



XXI ENANCIB

Encontro Nacional de Pesquisa em Ciência da Informação

50 anos de Ciência da Informação no Brasil:
diversidade, saberes e transformação social

Rio de Janeiro • 25 a 29 de outubro de 2021

XXI Encontro Nacional de Pesquisa em Ciência da Informação – XXI ENANCIB

GT-7 – Produção e Comunicação da Informação em Ciência, Tecnologia & Inovação

IDENTIFICAÇÃO DE POTENCIAIS CONFLITOS DE INTERESSE EM AVALIAÇÕES CIENTÍFICAS POR MEIO DE REDES COLABORATIVAS DE PESQUISA

IDENTIFICATION OF POTENTIAL CONFLICTS OF INTEREST IN SCIENTIFIC EVALUATIONS USING COLLABORATIVE RESEARCH NETWORKS

Wellington B. Rodrigues* – Universidade Federal do ABC
Sônia Vasconcelos – Universidade Federal do Rio de Janeiro
Jesús P. Mena-Chalco – Universidade Federal do ABC

Modalidade: Trabalho Completo

Resumo: A avaliação científica isenta de influência de potenciais conflitos de interesse (Col) é essencial para a qualidade e confiança nos resultados da pesquisa científica. Porém, Cols são um desafio cada vez mais evidente, especialmente em cenários de estreita colaboração científica. Neste trabalho é apresentado um comparativo entre duas estratégias de identificação de Col baseadas em análise de coautoria em redes colaborativas de pesquisa. São apresentadas duas estratégias - normalizada e binária. Na estratégia binária é mais difícil identificar Avaliadores sem Col e Avaliadores com uma rede de colaboração grande tem menor chances de serem Elegíveis. Na estratégia normalizada é possível identificar aumento significativo na quantidade de Avaliadores Elegíveis.

Palavras-Chave: Conflito de interesse; Redes colaborativas de pesquisa; Coautoria; Avaliações acadêmicas; Revisão por pares.

Abstract: Unbiased scientific assessment, free of the influence of potential conflicts of interest (Col) is essential for the quality of and trust in the results of scientific research. However, Cols have shown to be increasingly challenging, especially for collaborative research networks. This work presents a comparison between two strategies for identifying Cols based on analysis of coauthorship networks. - normalized and binary. Using a binary strategy makes it more difficult to identify Reviewers without Col and Reviewers within a large collaborative network are less likely to be Eligible. Using a normalized strategy, it is possible to identify a significant increase in the number of Eligible Reviewers.

Keywords: Conflict of interest; Coauthorship; Research Collaborative research networks; Academic evaluations; Peer review.

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

1 INTRODUÇÃO

O principal meio utilizado no processo avaliativo da ciência moderna é a avaliação por especialistas (*peer review*) da área a ser avaliada. Independente do modelo de *peer review* adotado, é esse respaldo de especialistas que norteia boa parte da seleção de projetos de pesquisa e de publicações em periódicos científicos (PATRUS; DANTAS; SHIGAKI, 2016). Esse modelo avaliativo também se reflete na formação de bancas para definir a progressão na carreira de professores em universidades. Reconhecendo limitações associadas à revisão por pares, Zuckerman e Merton (1971) indicam a importância desse sistema para o desenvolvimento da ciência, na visão dos próprios cientistas.

Porém, esse processo avaliativo, com revisão por pares, envolve diversos fatores que podem ser fonte de vieses; por exemplo, a competição entre cientistas (BUCCHI, 2014). Essa atenção sobre o processo avaliativo por pares, especialmente em publicações, vem ganhando cada vez mais atenção na literatura (SQUAZZONI et al., 2021).

Para aumentar a isenção e o escrutínio científico, potenciais conflitos de interesse (Col) (RESNIK; ELMORE, 2018) são considerados na seleção de revisores, em periódicos e agências de financiamento, por exemplo. Como descrito por ABDOL et al. (2012):

Col's não financeiros são frequentemente listados por editores e cientistas como vieses importantes ... No entanto, até onde sabemos, sua presença e natureza raramente tem sido avaliada; a maioria dos relatos abordaram Col's são de natureza industrial ou financeiros, particularmente na revisão por pares por periódicos... (ABDOL et al., 2012).

No grupo desses conflitos não financeiros, incluem-se relações profissionais, pessoais ou acadêmicas, que podem influenciar em decisões de avaliação mais subjetivas, baseadas, por exemplo, em um histórico pessoal, mesmo que involuntariamente. Considerando o reconhecimento desse problema na literatura científica, uma questão ainda pouco explorada, é a identificação de relações entre pesquisadores no momento de seleção de membros de bancas avaliativas, por exemplo. Devem ser evitados pareceres enviesados, uma vez que trazem potenciais prejuízos para a atividade científica e para a carreira de pesquisadores, além de fragilizar a credibilidade da avaliação.

Com a evolução da tecnologia e a maior quantidade de produções científicas (BORNMANN; MUTZ, 2015), existe uma demanda elevada por avaliadores. De acordo com Da Silva, Moreiro-Gonzalez e Mueller (2016), o processo mais difícil para um editor de

periódicos acadêmicos é justamente identificar revisores adequados para os artigos, pois existem muitos fatores a serem considerados. Para realizar essa tarefa, recursos computacionais podem ajudar a otimizar a identificação de Col's entre avaliadores e avaliados, bem como outras tarefas (CHECCO et al. 2021).

Uma identificação em larga escala da relação existente entre pesquisadores pode ser viabilizada por meio de análises em redes de colaboração (LIU et al. 2005), montada com base em trabalhos anteriores de cada pesquisador. Com uma rede unificada das relações de colaboração de todos os pesquisadores de um determinado grupo é possível identificar quais pesquisadores têm relações diretas e indiretas com outros.

Em processos avaliativos, a forma mais aceita, atualmente, para identificação de Col é a autodeclaração por parte dos avaliadores (FELDMAN et al. 2018). Porém, podem ser identificadas relações que o próprio avaliador não perceba. Pesquisadores que tenham grande volume de publicações e/ou uma rede de colaboradores muito extensa podem não perceber que um de seus coautores está entre os avaliados, se dando conta desse fato somente no momento da avaliação ou, até mesmo, posteriormente.

Uma relação de coautoria entre pesquisadores pode acontecer de formas diferentes, onde o grau de interação (e.g. planejamento, análise, escrita e comunicação) varia pela quantidade de pessoas envolvidas no trabalho. A relação entre dois pesquisadores, sendo um com mais publicações e outro com menos, deve ser analisada de forma distinta para cada parte. Cada publicação do pesquisador com mais publicações representa uma menor parcela em relação aos seus trabalhos como um todo, se comparado com um pesquisador que tenha menos publicações. Também pode-se assumir que um trabalho em que somente dois autores participaram gera uma relação mais forte entre esses dois pesquisadores do que um trabalho em que dez pesquisadores participaram. O contato entre cada pesquisador tende a ser menos intenso quando se tem mais colaboradores.

Neste artigo, apresentamos uma ampliação de análise anterior sobre relações colaborativas entre pesquisadores no Brasil (RODRIGUES; VASCONCELOS; MENA-CHALCO, 2020), mostrando a diferenciação entre identificar um Col em uma relação colaborativa binária e uma normalizada. Para efeito de demonstração dessa diferenciação, utilizamos como estudo de caso os bolsistas de produtividade em pesquisa (PQ) do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), ativos em 2020.

Foram analisados os currículos, obtidos através da Plataforma Lattes, para coletar as publicações e coautorias e suas áreas identificadas através da bolsa produtividade em pesquisa. Neste artigo temos como objetivo: (i) Identificar as diferenças na relação de coautoria entre pesquisadores das áreas distintas através de duas estratégias diferentes, e (ii) apresentar uma proposta de identificação de Col através de redes de coautoria.

2 DESENVOLVIMENTO

Neste artigo apresentamos uma análise exploratória do tipo quantitativa para obtenção dos resultados; inicialmente foram consideradas as informações de todos os PQ's, com bolsa ativa em outubro de 2020. Os PQ's foram escolhidos para representar suas respectivas áreas devido ao reconhecimento do CNPq em relação à atuação desse grupo na ciência. O tipo de bolsa pode variar de nível para cada PQ, sendo as seguintes possibilidades - do maior nível para o menor: SR, 1A, 1B, 1C, 1D e 2. No contexto deste artigo os PQ's de nível SR e todos os níveis 1 são tratados por Avaliadores e os de nível 2 Avaliados.

Os dados para esta análise foram extraídos em fevereiro de 2021, através do Currículo Lattes, onde tiveram as coautorias entre os PQ's identificadas, sendo considerado quatro tipos de produção: (i) Artigos em periódicos, (ii) Artigos em eventos, (iii) Livros e (iv) Capítulos de livros. O período das publicações considerado foi de 10 anos (2011 a 2020) e todas as publicações duplicadas foram removidas. A deduplicação foi baseada na comparação aproximada entre os títulos das produções bibliográficas, similar ao método usado no *scriptLattes* (MENA-CHALCO; CESAR JUNIOR, 2009). Em seguida foi construída uma rede de colaboração entre os PQ's. Após a criação da rede de colaboração contendo todos os PQ's, a rede foi dividida em 85 subredes, uma para cada área de pesquisa do CNPq. Apenas consideramos áreas compostas por mais de 10 PQ's, resultando em um total de 14962 pesquisadores distribuídos em 80 áreas, as áreas com menos de 10 pesquisadores foram Teologia (9), Museologia (6), Biologia Geral (3), Química Industrial (2) e Economia Doméstica (2).

2.1 Métricas para conflito de interesse

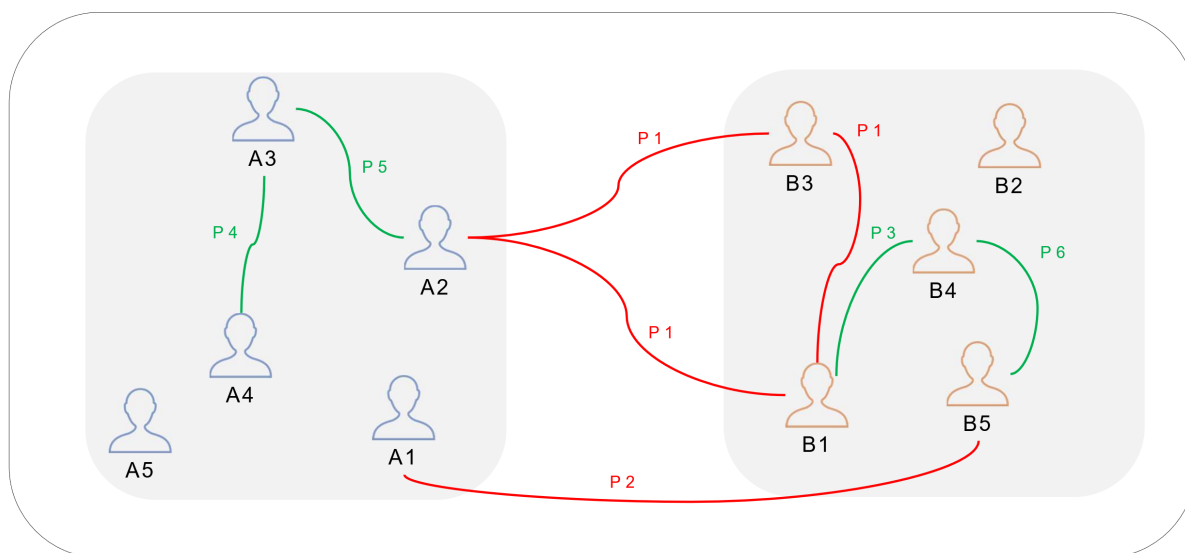
No contexto deste artigo será considerado Col como a relação direta entre um pesquisador do grupo Avaliadores e um pesquisador do grupo Avaliados. Tal relação pode ser

obtida através da análise da rede de colaboração, permitindo a extração de métricas da rede. Essas métricas podem ser utilizadas para identificar e mensurar Col's.

Para efeito de análise, iremos adotar duas estratégias de identificação de Col entre dois pesquisadores. A primeira forma será a relação binária, onde somente o fato de um pesquisador ter alguma coautoria com algum pesquisador do outro grupo já é considerado um potencial Col, ou seja, só há duas possibilidades, ou há Col ou não há Col.

Na segunda estratégia, relação normalizada, o peso da relação também é considerado, como forma de identificar relações em que um pesquisador tem pouca influência sobre outro pesquisador. Com isso, pode-se considerar ou não a existência de Col na relação entre eles. As arestas tiveram seus pesos definidos de acordo com a estratégia de peso normalizado proposta por LIU et al. (2005). Nessa estratégia citada, é considerada a exclusividade do artigo entre os pesquisadores que colaboraram como coautores, a frequência com que os coautores colaboram entre si e a quantidade total de artigos de cada pesquisador para definição de arestas direcionadas com o peso normalizado.

Figura 1 - Exemplo de Col entre Avaliador e Avaliado.



Fonte: Elaborado pelos autores (2021).

Na **Figura 1** é possível observar uma rede representando dois grupos, cada um com 5 integrantes, sendo o grupo A de Avaliadores e o grupo B de Avaliados. As arestas entre os pesquisadores são nomeadas de acordo com a publicação à qual elas representam, sendo

que, caso duas arestas tenham o mesmo nome representam a mesma publicação - nesse caso temos 8 arestas, representando 6 publicações.

Na relação binária, se a publicação representar um Col ela estará na cor vermelha e, caso não represente um Col, estará em verde. Dessa forma, publicações entre o mesmo grupo não representam Col; porém, publicações entre grupos distintos serão consideradas Col (e.g., P1 e P2 envolvem membros do grupo de Avaliadores e do grupo de Avaliados). Como uma publicação pode gerar mais de 1 aresta, caso aconteça de essa publicação gerar arestas entre os grupos e ao mesmo tempo no mesmo grupo, todas as arestas serão consideradas como Col, conforme pode ser observado no caso da P1. Os participantes do grupo de Avaliadores que não tenham nenhuma relação com o grupo de Avaliados serão tratados por Avaliadores Elegíveis; os que tiverem relações serão Avaliadores Inelegíveis.

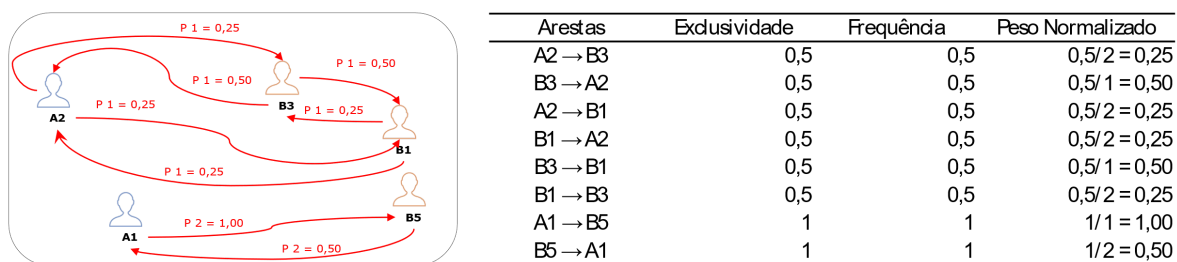
A relação binária, por ser menos flexível, pode gerar dificuldades no processo de identificação de Avaliadores Elegíveis. De fato, a relação entre pesquisadores é algo muito comum e o número de autores por publicação junto com a quantidade de publicações vem aumentando (BORNMANN; MUTZ, 2015) consideravelmente nos últimos anos (WAGNER; PARK; LEYDESDORFF, 2015). Identificar quais relações geram real proximidade entre os pesquisadores e, assim, identificando Col, pode auxiliar no aprimoramento da qualidade desse processo avaliativo. Utilizando a estratégia de arestas com peso normalizado é possível definir um valor no qual indique se há ou não Col. É adotado o uso de arestas direcionadas, pois o peso da aresta de saída de um vértice é ponderado pela quantidade total de publicações daquele vértice.

Para calcular o peso da aresta de saída é necessário identificar a exclusividade de um artigo. Um artigo que tenha apenas 2 autores terá a exclusividade 1 para cada autor. O segundo cálculo realizado é a frequência, que é dada pelo somatório de todas as exclusividades entre 2 autores. Por fim, o peso normalizado é calculado, onde se pondera a frequência pela quantidade total de artigos que um determinado pesquisador tem. Com essa medida, é possível identificar a dependência que um pesquisador tem do outro pesquisador.

Exclusividade	Frequência	Peso normalizado
$g_{i,j,k} = \frac{1}{(f(a_k)-1)}$	$c_{ij} = \sum_{k=1}^m g_{i,j,k}$	$w_{ij} = \frac{c_{ij}}{\sum_{k=1}^n c_{ik}}$

Para exclusividade os autores v_i e v_j são coautores em um artigo a_k e $f(a_k)$ é a quantidade total de autores no artigo, com isso $g_{i,j,k}$ representa o grau de exclusividade que cada autor tem com relação a esse artigo específico. A frequência, representada por c_{ij} , de colaboração entre dois autores é dada pelo somatório de todas as exclusividades das colaborações entre esses autores. Já w_{ij} é o peso normalizado na relação entre os dois autores, no qual é calculado através da exclusividade ponderada pela quantidade total de artigos de cada autor, assim, gerando um valor diferente para cada um, de maneira que o peso da relação de v_i para v_j é diferente do peso da relação de v_j para v_i .

Figura 2 - Exemplo de cálculo do peso normalizado.



Fonte: Elaborado pelos autores (2021).

A **Figura 2** apresenta a forma de cálculo do peso das arestas criadas por P1 e P2, únicas publicações que podem gerar Col. Ao lado esquerdo são apresentadas as arestas direcionadas e os pesquisadores envolvidos nas publicações P1 e P2, cada aresta direcionada é identificada por sua respectiva publicação e o peso normalizado dela. Ao lado direito está a tabela com todo o cálculo.

Fazendo uso da estratégia de peso normalizado é possível definir um valor arbitrário para ser considerado Col, que chamaremos de discriminante de conflito. O discriminante de conflito pode ser um valor variável e deve ser escolhido de maneira a se adequar à necessidade e características da rede a ser analisada. Em nossa rede de exemplo, de maneira ilustrativa, podemos adotar 0,6 como discriminante de conflito para as arestas de saída. Nesse caso, só será considerado Col caso alguma das arestas de saída do Avaliador para um Avaliado tenha peso superior a 0,6, logo, o Avaliador A2 ainda estará em condição de Elegível e somente o Avaliador A1 estará Inelegível para esse exemplo.

Para efeito de comparativo das duas estratégias adotadas na identificação de Col's, foram utilizados 2 indicadores, Coeficiente de Conflito (CC) e Coeficiente de Avaliação (CA). Cada um deles foi aplicado com ambas as estratégias de identificação de Col, gerando duas variações novas dos indicadores: Coeficiente de Conflito Normalizado (CCN) e Coeficiente de Avaliação Normalizado (CAN).

O CC tem por objetivo identificar a quantidade de conflitos (relações entre grupos distintos) existente na área e esse valor é ponderado pela quantidade de Avaliados, tendo como objetivo identificar a densidade de colaboração entre os grupos. Enquanto o CA é calculado através do total de Avaliados dividido pela quantidade de Avaliadores Elegíveis. Esse indicador mostra a quantidade de Avaliadores Elegíveis que tem para cada Avaliado, possibilitando a identificação de áreas em que seria muito difícil, ou até mesmo impossível, conseguir um Avaliador que não tenha relação com nenhum Avaliado de sua mesma área.

3 RESULTADOS

A **Tabela 1** apresenta as áreas com o nome, a quantidade de Avaliadores e Avaliados em cada uma delas; na sequência é apresentada a quantidade de Avaliadores Elegíveis através da estratégia Binária (AEB) seguida pelos resultados dos indicadores CC e CA. Posteriormente, a coluna AEN indica a quantidade de Avaliadores Elegíveis, quando utilizada a estratégia Normalizada e os CCN e CAN calculados a partir dos dados dos PQ's, considerando o período de 2011 à 2020.

Como discriminante de conflito foram utilizados os pesos normalizados das arestas de saída e de entrada das relações entre pesquisadores do grupo de Avaliadores e do grupo de Avaliados. Para definição do valor a ser utilizado, neste estudo de caso, como discriminante de conflito foram identificados os valores fora do padrão, através da divisão em quartis de todas as arestas da rede (LIMA; MAROLDI; DA SILVA, 2013). Como resultado, foi identificado que valores acima de 0,035 eram considerados *outliers*, logo, o valor utilizado como discriminante de conflitos para aresta de saída foi 0,035 (valor correspondente a aproximadamente 90% do total de relações); para as arestas de entrada foi utilizado o dobro desse valor, 0,07. Foi considerado que a relação do Avaliador com o Avaliado tem maior relevância do que o inverso, porém, não podemos desconsiderar o segundo caso.

Adicionalmente, as últimas duas colunas da **Tabela 1** apresentam a razão entre os indicadores calculados com a estratégia binária e com a estratégia de peso normalizado. Essa

razão tem por objetivo mostrar a diferença entre o uso de cada uma das estratégias através de uma escala entre [0,1]. Nos casos em que não houve alterações nos valores, a razão será igual a 1, como pode ser observado nas áreas de Eng. Mecatrônica e Eng. Naval e Oceânica para a Razão CC e CCN (RCC) e Razão CA e CAN (RCA); na área de Desenho Industrial somente para RCA é igual a 1. A **Tabela 1** está ordenada pela RCA.

Tabela 1 - Comparativo do CC e CA com CCN e CAN.

Área	Avaliadores	Avaliados	AEB	CC	CA	AEN	CCN	CAN	RCC	RCA
Odontologia	96	122	1	5,72	122,00	43	0,60	2,84	0,10	0,02
Botânica	82	131	2	6,65	65,50	40	0,47	3,28	0,07	0,05
Ciência e Tecno. de Alimentos	69	114	2	2,84	57,00	37	0,35	3,08	0,12	0,05
Saúde Coletiva	98	120	4	5,10	30,00	60	0,42	2,00	0,08	0,07
Agronomia	358	436	13	5,15	33,54	177	0,61	2,46	0,12	0,07
Medicina	254	291	13	4,30	22,38	173	0,34	1,68	0,08	0,08
Medicina Veterinária	141	154	6	5,01	25,67	77	0,51	2,00	0,10	0,08
Farmácia	76	88	4	2,93	22,00	51	0,30	1,73	0,10	0,08
Química	291	421	14	4,24	30,07	144	0,46	2,92	0,11	0,10
Astronomia	48	66	2	6,24	33,00	19	0,70	3,47	0,11	0,11
Bioquímica	129	114	11	3,49	10,36	88	0,43	1,30	0,12	0,12
Ecologia	102	99	8	7,17	12,38	61	0,53	1,62	0,07	0,13
Eng. Química	64	100	4	2,48	25,00	30	0,50	3,33	0,20	0,13
Zootecnia	101	165	6	4,73	27,50	42	0,53	3,93	0,11	0,14
Oceanografia	56	72	4	3,46	18,00	27	0,53	2,67	0,15	0,15
Eng. de Materiais e Metalúrgica	132	216	10	2,27	21,60	67	0,49	3,22	0,22	0,15
Nutrição	39	50	4	2,24	12,50	25	0,36	2,00	0,16	0,16
Enfermagem	77	102	7	3,06	14,57	40	0,50	2,55	0,16	0,17
Imunologia	94	81	12	4,07	6,75	68	0,42	1,19	0,10	0,18
Fisio. e Terapia Ocupacional	22	46	2	1,63	23,00	11	0,28	4,18	0,17	0,18
Parasitologia	75	69	10	2,99	6,90	53	0,35	1,30	0,12	0,19
Zoologia	106	123	14	4,55	8,79	70	0,42	1,76	0,09	0,20
Fisiologia	97	98	12	2,83	8,17	57	0,52	1,72	0,18	0,21
Geociências	214	253	18	3,51	14,06	85	0,71	2,98	0,20	0,21
Psicologia	125	188	14	2,30	13,43	66	0,45	2,85	0,20	0,21
Administração	45	160	4	0,90	40,00	18	0,21	8,89	0,24	0,22
Genética	140	127	22	2,76	5,77	98	0,39	1,30	0,14	0,22
Arqueologia	19	32	3	0,69	10,67	13	0,22	2,46	0,32	0,23
Ciência da Computação	142	275	10	2,60	27,50	43	0,67	6,40	0,26	0,23
Rec. Pesqueiros e Eng. de Pesca	47	56	6	2,43	9,33	25	0,52	2,24	0,21	0,24
Farmacologia	107	92	19	2,95	4,84	79	0,35	1,16	0,12	0,24
Educação Física	41	53	5	3,62	10,60	20	0,57	2,65	0,16	0,25
Microbiologia	101	102	18	2,56	5,67	71	0,36	1,44	0,14	0,25
Rec. Florestais e Eng. Florestal	54	92	7	2,72	13,14	27	0,42	3,41	0,16	0,26
Eng. de Produção	52	111	6	1,30	18,50	23	0,42	4,83	0,33	0,26
Física	449	523	53	3,98	9,87	203	0,71	2,58	0,18	0,26
Ciências Ambientais	30	102	7	0,50	14,57	25	0,07	4,08	0,14	0,28
Eng. Agrícola	57	82	9	2,18	9,11	30	0,46	2,73	0,21	0,30
Ciência da Informação	16	43	4	0,91	10,75	12	0,12	3,58	0,13	0,33
Eng. Sanitária	46	64	12	1,25	5,33	33	0,27	1,94	0,21	0,36
Biotecnologia	41	77	13	1,09	5,92	35	0,12	2,20	0,11	0,37
Fonoaudiologia	17	35	3	1,09	11,67	8	0,34	4,38	0,32	0,37
Eng. Elétrica	155	138	33	2,18	4,18	79	0,75	1,75	0,34	0,42
Comunicação	49	89	18	0,56	4,94	43	0,09	2,07	0,16	0,42
Biofísica	41	52	15	0,98	3,47	35	0,15	1,49	0,16	0,43
Morfologia	56	70	15	1,46	4,67	35	0,36	2,00	0,25	0,43
Eng. Civil	129	155	33	1,68	4,70	77	0,48	2,01	0,29	0,43
Eng. Mecânica	131	157	33	1,46	4,76	70	0,48	2,24	0,33	0,47
Microeletrônica	34	42	9	1,40	4,67	19	0,50	2,21	0,36	0,47
Eng. Nuclear	45	34	16	1,53	2,13	33	0,44	1,03	0,29	0,48
Direito	19	65	8	0,37	8,13	16	0,05	4,06	0,12	0,50
Turismo	4	15	1	0,33	15,00	2	0,13	7,50	0,40	0,50
Arquitetura e Urbanismo	33	83	13	0,54	6,38	25	0,12	3,32	0,22	0,52
Educação	187	230	82	1,09	2,80	153	0,16	1,50	0,15	0,54
Antropologia	64	110	27	0,73	4,07	50	0,15	2,20	0,21	0,54
Sociologia	103	118	43	1,02	2,74	79	0,25	1,49	0,24	0,54
Ciência Política	44	100	17	0,46	5,88	30	0,16	3,33	0,35	0,57
Economia	71	151	18	1,05	8,39	31	0,41	4,87	0,39	0,58
Eng. Aeroespacial	22	39	6	0,95	6,50	10	0,33	3,90	0,35	0,60
Linguística	101	134	49	0,77	2,73	80	0,21	1,68	0,27	0,61
Demografia	15	26	5	1,38	5,20	8	0,38	3,25	0,28	0,63
Serviço Social	31	53	14	0,79	3,79	22	0,17	2,41	0,21	0,64
Divulgação Científica	5	20	2	0,25	10,00	3	0,15	6,67	0,60	0,67
Eng. de Energia	14	23	6	0,43	3,83	9	0,22	2,56	0,50	0,67
História	99	167	59	0,38	2,83	85	0,10	1,96	0,27	0,69
Eng. de Transportes	25	26	5	1,62	5,20	7	0,92	3,71	0,57	0,71
Letras	106	133	70	0,46	1,90	97	0,08	1,37	0,16	0,72
Geografia	45	50	29	0,50	1,72	40	0,10	1,25	0,20	0,72
Planejamento Urbano e Regional	36	54	23	0,31	2,35	30	0,11	1,80	0,35	0,77
Biomedicina	11	25	8	0,20	3,13	10	0,04	2,50	0,20	0,80
Eng. de Minas	9	13	4	0,92	3,25	5	0,31	2,60	0,33	0,80
Probabilidade e Estatística	31	33	12	1,30	2,75	15	0,70	2,20	0,53	0,80
Rob. Mecatrônica e Automação	8	9	4	0,56	2,25	5	0,33	1,80	0,60	0,80
Eng. Biomédica	42	17	29	1,06	0,59	36	0,41	0,47	0,39	0,81
Filosofia	80	83	56	0,40	1,48	67	0,17	1,24	0,42	0,84
Artes	47	74	36	0,23	2,06	42	0,07	1,76	0,29	0,86
Matemática	164	169	73	1,05	2,32	83	0,73	2,04	0,70	0,88

Fonte: Elaborado pelos autores (2021).

Para melhor compreensão dos valores, podemos destacar a área de Ciência da Informação. Na **Tabela 1** essa área está na posição 39 das 80 áreas contempladas, logo, pode ser considerada uma área em situação mediana ao analisar o RCA. A Ciência da Informação possui 16 Avaliadores e 43 Avaliados, totalizando 59 pesquisadores. Dos 16 Avaliadores, somente 4 não tiveram nenhum relacionamento com ao menos um dos Avaliados, com isso somente esses 4 estão como Elegíveis ao adotar a estratégia binária. Quando calculamos os indicadores baseados na estratégia binária teremos $CC = 0,91$ (existem 150 relações na área, das quais 39 são entre Avaliados e Avaliadores, logo, o cálculo do CC é: $CC = 39/43$) e $CA = 10,75$ ($CA = 43/4$). Olhando para os indicadores na estratégia binária podemos ver que temos uma densidade de relações que podem indicar Col de 0,91 e que cada Avaliador Elegível teria que avaliar 10,75 Avaliados, o que normalmente é inviável.

Já com o uso da estratégia normalizada, somente as relações de saída de um Avaliador para um Avaliado com peso superior à 0,035 ou relações de entrada com o peso acima 0,07 foram consideradas Col. O total de Avaliadores Elegíveis é 12, indicando que 8 dos Avaliadores, que na estratégia binária ficaram Inelegíveis, tinham relações de baixo peso com o grupo de Avaliados. Assim, temos o $CCN = 0,12$ (apenas 5 das relações entre os grupos estavam acima do discriminante de conflito, logo, $CCN = 5/43$) e $CAN = 3,58$ ($CAN = 43/12$). Com a estratégia normalizada, é notável que a densidade de relações que podem indicar Col reduziu, ficando em 0,12, e cada Avaliador Elegível, nesse caso, teria que avaliar 3,58 Avaliados, uma situação normalmente viável.

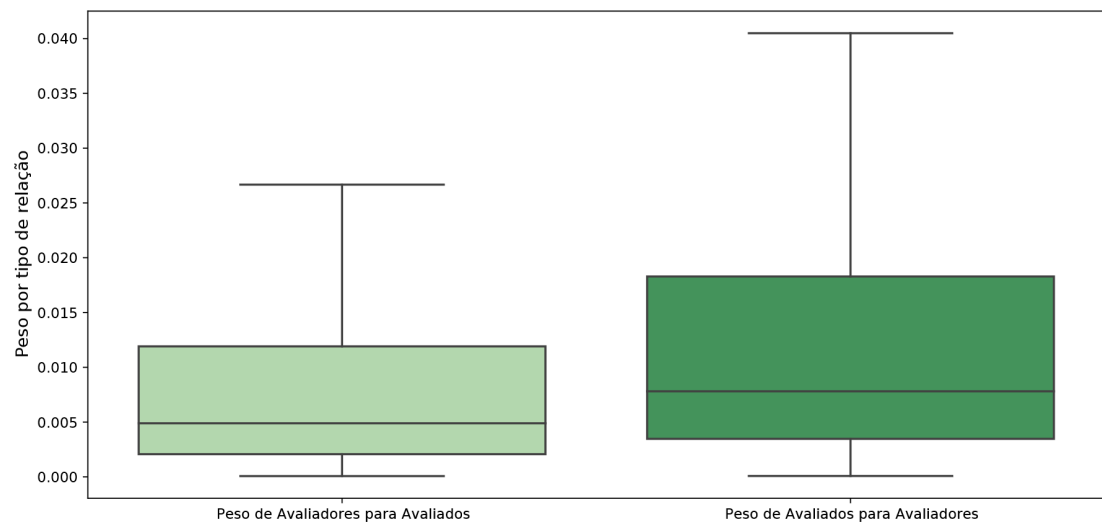
Fazendo uso dos indicadores calculados através das duas estratégias, é possível calcular a razão entre eles, sendo, $RCC = 0,13$ ($RCC = 0,12/0,91$) e $RCA = 0,33$ ($RCA = 3,58/10,75$). Onde, tanto o RCC quanto o RCA, quanto mais próximos à 0 maior foi a alteração entre os indicadores calculados com ambas as estratégias.

A área que teve maior alteração de RCA é Odontologia, tendo $RCA=0,02$, evidenciando que a área tem grande colaboração entre os grupos; porém, poucas colaborações com grande peso. Com o uso da estratégia binária para identificação de Col somente 1 dos 122 Avaliadores estaria em condição de Elegível. Porém, com o uso do peso normalizado, 43 Avaliadores ficaram Elegíveis. As áreas de Botânica e Ecologia ficaram com $RCC=0,07$, áreas com maior alteração na taxa de conflitos.

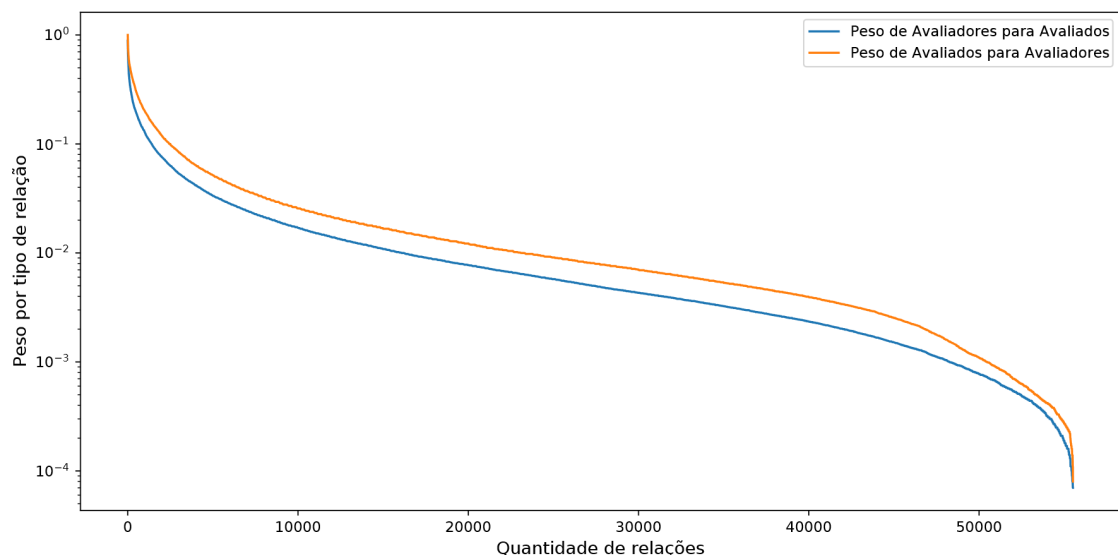
No comparativo entre CA e CAN, nota-se que 52 áreas apresentam $CA \geq 5$, enquanto apenas 6 áreas apresentam $CAN \geq 5$; é indicativo que, em caso de identificação de Col através da estratégia binária, em 52 áreas cada Avaliador teria que avaliar 5 ou mais Avaliados. Porém, com a estratégia normalizada, somente 5 áreas teriam esse problema. O CA médio das áreas foi de 13,26 e o CAN médio é 2,99.

Considerando o CC e o CCN, nenhuma área teve $CCN \geq 1$, logo, todas as áreas possuem menos de 1 conflito por Avaliado com a estratégia de peso normalizado. Já na estratégia binária, 52 áreas apresentam $CC \geq 1$. Note que, considerando todas as áreas o CC médio é 2,06, enquanto o CCN é de 0,37.

Figura 3 - Exemplo de Col entre Avaliador e Avaliado.



(a)



(b)

Fonte: Elaborado pelos autores (2021).

Os gráficos da **Figura 3** indicam a incidência e distribuição das arestas, entre os pesquisadores de grupos distintos, por peso. No gráfico **(a)**, *boxplot*, são apresentadas duas caixas, sendo a verde clara, à esquerda, referente ao peso das relações dos Avaliadores para os Avaliados e a caixa da direita, verde escuro, referente ao peso das relações dos Avaliados para os Avaliadores. É possível identificar que as relações de Avaliadores para Avaliados possuem mediana e limite superior menores do que no caso das relações de Avaliados para Avaliadores, indicando que, de maneira geral, os Avaliadores têm maior influência sobre os Avaliados do que o inverso.

Já no gráfico **(b)**, diagrama de linhas, é apresentada a quantidade de relações por peso, onde o eixo vertical apresenta o peso em escala logarítmica e o eixo horizontal apresenta a quantidade de relações. Podemos notar que, em ambos os casos, o formato da distribuição dos pesos se mantém, porém, com valores menores para as relações de Avaliadores para Avaliados.

4 CONSIDERAÇÕES FINAIS

Neste trabalho foram apresentadas duas propostas para identificação de Col através da análise de redes colaborativas de pesquisa. A estratégia binária é mais rígida e desconsidera que algumas relações entre pesquisadores são muito comuns e não geram fortes vínculos entre os participantes; dessa forma, pode-se reduzir a quantidade de Avaliadores Elegíveis e excluir pesquisadores com uma rede muito extensa, mesmo que não haja forte vínculo entre os grupos. Com a adoção da estratégia normalizada e uso de um discriminante de conflito é possível identificar as relações entre Avaliadores e Avaliados que estejam acima dos padrões de outras relações; sendo assim, expressam maiores indícios de Col e merecem maior atenção. Com o uso dessa estratégia mais flexível que a estratégia binária para identificação de Col's é possível diminuir o número de conflitos aparentes na análise da rede de colaboração e aumentar o número de Avaliadores Elegíveis através de parâmetros definidos no momento da análise.

A análise dos dados revela que as relações entre os PQ's sofrem grandes variações entre as áreas. Entretanto, mesmo em áreas com grande densidade de colaboração e muita colaboração entre Avaliadores e Avaliados, a maioria dos PQ's ainda não apresentam alto

grau de dependência direta com apenas um pesquisador. Em sua maioria, as relações são distribuídas entre os diversos pesquisadores. De uma maneira geral, os Avaliados têm relações de maior intensidade com os Avaliadores do que o inverso, apresentando média e mediana mais altas do que a das relações de Avaliadores para os Avaliados.

Com o uso das duas estratégias, binária e normalizada, foi possível identificar que cada área possui suas próprias características. Sendo que houve áreas onde ao adotar estratégias distintas não sofreram alteração em seus indicadores, é o caso das áreas de Eng. Mecatrônica e Eng. Naval e Oceânica, quando considerado RCC, e da Desenho Industrial, quando considerado o RCA. Enquanto outras áreas tiveram mudanças significativas, como foi o caso da Odontologia, que ao utilizar a estratégia binária possuía apenas 1 Avaliador Elegível, enquanto que, com o uso da estratégia normalizada, passou a ter 43 Avaliadores Elegíveis.

Neste trabalho, foi analisado um grupo de pesquisadores referência, não refletindo, necessariamente, os moldes de toda a ciência brasileira. Como próximos passos, pretendemos realizar experimentos com outros grupos, como a base completa dos doutores cadastrados na Plataforma Lattes e pesquisadores cadastrados na base internacional DBLP. Outras formas de abordagem também podem ser consideradas em trabalhos futuros, como por exemplo: considerar a idade acadêmica do pesquisador para ponderar suas relações; identificar os padrões existentes para cada área do conhecimento e com isso personalizar o discriminante de conflito de maneira mais adequada; priorizar Avaliadores que não estejam sobrecarregados no momento de indicação; ou análise de redes de bancas ao invés de redes de colaboração em publicações.

É importante ressaltar que esta rede de colaboração é composta, apenas, por um tipo de relação e ainda só considerando as relações que obtiveram sucesso entre os pesquisadores. Não há elementos suficientes para identificar todos os tipos de Col's possíveis, até porque, muitos desses Col's podem acontecer de maneira subjetiva, onde somente os envolvidos poderiam identificar (e.g., participações em organização de eventos e/ou workshops).

O processo de identificação de avaliadores é uma tarefa complexa e, com o crescente volume de produções, estratégias que possam reduzir os Col's são necessárias. A identificação de relações entre pesquisadores para grandes volumes de avaliações é inviável

de ser realizada através de uma análise exclusivamente humana, embora o aval final sempre dependa de indivíduos.

Esta abordagem apresentada amplia o escopo de análise e indica formas de direcionar os esforços para os principais pontos de atenção através de relações com maior peso, considerando a relevância e potencial de viés advindos de Col's na avaliação científica, possibilitando uma melhor compreensão sobre algumas peculiaridades, relativas às áreas, por exemplo na área de Administração, que ao adotar a estratégia de binária cada Avaliador teria que avaliar 40 Avaliados, com a estratégia normalizada esse número reduz para 8,89, com isso mostrando que há uma quantidade significativa de relações com peso acima do discriminante de conflito. Por outro lado, na Matemática cada Avaliador iria precisar avaliar 2,32 Avaliados na estratégia binária, enquanto na estratégia normalizada seria 2,04, indicando que a maioria das relações tem peso inferior ao discriminante de conflito. Desta forma pode-se notar que cada área tem um comportamento em relação às suas colaborações entre pesquisadores.

Os resultados indicam uma possibilidade de redução de tarefas operacionais e melhoria em processos de avaliação e, com isso, somando-se aos vários esforços em curso para aprimorar a confiabilidade e a integridade da ciência, de forma mais ampla. De acordo com Ethique et al. (2019) “Em muitos países, os pesquisadores são cada vez mais chamados individualmente ou coletivamente para avaliar seus colegas. O risco de que os avaliadores de pares se encontrem em uma situação de conflito de interesses está aumentando proporcionalmente”, o que nos remete de maneira mais incisiva à necessidade de apoio de recursos computacionais para auxiliar na tarefa de identificação de Col em grandes volumes de avaliações.

REFERÊNCIAS

ABDOUL, H.; PERREY, C.; TUBACH, F.; AMIEL, P.; DURAND-ZALESKI, I.; ALBERTI, C. Non-financial conflicts of interest in academic grant evaluation: a qualitative study of multiple stakeholders in France. **PLoS One**, v. 7, n. 4, p. e35247, 2012.

BORNMANN, L.; MUTZ, R. Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. **Journal of the Association for Information Science and Technology**, v. 66, n. 11, p. 2215-2222, 2015.

BUCCHI, M. Norms, competition and visibility in contemporary science: The legacy of Robert K. Merton. **Journal of Classical Sociology**, v. 15, n. 3, p. 233-252, 2015.

CHECCO, A.; BRACCIALE, L.; LORETI, P.; PINFIELD, S.; BIANCHI, G. AI-assisted peer review. **Humanities and Social Sciences Communications**, v. 8, n. 1, p. 1-11, 2021.

DA SILVA, C.; MOREIRO-GONZALEZ, J.; MUELLER, S. A revisão por pares a partir da percepção dos editores: um estudo comparativo em revistas brasileiras, espanholas e mexicanas. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, v. 14, n. 1, p. 126-143, 2016.

ETHIQUE, C.; ASKENAZY, P.; DIDIER, G.; MICHÈLE, L.; LUCIENNE, L.; JEAN-PIERRE, P.; GABRIEL, G. **Comets opinion 2019-39-interests and conflicts of interest in public research**. 2019.

FELDMAN, R.; DEVITO, J.; MENDEL, J.; CARROLL, D.; GOLDACRE, B. A cross-sectional study of all clinicians' conflict of interest disclosures to NHS hospital employers in England 2015-2016. **BMJ open**, v. 8, n. 3, 2018.

MENA-CHALCO, J.; CESAR JUNIOR, R. ScriptLattes: an open-source knowledge extraction system from the Lattes platform. **Journal of the Brazilian Computer Society**, v. 15, n. 4, p. 31-39, 2009.

LIMA, L.; MAROLDI, A.; DA SILVA, D. Outlier (s) nos cálculos bibliométricos: primeiras aproximações. **Liinc em revista**, v. 9, n. 1, 2013.

LIU, X.; BOLLEN, J.; NELSON, M.; SOMPEL, H. Co-authorship networks in the digital library research community. **Information processing & management**, v. 41, n. 6, p. 1462-1480, 2005.

PATRUS, R.; DANTAS, D.; SHIGAKI, H. Pesquisar é preciso. Publicar não é preciso: história e controvérsias sobre a avaliação por pares. **Avaliação: Revista da Avaliação da Educação Superior (Campinas)**, v. 21, n. 3, p. 799-820, 2016.

RESNIK, D.; ELMORE, S. Conflict of interest in journal peer review. **Toxicol Pathol**, v. 46, n. 2, p. 112-114, 2018.

RODRIGUES, W.; VASCONCELOS, S.; MENA-CHALCO, J. DESAFIOS ÉTICOS NO ÂMBITO DA AVALIAÇÃO DE PROJETOS: um estudo de caso com foco em conflitos de interesse. **Encontro Brasileiro de Bibliometria e Cientometria**, p. 239-247, 2020.

SQUAZZONI, F.; BRAVO, G.; FARJAM, M.; MARUSIC, A.; MEHMANI, B.; WILLIS, M.; GRIMALDO, F. Peer review and gender bias: A study on 145 scholarly journals. **Science Advances**, v. 7, n. 2, p. eabd0299, 2021.

WAGNER, C.; PARK, H.; LEYDESDORFF, L. The continuing growth of global cooperation networks in research: A conundrum for national governments. **PloS one**, v. 10, n. 7, p. e0131816, 2015.

ZUCKERMAN, H.; MERTON, R. Patterns of evaluation in science: Institutionalisation, structure and functions of the referee system. **Minerva**, p. 66-100, 1971.