

Explorando a Dinâmica Científica em Eventos: Um Estudo Bibliométrico com Auxílio de Técnicas de Ciência de Dados

Resumo

A produção científica brasileira encontra nos eventos acadêmicos um canal estratégico de disseminação, validação e interlocução de resultados de pesquisa. Apesar disso, os artigos publicados em anais de eventos permanecem pouco explorados nos estudos bibliométricos em larga escala. Este artigo apresenta uma análise dessa produção, utilizando dados extraídos da Plataforma Lattes, a partir de currículos de pesquisadores. Aplicando métricas clássicas da bibliometria, o estudo descreve padrões temporais, geográficos e temáticos da publicação científica em eventos. O trabalho utiliza-se de diversas técnicas de Ciência de Dados e Inteligência Artificial, ao combinar mineração de dados, modelagem de indicadores e visualização de tendências. Os resultados destacam a importância dos eventos para diversas áreas do conhecimento, reforçam a relevância da integração e tratamento de grandes volumes de dados e oferecem subsídios para o desenvolvimento de políticas públicas voltadas à ciência aberta e à valorização da produção científica brasileira.

Palavras-chave: Anais de eventos; Produção científica; Bibliometria; Plataforma Lattes; Ciência aberta.

1. Introdução

A produção científica é um dos principais indicadores do desenvolvimento de um país e se materializa por meio de diversos canais de comunicação, entre os quais os eventos acadêmicos ocupam lugar de destaque. Congressos, simpósios, seminários e conferências funcionam como arenas privilegiadas para a apresentação de resultados de pesquisas em andamento, debates sobre avanços científicos e formação de redes de colaboração (Mello, 1996). Os anais de eventos, por sua vez, registram essa produção, funcionando como fontes de conhecimento acessíveis e atualizadas (Campello et al. 2007).

A motivação para este trabalho reside na percepção de que, embora os eventos científicos sejam largamente utilizados por pesquisadores brasileiros, especialmente nas áreas aplicadas, os estudos que analisam a produção oriunda desses espaços são ainda incipientes. Em muitos campos, como informática e engenharia, os artigos publicados em anais têm status semelhante ao de periódicos, sendo reconhecidos por sua originalidade, impacto e relevância prática.

A Plataforma Lattes, criada e mantida pelo CNPq, é uma base nacional de currículos acadêmicos que concentra dados atualizados sobre formação, atuação e produção científica. Dada sua abrangência e padronização, é uma fonte estratégica para o estudo em larga escala da produção científica nacional. No entanto, a recuperação de dados estruturados, especialmente sobre publicações em eventos, exige a utilização de ferramentas de extração e tratamento específicos.

Assim, este artigo tem como objetivo analisar a produção científica brasileira em eventos, com base em dados extraídos da Plataforma Lattes, concentrando-se em autores com doutorado concluído. A análise contempla aspectos como volume de publicações, áreas do conhecimento, distribuição geográfica, editoras, uso de identificadores persistentes (como DOI) e padrões de autoria e colaboração. O trabalho se ancora nas ferramentas da bibliometria e utiliza técnicas de

ciência de dados para explorar um conjunto expressivo de informações até então pouco analisado em sua totalidade.

Além de preencher uma lacuna nos estudos sobre comunicação científica, esta pesquisa está inserida no contexto mais amplo da transformação digital e da ciência aberta. Ao estruturar e visualizar informações dispersas, ela contribui para a transparência, a avaliação e a valorização do conhecimento gerado em eventos científicos no Brasil.

2. Fundamentação Teórica

A fundamentação teórica deste trabalho está ancorada em três pilares principais: a bibliometria como ferramenta de análise científica, a Plataforma Lattes como repositório de dados curriculares e os eventos científicos como canal de comunicação do conhecimento.

A bibliometria surgiu como campo de estudo que aplica métodos estatísticos e matemáticos para mensurar a produção, a disseminação e o uso da informação científica. As principais leis bibliométricas utilizadas são: a Lei de Lotka, que descreve a produtividade dos autores e aponta que poucos autores concentram grande parte das publicações; a Lei de Bradford, que analisa a dispersão da produção entre fontes de publicação; e a Lei de Zipf, que investiga a frequência de termos em textos científicos, permitindo a identificação de tópicos recorrentes.

O uso dessas métricas é cada vez mais comum para compreender padrões em grandes conjuntos de dados, como currículos, bases de periódicos e repositórios institucionais. A aplicação conjunta dessas leis permite uma visão mais completa e integrada dos fenômenos relacionados à ciência e sua comunicação.

A Plataforma Lattes foi criada em 1999 como iniciativa do CNPq para unificar a base curricular dos pesquisadores brasileiros. Com mais de 9 milhões de currículos registrados (maio de 2025), ela se tornou referência nacional e internacional como repositório de dados acadêmicos. A padronização de categorias como tipo de publicação, natureza do trabalho, área de atuação e instituição de vínculo faz dela uma fonte estratégica para estudos de políticas públicas e avaliação da ciência.

Eventos científicos, por sua vez, representam um componente essencial da dinâmica científica. Segundo Meadows e Lemos (1999), esses encontros constituem o protótipo da comunicação científica informal, mas formalizam-se por meio dos anais. Población et al. (1996) classificam esses documentos como literatura cinzenta, dada sua circulação restrita e informal. No entanto, em muitas áreas do conhecimento, especialmente nas engenharias e na ciência da computação, os anais têm importância equivalente à dos periódicos científicos.

A relevância dos eventos também está associada à sua capacidade de estimular a interlocução entre pares, o surgimento de colaborações e a rápida disseminação de resultados. Por isso, os anais se configuram como fontes valiosas para a identificação de tendências emergentes, tópicos inovadores e padrões de interação científica.

3. Procedimentos Metodológicos

Este estudo foi conduzido a partir da análise de dados extraídos da Plataforma Lattes, com enfoque nos currículos de pesquisadores com doutorado concluído. A escolha por esse grupo se

deve ao seu alto grau de atualização e à densidade da produção científica registrada. Utilizou-se como ferramenta principal o LattesDataXplorer (Dias, 2016), arcabouço desenvolvido para automação da coleta, tratamento e visualização de dados da Plataforma Lattes.

A etapa de aquisição envolveu o download dos currículos em formato XML disponibilizados pelo CNPq. Foram coletados mais de 8 milhões de currículos, dos quais 360.888 correspondiam a pesquisadores com doutorado concluído. Desses, identificou-se um total de 11.416.655 artigos classificados como “Trabalhos completos ou resumos publicados em anais de eventos”.

O tratamento dos dados incluiu a padronização de nomes de eventos, cidades, idiomas e editoras. Também foram aplicadas regras para identificar o tipo de trabalho (resumo simples, resumo expandido ou artigo completo) e presença de identificadores persistentes, como DOI. A curadoria foi realizada com scripts em Python, utilizando expressões regulares e técnicas de processamento de linguagem natural para normalização de dados textuais.

Para fins de análise, os dados foram organizados em categorias por área do conhecimento (segundo classificação do CNPq), região geográfica, tipo de instituição e período de publicação. Em seguida, aplicaram-se as Leis de Bradford, Lotka e Zipf para extrair padrões de dispersão, concentração e frequência de temas.

A análise quantitativa foi complementada por visualizações gráficas que demonstram a evolução temporal da produção, a distribuição geográfica dos eventos e a natureza dos trabalhos apresentados.

4. Resultados

A análise revelou que aproximadamente 84% dos doutores cadastrados na Plataforma Lattes possuem pelo menos uma publicação em anais de eventos. O conjunto analisado corresponde a mais de 11 milhões de registros, representando uma importante base de artigos em eventos mapeadas no Brasil.

Ao analisar a evolução anual da quantidade de publicações em eventos, observa-se um crescimento acentuado nos anos 2000, com ápice em 2011. A partir de então, verifica-se uma queda, atribuída possivelmente à não atualização dos currículos por parte dos pesquisadores.

As áreas com maior número de publicações foram as Engenharias, Ciências Sociais Aplicadas e Ciências Humanas. A distribuição regional mostrou predominância dos estados do Sudeste, especialmente São Paulo, Minas Gerais e Rio de Janeiro (Figura 1).

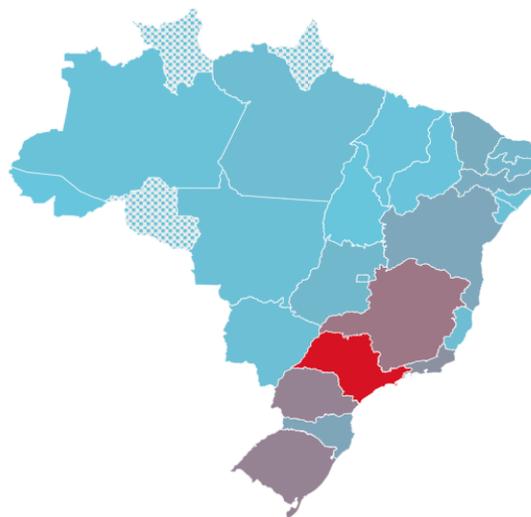


Figura 1 – Mapa de calor com o quantitativo de eventos por estados da Federação.

A aplicação da Lei de Lotka indicou que um pequeno grupo de autores concentra grande parte da produção em eventos. A Lei de Bradford revelou que há um núcleo de eventos altamente representativos, enquanto a Lei de Zipf identificou termos recorrentes como “educação”, “tecnologia”, “inovação” e “ambiente”.

Apesar da expressiva produção, observou-se que apenas uma pequena parcela dos artigos possui identificador persistente (DOI), o que dificulta a rastreabilidade e a indexação em bases acadêmicas.

5. Considerações Finais

Os resultados deste estudo confirmam a importância dos eventos científicos como espaço de difusão da produção acadêmica brasileira. A análise bibliométrica dos dados extraídos da Plataforma Lattes permitiu mapear, com alta granularidade, os padrões de autoria, temas, distribuição geográfica e evolução temporal da produção em anais de eventos.

Em termos de políticas públicas, o trabalho reforça a necessidade de valorização da literatura cinzenta representada pelos anais. A padronização de formatos, a adoção de identificadores persistentes e a inclusão desses documentos em repositórios institucionais devem ser estimuladas.

Referências

CAMPELLO, B. S.; CENDÓN, B. V.; KREMER, J. M. Fontes de informação para pesquisadores e profissionais. Belo Horizonte: Editora UFMG, 2007. 319 p.

DIAS, T. M. R. Um Estudo da Produção Científica Brasileira a partir de Dados da Plataforma Lattes. 181 p. Tese (Doutorado em Modelagem Matemática e Computacional). Centro Federal de Educação Tecnológica de Minas Gerais, Setembro 2016.

MEADOWS, A. J.; LEMOS, A. A. B. de L. A comunicação científica. [S.l.]: Briquet de Lemos/livros, 1999.

MELLO, L. L. C. C. Os anais de encontros científicos como fonte de informação: relato de pesquisa. Revista de biblioteconomia de Brasília, v. 20, n. 1, p. 53–68, 1996.