**. p. 49-51. 1962.

LARIVIÈRE, V.; GINGRAS, Y. Averages of ratios vs. ratios of averages: An empirical analysis of four levels of aggregation. **Journal of informetrics**, v. 5, n. 3, p. 392-399, 2011.

LEYDESDORFF, L. On the normalization and visualization of author co-citation data: Salton's Cosine versus the Jaccard index. **Journal of the American Society for Information Science and Technology**, v. 59, n. 1, p. 77-85, 2007.

LEYDESDORFF, L.; TEKLES, A.; BORNMANN, L. "A proposal to revise the disruption index". **Profesional de la información**, v. 30, n. 1, e300121. 2021.

LUCAS, E. O.; ZORITA, J. C. G. Produção científica sobre capital social: estudo por acoplamento bibliográfico. **Em Questão**, v. 20, n. 3, p. 27-42, 2014.

LV, R.; CHANG, H. Bibliometric-based Study of Scientist Academic Genealogy. **Journal of Data and Information Science**, v. 6, n. 3, p. 146, 2021.

MAAR, J. H. Justus von Liebig, 1803-1873. Parte 1: vida, personalidade, pensamento. **Química Nova**, v. 29, p. 1129-1137, 2006.

MENA-CHALCO, J. P. Genealogia acadêmica: uma ferramenta para investigar a origem, evolução e disseminação de áreas do conhecimento. **Encontro Brasileiro de Bibliometria e Cientometria**, v. 5, p. 1-90, 2016.

MIYAHARA, E. K. **Genealogia Acadêmica Lattes**. Monografia (Bacharelado em Ciência da Computação) – Universidade de São Paulo, Instituto de Matemática e Estatística, 2011.

MOED, H. F. **Applied evaluative informetrics**. Berlin: Springer International Publishing, 2017.

MONTOYE, H. J.; WASHBURN, R. Research Quarterly contributors: An academic genealogy. **Research Quarterly for Exercise and Sport**, v. 51, n. 1, p. 261-266, 1980.

MOREIRA, T. H. J; DIAS, T. M. R; MOITA, G. F. Genealogia Acadêmica da Relação Orientador-Orientado na Área de Ciência da Computação. **Anais do Computer on the Beach**, p. 060-069, 2017.

MOURA, V. P. A; FARIA, L. I. L. A contribuição da genealogia acadêmica para a construção de indicadores bibliométricos. **Em Questão**, v. 27, n. 1, p. 336-360, 2021.

MUGNAINI, R.; JANNUZZI, P. M. Indicadores bibliométricos da produção científica brasileira: uma análise a partir da base Pascal. **Ciência da informação**, v. 33, n. 2, 2004.

OLIVEIRA, E. F. T.; ALVES, B. H. Cosseno de Salton, Índice de Jaccard e Correlação de Pearson: comparando índices normalizados e absolutos em análise de cocitação de autores. **Em Questão**, v. 23, 2017.

PERIANES-RODRIGUEZ, A.; WALTMAN, L.; VAN ECK, N. J.. Constructing bibliometric networks: A comparison between full and fractional counting. **Journal of Informetrics**, v. 10, n. 4, p. 1178-1195, 2016.

PINTO, E. R. Estudo da dinâmica de epidemias em redes complexas. **Dissertação de mestrado** – Universidade Estadual Paulista Campus Botucatu. 2018.

PRADO, Marcos Aparecido Rodrigues; CASTANHA, Renata Cristina Gutierrez. INDICADORES: CONCEITOS FUNDAMENTAIS E IMPORTÂNCIA EM CT&I. In: **Tópicos da Bibliometria para bibliotecas universitárias**, p. 50, 2020.

REAL, R.; VARGAS, J. M. The probabilistic basis of Jaccard's index of similarity. **Systematic biology**, v. 45, n. 3, p. 380-385, 1996.

ROSSI, Luciano; MENA-CHALCO, Jesús P. Caracterização de árvores de genealogia acadêmica por meio de métricas em grafos. In: **Anais do III Brazilian Workshop on Social Network Analysis and Mining – SBC**. p. 21-32. 2014.

ROSSI, L. et al. Genealogia acadêmica dos doutores atuantes em matemática: um mapeamento macro na ciência brasileira. **Proceeding Series of the Brazilian Society of Computational and Applied Mathematics**, v. 6, n. 1, 2018.

ROSSI, L.; DAMACENO, R.; MENA-CHALCO, J. P. Genealogia acadêmica: um novo olhar sobre impacto acadêmico de pesquisadores. **Parcerias Estratégicas**, v. 23, n. 47, p. 197-212, 2018.

SACHTLEBEN, R. Nobel prize winners" descended" from Liebig: A table of academic geneology (Tr. by Ralph E. Oesper.). **Journal of Chemical Education**, v. 35, n. 2, p. 73, 1958.

SAFER, M. A.; TANG, R. The psychology of referencing in psychology journal articles. **Perspectives on Psychological Science**, v. 4, n. 1, p. 51-53, 2009.

SANTOS, R. N. M.; KOBASHI, N. Y. Bibliometria, Cientometria, infometria: conceitos e aplicações. **Anais Encontro Nacional de Pesquisa em Ciência da Informação. ENANCIB**. p. 155-172. 2009.

SEN, S. K.; GAN, S. K. A mathematical extension of the idea of bibliographic coupling and its applications. **Annals of Library Science and Documentation**. v.30, n.2, p.78-82. 1983.

SHEN, S. et al. A refined method for computing bibliographic coupling strengths. **Journal of Informetrics**, v. 13, n. 2, p. 605-615, 2019.

SILVEIRA, M. A. A.; CAREGNATO, S. E.; BUFREM, L. S. Estudo das razões das citações na ciência da informação. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 7, n. 2, 2014.

SMALL, H. Co-citation in the scientific literature: A new measure of the relationship between two documents. **Journal of the American Society for information Science**, v. 24, n. 4, p. 265-269, 1973.

SUGIMOTO, C. R. Academic Genealogy. In: CRONIN, B.; SUGIMOTO, C.R.(Eds.). **Beyond bibliometrics: harnessing multidimensional indicators at scholarly impact**. Cambridge: MIT Press, 2014.

SUN, F. et al. Citation genetic genealogy: A new perspective for citation analysis in scientific literature. In: **Proceedings of ISSI**. 2011. p. 817-828.

SUN, F.; ZHU, L. Citation genetic genealogy: a novel insight for citation analysis in scientific literature. **Scientometrics**, v. 91, n. 2, p. 577-589, 2012.

THIJS, B.; ZHANG, L.; GLÄNZEL W. Bibliographic coupling and hierarchical clustering for the validation and improvement of subject-classification schemes. **Scientometrics**, v. 105, n. 3, p. 1453-1467, 2015.

VANZ, S. A. S.; STUMPF, I. R. C. Procedimentos e ferramentas aplicados aos estudos bibliométricos. **Informação & Sociedade: estudos. João Pessoa, PB. Vol. 20, n. 2 (maio/ago. 2010), p. 67-75**, 2010.

VIANA, L. C. S. et al. Genealogia acadêmica e sua relação com a trajetória do conhecimento na Escola Nacional de Saúde Pública Sergio Arouca, Fundação Oswaldo Cruz. **Transinformação**, v. 31, 2019.

VLADUTZ, G.; COOK, J. Bibliographic coupling and subject relatedness. **Proceedings of the American Society for Information Science**, v. 21, p. 204-207, 1984.

WANG, Z.; GLÄNZEL, W.; CHEN, Y. The impact of preprints in Library and Information Science: an analysis of citations, usage and social attention indicators. **Scientometrics**, v. 125, n. 2, p. 1403-1423, 2020.

WHITE, Howard D. Cocited author retrieval online: an experiment with the social indicators literature. **Journal of the American Society for Information Science**, v. 32, n. 1, p. 16-21, 1981.

WHITE, Howard D.; GRIFFITH, B. Author cocitation: A literature measure of intellectual structure. **Journal of the American Society for information Science**, v. 32, n. 3, p. 163-171, 1981.

WHITE, Howard D. Authors as citers over time. **Journal of the American Society for Information Science and Technology**, v. 52, n. 2, p. 87-108, 2001.

WU, Lingfei; WANG, Dashun; EVANS, James A. Large teams develop and small teams disrupt science and technology. **Nature**, v. 566, n. 7744, p. 378-382, 2019.

ZHAO, D.; STROTMANN, A. Evolution of Research Activities and Intellectual Influences in Information Science 1996–2005: Introducing Author Bibliographic-Coupling Analysis. **Journal of the American Society for Information Science and Technology**, v. 59, n. 13, p. 2070-2086, 2008.

ZHAO, D; STROTMANN, A. The knowledge base and research front of information science 2006–2010: An author cocitation and bibliographic coupling analysis. **Journal of the Association for Information Science and Technology**, v. 65, n. 5, p. 995-1006, 2014.

ZHAO, D; STROTMANN, A. Mapping knowledge domains on Wikipedia: an author bibliographic coupling analysis of traditional Chinese medicine. **Journal of Documentation**, 2021.

APÊNDICE 2 – Código em R do The Genealogical Coupler

```
#Pacotes
```

```
library(dplyr)
library(RVenn)
library(igraph)
library(shiny)
library(shinydashboard)
library(flexdashboard)
library(visNetwork)
library(networkD3)
```

```
# UI
```

```
ui <- fluidPage(
  dashboardPage(
    title = 'the_coupler: Academic Genealogy',
    skin = "purple",
    dashboardHeader(
      title = img(src = "gc.png"),
      dashboardSidebar(
        fileInput("file1", "Selecione o arquivo:", accept = ".txt"),
        selectInput("sep", "Separador:", c(" ", "Vírgula = ", "Ponto_Vírgula = ";", "Tabulado = "\t")),
        selectInput("normalization", "Normalizações:", c("sem normalização", "Cosseno de Salton", "Índice de Jaccard", "CABG")),
        fluidPage(h5(" ")),
        fluidPage(
          tags$a(href = "https://github.com/rafaelcastanha/The-Genealogic-Coupler", "Instruções e Código (GitHub)", style = "margin-left: 15px;"),
          fluidPage(h5(" ")),
          fluidPage(
            tags$a(href = "https://zenodo.org/record/7130614#.YzdGCnbMLIU", "Arquivos para testes", style = "margin-left: 15px;"),
            fluidPage(h5(" ")),
            fluidPage(
              tags$a(href = "https://zenodo.org/record/7275086#.Y2LqIHbMLIU", "Sobre a métrica CABG", style = "margin-left: 15px;"),
              fluidPage(h5(" ")),
              fluidPage(
                tags$a(href = "http://lattes.cnpq.br/4834832439175113", "Currículo Lattes", style = "margin-left: 15px;"),
                fluidPage(h5(" ")),
                fluidPage(
                  tags$a(href = "https://www.researchgate.net/profile/Rafael-Gutierrez-Castanha-2", "ResearchGate", style = "margin-left: 15px;"),
                  fluidPage(h5(" ")),
                  fluidPage(
                    tags$a(href = "https://rafaelcastanha.shinyapps.io/thecoupler/", "The Coupler (Original)", style = "margin-left: 15px;"),
                    fluidPage(h5(" ")),
                    out = fluidPage(p("Desenvolvido por:", style = "margin-left: 15px;")),
                    out = fluidPage(p("Rafael Gutierrez Castanha", style = "margin-top: -8px; margin-left: 15px;")),
                    out = fluidPage(p("rafael.castanha@unesp.br", style = "margin-top: -10px; margin-left: 15px;")),
                    actionButton("runmodel", "Coupling!"),
                    dashboardBody(
                      tags$head(
                        tags$style(HTML('.content-wrapper {background-color: #E2DDF0;}')),
                        tags$head(
                          tags$style(HTML('.container-fluid {padding-right: 0px; padding-left: 0px;}')),
                          fluidRow(
                            width = 12,
                            height = NULL,
                            tabsetPanel(
                              type = "tab",
                              tabPanel(
                                title = h5("Rede de Genealogia Acadêmica", style = "color: black;"),
                                column(
                                  textOutput("erro"),
                                  width = 12,
                                  visNetworkOutput(
                                    outputId = "PlotCoupling",
                                    width = "100%",
                                    height = 540
                                  )
                                )
                              ),
                              tabPanel(
                                title = h5("Frequências de Acoplamento", style = "color: black;"),
                                dataTableOutput(
                                  outputId = "DataFrameCoupling",
                                  downloadButton("dlfreq", "Download Data")
                                )
                              ),
                              tabPanel(
                                title = h5("Unidades de Acoplamento", style = "color: black;"),
                                style = "overflow-x: scroll",
                                dataTableOutput(
                                  outputId = "DataFrameUnits",
                                  downloadButton("dlunits", "Download Data")
                                )
                              ),
                              tabPanel(
                                title = h5("Matriz de Citação", style = "color: black;"),
                                style = "overflow-x: scroll",
                                dataTableOutput(
                                  outputId = "DataFrameCit",
                                  downloadButton("dlcit", "Download Data")
                                )
                              ),
                              tabPanel(
                                title = h5("Matriz de Acoplamento", style = "color: black;"),
                                style = "overflow-x: scroll",
                                dataTableOutput(
                                  outputId = "DataFrameMatrix",
                                  downloadButton("dlaba", "Download Data")
                                )
                              ),
                              tabPanel(
                                title = h5("Matriz de Cocitação", style = "color: black;"),
                                style = "overflow-x: scroll",
                                dataTableOutput(
                                  outputId = "DataFrameCocit",
                                  downloadButton("dlcocit", "Download Data")
                                )
                              )
                            )
                          )
                        )
                      )
                    )
                  )
                )
              )
            )
          )
        )
      )
    )
  )
)
```

```

#Server

server <- function(input, output){
observe({

#Arquivo

input$runmodel
if (input$runmodel==0)
return()
else
isolate({
r<-reactive({input$file1})
req(r())
})

corpus<-isolate(read.table(r)$datapath, header = FALSE, sep = input$sep, quote="\")
colnames(corpus)<-corpus[1,]
corpus<-corpus[(-1),]
hd<-gsub("\\.$", "", names(corpus))
colnames(corpus)<-hd

#Corpus para dataframe

corpus<-as.data.frame(corpus)
#remover espaços e vazios

corpus<-corpus
corpus[corpus=="|corpus==" "|corpus==" ]<-NA
empty_columns<-sapply(corpus, function(x) all(is.na(x)|x==""))
corpus<-corpus[!empty_columns]

#Contagem de itens citados por lista

citados<-function(x){return(length(which(!is.na(x))))}
itens_citados<-apply(X=corpus,FUN=citados,MARGIN=c(1,2))
df1<-as.data.frame(itens_citados, header=TRUE)
df2<-colSums(df1)
df2<-as.data.frame(df2, header=TRUE)
df2<-tibble::rownames_to_column(df2, "VALUE")
colnames(df2)[1]<-"units"
colnames(df2)[2]<-"refs"
referencias<-df2

#Transformação em objeto Venn

corpus_aba<-Venn(corpus)

#Intersecção Pareada: identificação das unidades de acoplamento

ABA<-overlap_pairs(corpus_aba)
ABA<-ABA[1:length(corpus)-1]
ABA_GA<-ABA

#Unidades por acoplamento

unit_aba<-na.omit(stack(ABA))
units_coupling<-unit_aba %>% group_by(ind) %>% summarise(valeus=(paste(values, collapse="; ")))
units_final<-as.data.frame(units_coupling)
colnames(units_final)[1]<-"Unidades de Análise"

```

```

colnames(units_final)[2]<-"Unidades de Acoplamento"

#Intensidades de ABA

df<-as.data.frame(table(stack(ABA)))

#Erro em caso de acoplamentos nulos

if (nrow(df)==0) {
nodes <- data.frame(id = 1, shape = "icon", icon.face = 'Ironics',
icon.code = c("f100"))
edges <- data.frame(from = c(1))
output$PlotCoupling<-renderVisNetwork({
visNetwork(nodes=nodes, edges=edges,shape = "icon",
main="WARNING: No couplings between units (ALERTA: Não há acoplamento entre as unidades!)" %>%
addIronics()
})
stop
}

else {

#Intensidades de ABA

int_aba<-aggregate(Freq ~ ind, data = df, FUN = sum)
Freq_ABA<-data.frame(do.call("rbind",strsplit(as.character(int_aba$ind),"...",fixed=TRUE)))
Freq_ABA["Coupling"]<-int_aba$Freq
colnames(Freq_ABA)[1]<-"Orientador"
colnames(Freq_ABA)[2]<-"Orientandos"
m2=merge(Freq_ABA,df2,by.x="Orientandos",by.y="units",all.x=TRUE)
m1=merge(m2,df2,by.x="Orientador",by.y="units",all.x=TRUE)
colnames(m1)[4]<-"L2"
colnames(m1)[5]<-"L1"
Freq_ABA<-m1 %>% select("Orientador","Orientandos","L1","L2","Coupling")

#Normalizações

novacoluna<-c("Saltons_Cosine")
Freq_ABA[,novacoluna]<-Freq_ABA$Coupling/sqrt(Freq_ABA$L1*Freq_ABA$L2)
novacoluna_2<-c("Jaccard_Index")
Freq_ABA[,novacoluna_2]<-Freq_ABA$Coupling/(Freq_ABA$L1+Freq_ABA$L2-Freq_ABA$Coupling)
novacoluna_3<-c("CABG")
Freq_GA<-Freq_ABA
Freq_GA[,novacoluna_3]<-Freq_ABA$Coupling/references[1,2]
Freq_GA<-Freq_GA[1:(length(corpus)-1),]
Freq_ABA<-Freq_GA

#Rede de acoplamento bibliográfico

net_list<-filter(Freq_ABA, Coupling>0)
links<-data.frame(source=c(net_list$Orientador), target=c(net_list$Orientandos))
network_ABA<-graph_from_data_frame(d=links, directed=F)
edge_ABA<-net_list$Coupling
edge_CS<-net_list$Saltons_Cosine
edge_IJ<-net_list$Jaccard_Index
edge_CABG<-net_list$CABG
node1<-as.data.frame(references$units)
node2<-as.data.frame(references$units)
colnames(node1)[1]<-"id"
colnames(node2)[1]<-"label"

```

```

node<-data.frame(node1, node2)

ED_coupling<-mutate(links, width = edge_ABA)
colnames(ED_coupling)[1]<-"from"
colnames(ED_coupling)[2]<-"to"

ED_salton<-mutate(links, width = edge_CS*10)
colnames(ED_salton)[1]<-"from"
colnames(ED_salton)[2]<-"to"

ED_jaccard<-mutate(links, width = edge_IJ*10)
colnames(ED_jaccard)[1]<-"from"
colnames(ED_jaccard)[2]<-"to"

ED_CABG<-mutate(links, width = edge_CABG*10)
colnames(ED_CABG)[1]<-"from"
colnames(ED_CABG)[2]<-"to"

#Matrizes Adjacencia

dt<-stack(corpus)
dt2<-table(dt$values[row(dt[-1])], unlist(dt[-1]))
mtx_cit<-t(dt2)
mtx_cocit<-(t(mtx_cit) %*% mtx_cit)
mtx_cocit<-as.table(mtx_cocit)

#Matriz Acoplamento

mtx_aba<-(mtx_cit %*% t(mtx_cit))
diag(mtx_aba)<-0
mtx_aba<-as.data.frame(mtx_aba)
mtx_aba<-tibble::rownames_to_column(mtx_aba, " ")

#matriz de Citação

mtx_cit<-as.data.frame.matrix(mtx_cit)
mtx_cit<-tibble::rownames_to_column(mtx_cit, " ")

#Matriz de Cocitação

mtx_cocit_df<-as.data.frame(mtx_cocit)
links_cocit<-data.frame(source=c(mtx_cocit_df$Var1), target=c(mtx_cocit_df$Var2))
network_cocit<-graph_from_data_frame(d=links_cocit, directed=T)
E(network_cocit)$weight<-mtx_cocit_df$Freq
mtx_adj_cocit<-as_adjacency_matrix(network_cocit, attr="weight")
mtx_adj_cocit_df<-as.data.frame(as.matrix(mtx_adj_cocit))
mtx_adj_cocit_df<-tibble::rownames_to_column(mtx_adj_cocit_df, " ")

output$erro<-renderText({
input$runmodel
if (nrow(df)!=0) {print(" ")}
})

#Plots

output$PlotCoupling <- renderVisNetwork({
input$runmodel
if ("sem normalização" %in% input$normalization){
visNetwork(node, ED_coupling) %>%
visOptions(highlightNearest = TRUE, nodesIdSelection = TRUE) %>%

```

```

visEdges(arrows = "to") %>%
visHierarchicalLayout(sortMethod="directed")}
else{
if ("Cosseno de Salton" %in% input$normalization){
visNetwork(node, ED_salton) %>%
visOptions(highlightNearest = TRUE, nodesIdSelection = TRUE) %>%
visEdges(arrows = "to") %>%
visHierarchicalLayout(sortMethod="directed")}
else{
if ("Índice de Jaccard" %in% input$normalization){
visNetwork(node, ED_jaccard) %>%
visOptions(highlightNearest = TRUE, nodesIdSelection = TRUE) %>%
visEdges(arrows = "to") %>%
visHierarchicalLayout(sortMethod="directed")}
else{
if ("CABG" %in% input$normalization)
visNetwork(node, ED_CABG) %>%
visOptions(highlightNearest = TRUE, nodesIdSelection = TRUE) %>%
visEdges(arrows = "to") %>%
visHierarchicalLayout(sortMethod="directed")
}
}
}
}}

```

#Outputs tabelas

```

output$DataFrameCoupling <- renderDataTable(Freq_ABA)
output$DataFrameUnits <- renderDataTable(units_final)
output$DataFrameCit <- renderDataTable(mtx_cit)
output$DataFrameMatrix <- renderDataTable(mtx_aba)
output$DataFrameCocit <- renderDataTable(mtx_adj_cocit_df)

```

#Downloads

```

output$dlfreq <- downloadHandler(
filename = function(){
paste("Frequências de Acoplamento", "txt", sep=".")
},
content = function(file){
write.table(Freq_ABA, file, sep="\t", row.names = F, col.names = TRUE)
})
output$dlunits <- downloadHandler(
filename = function(){
paste("Unidades de Acoplamento", "txt", sep=".")
},
content = function(file){
write.table(units_final, file, sep="\t", row.names = F, col.names = TRUE)
})
output$dlcit <- downloadHandler(
filename = function(){
paste("Matriz de Citacao", "txt", sep=".")
},
content = function(file){
write.table(mtx_cit, file, sep="\t", row.names = F, col.names = TRUE)
})
output$dlaba <- downloadHandler(
filename = function(){
paste("Matriz de Acoplamento", "txt", sep=".")
},

```

```
content = function(file){
write.table(mtx_adj, file, sep="\t", row.names = F, col.names = TRUE)
})
output$dlcocit <- downloadHandler(
filename = function(){
paste("Matriz de Cocitacao", "txt", sep=".")
},
content = function(file){
write.table(mtx_adj_cocit_df, file, sep="\t", row.names = F, col.names = TRUE)
})
}
})
}
```

#Rodar The Genealogical Coupler

```
shinyApp(ui, server)
```