

Análise da regionalidade do conjunto de artigos publicados em eventos científicos

Analysis of the regionality of the set of articles published in scientific events

Fernanda Silva Coimbra¹, Thiago Magela Rodrigues Dias², Ronaldo Ferreira de Araújo³

(1) CEFET- MG, coimbra.sfernanda@gmail.com

(2) CEFET-MG, thiagomagela@cefetmg.br

(3) UFAL, ronaldo.araujo@ichca.ufal.br

Resumo:

A maioria dos trabalhos que avaliam a produção científica em eventos científicos possui repositórios específicos como fonte de dados, muitas vezes restritos a algumas áreas do conhecimento. Embora tais trabalhos apresentem resultados interessantes, não foram encontrados estudos macro que englobam os eventos em geral, com grande número de publicações ou que abordam diversas áreas. Nesse contexto, este trabalho tem como objetivo verificar características da produção científica publicada em eventos, compreendendo a regionalidade destes eventos, utilizando métricas bibliométricas nos dados extraídos dos currículos cadastrados na Plataforma Lattes de um grupo de 360.888 doutores, ao todo foram analisados 11.416.655 trabalhos publicados em anais de eventos. Os resultados demonstram uma análise da regionalidade dos eventos no âmbito de países, estados e cidades. Possibilitando verificar que a maioria dos trabalhos são em eventos no Brasil, predominantemente na região Sudeste e Sul, também é possível compreender as cidades que tiveram mais eventos científicos.

Palavras-chave: Produção científica; eventos científicos; anais; Plataforma Lattes.

Abstract:

Most works that evaluate scientific production in scientific events have specific repositories as a source of data, often restricted to some areas of knowledge. Although such works present interesting results, no macro studies were found that encompass events in general, with a large number of publications or that address several areas. In this context, this work aims to verify characteristics of scientific production published in events, including the regionality of these events, using bibliometric metrics in the data extracted from the curricula registered in the Lattes Platform of a group of 360.888 doctors, in all 11.416.655 works were analyzed published in annals of events. The results demonstrate an analysis of the regionality of the events in the scope of countries, states and cities. Making it possible to verify that most of the works are in events in Brazil, predominantly in the Southeast and South regions, it is also possible to understand the cities that had more scientific events.

Keywords: Scientific production; scientific events; annals; Lattes Platform.

1 Introdução

A produção científica é propagada de diversas formas e através da *internet* é possível ter acesso aos trabalhos científicos. Ao longo dos últimos anos vários autores têm se interessado pela análise da produção científica, aumentando assim o volume de estudos que visam entender como a ciência evolui e de qual forma acontece a colaboração científica (Dias, 2016).

As produções científicas são partes integrantes do processo de produção de conhecimento do indivíduo, no qual pode disponibilizar esse conhecimento adquirido através diversas formas (Domingues, 2014). Um exemplo de meio de propagação do conhecimento são os eventos científicos, nestes eventos são gerados documentos,

popularmente conhecidos como anais. As publicações geradas nos eventos são vistas por alguns estudos, como as produções acadêmicas mais atuais (Carmona; Pereira, 2018).

Analisar como as publicações neste meio de divulgação vem sendo realizadas, se apresenta como um importante mecanismo para a compreensão da evolução dos eventos científicos em um contexto geral ou em determinadas áreas do conhecimento.

Porém, as informações relacionadas à produção científica por meio de eventos científicos estão presentes em inúmeros repositórios de dados, dificultando assim a recuperação e análise dos dados de forma ampla. Representando desta forma um

grande desafio a ser enfrentado para este tipo de análise.

Neste contexto, no Brasil, a Plataforma Lattes do CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) se tornou um padrão para registro de dados curriculares da comunidade científica. A Plataforma Lattes se transformou no local onde os estudantes e pesquisadores do país podem registrar sua vida pregressa e atual, ou seja, um padrão nacional, com uma riqueza de informações, confiabilidade e abrangência (Plataforma Lattes, 2022). A Plataforma Lattes tem como premissa que os indivíduos realizem a inserção de todas suas informações curriculares na sendo que, após a inclusão, todos estes dados estão disponíveis em acesso aberto na internet, podendo ser editados no momento que o indivíduo optar. É um rico repositório, contemplando o registro da trajetória profissional, acadêmica e produção científica, no qual possibilita diversas e diferentes análises, justificando assim, a escolha desse repositório como fonte de dados para este trabalho. Porém, não é possível recuperar todos os currículos de uma única vez.

Desse modo, a Plataforma Lattes surge como uma ótima alternativa para coleta de dados do conjunto de dados sobre publicações científicas em anais de eventos.

Logo, com o conjunto de dados sobre os artigos publicados em anais de eventos registrados nos currículos da Plataforma Lattes, diversas métricas bibliométricas poderão ser aplicadas para verificar características da regionalidade da produção científica brasileira neste meio de comunicação.

Logo, o principal objetivo deste trabalho é a análise do conjunto de dados curriculares da Plataforma Lattes, no intuito de se apresentar informações inéditas sobre este meio de publicação no Brasil.

2 Procedimentos Metodológicos

A metodologia utilizada neste estudo teve como base a análise bibliométrica concomitante com metodologia quantitativa. Foi utilizado o *LattesDataExplorer* (Dias, 2016), esse arcabouço permite realizar o processo de extração e seleção dos dados curriculares da Plataforma Lattes, que

envolve um conjunto de técnicas e métodos, que possibilita a coleta, seleção, tratamento e análise dos dados. Para este trabalho, foi utilizado apenas os módulos de coleta e seleção do *LattesDataExplorer*, com o intuito de extrair e selecionar os dados curriculares da Plataforma Lattes. Deste modo, o módulo Coleta, foi realizado em etapas, sendo elas: coleta de URLs (*Uniform Resource Locator*), coleta de identificadores e extração dos currículos.

O extrator *LattesDataExplorer* foi utilizado em janeiro de 2021, para realizar a coleta de todos os currículos, cerca de 7 milhões de registros. O formato dos currículos é XML (*eXtensible Markup Language*), tal formato permite delimitações e é apropriado para o processamento automático, possibilitando uma melhor manipulação dos dados.

Para obter uma análise acurada dos artigos publicados em anais de eventos, preferiu-se estabelecer um conjunto de indivíduos através do nível de formação acadêmica/titulação que possuem o nível de formação doutorado concluído. A escolha é feita mediante ao que Dias (2016) menciona em seu trabalho: os doutores são responsáveis por 74,51% dos artigos publicados em periódico e 64,67% dos artigos publicados em anais de eventos, além de possuir em geral data de atualização de seus currículos recente e notadamente são responsáveis pelo mais alto nível de formação. Com o intuito de auxiliar as análises dos dados, após a utilização do arcabouço, foi necessário criar métodos para a seleção, tratamento e visualização conforme Figura 1 (Apêndice 01).

A etapa de tratamento baseou-se em analisar o conjunto de arquivos XML. Para cada currículo foi extraído um arquivo em formato XML. O currículo XML extraído da Plataforma Lattes apresenta sua estrutura o elemento raiz, denominado como "Currículo"; e possui cinco elementos filhos que possuem seus próprios elementos e atributos. Cada currículo é único e possui informações próprias; esses dados agregam informações sobre grandes áreas, formação acadêmica, orientações, produções, entre outras. Nesta etapa da análise é realizada a divisão das informações do currículo, posteriormente acessar as informações de interesse e

desprezar algumas informações que para este trabalho são irrelevantes. Após realizar o tratamento do XML, foi realizada a etapa de visualização em que é realizada a caracterização dos dados, permitindo a análise dos dados inseridos na Seção de Trabalhos em Eventos, seção que contém trabalhos publicados em anais de eventos, caso o indivíduo tenha informado em seu currículo.

Diversos são os desafios enfrentados nas etapas de tratamento e transformação dos dados, em especial, a falta de padrão no registro dos títulos dos eventos científicos. Para contornar tal desafio, a utilização de algumas técnicas de Processamento de Linguagem Natural foram adotadas, como por exemplo a identificação de similaridades entre *strings*, e ainda, a adoção de dicionários controlados para validação dos dados.

Através da caracterização dos dados foi possível obter indicadores gerais como: total de artigos publicados em anais de eventos, ano de publicação dos artigos, totais de indivíduos que possuem artigos publicados em anais de eventos, publicações por grande área, indivíduos sem publicações, publicações com identificadores persistentes; tais indicadores são apresentados na seção de Resultados.

3 Resultados

A caracterização inicial resultou em um conjunto de 360.888 currículos, este quantitativo representa aproximadamente 5% de currículos cadastrados na Plataforma Lattes (dados de janeiro de 2021). Dentre esses currículos foi verificado que 57.403 currículos não possuíam nenhum artigo informado na seção de trabalhos publicados em anais de eventos, correspondendo a 16% do conjunto de doutores. Após as análises dos dados foi possível compreender algumas características no que tange a regionalidade dos eventos científicos.

3.1 Regionalidade dos eventos científicos

Em todos os trabalhos inseridos na Plataforma Lattes é possível informar o “País do Evento”, se tornando uma informação relevante para as análises dos eventos científicos. A produção científica brasileira

também pode ser publicada em eventos internacionais. Foram encontrados 185 países vinculados aos trabalhos analisados. Devido a este volume de países, foi realizado o ranking com os 10 países que possuem maior número de trabalhos publicados em anais de eventos, conforme Figura 2 (Apêndice 02).

Pode-se observar que o Brasil é o país que possui maior número de publicações. A falta de investimento para tais publicações internacionais, que em geral, necessitam de recursos para viagens, pode ser uma hipótese. Segundo Serra, Fiates e Ferreira (2008), os investimentos das instituições de ensino superior para enviar seus docentes a eventos internacionais vêm diminuindo rapidamente. Os critérios de participação entre congressos nacionais e internacionais são de suma importância para a melhoria da produção acadêmica, devendo ser olhado como investimento e não como custo.

Através das análises realizadas, foi possível constatar que na Plataforma Lattes, possui trabalhos publicados em anais de eventos em 185 países, abrangendo praticamente todos os continentes: África, América do Norte, América do Sul, Ásia, Europa e Oceania. Mesmo sendo predominante os eventos serem no Brasil, pode-se observar, que a produção científica brasileira é propagada em inúmeros países, possibilitando visibilidade dos trabalhos em outros países. Além dessa distribuição por países, também foi possível identificar a sua distribuição por estados da federação.

Classificação	Estado	Total
1	São Paulo	25,36%
2	Minas Gerais	11,62%
3	Rio Grande do Sul	9,88%
4	Paraná	9,83%
5	Rio de Janeiro	7,78%
6	Santa Catarina	4,91%
7	Bahia	4,57%
8	Ceará	3,95%
9	Pernambuco	3,81%
10	Rio Grande do Norte	2,64%
11	Paraíba	2,50%
12	Goiás	2,14%
13	Distrito Federal	2,11%
14	Pará	1,48%
15	Espírito Santo	1,38%

16	Mato Grosso do Sul	1,17%
17	Mato Grosso	1,00%
18	Alagoas	0,94%
19	Sergipe	0,93%
20	Amazonas	0,64%
21	Maranhão	0,51%
22	Piauí	0,46%
23	Tocantins	0,19%
24	Acre	0,08%
25	Rondônia	0,05%
26	Roraima	0,05%
27	Amapá	0,03%

Tabela 1 – Eventos por estados da federação.

Fonte: Elaborado pelo autor (2022).

Conforme pode ser observado, os dois estados que possuem maior concentração de eventos estão localizados na região Sudeste: São Paulo e Minas Gerais. Seguidamente de estados na região Sul: Rio Grande do Sul e Paraná. Já os 3 estados que apresentaram menor volume de eventos foram: Roraima, Rondônia e Amapá. A distribuição apresentada pode estar relacionada ao fato de que a maior parte dos estados estão localizados na região Sudeste e Sul, onde estão as maiores instituições de ensino do país. De acordo com Dias (2016) a região Sudeste- Sul concentra 62,6% dos indivíduos com endereço profissional cadastrado em seus currículos. Em vista disso, faz sentido que a maioria dos eventos sejam nessas regiões, propiciando assim, maior facilidade da participação dos indivíduos. Além da maior concentração de indivíduos, essas regiões também podem estar relacionadas à concentração das principais instituições de ensino públicas do país, onde se agrupam os principais cursos de pós-graduação.

Partindo para uma análise mais específica a respeito da regionalidade dos eventos, foi realizada a caracterização dos eventos por cidades, totalizando cerca de 6.000 nomes de cidades encontradas após curadoria de dados. Também pôde-se verificar que existem 1.217.771 trabalhos publicados em anais de eventos não possuem cidade vinculada. A Tabela 2 apresenta a classificação das 30 cidades que possuem maior volume de trabalhos em eventos.

Classificação	UF	Cidade	Total
1	SP	São Paulo	630.885
2	RJ	Rio De Janeiro	479.965
3	CE	Fortaleza	304.453
4	BA	Salvador	282.003
5	SC	Florianópolis	281.386
6	RS	Porto Alegre	269.656
7	PR	Curitiba	260.442
8	MG	Belo Horizonte	256.105
9	SP	Águas De Lindóia	246.934
10	SP	Campinas	229.971
11	RN	Natal	199.188
12	PR	Foz Do Iguaçu	183.014
13	DF	Brasília	173.070
14	MG	Caxambu	160.960
15	SP	Ribeirão Preto	145.843
16	PB	João Pessoa	143.686
17	GO	Goiânia	137.544
18	RS	Gramado	135.260
19	PA	Belém	120.145
20	PR	Londrina	111.026
21	RS	Santa Maria	108.941
22	MG	Uberlândia	99.975
23	PR	Maringá	98.163
24	SP	São Carlos	90.252
25	RS	Pelotas	85.645
26	ES	Vitoria	84.173
27	AL	Maceió	74.510
28	SP	Bauru	72.937
29	MG	Viçosa	68.569
30	MT	Cuiabá	66.653

Tabela 2 – Total de trabalhos por cidade.

Fonte: Elaborado pelo autor (2022).

Dentre as 30 cidades que possuem mais trabalhos publicados em eventos é perceptível a diferença nos totais, a diferença para a 1ª posição (São Paulo) e a 2ª posição (Rio de Janeiro) é de 150.920 trabalhos. Ao se comparar os resultados apresentados na Tabela 1, percebe-se que houve algumas diferenças entre as duas cidades do Nordeste entre as 5 primeiras classificadas: Fortaleza, na posição 3 e Salvador, na posição 4.

Ou seja, não são cidades da região Sudeste-Sul, podendo assim, serem regiões que possam ganhar destaques futuramente, hipoteticamente por sua localização geográfica estar em uma área litorânea. Outra característica é que 14 cidades são capitais dos seus respectivos estados,

demonstrando assim a importância de os eventos serem em cidades com maior facilidade de acesso, mediante a questões de transporte como aeroportos.

4 Considerações Finais

A partir dos resultados e análises aqui apresentados, observou-se que os dados extraídos da Plataforma Lattes são uma excelente fonte para compreender como acontece a produção científica brasileira em anais de congresso, visto ser complexo encontrar um repositório com todos os dados de eventos científicos. Através das análises realizadas pode-se observar algumas características gerais no que tange a regionalidade dos eventos, sendo um fator importante, possibilitando compreender onde ocorrem os eventos nacionais e internacionais. Quais são as regiões do Brasil que possuem maior representatividade de eventos e cidades. Sendo possível constatar que existem regiões e cidades predominantes.

Importante destacar que devido ao preenchimento manual realizado pelos indivíduos no ato de atualização de um currículo, diversos erros, principalmente de digitação em geral leva a não consideração de um determinado registro.

Referências

CARMONA, Ingrid Valadares; PEREIRA, Vinicius Pereira. Ciência, tecnologia e sociedade e educação ambiental: uma revisão bibliográfica em anais de eventos científicos da área de ensino de ciências. *Revista Ciências & Ideias*, v. 8, n. 3, p. 94–114, 2018.

DIAS, Thiago Magela. Rodrigues. 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.

DOMINGUES, I. O sistema de comunicação da ciência e o taylorismo acadêmico: questionamentos e alternativas. *Estudos*

avançados, *SciELO Brasil*, v. 28, n. 82, p. 225–250, 2014.

PLATAFORMA LATTES. História do surgimento da Plataforma Lattes. Brasil, 2022.

SERRA, F. A. R.; FIATES, G. G.; FERREIRA, M. P. Publicar é difícil ou faltam competências? o desafio de pesquisar e publicar em revistas científicas na visão de editores e revisores internacionais. *RAM. Revista de Administração Mackenzie, SciELO Brasil*, v. 9, n. 4, p. 32–55, 2008.

Apêndice 01



Figura 1 – Processo de seleção, tratamento e visualização dos dados.
Fonte: Elaborado pelo autor (2022).

Apêndice 02



Figura 2 – Ranking - Países com maior publicação de artigos em anais de eventos.
Fonte: Elaborado pelo autor (2022).