



COEFICIENTE DE ACOPLAMENTO BIBLIOGRÁFICO GENEALÓGICO PARA MENSURAR A INTENSIDADE DA PRESERVAÇÃO DAS CORRENTES TEÓRICO-METODOLÓGICAS NAS REDES DE GENEALOGIA ACADÊMICA DA ÁREA DE MATEMÁTICA

Informações:

Nome: Rafael Gutierrez Castanha

Instituição: Universidade Estadual Paulista (UNESP). Campus Marília/SP

E-mail: rafael.castanha@unesp.br

Orientador/a: Maria Cláudia Cabrini Grácio

Data prevista para defesa: Fevereiro de 2023

Título (provisório) da tese: Coeficiente de acoplamento bibliográfico genealógico para mensurar a intensidade da preservação das correntes teórico-metodológicas nas redes de genealogia acadêmica da área de Matemática

Resumo: Objetivo: Propõe um coeficiente de perspectiva relacional de citação e recitação, denominado Coeficiente de Acoplamento Genealógico, para mensurar a intensidade das proximidades teórico-metodológicas entre pesquisadores no escopo das relações de genealogia acadêmica no campo da Matemática no Brasil. 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 a 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. Resultados: Demonstra que o coeficiente proposto é correlacionado, mas não concordante com as demais medidas comparadas apresentando-se como indicador alternativo e confiável para análises de acoplamento bibliográfico entre pesquisadores com relação de genealogia acadêmica.

Palavras-Chave: Acoplamento Bibliográfico. Genealogia Acadêmica. Medidas de Similaridade.

1 INTRODUÇÃO E PROBLEMATIZAÇÃO

Estudar a genealogia acadêmica (GA) de determinado grupo de pesquisadores ou de um pesquisador individual visa compreender a herança intelectual passada de orientador para orientando, possibilitando compreender como se constitui a relação orientador-orientando e de que maneira ocorre a transmissão e difusão do conhecimento e a herança intelectual entre eles (SUGIMOTO, 2014). Comumente, as análises de GA são baseadas em grafos (redes)

hierarquicos e Rossi *et al.* (2018) indicam que várias são as métricas topológicas passíveis de aplicabilidade no âmbito das redes de GA, em que a maioria destas métricas contemplam a evolução das redes e dá enfoque à visualização das relações existentes e da inserção de novos pesquisados (e gerações), entretanto, essas análises não contemplam os valores de conectividade das arestas das redes de GA.

Castanha e Grácio (2020) apontam a valoração das arestas em redes de GA, por meio da intensidade de acoplamento bibliográfico entre autores (ABA), como potencial indicador para avaliar a herança intelectual perpassada de orientador para orientando, uma vez que, ao citar os mesmos referentes, dois pesquisadores suscitam proximidades teórico-metodológicas. Ainda, a conectividade entre dois pesquisadores com relação de GA, representada pela intensidade de acoplamento, torna-se mais expressiva se consideradas a sobreposição entre identidades de citação. A identidade de citação, segundo White (2001), é representada pelo autores recitados diacronicamente (citados em 2 ou mais documentos) na lista de referentes de determinado pesquisador.

Nesse contexto, esta tese pretende responder ao seguinte problema: Existe um coeficiente do tipo ABA adequado para a mensuração da intensidade da transmissão das correntes teórico-metodológicas nas redes de genealogia acadêmica, e da manutenção das correntes científica/intelectuais em grupos de pesquisadores no escopo das relações de genealogia acadêmica do campo da Matemática? O estudo tem como hipótese: Há um Coeficiente de Acoplamento derivado do coeficiente normalizado de ABA 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.

Desse modo, esta tese propõe um coeficiente de perspectiva relacional de recitação baseado no ABA, denominado Coeficiente de Acoplamento Genealógico (CAG), para mensurar a intensidade das proximidades teórico-metodológicas entre pesquisadores no escopo das relações de genealogia acadêmica no campo da Matemática no Brasil. De maneira específica: i) fundamentar o Coeficiente de Acoplamento Genealógico; ii) comparar três formas de normalização do Acoplamento Bibliográfico de Autores: Índice de Jaccard, Cosseno de Salton e Coeficiente de Acoplamento 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.

2 PROCEDIMENTOS METODOLÓGICOS

A tese adota como universo de pesquisa os bolsistas PQ Sênior da área de Matemática vinculados ao IMPA que têm sua bolsa em vigência e seus respectivos orientandos egressos de doutorado com produções científicas (somente artigos) indexadas na base de dados e Web of Science (WoS) e que atendam ao critério de autoria: o orientador é primeiro autor das publicações e o orientando não é coautor da publicação; o orientando é primeiro autor das publicações e o orientador não é coautor da publicação. A partir da identificação dos pesquisadores, dos respectivos orientandos e de toda produção científica indexada, foi constatado 5 bolsistas PQ-Sr do IMPA e 62 orientandos egressos de doutorado com produções que atendem ao critério de autoria estabelecido. Após esta identificação esta pesquisa se desenvolve nas seguintes etapas: i) Extração dos referentes citados e recitados por cada pesquisador analisado; ii) Organização e aplicação das medidas de similaridades baseadas no ABA para cada par de orientação; iii) Comparação entre o coeficiente de acoplamento genealógico, Índice de Jaccard e Cosseno de Salton baseados ABA; iv) Mapeamento da estrutura intelectual da floresta de genealogia acadêmica composta pelos bolsistas PQ-Sr e seus orientandos.

Considerando que o desenvolvimento do CAG, bem como a comparação da nova medida com o Cosseno de Salton e Índice de Jaccard, é parte fundamental desta tese, se esclarece que a escolha das outras duas medidas se dá pela consolidação e uso ao longo dos anos em diversos estudos, pioneiros e contemporâneos, destas medidas. No que diz respeito à formulação do CAG, tem-se: Sejam $L1$ e $L2$, as listas de autores recitados diacronicamente por orientador e orientando que atendam ao critério de autoria respectivamente, o CAG é dado pela Equação 1:

$$CAG = \frac{(L1 \cap L2)}{L1} \quad [1]$$

Em que o denominador $L1$ representa a cardinalidade da lista de autores recitados do orientador, e a interseção $L1 \cap L2 = ABA(L1, L2)$. Neste caso, o valor do ABA é uma medida pautada na citação de autores em comum entre orientador e orientando, isto é, a quantidade de autores em comum que compõe a identidade científica de citação dos pesquisadores acoplados. Dado o fato de $L1 \geq (L1 \cap L2)$, tem-se que $0 \leq CAG \leq 1$. Dessa maneira, o Cosseno

de Salton (*CS*) e o Índice de Jaccard (*IJ*) a serem comparados com o CAG são formulados nas Equações 2.

$$CS = \frac{ABA(L1, L2)}{\sqrt{L1 \times L2}} \quad e \quad IJ = \frac{ABA(L1, L2)}{L1 \cup L2} = \frac{ABA(L1, L2)}{L1 + L2 - L1 \cap L2} \quad [2]$$

Para comparação do CAG com as demais medidas, foram realizadas análise de correlação entre as medidas via Coeficiente de Correlação Intraclasse (ICC) seguido de análise de concordância por meio do método de *Bland-Altman*. Este último aplica o teste T de *Student* entre as diferenças entre as medidas e o valor de referência zero, e, gera um gráfico de dispersão contendo a diferença das medidas e a média das diferenças. O teste T aponta se há diferenças significativas (a nível de 5%) entre os valores obtidos pelas medidas e o gráfico de dispersão ilustra se estas diferenças.

RESULTADOS PARCIAIS

A fim de cunhar um novo indicador relacional baseado no ABA, constatou-se que, ao CAG atingir o valor zero, significa que não há qualquer proximidade entre os autores recitados pelo orientador e pelo orientando, e 1 (ou 100%) 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 de modo que $L1 \subseteq L2$. Assim, a principal característica do CAG é de, ao fixar o denominador em *L1*, o CAG atinge o valor máximo (igual a 1) sem necessariamente *L1* ser igual a *L2*, justamente por $L1 \geq (L1 \cap L2)$. Esta característica não se encontra nas outras medidas (*CS* e *IJ*) que atingem valor máximo somente se os conjuntos a serem comparados forem idênticos (mesmos elementos e mesma cardinalidade).

Como demais características do indicador proposto tem-se: i) $L1=L2 \rightarrow CS=CAG$; ii) $(L1 \cap L2)=L2 \rightarrow IJ=CAG$; iii) CAG é necessariamente maior ou igual ao IJ ($CAG \geq IJ$); iv) $CAG > CS \rightarrow L2 > L1$; v) $Se L1 \cap L2 = 0 \text{ ou } L1 = L2 \rightarrow CS = IJ = CAG$.

Ressalta-se que, o tamanho da lista de autores recitados do orientador não é obrigatoriamente maior que a de seu orientando, e dessa maneira, ao fixar o denominador do CAG em *L1*, mesmo que o orientando tenha recitado mais autores que seu orientador, o CAG poderá atingir acoplamento perfeito (igual a 1 ou 100%), pois o indicador guarda consigo a principal ideia dos estudos de GA: a herança intelectual herdada pelo orientando perante seu orientador, neste caso representada pela lista de autores recitados pelo orientador (*L1*).

Ao comparar as medidas (CAG, CS e IJ), aplicadas ao conjunto das 62 orientações, entre si, via ICC e método de *Bland-Altman*, foi verificada forte correlação entre CAG e CS, e, entre CAG e IJ. O ICC pode ser entendido como um descritor de confiabilidade entre métodos. Sabendo que as medidas de Jaccard e Salton são reconhecidas e bem estabelecidas na literatura especializada no campo de Ciência da Informação, o CAG estar correlacionado a estas medidas, indica que está "calibrado" com elas. Em suma, o CAG é capaz de captar as variabilidades tão bem quanto Jaccard e Salton, mas guarda distinções em seus resultados, pelas características descritas anteriormente.

Contudo, este resultado não aponta necessariamente concordância entre os métodos. Ao submeter as medidas a análise de concordância via teste T de *Student*, foi verificado que existem diferenças numéricas significativas entre CAG e CS, e CAG e IJ, isto é, uma não concordância. A não concordância numérica está relacionada justamente as características elencadas anteriormente. Assim, o CAG se apresenta como indicador calibrado as demais medidas mas numericamente diferente devido sua formulação baseada no conceito de heranças intelectual, matemática representado pelo denominador fixo em *LI*.

CONSIDERAÇÕES PARCIAIS

Ao propor o CAG, constatou-se que, tanto o IJ, quanto o CS, atingem seu valor máximo apenas se os conjuntos analisados forem idênticos, enquanto o CAG, ao admitir a possibilidade de alcançar um acoplamento perfeito mesmo comparando conjuntos de tamanhos diferentes, se apresenta como medida de similaridade ajustada para ao ABA com relação de GA, pois, um acoplamento perfeito nesta perspectiva, aponta que o orientando herdou toda lista de autores recitados de seu orientador (recita os mesmos autores).

Sob a ótica dos estudos de GA, um acoplamento perfeito, segundo o CAG, 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). Por fim, o CAG se anuncia como uma medida de similaridade inédita, fundamentada e concebida a partir dos conceitos teóricos das análises de GA, objetiva e alternativa às demais presentes na literatura relacionada ao acoplamento bibliográfico.

REFERÊNCIAS

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**; [s.l.], v. 24, n. 2, p. e020039, 2020.

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**, [s.l.], v. 6, n. 1, 2018.

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

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