

WINEINFORMATICS: UMA ANÁLISE BIBLIOMÉTRICA.

Patricia Sanches Jonas y Luiz Gustavo Lovato.

Cita:

Patricia Sanches Jonas y Luiz Gustavo Lovato (2025).
WINEINFORMATICS: UMA ANÁLISE BIBLIOMÉTRICA. XIII Jornada de Produção Científica e Tecnológica, XVI Ciclo de Palestras Tecnológicas, I Semana da Pedagogia e X Semana da Biologia. Instituto Federal de São Paulo - Câmpus São Roque, São Roque.

Dirección estable: <https://www.aacademica.org/jpctifpsrq/55>

ARK: <https://n2t.net/ark:/13683/paWp/k4b>



Esta obra está bajo una licencia de Creative Commons.
Para ver una copia de esta licencia, visite
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>.

Acta Académica es un proyecto académico sin fines de lucro enmarcado en la iniciativa de acceso abierto. Acta Académica fue creado para facilitar a investigadores de todo el mundo el compartir su producción académica. Para crear un perfil gratuitamente o acceder a otros trabajos visite:
<https://www.aacademica.org>.

WINEINFORMATICS: UMA ANÁLISE BIBLIOMÉTRICA

Patricia Sanches Jonas

Luiz Gustavo Lovato, luiz.lovato@ifsp.edu.br

Resumo

Este trabalho tem como objetivo mostrar o estado atual da pesquisa científica sobre padrões encontrados nas avaliações de vinhos feitas por experts e consumidores. Para tanto, foi realizado um estudo bibliométrico das publicações inseridas nas bases de dados Scopus e Web of Science (WoS). Analisou-se a correlação entre cobertura, sobreposição e concentração dos documentos indexados nas duas bases de periódicos. Durante o processo de busca, foi obtido um conjunto de 125 artigos. Com base nos resultados do estudo comparativo, concluímos que as bases de dados WoS e Scopus diferem no volume de artigos publicados e fontes de artigos únicos. Portanto, as bases são complementares e não mutuamente excludentes. A base Scopus cobre mais a área pesquisada, incluindo o maior número de artigos e periódicos. Ademais, foi identificada uma tendência de crescimento de estudos publicados na área desde 2018, conformando uma nova área de pesquisa que abarca ciência de dados e estudos de comportamento do consumidor de vinhos. A esse campo é dado o nome de *wineinformatics*.

Palavras-chave: Bibliometria, Experts, Consumidores, Vinho, Mineração de texto

Modalidade: Resumo Expandido

Apresentação

Com o advento das redes sociais e o acesso facilitado a smartphones, a produção e disseminação de opiniões pessoais sobre todo e qualquer tipo de produto, inclusive o vinho, parece ter diminuído a relevância dos guias de vinhos, que geralmente produzem suas avaliações baseados em critérios objetivos (parâmetros visuais, olfativos e gustativos) (SMITH, 2019). Atualmente, aplicativos de avaliações de vinhos têm atuado como uma ferramenta de benchmarking para consumidores com variados níveis de conhecimento. Ao mesmo tempo em que funcionam como uma biblioteca/adega digital, o compartilhamento massivo e difuso de notas qualitativas e quantitativas levanta a discussão sobre a credibilidade e, portanto, validade daquela avaliação que, potencialmente, influencia a decisão de compra.

Da mesma forma, guias de vinhos aclamados internacionalmente ainda ocupam esse lugar de detenção do conhecimento sobre a descrição, avaliação, hierarquização e sugestão do que julgam ser bons vinhos para consumidores neófitos e enófilos experientes. Alguns exemplos: a revista estado-unidense *Wine Spectator*, a publicação britânica *Decanter* e os mundialmente famosos guias pessoais de Robert Parker (*Wine Advocate*) e Jancis Robinson (CHEN et al., 2018; HONORÉ-CHEDOZEAU; OTHEGUY; VALENTIN, 2024). Os quais com suas escalas de pontuação, influenciaram e ainda influenciam não só o consumidor final, mas também, e sobretudo, os produtores, que passaram a elaborar vinhos para agradar ao gosto de um determinado guia especializado, gerando, inclusive, termos como “vinhos parkerizados”, atribuído a vinhos tintos encorpados e com aromas intensos de carvalho, que teriam a predileção de Robert Parker (JAMES, 2018; SACKRIS, 2019)

Mas afinal o gosto é algo compartilhado? É possível evidenciar convergências e divergências entre os julgamentos derivados de impressões pessoais e publicações especializadas sobre um mesmo tópico? Alguns estudos já investigaram as tentativas de padronização de métodos de degustação de vinhos e transformação de descrições subjetivas em conhecimento objetivo (PHILLIPS, 2016), o efeito de opiniões de experts sobre o preço de vinhos (SPENCE, 2024), a correlação entre a preferência e a qualidade percebida tanto para

consumidores quanto para experts (HOPFER; HEYMANN, 2014), e o desenvolvimento de preditores de qualidade de vinhos baseados nos atributos do produto e avaliações de consumidores (MA et al., 2024).

Devido à crescente relevância do tema, este trabalho se propôs a analisar e compilar a evolução da bibliografia existente sobre a área nos últimos anos. Dessa forma, realizou-se um estudo bibliométrico, onde foram pesquisados artigos sobre técnicas e metodologias de análise de padrões qualitativos e quantitativos nas avaliações de vinhos feitas por experts e por consumidores. Além disso, analisou-se as características da produção científica com base nas seguintes variáveis: ano, autores, país de origem, relevância das bases de indexação de periódicos e principais periódicos.

Materiais e métodos

Esta seção descreve o processo de elaboração da análise bibliométrica de artigos científicos cujo tópico principal é a análise da avaliação de vinhos feitas por experts e consumidores utilizando mineração de texto e outras metodologias de análises de dados qualitativos e quantitativos.

O objetivo não é conduzir uma revisão sistemática do conteúdo dos artigos encontrados nas bases de dados, mas sim realizar uma análise descritivo-quantitativa da evolução desse tópico na produção científica (ÖZTÜRK et al., 2024).

A análise bibliométrica teve início com a seleção criteriosa das bases de dados científicas, etapa essencial para garantir a qualidade dos resultados. Essas bases são sistemas estruturados logicamente, que organizam informações de forma acessível, permitindo sua atualização constante e integração com diferentes programas computacionais, conforme critérios específicos (LUQUE, 1995). Elas são fundamentais para a investigação bibliométrica, pois possibilitam o levantamento da produção científica de autores, instituições, regiões e países, além de revelar tendências e áreas emergentes do conhecimento. A profundidade e a validade da análise dependem diretamente da abrangência e representatividade da base escolhida em relação ao campo de estudo. Uma escolha inadequada pode comprometer a interpretação dos dados, gerando resultados limitados ou distorcidos. Por isso, é indispensável compreender o escopo, as características e as limitações de cada base, assegurando que a análise seja precisa, relevante e confiável.

Após algumas buscas pelas bases Scielo, Portal de Periódicos da Capes, Scopus e Web of Science, optou-se por realizar a pesquisa com as duas últimas, por apresentarem o maior número de artigos sobre o tema e pela importância que essas ferramentas têm para os pesquisadores como fonte bibliográfica para fundamentar seus trabalhos.

Encontramos dificuldades nas diferenças das políticas de indexação de documentos que cada base de periódicos segue; enquanto algumas transferem todas as fontes, outras o fazem seletivamente, o que exige um trabalho manual para comparar os bancos de dados.

De toda forma, o passo-a-passo seguiu a seguinte ordem (Figura 1) (ÖZTÜRK et al., 2024):

1. Teste das palavras-chave nas bases de periódicos;
2. Definição das palavras-chave e operadores booleanos.
3. Definição das bases de periódicos;
4. Definição do período de análise;
5. Comparação das bases por meio do Índice de Meyer, sobreposição e sobreposição relativa;

XIII Jornada de Produção Científica e Tecnológica

XVI Ciclo de Palestras Tecnológicas

I Semana da Pedagogia

X Semana da Biologia



6. Análise da evolução temporal das publicações, autores, países que publicam, periódicos que contemplam o tema;
7. Exclusão dos artigos repetidos;
8. Exclusão dos artigos não relacionados ao tema;
9. Lista final dos artigos científicos sobre o tópico nas duas bases;
10. Análise quantitativa da lista final dos artigos.

A *string* final de busca foi definida com base em testes anteriores. A expressão final utilizada foi:

"wine" AND "tasting" OR review OR "vivino" OR "recommendation" OR "opinion" OR "expert" OR "consumer" AND "text mining" OR "machine learning" OR "artificial intelligence".

O período analisado foi de 1987 (ano do primeiro artigo identificado sobre o tema) até 28 de abril de 2025, data de corte da pesquisa.

Índice de Meyer (IM)

Para avaliar ou comparar o grau de cobertura temática de uma base de dados, utiliza-se o Índice de Meyer, que mensura a singularidade de uma base em relação a outras (SÁNCHEZ et al., 2017). Esse índice considera o número de documentos únicos presentes em uma base, atribuindo maior peso a conteúdos exclusivos (peso igual a 1,0) e reduzindo progressivamente o valor de documentos duplicados (peso igual 0,5) ou triplicados (peso igual a 0,3) conforme aparecem em múltiplas bases. A ideia central é que quanto maior o índice, maior a contribuição original da base para o tema analisado. O cálculo é feito a partir da razão entre o somatório dos documentos ponderados e o total de documentos da base:

$$\text{Índice de Meyer (IM)} = \frac{\text{Somatório de Documentos} \times \text{Peso}}{\text{Número total de Documentos}}$$

Sobreposição

Para analisar o grau de similaridade entre duas bases, pode-se aplicar a Sobreposição Tradicional, que expressa o percentual de documentos compartilhados entre elas (SÁNCHEZ et al., 2017). Esse indicador revela o quanto os conteúdos se repetem, permitindo identificar redundâncias ou complementaridades. A fórmula considera a interseção dos documentos das duas bases em relação ao total combinado:

$$\text{Sobreposição} = 100 \times (|A \cap B| / |A \cup B|)$$

Sobreposição Relativa

A Sobreposição Relativa permite observar o quanto uma base está contida na outra, oferecendo uma visão assimétrica da cobertura. Esse indicador é útil para entender a abrangência de uma base em relação à outra, especialmente em estudos comparativos. Os cálculos são feitos separadamente para cada base:

$$\text{Sobreposição Relativa em A} = 100 \times (|A \cap B| / |A|)$$

$$\text{Sobreposição Relativa em B} = 100 \times (|A \cap B| / |B|)$$

Esses três índices auxiliam estudos bibliométricos e análises de cobertura informacional, permitindo identificar quais bases oferecem maior originalidade, quais se complementam e quais apresentam maior redundância.

Resultados

Em ambas as bases de dados, a forma de comunicação mais utilizada é o artigo científico. Portanto, não foram incluídos trabalhos de conferências, artigos de jornais, capítulos de livros, entre outros tipos de documentos. Um quadro-resumo dos resultados é apresentado na Tabela 1.

Após a busca por artigos relacionados ao tema, 66 artigos foram selecionados na WoS e 106 na Scopus, estes compõem os dados de entrada para o estudo bibliométrico comparativo.

A base de dados Scopus foi a primeira a indexar artigos sobre o tema, registrando um artigo de 1987, consolidando-se como a principal fonte de registros na área. Observa-se uma expressiva concentração da produção científica nos últimos dez anos, período que abrange 95% dos artigos publicados, sendo que 75% foram produzidos apenas nos últimos cinco anos. Esses dados evidenciam a recente ascensão e crescente interesse do tema na comunidade acadêmica. A Figura 2 apresenta a distribuição temporal da produção científica entre 1987 e 2025, destacando essa tendência de crescimento.

A Figura 3 apresenta a distribuição geográfica da produção científica sobre o tema, destacando os países com maior número de publicações. Observa-se que Estados Unidos, China, Itália, Espanha e Portugal lideram em volume de artigos indexados, evidenciando o protagonismo dessas nações na consolidação e avanço das pesquisas na área.

Na base Scopus foram identificados 160 autores, com uma média de 1,5 autores por artigo. Já na Web of Science (WoS), esse número foi significativamente maior, totalizando 270 autores e uma média de 4 autores por artigo, o que evidencia uma predominância de trabalhos colaborativos em ambas as bases. O autor com maior número de publicações na Scopus foi Bernard Chen, com quatro artigos, seguido por S. Fuentes, C.G. Viejo e Tongson, cada um com três publicações. Na WoS, os dez autores mais produtivos apresentaram uniformidade, com dois artigos publicados cada.

A elaboração da lista de autores e das estatísticas associadas apresentou desafios metodológicos, devido às diferenças nos critérios de indexação entre as bases. Observou-se inconsistência na padronização das iniciais dos nomes, o que dificultou a identificação precisa de autores recorrentes. Por exemplo, o mesmo autor pode ser referenciado como "Chen, B." na Scopus e "Chen, B.R." na WoS, comprometendo a uniformidade dos dados e exigindo tratamento manual para evitar duplicidades.

Correlação entre Scopus e WoS

A pesquisa com o conjunto de palavras-chave selecionado resultou em 106 artigos científicos na base de dados da Scopus e 66 na Web of Science (WoS), entretanto verificou-se que 47 artigos (38%) estavam presentes em ambas as bases ($Scopus \cap WoS$), totalizando 125 artigos distintos ($Scopus \cup WoS$). Dentre esses, 78 documentos (62%) eram exclusivos de uma única base: 59 apenas na base da Scopus (56%) e 19 apenas na WoS (29%).

A análise cruzada revelou que 71% dos documentos da WoS também estavam indexados na Scopus, enquanto 44% dos documentos da Scopus estavam presentes na WoS. Esses dados evidenciam uma sobreposição parcial entre as bases, reforçando a importância da análise conjunta para uma representação mais abrangente da produção científica sobre o tema.

A sobreposição calculada foi de 38%. Esse resultado, indica uma similaridade de 38% entre as bases. Ou uma divergência de 62% entre as bases de periódicos, evidenciando a complementaridade entre os repositórios.

Para a sobreposição relativa, os resultados foram:

Sobreposição Relativa da Scopus na WoS = 72%;

XIII Jornada de Produção Científica e Tecnológica

XVI Ciclo de Palestras Tecnológicas

I Semana da Pedagogia

X Semana da Biologia



Sobreposição Relativa da WoS na Scopus = 44%.

Isso significa que a Scopus cobre 72% dos artigos da WoS sobre o tema, enquanto a WoS cobre 44% dos artigos da Scopus, demonstrando maior abrangência da primeira.

Cálculo do Índice de Meyer (IM)

Índice de Meyer Scopus = 0,77;

Índice de Meyer WoS = 0,64.

Quanto maior o índice, maior a unicidade da base de dados, ou seja, ela contém maior número de documentos únicos, a Scopus possui maior número de documentos únicos (IM 0,77) com relação a WoS (0,64). Na Tabela 1 é possível observar os resumos de todos os resultados estatísticos encontrados durante o estudo.

Dos 125 artigos identificados nas bases Scopus e Web of Science, foram selecionados 24 documentos para leitura e análise aprofundada, com base na leitura do título, resumo e palavras-chave (Tabela 3). O critério para exclusão baseou-se em eliminar estudos que não utilizaram algoritmos de inteligência artificial para verificar padrões em avaliações de vinhos feitas por experts ou consumidores. Boa parte dos artigos excluídos tem utilizado técnicas de bioinformática para analisar grandes quantidades de dados relacionados a compostos voláteis e fenólicos, ou ainda, genótipos expressos em plantas do gênero *Vitis*. A Figura 4 apresenta a evolução temporal das 24 publicações que passaram pelo filtro, evidenciando uma concentração nos últimos sete anos, o que sugere que o tema é emergente e oferece espaço promissor para novas investigações.

Na Figura 5, destaca-se o autor Chen, B., com o maior número de publicações entre os artigos analisados. Sua produção científica, em especial o livro *Wineinformatics*, tem sido utilizada como base para o desenvolvimento do projeto de iniciação científica “Gosto: cada um tem o seu?”, que propõe uma abordagem de análise quantitativa e qualitativa de avaliações de vinhos utilizando técnicas de *machine learning*.

Dentre os 24 artigos, foram identificados 83 autores, resultando em uma média de 3,45 autores por artigo, o que evidencia o caráter colaborativo da produção científica na área. A Figura 6 apresenta a distribuição geográfica dos artigos, com destaque para o Brasil e os Estados Unidos, ambos com cinco publicações. Em seguida, aparecem Portugal, China e Índia, reforçando o caráter internacional da pesquisa sobre o tema.

Na segunda etapa de filtragem, foi realizada a identificação dos periódicos que publicaram os artigos selecionados (ver Tabela 2). Observa-se uma diversidade de áreas temáticas, com publicações tanto em revistas especializadas em vinhos quanto em periódicos voltados para informática, estatística, ciência dos alimentos e áreas correlatas. Essa variedade reforça o caráter interdisciplinar da temática, que transita entre o universo enológico e as tecnologias emergentes de análise de dados.

Considerações finais

O estudo revela que o tema da avaliação de vinhos por experts e consumidores é recente, com forte crescimento nos últimos anos e grande potencial para novas investigações. A análise bibliométrica comparativa mostra que Scopus e WoS são bases complementares, com cobertura parcial sobreposta, mas alta produção exclusiva em cada uma.

A predominância de estudos com abordagens em *machine learning*, análise de sensorial, mineração de texto e recomendações digitais aponta para uma transformação da enologia sensorial, aproximando-a de áreas como ciência de dados e inteligência artificial.

Finalmente, destaca-se a importância da correta escolha da base de dados para assegurar a abrangência e qualidade das pesquisas bibliográficas. Os resultados obtidos podem apoiar futuras investigações acadêmicas e práticas no campo da enologia, marketing sensorial e análise preditiva.

Referências

CHEN, Bernard *et al.* Wineinformatics: A Quantitative Analysis of Wine Reviewers. **Fermentation**, v. 4, n. 4, p. 82, dez. 2018.

HONORÉ-CHEDOZEAU, Carole; OTHÉGUY, Méven; VALENTIN, Dominique. Tell us how you taste wine, and we will tell you what kind of expert you are! **Food Research International**, v. 178, p. 113899, 1 fev. 2024.

HOPFER, Helene; HEYMANN, Hildegard. Judging wine quality: Do we need experts, consumers or trained panelists? **Food Quality and Preference**, v. 32, p. 221–233, 1 mar. 2014.

JAMES, Andrew. How Robert Parker's 90+ and Ann Noble's Aroma Wheel Changed the Discourse of Wine Tasting Notes. **ILCEA. Revue de l'Institut des langues et cultures d'Europe, Amérique, Afrique, Asie et Australie**, n. 31, 1 mar. 2018.

LUQUE MARTÍNEZ, T. Líneas de investigación y bases de datos para la investigación. **Investigaciones Europeas de Dirección y Economía de la Empresa**, v. 1, n. 2, p. 35–50, 1995.

MA, Rui *et al.* From vineyard to table: Uncovering wine quality for sales management through machine learning. **Journal of Business Research**, v. 176, p. 114576, 1 abr. 2024.

ÖZTÜRK, Oğuzhan; KOCAMAN, Rıdvan; KANBACH, Dominik K. How to design bibliometric research: an overview and a framework proposal. **Review of Managerial Science**, v. 18, n. 11, p. 3333–3361, nov. 2024.

PHILLIPS, Christopher J. The taste machine: Sense, subjectivity, and statistics in the California wine world. **Social Studies of Science**, v. 46, n. 3, p. 461–481, 2016.

SACKRIS, David C. What Jancis Robinson Didn't Know May Have Helped Her. **Erkenntnis**, v. 84, n. 4, p. 805–822, 1 ago. 2019.

SÁNCHEZ, Amador Durán; DE LA CRUZ DEL RÍO RAMA, María; GARCÍA, José Álvarez. Bibliometric analysis of publications on wine tourism in the databases Scopus and WoS. **European Research on Management and Business Economics**, v. 23, n. 1, p. 8–15, jan. 2017.

SMITH, Barry C. Getting More Out of Wine: wine experts, wine apps and sensory science. **Current Opinion in Food Science**, Sensory Science and Consumer Perception • Food Physics & materials Science. v. 27, p. 123–129, 1 jun. 2019.

SPENCE, Charles. Cognitive influence on the evaluation of wine: The impact and assessment of price. **Food Research International**, v. 187, p. 114411, 1 jul. 2024.

Apêndice

Tabela 1 - Resumo dos resultados

Plataforma	WoS	Scopus
Total de Documentos	66	106
Doc. Únicos	19	59
% Doc. Único	29%	56%
Doc. Repetidos	47	47
% de repetidos	71%	44%
Sobreposição	38%	-
Sobreposição Rel. em Scopus	44%	-
Sobreposição Rel. em WoS	-	72%
Índice de Meyer	0,64	0,77

Tabela 2 - Periódicos que mais publicaram sobre o assunto

Periódico	Qtd.
Fermentation	4
International Journal of Wine Business Research	3
Big Data and Cognitive Computing	2
Artificial Intelligence in Agriculture	1
British Food Journal	1

Tabela 3 – Lista dos 24 artigos selecionados.

Nº	Título
1	Assessing the nexus of label content and consumer reviews: insights from premium wines on Vivino
2	Big data analysis of Spanish wine consumers reviews
3	Wineinformatics: Wine Score Prediction with Wine Price and Reviews
4	A machine learning and linear programming aided approach to wine ranking and selection
5	Flavoromic analysis of wines using gas chromatography mass spectrometry and sensory techniques
6	From vineyard to table: Uncovering wine quality for sales management through machine learning
7	Using Machine Learning to Predict Wine Quality and Prices: A Demonstrative Case Using a Large Tabular Database
8	Application of Linear Discriminant Analysis and k-Nearest Neighbors Techniques to Recommendation Systems
9	Applying Neural Networks in Wineinformatics with the New Computational Wine Wheel
10	Complementing human effort in online reviews: A deep learning approach to automatic content generation and review synthesis
11	X-Wines: A Wine Dataset for Recommender Systems and Machine Learning

XIII Jornada de Produção Científica e Tecnológica

XVI Ciclo de Palestras Tecnológicas

I Semana da Pedagogia

X Semana da Biologia

- | | |
|----|---|
| 12 | soMLier: a South African wine recommender system |
| 13 | Classifications on wine informatics using PCA, LCA and supervised machine learning techniques |
| 14 | Predicting wine prices based on the weather: Bordeaux vineyards in a changing climate |
| 15 | Wine recommendation algorithm based on partitioning and stacking integration strategy for Chinese wine consumers |
| 16 | Enhancing the wine tasting experience using greedy clustering wine recommender system |
| 17 | Wineinformatics: Can wine reviews in bordeaux reveal wine aging capability? |
| 18 | Selection of features and prediction of wine quality using artificial neural networks |
| 19 | Wine ontology influence in a recommendation system |
| 20 | A generalized wine quality prediction framework by evolutionary algorithms |
| 21 | Predictive modeling for wine authenticity using a machine learning approach |
| 22 | Using content analysis to characterise the sensory typicity and quality judgements of Australian Cabernet Sauvignon wines |
| 23 | Online wine ecosystem: the digital narrative of Sangiovese |
| 24 | Wineinformatics: Regression on the grade and price of wines through their sensory attributes |

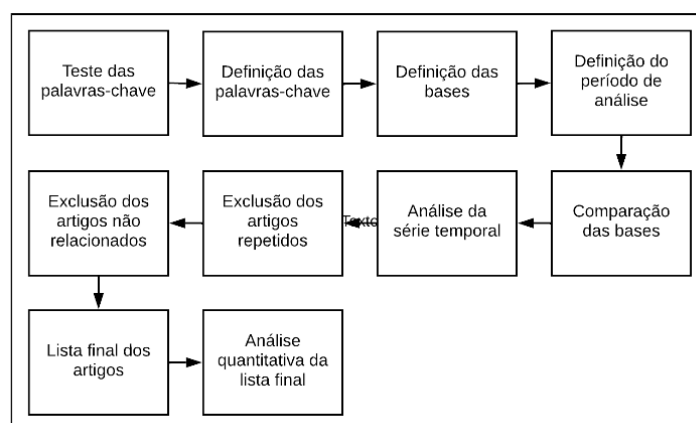


Figura 1 – Fluxograma da análise bibliométrica

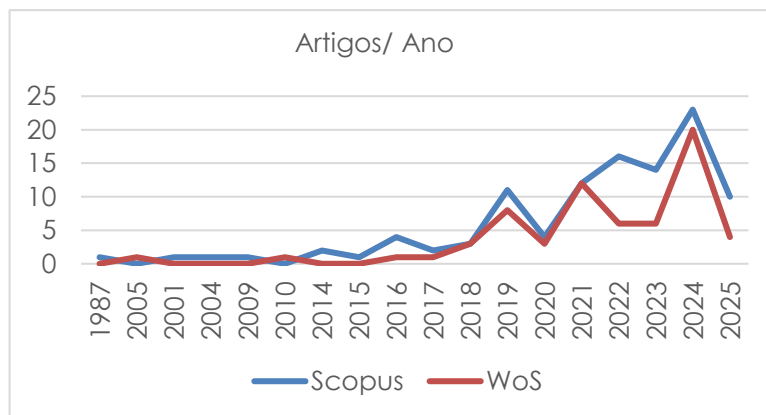


Figura 2 – Evolução do número de artigos publicados nas bases Scopus (n=106) e Web of Science (WoS)

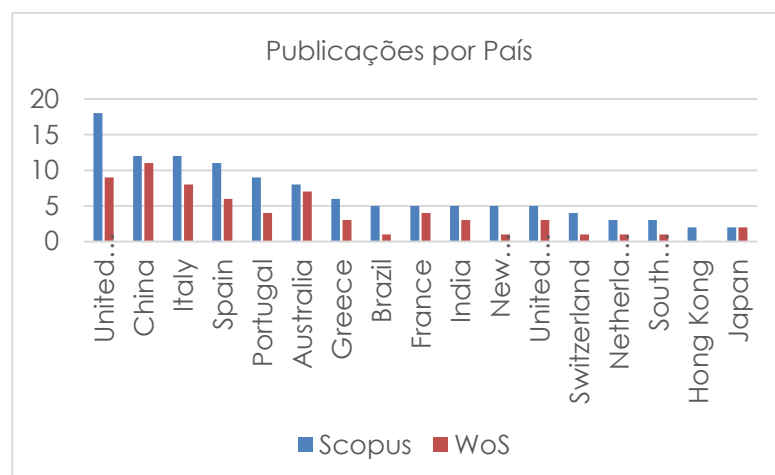


Figura 3 – Evolução de artigos publicados por país nas bases Scopus (n=106) e Web of Science

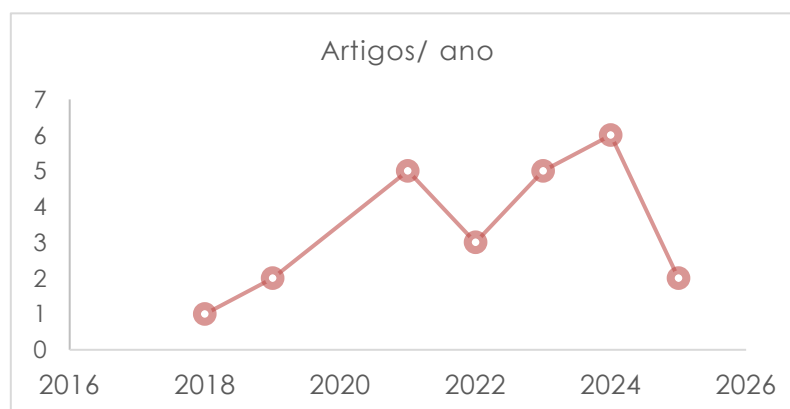


Figura 4 – Evolução temporal dos 24 artigos

XIII Jornada de Produção Científica e Tecnológica

XVI Ciclo de Palestras Tecnológicas

I Semana da Pedagogia

X Semana da Biologia

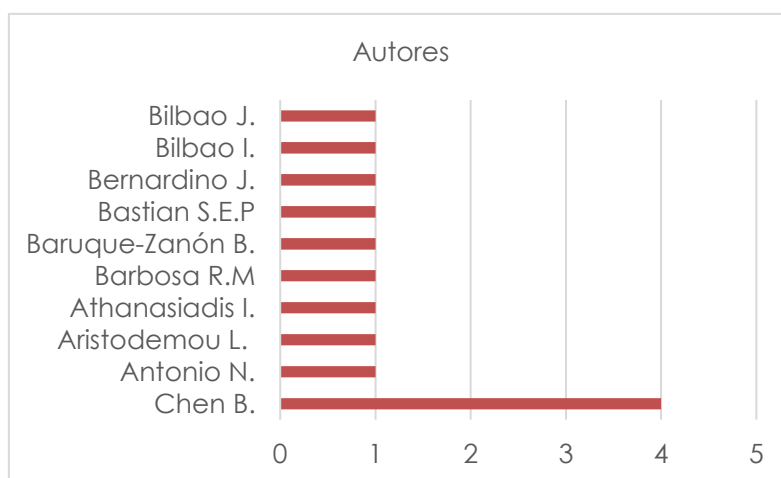


Figura 5 – Autores dos 24 artigos selecionados

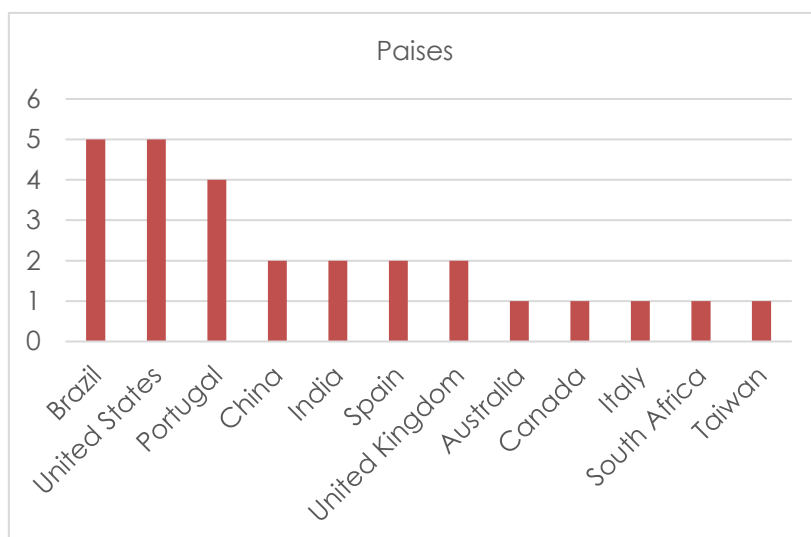


Figura 6 – Países de origem dos autores artigos 24 selecionados