

A PRODUÇÃO BIBLIOGRÁFICA SOBRE A AMAZÔNIA: UMA ANÁLISE DE *CLUSTER*¹

Fernanda De Negri²

SINOPSE

O texto analisa a produção científica sobre a Amazônia, publicada em periódicos indexados pela Web of Science no período de 2000 a 2023. Mais de 30 mil artigos científicos sobre a Amazônia foram agrupados em seis diferentes *clusters*, a partir da similaridade semântica entre eles, método que foi capaz de agrupar esses artigos segundo os principais temas estudados. Os *clusters* 1 e 6 estão mais focados em aspectos ambientais e de uso da terra, sendo que o *cluster* 1 se concentra mais no solo, e o *cluster* 6, no desmatamento e na conservação. O *cluster* 2 destaca-se pela ênfase em estudos de biodiversidade e questões de saúde, como a malária. O *cluster* 3 é focado em aspectos hidrológicos e climáticos da bacia do Amazonas. O *cluster* 4 apresenta uma abordagem mais voltada para questões sociais e indígenas, além de análises ambientais. O *cluster* 5 é único no seu enfoque em taxonomia e na descrição de novas espécies. Essas parecem ser as principais questões que mobilizam a comunidade científica nacional e internacional nos seus estudos sobre a Amazônia.

Palavras-chave: Amazônia; produção científica; bibliometria; Web of Science; análise de *cluster*; processamento de linguagem natural.

1 INTRODUÇÃO

No *Radar* nº 75, publicou-se uma análise bibliométrica da pesquisa mundial sobre a Amazônia (De Negri, 2024). A partir do levantamento de palavras-chave nas publicações científicas internacionais indexadas na Web of Science, analisaram-se mais de 30 mil artigos publicados sobre o bioma no período de 2000 a 2023.

Naquele artigo, apresentou-se o crescimento expressivo da produção científica sobre a Amazônia no mundo durante essas duas décadas, que chegou a mais de 3 mil artigos por ano no final da década de 2010. Mostrou-se que as instituições brasileiras como a Universidade de São Paulo (USP), a Universidade Federal do Pará (UFPA), o Instituto Nacional de Pesquisas da Amazônia (Inpa) lideram a produção científica mundial sobre a Amazônia. Também foi possível evidenciar as áreas científicas mais relevantes dos artigos sobre o bioma, que são as ciências ambientais/ecologia, a zoologia e a geologia.

Este artigo complementa aquele do *Radar* nº 75 com uma análise de *cluster* sobre os principais temas e assuntos abordados pela produção científica internacional a respeito da Amazônia. Essa abordagem está menos orientada para áreas científicas e mais preocupada em mapear as principais questões que têm despertado o interesse da comunidade científica nacional e internacional sobre a Amazônia.

1. DOI: <https://dx.doi.org/10.38116/radar78art2>

2. Diretora de Estudos e Políticas Setoriais, de Inovação, Regulação e Infraestrutura do Instituto de Pesquisa Econômica Aplicada (Diset/Ipea). E-mail: fernanda.denegri@ipea.gov.br.

2 METODOLOGIA

Além da tradicional análise bibliométrica, há interesse em compreender melhor quais assuntos ou temas de pesquisa mais despertam a atenção dos cientistas quando o assunto é Amazônia. Três campos disponíveis nos metadados dos artigos foram utilizados para uma análise mais aprofundada e específica sobre os assuntos estudados em cada um desses artigos: título, palavras-chave e resumo, todos eles em língua inglesa.³ Esses três campos foram submetidos a métodos de processamento de linguagem natural para a clusterização dos artigos científicos segundo assuntos similares,⁴ o que permitirá ir além da tradicional classificação dos artigos por área do conhecimento e entender as perguntas ou preocupações centrais dos cientistas.

O pré-processamento dos três campos de texto consistiu, em um primeiro momento, na junção desses campos e na fragmentação das *strings* de texto em unidades menores, palavras ou expressões, ou seja, a tokenização desses campos. Depois disso, foram excluídas palavras comuns sem conteúdo semântico (chamadas de *stop words*), tais como artigos, preposições e conjunções. A seguir, os textos síntese dos artigos publicados foram vetorizados, ou seja, transformados em vetores numéricos utilizando a técnica *term frequency-inverse document frequency* (TF-IDF), que é uma medida estatística usada para avaliar a importância de uma palavra em um documento em relação a um conjunto de documentos. Ela considera a frequência de uma palavra no documento (ou texto) específico (TF), bem como a frequência dessa palavra em todos os documentos (ou textos) analisados (IDF). A pontuação TF-IDF de um termo em um documento é então obtida multiplicando os valores de TF e IDF. Assim, termos que aparecem frequentemente em um documento, mas raramente no conjunto de documentos, recebem uma pontuação alta, destacando sua importância única naquele documento.

A medida TF-IDF foi então utilizada como base para a clusterização dos artigos, ou seja, para o agrupamento desses artigos em grupos semanticamente similares, utilizando o método *k-means*, em que *k* é o número de *clusters* escolhidos. Os artigos foram agrupados em seis *clusters* diferentes, com base em uma análise visual da sua distribuição (método do cotovelo). Dessa forma, os documentos que compartilham temas ou características semelhantes (segundo as representações TF-IDF) são agrupados nos mesmos *clusters*.

3 RESULTADOS

A aplicação dessa metodologia de clusterização aos mais de 30 mil artigos mapeados sobre a Amazônia resultou em seis grupos ou *clusters* agrupados por similaridade semântica e temática. Para cada *cluster* de artigos, também são apresentadas nuvens de palavras⁵ que auxiliam na visualização dos temas mais relevantes em cada um deles.

3. A própria base da Web of Science disponibiliza esses campos em inglês. Dessa forma, não foi necessário utilizar nenhum algoritmo ou rotina própria de tradução para compatibilizar o idioma dos diferentes artigos.

4. A aplicação desse método de clusterização contou com o auxílio da ferramenta ChatGPT, operando a linguagem Python, a partir de instruções da autora.

5. As nuvens de palavras foram criadas a partir dos campos da Web of Science utilizados na análise de *cluster*, que estão em inglês, por isso essas figuras estão nesse idioma.

O segundo *cluster* é o maior entre eles, concentrando mais de 31% de todos os artigos sobre o bioma. As pesquisas desse grupo são diversas, o que sugere que esse *cluster* poderia ser subdividido em pelo menos três, pois abrange estudos sobre a biodiversidade da Amazônia, com um foco significativo em estudos ictiológicos, bem como pesquisas sobre saúde, especialmente malária.

Entre os cerca de 10 mil artigos desse *cluster*, a maior parte (quase 7 mil) são estudos sobre biodiversidade, especialmente de peixes, comportamento de espécies e ecologia. Um segundo *subcluster*, com mais de 2 mil estudos, compreende pesquisas sobre questões ambientais, como deposição de mercúrio na floresta amazônica e seus impactos. Por fim, o terceiro *subcluster*, com menos de 1.000 artigos, compreende as pesquisas na área de saúde, com foco significativo na malária e em outras questões de saúde pública.

Entre as principais instituições responsáveis pelos estudos desse *cluster* estão a UFPA, a USP, a Fundação Oswaldo Cruz (Fiocruz) e o Inpa.

FIGURA 3

Cluster 3: hidrologia, clima, precipitação, bacia do rio Amazonas



Elaboração da autora.

Obs.: A ilustração não pôde ser padronizada nem revisada em virtude das condições técnicas dos originais (nota do Editorial).

Os artigos do *cluster 3* representam 21%, ou pouco mais de 6 mil, de todos os artigos da amostra analisada e investigam os sistemas fluviais e climáticos da Amazônia, com ênfase especial no rio Amazonas. Esses estudos envolvem a análise de dados sobre hidrologia e precipitações, sobre o ciclo da água, a deposição de carbono e os impactos das mudanças climáticas, utilizando modelagens avançadas para prever futuras transformações e seus efeitos sobre os ecossistemas locais e globais.

Interessante notar que esse *cluster* representava cerca de 30% dos artigos publicados sobre a Amazônia no início da série e teve sua participação reduzida para algo entre 15% e 20% nos anos finais dessa série. Pode ser evidência de que estudos sobre o clima, a hidrologia e a precipitação na Amazônia chegaram a um estágio mais maduro ou simplesmente o reflexo do crescimento de outros temas de pesquisa.

Entre as instituições mais frequentes nesse *cluster* estão novamente os franceses CNRS e IRD, a Universidade da Califórnia e a USP.

No *cluster 5*, em que ficam 11% do total dos artigos, a taxonomia e a identificação de novas espécies são predominantes, refletindo o constante descobrimento da biodiversidade na Amazônia. Esses estudos são fundamentais para a ciência biológica, contribuindo para o conhecimento global de espécies neotropicais e ajudando a orientar esforços de conservação baseados no entendimento das diversas formas de vida que habitam a região.

Entre as principais instituições dos autores desses estudos estão a UFPA, o Inpa, a USP e o Museu Paraense Emílio Goeldi.

FIGURA 6

Cluster 6: desmatamento, floresta, uso da terra, biodiversidade



Elaboração da autora.

Obs.: A ilustração não pôde ser padronizada nem revisada em virtude das condições técnicas dos originais (nota do Editorial).

Finalmente, o *cluster 6* (14% do total de artigos) aborda a interação entre o ser humano e a floresta, com um forte enfoque no desmatamento, na conservação e no uso da terra. Os trabalhos aqui discutem as consequências da exploração florestal, estratégias para mitigar a perda de hábitat e a importância de políticas sustentáveis para preservar as florestas tropicais e a biodiversidade da Amazônia.

Entre as instituições que mais se destacam nesse *cluster* estão o Inpa, a Embrapa e as Universidades da Flórida e da Califórnia.

4 CONCLUSÕES

Em síntese, os *clusters 1* e *6* estão mais focados em aspectos ambientais e de uso da terra, enquanto o *cluster 1* se concentra mais no solo e o *cluster 6*, no desmatamento e conservação. O *cluster 2* destaca-se pela ênfase em estudos de biodiversidade e questões de saúde, como a malária. O *cluster 3* é distintamente focado em aspectos hidrológicos e climáticos da bacia do Amazonas. O *cluster 4* apresenta uma abordagem mais voltada para questões sociais e indígenas, além de análises ambientais. O *cluster 5* é único no seu enfoque em taxonomia e na descrição de novas espécies.

A elevada participação de instituições internacionais, especialmente nos *clusters 1* e *3*, mostra o interesse da comunidade científica internacional na interação e nos impactos da Amazônia (uso do solo, da hidrologia e precipitação) sobre o clima global. Em contrapartida, estudos mais centrados na taxonomia e identificação de novas espécies, assim como em aspectos de saúde e sociais da região, tendem a ser uma preocupação mais relevante das instituições brasileiras na região amazônica.

Essa nova classificação dos artigos complementa a análise por área científica e dá mais concretude às principais questões de pesquisa abordadas pelos cientistas que se debruçam sobre a Amazônia. Além disso, também pode ser utilizada para identificar lacunas e tendências emergentes na pesquisa sobre o bioma. Estudos sobre deposição de mercúrio na floresta, por exemplo, eram algumas dezenas no início dos anos 2000 e saltaram para a casa das centenas no final do período analisado, refletindo a crescente preocupação científica sobre os impactos do aumento da atividade mineradora na região. Também diminuíram, em termos de participação percentual, os estudos voltados exclusivamente ao clima da região, ao passo que cresceram os estudos que analisam as interações do clima com as populações locais, evidenciando a maior importância dada pela comunidade científica aos aspectos sociais da questão ambiental. Contudo, estudos sobre o monitoramento de queimadas ou sobre os efeitos de secas extremas, eventos que recentemente assolaram a região, ainda não entraram no radar da comunidade científica.

REFERÊNCIA

DE NEGRI, F. A pesquisa sobre Amazônia no mundo: uma análise bibliométrica. **Radar**, n. 75, abr. 2024. Disponível em: https://repositorio.ipea.gov.br/bitstream/11058/13656/9/Radar_75_Art5_A_pesquisa_sobre_amazonia.pdf.

