

# **Modelo de Agrupamento de Pesquisadores e seu Coeficiente de Similaridade: Um Estudo de Caso com Docentes das Universidades Estaduais da Bahia**

**Aldo M. Barbosa <sup>1</sup>, Gleidson M. Costa <sup>2</sup>, Cíntia S. da Conceição <sup>2</sup>, Victor Hugo de J. Oliveira <sup>3</sup>, Rafael do N. Argôlo <sup>1</sup>, João Victor Cafe dos R. Batista <sup>1</sup>, Leandro B. Santos <sup>2</sup>, Eduardo Manuel de F. Jorge <sup>1</sup>**

<sup>1</sup> Departamento de Ciência Exatas e da Terra – Universidade do Estado da Bahia (UNEB)  
– Salvador – BA – Brasil

<sup>2</sup> Universidade Federal do Recôncavo da Bahia (UFRB)  
– Feira de Santana, BA – Brasil

<sup>3</sup> Universidade Federal de Minas Gerais (UFMG)  
– Belo Horizonte, MG – Brasil

ambarbosa@uneb.br, geucosta@outlook.com,  
cintiaconceicao@aluno.ufrb.edu.br,  
victorhugodejesusoliveira@gmail.com, rafaelnargolo@outlook.com,  
batistajv2012@gmail.com, lbsantos@ufrb.edu.br, ejorge@uneb.br

**Abstract.** *The evaluation of researchers is essential for the effective management of academic institutions. However, traditional bibliometric indicators present limitations, such as the difficulty of comparing researchers from different areas and the complexity in interpreting the results. This study proposes a classification model and grouping of researchers that considers multiple criteria, including graduation time, scientific production and student orientation. Based on these data from two researchers from state universities in Bahia, we seek to identify standards that validate scientific production and assist in the strategic decisions of managers, providing concise feedback to researchers.*

**Resumo.** *A avaliação de pesquisadores é fundamental para a gestão eficaz de instituições acadêmicas. No entanto, os indicadores bibliométricos tradicionais apresentam limitações, como a dificuldade de comparar pesquisadores de diferentes áreas e a complexidade na interpretação dos resultados. Este estudo propõe um modelo de classificação e agrupamento de pesquisadores que considera múltiplos critérios, incluindo tempo de titulação, produção científica e orientação de alunos. A partir desses dados dos pesquisadores das universidades estaduais da Bahia, busca-se identificar padrões que avaliem a produção científica e auxiliem nas decisões estratégicas dos gestores, fornecendo feedback conciso aos pesquisadores.*

## **1. Introdução**

Os indicadores bibliométricos são amplamente utilizados na avaliação da produção científica, fornecendo métricas quantitativas que mensuram impacto, relevância e visibilidade de pesquisadores e publicações. Esses indicadores identificam autores e periódicos influentes, orientando decisões estratégicas em instituições de pesquisa. Contudo, há carência de soluções integradas para consulta e organização da produção científica. A multiplicidade de bases de dados dificulta a recuperação eficiente de informações, como apontado por Jorge et al. (2020) no caso da UNEB, cuja produção acadêmica é volumosa e dispersa.

A heterogeneidade de formatos e padrões das bases de dados agrava a complexidade do processo. A ausência de padronização nos metadados e o uso de vocabulários controlados distintos comprometem a interoperabilidade dos sistemas, limitando buscas integradas e a criação de repositórios unificados. Embora essenciais, os indicadores bibliométricos, quando utilizados isoladamente, apresentam limitações, como a incapacidade de capturar aspectos qualitativos da produção científica ou especificidades de áreas de pesquisa.

Com o aumento da competitividade por financiamentos, a qualidade das publicações tornou-se um critério fundamental para demonstrar relevância acadêmica. Indicadores como o Índice h, Journal Citation Reports (JCR) e o sistema Qualis Capes são essenciais no Brasil para políticas de alocação de recursos e avaliação institucional (Mugnaini, 2004). Eles também orientam a classificação de pesquisadores e a concessão de financiamentos por agências como Capes, Finep e CNPq.

A evolução tecnológica e o crescimento da produção científica exigem soluções flexíveis e escaláveis que garantam não apenas a recuperação organizada da informação, mas também sua preservação a longo prazo. Entretanto, o uso excessivo de métricas quantitativas pode distorcer comportamentos, levando os pesquisadores a priorizar números em detrimento da qualidade e inovação. Além disso, a padronização das métricas desconsidera variações entre disciplinas, demandando modelos de avaliação mais robustos e multicritérios.

Este artigo propõe um modelo de classificação e agrupamento de pesquisadores baseado em análise multicritério, considerando tempo de titulação, produção técnica e científica, e formação de recursos humanos na pós-graduação. Para validar o modelo, foi realizado um experimento com dados de pesquisadores das universidades estaduais da Bahia, utilizando informações dos últimos 10 anos extraídas do Sistema de Mapeamento de Competências Científicas da Bahia (SIMCC, 2023).

## **2. Índices bibliométricos**

A avaliação do impacto de pesquisadores e periódicos na ciência utiliza diversas métricas, sendo o índice-h e o Fator de Impacto (FI) duas das mais conhecidas, cada uma com suas limitações e especificidades.

O índice-h combina quantidade e relevância das publicações de um pesquisador, mas deve ser utilizado com cautela, especialmente quando comparado entre diferentes áreas científicas. A superestimação deste indicador pode levar a avaliações inadequadas, como observado em críticas à utilização exclusiva do índice-h para ranquear pesquisadores (Egghe, 2006). É necessário complementar essa métrica com outras, para

uma avaliação mais justa e abrangente. Além disso, pesquisas inovadoras, que recebem menos citações, podem ser subvalorizadas por este indicador.

O Fator de Impacto (FI), criado por Eugene Garfield, mede a relevância de periódicos científicos pela média de citações dos artigos dos dois anos anteriores (UNIVERSIDADE FEDERAL DE UBERLÂNDIA, 2024). Embora útil, o FI não reflete a qualidade individual dos artigos. O Journal Citation Report (JCR) no Portal Capes fornece o FI dos periódicos indexados na Web of Science (UNIVERSIDADE FEDERAL DE UBERLÂNDIA, 2024). Tanto o índice-h quanto o FI variam entre áreas e devem ser usados com outras métricas. O Eigenfactor, por exemplo, pondera citações pela relevância da revista e elimina auto-citações. A combinação dessas métricas oferece uma análise mais robusta da produção científica.

### 3. Metodologia

A Tabela 1 apresenta os parâmetros utilizados para classificar os pesquisadores em níveis que variam de E (mais baixo) a A+ (mais alto). Sendo realizada por meio de uma análise multicritério que considera o tempo desde a conclusão do doutorado, a produção acadêmica e técnica, além da experiência do pesquisador em programas de pós-graduação Stricto Sensu.

**Tabela 1. Parâmetros para classificar pesquisadores**

Nível	Tempo de Doutorado (em anos)	Produção Acadêmica*	Experiência PG (Orientação)	Regra Combinada
A+	≥ 10	≥ 2 A1	≥ 4 Doutorados	≥ 1 A1 + ≥ 1 patente
A	≥ 10	≥ 1 A1	≥ 2 Doutorados	≥ 1 patente
B+	≥ 8	≥ 2 A*	≥ 2 Mestrados ou ≥ 1 Doutorado	≥ 1 A* + (≥ 1 patente ou ≥ 3 softwares)
B	≥ 8	≥ 1 A*	≥ 2 Mestrados ou ≥ 1 Doutorado	≥ 1 patente ou ≥ 3 softwares
C+	≥ 6	≥ 2 A*	≥ 1 Mestrado ou ≥ 1 Doutorado	≥ 1 A* + (≥ 1 patente ou ≥ 3 softwares)
C	≥ 6	≥ 1 A*	≥ 1 Mestrado ou ≥ 1 Doutorado	≥ 1 patente ou ≥ 3 softwares
D+	≥ 3	≥ 1 A*	-	≥ 1 patente ou ≥ 3 softwares
D	≥ 3	≥ 1 B**	-	≥ 1 patente ou ≥ 3 softwares
E+	< 3	≥ 1 B**	-	≥ 1 A* ou ≥ 1 patente ou ≥ 3 softwares
E	Todos que não foram classificados			

\*Estrato Qualis A1, A2, A3 ou A4

\*\*Estrato Qualis B1, B2, B3 ou B4

A definição multicritério dos parâmetros da Tabela 1 baseou-se nos seguintes documentos: a) Documento de área da Capes para a área 45 (Interdisciplinar); b) Chamada pública CNPq 04/2021 para Bolsas de Produtividade em Pesquisa (PQ), que valoriza pesquisadores com produção científica e tecnológica de destaque; c) Chamada CNPq 05/2021 para Bolsas de Produtividade Sênior (PQ-Sr), que reconhece líderes em suas áreas; d) Chamada 04/2023 para Bolsas de Desenvolvimento Tecnológico e Extensão Inovadora (DT), que valoriza participação em atividades de desenvolvimento tecnológico; e) Referência do Grupo de Trabalho Produção Técnica da CAPES, criado em 2018, para desenvolver metodologia de avaliação da produção técnica e tecnológica.

A produção acadêmica dos pesquisadores, medida pela quantidade de artigos e qualificação no Qualis, considera os últimos quatro anos, correspondendo ao período de avaliação quadrienal da Capes. A análise está na ficha de avaliação da área interdisciplinar, que calcula o indicador de produção intelectual em periódicos (IndProdArt).

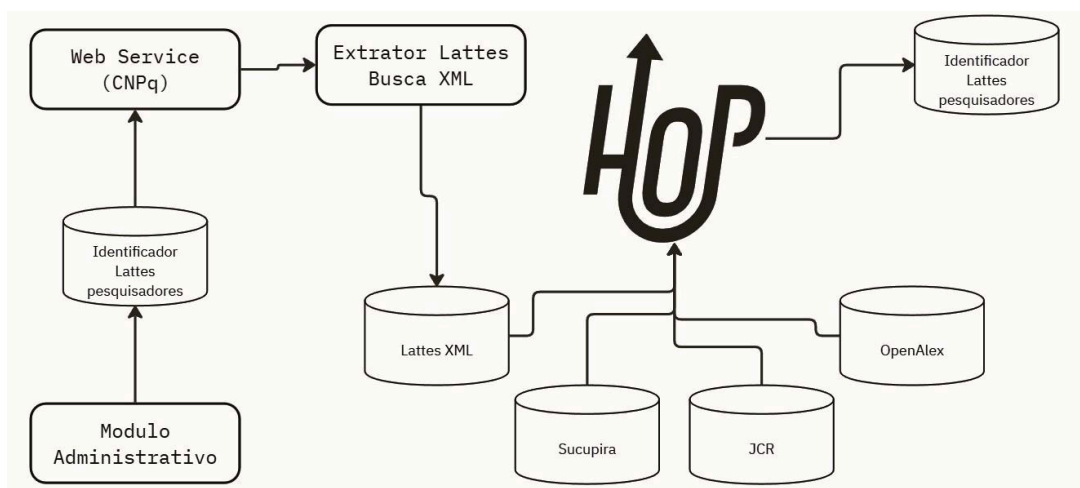
O tempo de doutorado, análise do estrato Qualis e experiência na pós-graduação Stricto Sensu seguem o Edital CNPq N° 04/2021, que exige 3 anos (Pesquisador 2), 8 anos (Pesquisador 1D) e 10 anos (Pesquisador 1C e 1B), com uma faixa intermediária de 3 a 6 anos. O edital também valoriza artigos bem avaliados no Qualis-Capes e exige pelo menos duas orientações concluídas para níveis mais elevados de bolsas.

Além desses critérios, há uma regra que permite substituir a produção acadêmica por produção técnica, como patentes e softwares. Se o pesquisador não atingir a produção acadêmica necessária, pode usar essa substituição. Nos níveis A+ e A, um artigo A1 pode ser trocado por uma patente; nos níveis B+, B, C+, C e D+, qualquer artigo pode ser substituído por uma patente ou três softwares.

#### **4. Sistema de Mapeamento de Competências Científicas da Bahia**

Para este estudo, utilizou-se a plataforma SIMMC (Sistema de Mapeamento de Competências Científicas da Bahia) como base de dados, apoiando gestores na análise da produção acadêmica de cerca de 5 mil docentes das quatro Universidades Estaduais da Bahia (Uebas) (SANTOS et al., 2024). A base inclui informações sobre doutorado, produção acadêmica, orientações e qualificação de periódicos, extraídas das Plataformas Lattes e Sucupira, sob gestão da Capes. A Plataforma Lattes organiza currículos de pesquisadores brasileiros, enquanto a Plataforma Sucupira avalia periódicos científicos. A Figura 1 mostra a integração de dados no SIMMC, onde informações dos currículos Lattes são processadas via ETL (Extração, Transformação e Carga) e inseridas no banco de dados do SIMMC, incluindo estratos Qualis da Plataforma Sucupira (SANTOS et al., 2024).

A integração e orquestração de dados, que combina informações de diferentes fontes para análise e apoio à tomada de decisões, foi implementada utilizando a ferramenta Apache Hop (The Hop Team, 2019).

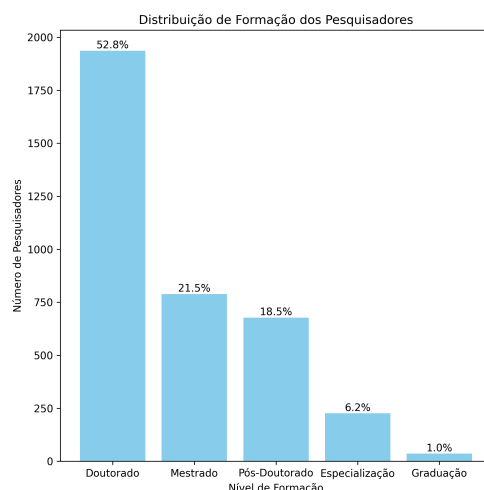


**Fig 1. Orquestração de Dados e Data Lake**

A próxima seção caracteriza o banco de dados dos pesquisadores das universidades estaduais da Bahia, destacando as características dos docentes, suas produções acadêmicas e a relação entre variáveis como orientações acadêmicas e participações em eventos, servindo como referência para o modelo de classificação.

#### 4.1. Caracterização da Base de Dados

A extração foi realizada em setembro de 2024, levando em conta a produção dos últimos 10 anos, e totalizou 3666 docentes distribuídos entre as universidades estaduais da Bahia, UNEB, UESB, UEFS e UESC. Nessa extração, o dataset possui pesquisadores enquadrados pelo critério de titulação/estágio pós-doutoral apresentado na Figura 2.

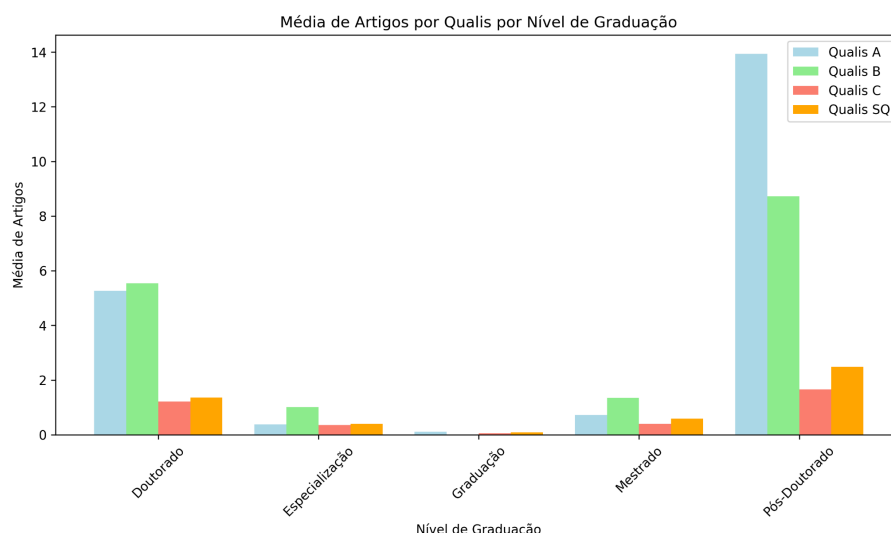


**Fig 2. : Distribuição de Formação dos Pesquisadores**

A distribuição dos pesquisadores por nível de formação destaca a predominância de doutores, com 1.937 indivíduos, representando 52,8% da amostra. Em seguida, 789

mestres (21,5%) e 678 pós-doutores (18,5%) compõem a base de dados. Além disso, 226 pesquisadores possuem especialização (6,2%), enquanto 36 têm graduação (1%).

Além da formação dos docentes, a análise revelou a produção acadêmica dos pesquisadores. Foram desenvolvidas 167 patentes, com 24 concedidas, e a produção literária inclui 154 livros e 545 capítulos. Também foram criados 22 softwares. Além disso, contabilizamos 29.315 artigos, que serão agrupados na Figura 3 conforme o nível de formação do pesquisador e classificados pelo Qualis.



**Fig 3. Média de Artigos por Qualis por Nível de Graduação**

A análise da média de artigos por Qualis revela padrões significativos na produção científica dos pesquisadores. Doutores destacam-se com uma média de 5,27 artigos no Qualis A, enquanto pós-doutores têm uma média ainda maior, de 13,93 artigos. Em contrapartida, mestres e especialistas apresentam médias modestas, com 0,72 e 0,38 artigos, respectivamente, e graduados têm a menor média, com apenas 0,11 artigos.

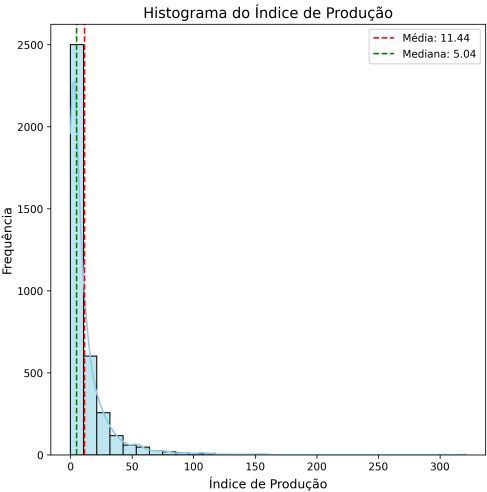
Muitos graduandos por estarem no início de suas carreiras de pesquisa, sua produção científica é limitada até que adquiram mais experiência e estabeleçam suas trajetórias acadêmicas. Essa fase inicial envolve o desenvolvimento de habilidades e a participação em projetos de pesquisa, levando tempo até se traduzir em publicações. Vale ressaltar que a presença de valores mínimos ou próximos de zero podem ser explicados também pela falta de atualização dos currículos, algo comum no início de suas jornadas e que afeta diretamente as métricas dado a natureza da coleta.

A Tabela 2 apresenta as estatísticas descritivas do índice de produção dos pesquisadores, calculado conforme o barema da CAPES descrito no Relatório de Avaliação da Área Interdisciplinar para programas de graduação multidisciplinares (BRASIL, 2022). O índice foi adaptado para análise individual, ponderando diferentes tipos de produções acadêmicas, como artigos, livros e capítulos, de acordo com os critérios da CAPES.

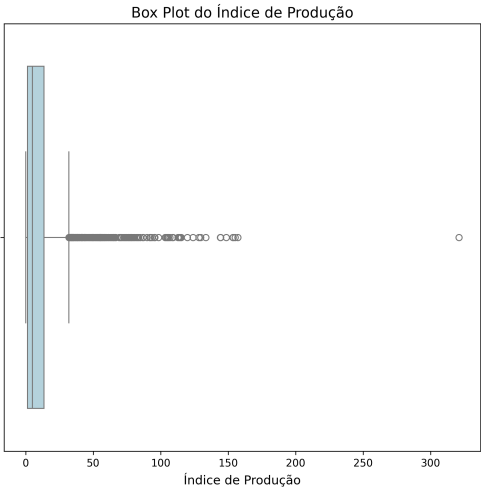
Os dados indicam uma distribuição altamente assimétrica à direita, com um coeficiente de assimetria de 4,37. Isso evidencia que a maioria dos pesquisadores apresenta índices de produção em níveis baixos, enquanto poucos pesquisadores alcançam índices muito altos, resultando em uma cauda pronunciada à direita da distribuição.

**Tabela 2. Estatísticas descritivas do índice de produção dos pesquisadores**

Estatística	Valor
Mínimo	0.00
1º Quartil	1.37
Média	11.44
Mediana	5.03
3º Quartil	13.62
Máximo	320.97
Coeficiente de Assimetria	4.37



**Fig 4. Histograma do índice de Produção**

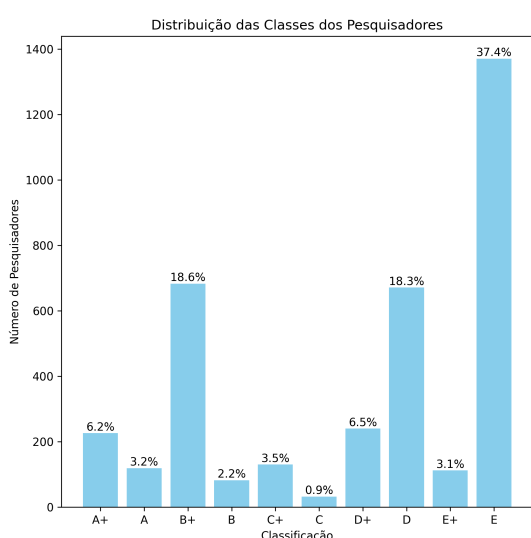


**Fig 5. Box Plot do Índice de Produção**

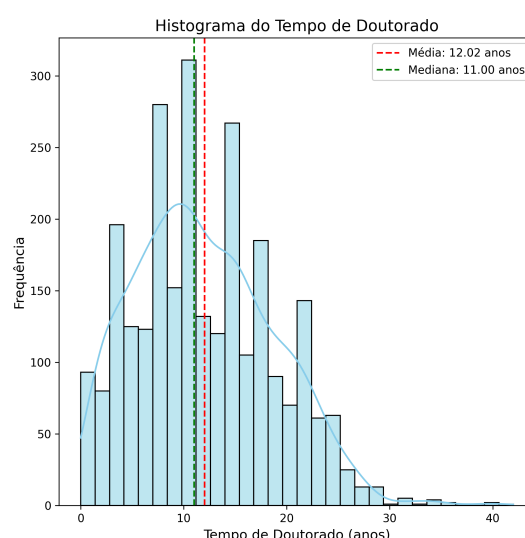
O valor máximo elevado (320,97) sugere a presença de outliers, o que eleva a média (11,44) em relação à mediana (5,03). O histograma (Figura 4) confirma a assimetria, enquanto o box plot (Figura 5) destaca que 75% dos pesquisadores têm um índice inferior a 13,625, e a mediana indica que metade possui um índice abaixo de 5,037.

## 5. Especificação e Aplicação do Modelo no Cenário dos Pesquisadores das UEBAS

A classificação dos pesquisadores revela a heterogeneidade na produção científica em diferentes áreas do conhecimento. Na Figura 6 é analisada a distribuição das classes mostrando uma concentração significativa nas classes D e E, a Figura 7 evidencia que o elevado número de pesquisadores classificados como E se deve, principalmente, à baixa produção científica. Apesar de possuírem doutorado, muitos ainda não conseguiram avançar para classificações mais altas, como C e B, indicando que a obtenção do título não garante uma trajetória produtiva na pesquisa.



**Fig 6. Distribuição das Classes dos Pesquisadores**

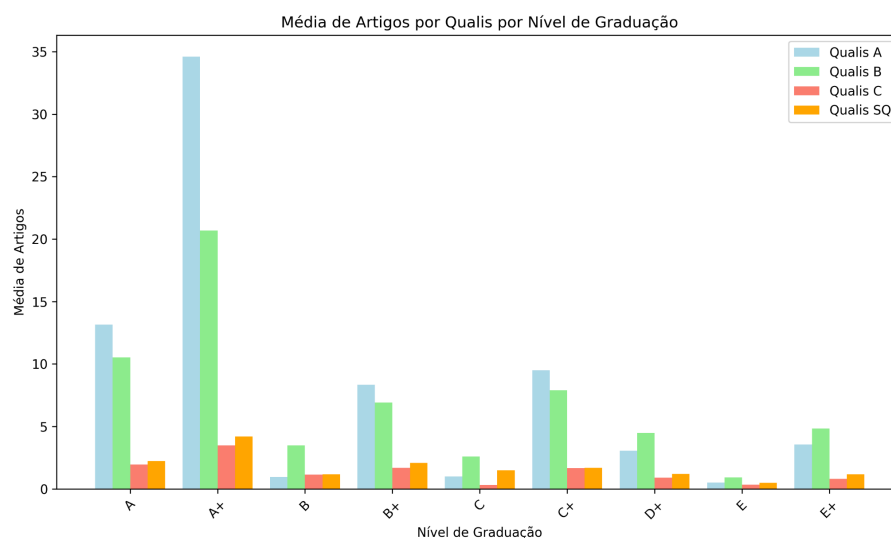


**Fig 7. Tempo de doutorado dos pesquisadores**

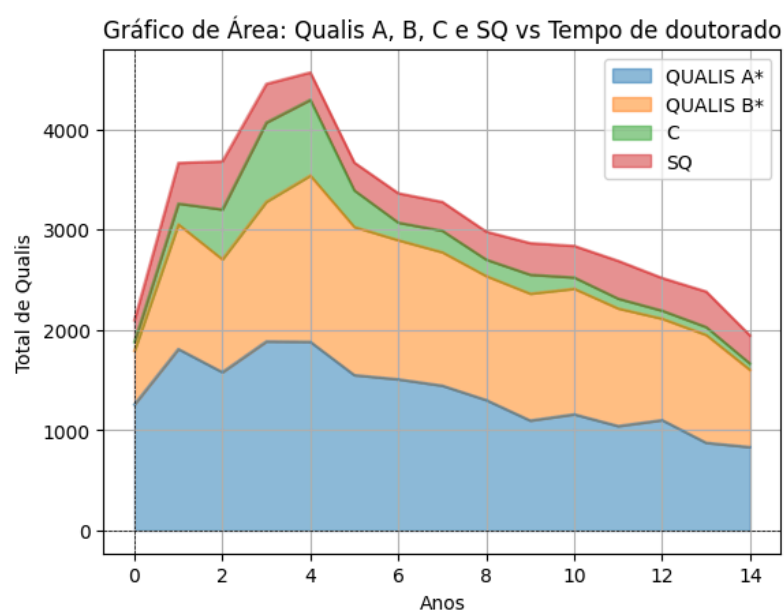
Por outro lado, os pesquisadores classificados como A e A+ se destacam pela produção científica de alta qualidade e impacto. A Figura 8 revela que esses indivíduos compartilham características comuns, como maior experiência, intensa atividade de pesquisa e publicações em periódicos de renome.

Após os primeiros passos na carreira, esses pesquisadores avançam significativamente em suas produções. Aqueles com mais tempo e experiência tendem a colaborar mais entre si, potencializando a quantidade e a qualidade de suas publicações. Sendo um fator crucial para o desenvolvimento da pesquisa como um todo.





**Fig 8. Quantidade de artigos por classificação**



**Fig 9. Qualis A, B, C e SQ vs Tempo de doutorado**

Essa análise identifica os fatores que distinguem pesquisadores de alto desempenho, destacando que aqueles classificados em categorias de excelência mantêm uma constância na produção de artigos com Qualis elevado. Em contrapartida, como mostrado na Figura 9, outras classificações podem gerar um número alto de publicações, mas apresentam um declínio significativo na qualidade ao longo do tempo.

Além disso, enfatiza a importância de políticas institucionais que promovam a formação contínua e a colaboração entre pesquisadores. Programas de mentoria e redes de pesquisa são essenciais para ajudar os pesquisadores em estágios iniciais a avançar em suas carreiras. Ao criar um ambiente de cooperação, as instituições podem melhorar a qualidade da produção científica e promover um avanço equitativo na carreira acadêmica.

## 6. Conclusões preliminares

Este estudo apresentou um modelo de classificação e agrupamento de pesquisadores que considera múltiplos critérios, como tempo de titulação e produção científica. A análise dos dados das universidades estaduais da Bahia destacou a relevância da experiência e da colaboração para aumentar a qualidade e a quantidade das publicações.

A discussão sobre índices bibliométricos, como o índice-h e o Fator de Impacto, evidenciou a necessidade de uma avaliação mais justa, que leve em conta as especificidades de cada área do conhecimento. A proposta de um sistema de mapeamento de competências científicas contribui para a gestão acadêmica e orienta pesquisadores em suas trajetórias profissionais.

Em síntese, a implementação deste modelo pode identificar pesquisadores de alto desempenho e promover um ambiente de pesquisa colaborativo. Para aprimorar essa abordagem, futuros estudos devem considerar novas métricas e metodologias, além de investigar ferramentas de análise de rede para mapear colaborações interinstitucionais, enriquecendo a compreensão da produção científica contemporânea.

## 7. Referências

- Egghe, L. Theory and practise of the g-index. *Scientometrics*, vol. 69, 131–152 (2006). <https://doi.org/10.1007/s11192-006-0144-7>.
- Barreto, M. L. et al. Diferenças entre as medidas do índice-h geradas em distintas fontes bibliográficas e engenho de busca. *Revista de Saúde Pública*, v. 47, n. 2, p. 231-238, 2013. DOI: 10.1590/S0034-8910.2013047004533.
- Chemistry World. Disponível em :<Chemistry news, research and opinions | Chemistry World>. Acesso em: 01 out. 2024.
- Critérios de Julgamento. Disponível em: <<https://www.gov.br/cnpq/pt-br/composicao/comites-de-assessoramento/criterios-de-julgamento>>. Acesso em: 01 out. 2024.
- SANTOS, M. S. dos; OLIVEIRA, V. H. de J.; JORGE, E. M. de F.; COSTA, G. de M. Solução para Mapeamento e Consulta das Competências dos Pesquisadores: uma arquitetura para extração, integração e consultas de informações acadêmicas. *Cadernos de Prospecção*, [S. l.], v. 17, n. 2, p. 671–688, 2024. DOI: 10.9771/cp.v17i2.56670. Disponível em: <https://periodicos.ufba.br/index.php/nit/article/view/56670>. Acesso em: 20 out. 2024.
- BRASIL. Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. Diretoria de Avaliação. Relatório de Avaliação da Área Interdisciplinar para programas de graduação multidisciplinares. Brasília, 2022. Disponível em: [https://www.gov.br/capes/pt-br/centrais-de-conteudo/documentos/avaliacao/19122022\\_45.INTE\\_RelatorioFinal.pdf](https://www.gov.br/capes/pt-br/centrais-de-conteudo/documentos/avaliacao/19122022_45.INTE_RelatorioFinal.pdf). Acesso em: 20 out. 2024.