

# UMA ANÁLISE TEMPORAL DOS PRINCIPAIS TÓPICOS DE PESQUISA DA CIÊNCIA BRASILEIRA A PARTIR DAS PALAVRAS-CHAVE DE PUBLICAÇÕES CIENTÍFICAS<sup>1</sup>

Email:  
jetherog@gmail.com  
thiagomagela@gmail.com

Jether Oliveira Gomes, Thiago Magela Rodrigues Dias, Gray Farias Moita

## *Resumo*

A produção de trabalhos científicos apresentou um crescimento importante nas últimas décadas, sendo a internet o principal fator de acesso e difusão destes. Conseqüentemente, nota-se um esforço global de todas as áreas do conhecimento quanto à estudos sobre dados da produção científica, a fim de conhecer o que tem sido pesquisado. Sendo que, tais estudos podem servir à diversos propósitos, como, fornecer base para construção de políticas de incentivo a pesquisa visando novos avanços na ciência. Neste contexto, este trabalho apresenta uma proposta de análise textual sobre o desenvolvimento científico brasileiro registrado ao longo da história na base curricular da Plataforma Lattes. Para tanto, foi desenvolvido um arcabouço de componentes com a finalidade de filtrar e tratar os dados dos currículos para padronização e definição das informações essenciais a serem analisadas, e, ao mesmo tempo, diminuir o processamento computacional para geração dos resultados desejados. Os primeiros resultados são apresentados a partir da aplicação de análises bibliométricas e técnicas baseadas em análises de redes sociais, sobre as palavras-chave de artigos publicados em anais de congressos e em periódicos. Assim, quantitativamente, foi possível apresentar uma caracterização geral das palavras-chave utilizadas, e, com isso, destacar os principais tópicos de pesquisas desenvolvidos ao longo dos últimos anos.

**Palavras-chave:** Palavras-chave. Tópicos de Pesquisa. Plataforma Lattes.

## *Abstract*

The production of scientific works presented a significant growth in the last decades, being the Internet the main factor of access and diffusion of these. Consequently, a global effort of all the areas of the knowledge regarding the studies on data of the scientific production is noted, in order to know what has been researched. These studies may serve several purposes, such as providing a basis for the construction of policies to stimulate research aiming at new advances in science. In this context, this paper presents a proposal for a textual analysis on the Brazilian scientific development recorded throughout history in the Lattes Platform curriculum. In order to do so, a framework of components was developed with the purpose of filtering and treating the data of the curricula for standardization and definition of the essential information to be analyzed, and at the same time, to reduce the computational processing to generate the desired results. The first results are presented through the application of bibliometric analyzes and techniques based on social network analyzes, on the keywords of articles published in congress proceedings and in periodicals. Thus, quantitatively, it was possible to present a general characterization of the keywords used, and, with this, to highlight the main topics of research developed over the last years.

**Keywords:** Keywords. Research Topics. Lattes Platform.

<sup>1</sup> Doutorado em Modelagem Matemática e Computacional (PPGMMC) – CEFET-MG.

## 1 INTRODUÇÃO

O grande número de informações disponibilizadas pela internet e a sociabilização da atividade científica por parte de redes de pesquisadores são os fatores essenciais para o atual desenvolvimento da ciência (BRITO et. al., 2016). Para Dias (2016), serviços como bibliotecas digitais e sítios para registro individual de produção científica são alguns exemplos de como a internet tem contribuído na quantidade de trabalhos publicados, permitindo que usuários não apenas acessem conteúdo disponível, mas também possam registrar a sua produção científica. Desta forma, trabalhos publicados e disponibilizados podem ser acessados instantaneamente contribuindo para a expansão do conhecimento.

Numa sociedade competitiva em escala mundial, implementar o conhecimento técnico e científico é tarefa primordial para alavancar o desenvolvimento econômico e social (CAVACINI, 2016), principalmente para os países em desenvolvimento que são consumidores deste conhecimento, de forma a identificar suas necessidades e peculiaridades (SAES, 2005). Porém, muitas das vezes, seu plano de implementação é condicionado à existência de recursos limitados e uma imposição cada vez maior de racionalidade e objetividade na aplicação dos poucos recursos disponíveis.

Neste contexto, pesquisadores de todos os domínios têm dedicado esforços com intuito de analisar a produção científica sobre diferentes perspectivas, como: análises de citações, indicadores de produtividade, colaboração científica e análise de tópicos. Dentre estes, esforços para analisar tópicos de pesquisas constituem formas de melhorar a compreensão do comportamento de indivíduos ou grupos de pesquisadores, bem como, evidenciar o que tem sido produzido acerca da ciência.

Os trabalhos que analisam tópicos também funcionam como uma boa revisão da literatura, o que permite, por exemplo, a verificação por parte do setor industrial se o que está sendo desenvolvido pela ciência contempla as necessidades da indústria (KHAN e WOOD, 2015). Para tanto, tais trabalhos geralmente exploram repositórios de artigos científicos, analisando seus títulos, resumos ou até todo o texto para extrair tópicos de pesquisas e analisá-los através de análises bibliométricas ou técnicas de análises de redes sociais. Entretanto, títulos e resumos de trabalhos podem não representar exatamente os assuntos abordados, devido a necessidade de se preocupar com a semântica e estrutura dos termos. Logo, uma abordagem interessante é a análise das palavras-chave de um conjunto de publicações científicas, pois são inseridas cuidadosamente por seus respectivos autores e fornecem uma possibilidade de descrever os assuntos principais que permeiam o trabalho de forma clara e objetiva sem se preocupar com questões semânticas (MCCLOSKEY, 1998; YI e CHOI, 2012). Com isso, no atual trabalho, palavras-chave de publicações científicas são igualmente referenciadas como tópicos de pesquisas.

Dentre os diversos trabalhos encontrados na literatura, nota-se um interesse global de todas as áreas de pesquisas em conhecer o que se tem feito acerca da ciência. Contudo, no contexto nacional, geralmente as análises realizadas utilizam repositórios de dados internacionais, que são específicas de uma determinada área ou periódico, e com isso, realizam estudos sobre um determinado assunto ou área de pesquisa. No entanto, justamente por tratarem de análises específicas e utilizarem repositórios de dados internacionais, não podem representar em totalidade o que é produzido no Brasil.

De acordo com Brito et. al. (2016), estudos desta natureza são considerados urgentes no Brasil e podem retratar o que é desenvolvido e publicado em ciência, tecnologia e inovação,

possibilitando gerar parâmetros que podem nortear esforços e investimentos para impulsionar resultados de pesquisa. Em Trucolo (2016), o autor destaca que muitas vezes os investimentos focam em áreas de pesquisas já consolidadas e populares, nas quais se acredita que haverá retorno, ou ainda, identificadas como tendências globais. Entretanto, num país com dimensões continentais como o Brasil, uma estratégia interessante seria investir nas áreas e tópicos de pesquisas com os maiores potenciais de crescimento, ampliando as chances do retorno da investigação científica e canalizando os recursos, que na maioria das vezes são reduzidos.

Para a realização de análises sobre o patamar científico brasileiro, o importante e abrangente repositório de dados da Plataforma Lattes, desenvolvida e mantida pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), é tido como um diferencial. Esse repositório é composto de dados referentes a grupos de pesquisas, instituições e de currículos de mais de 5.7 milhões de indivíduos, número este que tem crescido linearmente a cada ano (DIAS, 2016).

Assim sendo, é proposto neste trabalho uma análise temporal das principais palavras-chave existentes nos artigos científicos de todos os indivíduos com doutorado concluído que possuem os currículos cadastrados na Plataforma Lattes. Para tanto, inicialmente, as palavras-chave são extraídas e processadas, para posteriormente serem analisadas através de análises bibliométricas e técnicas de análises de redes sociais, para destacar os principais tópicos de interesses dos pesquisadores brasileiros de todas as grandes áreas do conhecimento, bem como a relação entre eles, ao longo dos últimos 55 anos de pesquisas registrados nos currículos da Plataforma Lattes.

## 2. TRABALHOS CORRELATOS

Nesta seção, são apresentados trabalhos que de alguma forma têm analisados termos a partir de um conjunto de publicações científicas para compreender como tem ocorrido a evolução científica, destacar os tópicos de interesse num determinado período e evidenciar quais são de investigação atual. De acordo com Khan e Wood (2015), estes estudos, geralmente, baseiam-se em popularidade (analisa a frequência de utilização de termos contidos nos títulos ou palavras-chave de artigos científicos) ou redes de citações (analisa as relações entre citações de artigos e autores). Adicionalmente, Choi et. al. (2011) destacam a importância em analisar as redes de palavras-chave para evidenciar os tópicos mais importantes baseado em medidas de centralidade.

Medeiros e Mena-Chalco (2013) analisaram 633.508 currículos cadastrados na Plataforma Lattes e 4.946.990 publicações para estudar a rede social dos indivíduos de áreas distintas. Adicionalmente, calcularam as frequências dos termos dos títulos para identificar quais são mais utilizados em cada época. Em continuação, utilizaram mapas de termos e árvores de mapas para as 200 palavras mais frequentes de cada área por período de tempo.

Em Trucolo e Digiampietri (2014a), os autores aplicaram regressões nas medidas de importância *TF-IDF* dos termos extraídos de 34.289 títulos de artigos publicados pelos doutores da Ciência da Informação entre 1991 e 2012, para identificar tendências de pesquisas. Assim, um termo era considerado tendência se apresentasse uma alta previsão de *TF-IDF*. Trucolo e Digiampietri (2014b) aplicaram as mesmas regressões nos termos de 57.501 títulos de artigos, publicados entre 1911 e 2011, pelos professores de Ciência da Computação. Em continuação, Trucolo (2016) considerou a fonte geradora da informação e, com isso, utilizou, também,

resultados de métricas de redes sociais para análises de tendências nos mesmos dados estudados em Trucolo e Digiampietri (2014b).

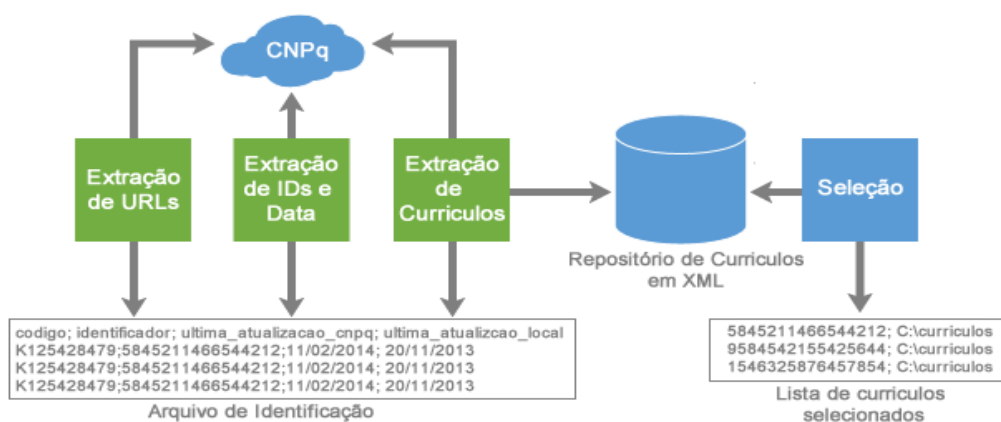
Por fim, vários outros trabalhos têm utilizado análises bibliométricas, técnicas de redes sociais e análises de tendências, para extração de conhecimento acerca do que tem sido desenvolvido nas mais variadas áreas de pesquisas com propósitos distintos, como: taxas de carbono (ZHANG et al., 2016), linhas de produtos de software (HERADIO et al. 2016), criatividade (ZHANG et al. 2015); educação matemática (FADIGAS et al., 2009); biomedicina (MADLOCK-BROWN,2014); epidemiologia na Alemanha (PETER et al., 2016); saúde no Brasil (PEREIRA et al., 2007); estrutura intelectual (RAVIKUMAR et al., 2015); e; pesquisas em universidades (CHEN e XIAU, 2016).

### 3. MATERIAIS E MÉTODOS

A motivação escolha da Plataforma Lattes como fonte de informação está relacionada basicamente a quatro fatores: (1) aos dados estarem disponíveis na internet e não terem sido amplamente analisados (DIGIAMPIETRI, 2015); (2) tratar da integração de dados de produções científicas de todas as áreas de C&T existentes na ciência brasileira ao longo de toda trajetória; (3) por não negligenciar os artigos publicados em periódicos nacionais que muitas vezes não são indexados, e também os artigos de anais de congresso (DIAS, 2016); e, (4) por ser uma poderosa fonte para fornecimento de dados de alta qualidade para medir e avaliar o desempenho acadêmico nacional (LANE, 2010).

A extração dos currículos dos doutores na versão XML foi realizada através do *LattesDataXplorer*, proposto e desenvolvido por Dias (2016) para coletar os dados curriculares do repositório da Plataforma Lattes (Figura 1).

**Figura 1** – Processo de aquisição dos dados *LattesDataXplorer*.



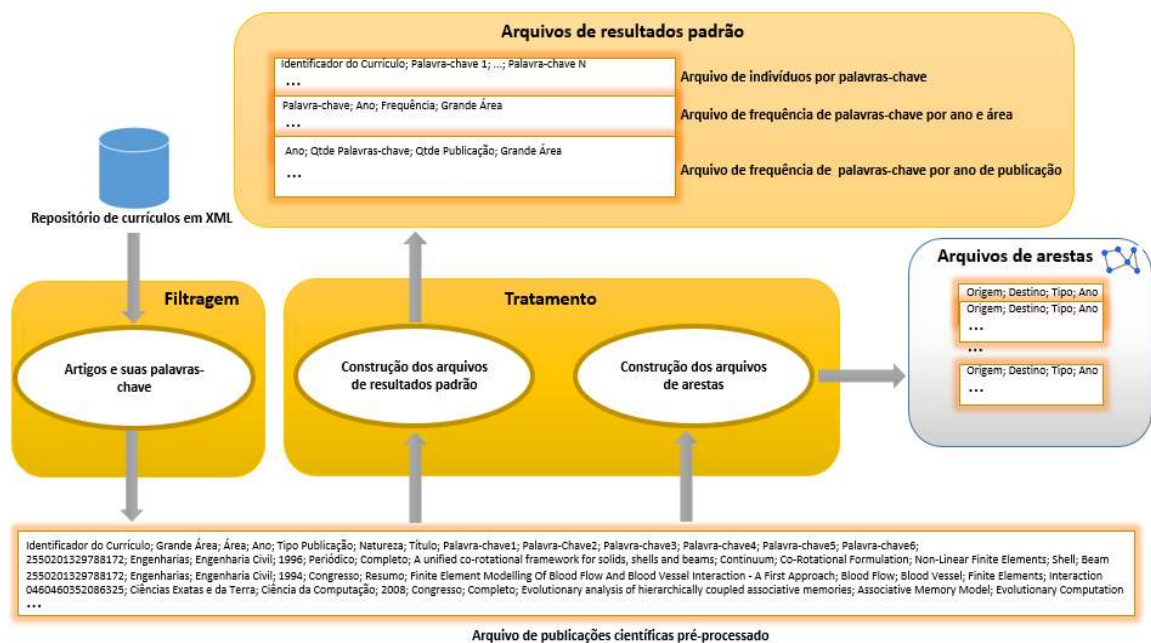
**Fonte:** Dias (2016)

Inicialmente, todos os currículos cadastrados na Plataforma Lattes são extraídos através da execução de três etapas pelo *LattesDataXplorer*, a saber: (1) extração de URLs, responsável por adquirir as referências únicas para os currículos cadastrados, (2) extração de Ids, para possibilitar o acesso a cada currículo e extrair seu identificador, e (3) extração dos currículos, responsável pelo

download e armazenamento em disco dos arquivos. Para, posteriormente, através do módulo de seleção, selecionar dentre estes, os currículos que possuem nível de capacitação doutorado concluído, definindo assim o conjunto principal de arquivos XML a serem processados.

Para o propósito deste trabalho, foi desenvolvido um arcabouço de componentes para suportar as análises desejadas (Figura 2), onde os componentes “filtragem” e “tratamento” são responsáveis pelo processo de seleção, tratamento e modelagem dos dados dos currículos que são fundamentais para as análises do atual trabalho, e, com isso, diminuir o tempo de processamento computacional.

**Figura 2** – Processo de filtragem e tratamento dos dados.



Fonte: Próprio Autor

O componente de “filtragem” realiza a etapa de mineração nos currículos para extrair as informações dos artigos, armazenando-as à parte num arquivo de publicações científicas, com isso, definindo o conjunto de dados centrais a serem estudados. As informações dos artigos incluem: identificador do currículo; grande área da publicação; ano de publicação; tipo de publicação; título e palavras-chave.

Por outro lado, o componente de “tratamento”, processa os dados do arquivo de publicações científicas para tratá-los e caracterizá-los, e, a partir destes, constrói um conjunto dos arquivos de resultados padrão e arquivos de arestas para facilitar as análises. Esse componente realiza basicamente três etapas, a saber: *limpeza dos dados*, *identificação de colaboração científica*, e *construção dos arquivos*.

A *limpeza dos dados* serve para realizar o processamento das palavras-chave de tal forma a excluir os termos que não representam tópicos de estudos e agrupar as palavras que possuem mesmo valor semântico. Para isso, inicialmente, o método desenvolvido obtém as palavras-chave de cada artigo analisado. Em seguida, cada uma das palavras-chave são associadas ao idioma cadastrado para o artigo, servindo de referência no processo de *radicalização*. O processo de

*lowercase* converte todas as palavras para minúsculo no intuito de padronizar o conjunto. No processo de *stopWords* são removidos os termos que não possuem valores semânticos. Posteriormente, no processo de *normalização* todos os acentos e pontuações são retirados das palavras-chave. Finalmente, o processo de *radicalização* é responsável pela redução da palavra-chave a seu radical. Contudo, em caso de palavras-chave compostas, este processo é executado em cada termo individualmente, e, concatenado formando uma única palavra.

Na etapa de *identificação de colaboração científica*, todos os artigos do arquivo de publicações científicas são processados para identificar e tratar os trabalhos realizados em colaboração. Esse procedimento é realizado para dados de um mesmo artigo científico não ser considerado mais de uma vez durante as análises. Para tanto, fez-se necessário uma adaptação no método *ISCooll* proposto por Dias e Moita (2015) para identificar colaborações científicas em grande volume de dados. Enquanto o *ISCooll* original utiliza um dicionário para vincular os artigos a seus autores, o *ISCooll* adaptado adota um dicionário para vincular os artigos ao conjunto união das palavras-chave. Outra diferença está na formação da chave do dicionário; enquanto *ISColl* original utiliza o título do artigo concatenado com o ano de publicação, o *ISCooll* adaptado adiciona também a grande área da publicação.

A *construção dos arquivos* serve para facilitar a extração de conhecimento durante as análises. Nesta etapa, foram construídos dois conjuntos de arquivos, a saber: arquivos de resultado padrão e arquivos de arestas. Os arquivos de resultados padrão formam o repositório principal que possibilitam as análises estatísticas sobre os tópicos de pesquisas desenvolvidos. Os arquivos de arestas permitem a construção das redes de palavras-chave. As redes de palavras-chave são baseadas no processo de adição de cliques por justaposição, onde cada artigo possui um conjunto de palavras-chave que formam uma clique. Ou seja, para cada artigo analisado uma clique é gerada, na qual palavras-chave se caracterizam como vértices e os vínculos entre estes são estabelecidos quando duas ou mais palavras-chave são utilizadas no mesmo artigo.

#### 4. RESULTADOS

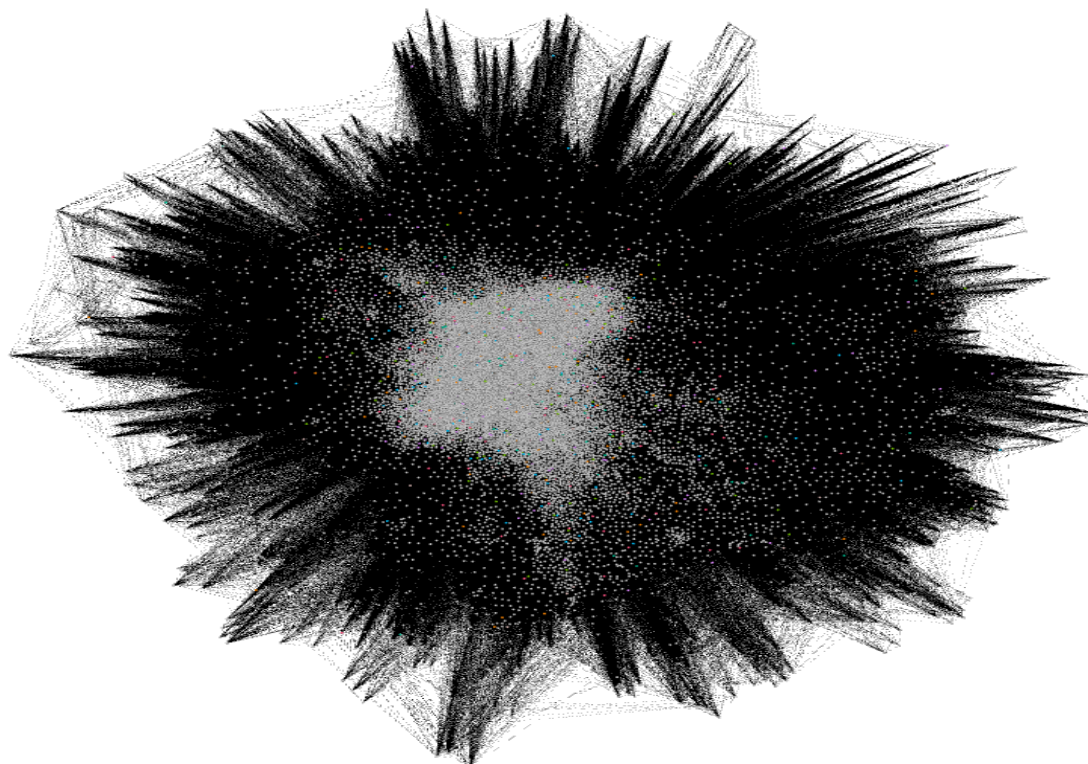
Apesar da base curricular da Plataforma Lattes contar com mais de 5.7 milhões currículos cadastrados, neste trabalho, foi considerado apenas os currículos cujos indivíduos possuem o nível de capacitação doutorado concluído. Entretanto, vale ressaltar que o número de doutores representa cerca de 5% do total de indivíduos com currículos cadastrados na Plataforma Lattes, contudo, estes são responsáveis cerca de 70% de todos os artigos cadastrados nos currículos; que, aliado a diversidade e contemporaneidade de atualização destas informações, pondera a validade dos dados para auxiliar na compreensão sobre a produção científica brasileira (DIAS, 2016).

Os dados foram coletados em abril de 2017, totalizando 265.170 currículos de indivíduos com doutorado concluído. Para as análises, foram considerados os artigos únicos (colaboração) publicados em anais de congressos e em periódicos referentes ao período de 1962 até 2016, totalizando 10.040.664 artigos e 24.256.312 palavras-chave.

Inicialmente, foi construída uma rede completa com base em todo conjunto de palavras-chave que os doutores associaram a seus artigos cadastrados nos currículos da Plataforma Lattes, totalizando 2.088.220 vértices e 20.093.922 arestas, onde sua componente gigante é responsável por 1.903.657 vértices e 20.046.249 arestas (Figura 3). Nesse caso, o cálculo de algumas métricas, necessitou do dispêndio de um alto tempo computacional, que, dado a limitação de tempo, tornou-

se inviável. Assim sendo, e, amparado pela análise da distribuição de quantidade de palavras-chave por frequência de utilização, onde 98,6% das palavras-chave únicas possuem frequência de utilização inferior a 100, tomou-se a decisão de reconfigurar o conjunto sem estas palavras-chave para facilitar os cálculos referentes aos resultados apresentados a seguir. Com isso, a nova rede é composta por 112.812 vértices e 8.779.131 arestas e sua componente gigante possui 112.799 vértices e 8.779.131 arestas, ou seja, 99% de abrangência desta rede.

**Figura 3** – Rede de palavras-chave completa (1962 – 2016).



**Fonte:** Próprio Autor

Em continuação, para conhecer os tópicos que se constituem como os principais assuntos de pesquisa contidos na rede, todos os vértices (palavras-chave processadas) foram ranqueados de acordo com seu número de grau, e, posteriormente, todas as palavras-chave processadas tiveram suas frequências calculadas. Para facilitar a comparação entre as palavras-chave classificadas pelas duas medidas de importâncias utilizadas, a Tabela 1 apresenta as quinze principais palavras-chave ranqueadas pela medida de frequência e as respectivas ordem de ranqueamento considerando o grau do vértice referente ao conjunto histórico dos dados.

**Tabela 1:** Comparativo entre o ranqueamento das palavras-chave entre 1962 e 2016.

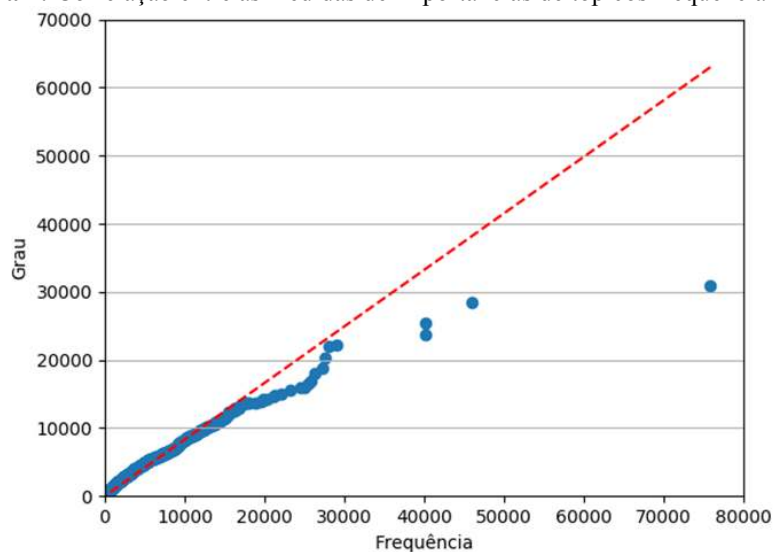
1962-2016	Frequência	Grau	R. Frequência	R. Grau
<b>Educação</b>	75.855	30.951	1	1
<b>Formação do Professor</b>	46.032	15.923	2	13
<b>Enfermagem</b>	40.228	13.710	3	25
<b>Epidemiologia</b>	40.158	25.398	4	3
<b>Crianças</b>	29.048	18.047	5	9

<b>Políticas Públicas</b>	28.100	15.929	6	12
<b>Amazônia</b>	27.626	23.800	7	4
<b>Bovinos</b>	27.236	16.883	8	10
<b>Idoso</b>	26.291	11.589	9	40
<b>Ensino</b>	25.760	14.268	10	18
<b>Ratos</b>	25.480	20.270	11	7
<b>Brasil</b>	24.981	28.527	12	2
<b>Diagnóstico</b>	24.520	22.172	13	5
<b>Educação Ambiental</b>	23.194	10.103	14	61
<b>Cultura</b>	22.123	15.026	15	15

Fonte: Próprio Autor

Como pode ser observado, apesar de existir um alto percentual de palavras-chave iguais entre as principais ranqueadas pela frequência e grau, praticamente a ordem de destaque destas não são a mesma, exceto para “Educação” e “Cultura”. Diante disto, para melhor compreender a relação entre os resultados encontrados, uma análise de correlação foi realizada. Para tal análise, como apresentado na *Figura 4*, foi calculado o *coeficiente de Spearman* considerando todo o conjunto de palavras-chave entre frequência e grau.

**Figura 4:** Correlação entre as medidas de importâncias de tópicos frequência e grau.



Fonte: Próprio Autor

Por fim, como pode ser observado, os resultados apresentam um coeficiente de correlação positivo de “0,9572” para frequência e grau. Pelo valor obtido, destaca-se a comprovação de uma relação direta entre a frequência das palavras-chave e medidas baseada em grau nas redes de palavras-chave, indicando que a medida de grau é uma alternativa interessante para identificação dos principais tópicos investigados.

## 5. CONSIDERAÇÕES FINAIS

Foi verificado que o grande aumento no número de publicações científicas disponível na web, tem tornado a bibliometria objeto de estudo para um crescente número de pesquisadores de

todas as áreas do conhecimento, interessados em compreender o que a comunidade científica tem produzido. Tal entendimento serve principalmente como auxílio a diversos níveis para tomada de decisão, como por exemplo: (1) em um nível operacional, o pesquisador que deseja iniciar uma nova pesquisa, pode escolher assuntos que estão em foco ou que tendem ganhar repercussão num curto espaço de tempo, e; (2) no nível estratégico, é primordial para os governos, negócios e fundações que devem decidir sobre investimentos para construção de políticas científicas, ter competências e insumos para analisar o patamar científico das nações e/ou grupos individualizados, seja para impulsionar resultados sobre determinada pesquisa ou iniciar estudos sobre novos temas.

Neste contexto, este trabalho apresenta o desenvolvimento de um primeiro estudo para identificar e analisar temporalmente (ano a ano) o comportamento dos principais tópicos de pesquisas de interesses da comunidade científica brasileira. Para tanto, utiliza as palavras-chave de todos os artigos científicos publicados pelos doutores em anais de congressos e periódicos cadastrados na base curricular da Plataforma Lattes. Estes artigos possuem datas de publicações entre 1962 e 2016, representando os últimos 55 anos de história de desenvolvimento da ciência brasileira.

Os resultados apresentados foram a partir de análises bibliométricas, baseado em frequência e técnicas para análises de redes sociais, aplicadas sobre as palavras-chave dos artigos. Comparando as 15 principais palavras-chave ranqueadas pelas duas medidas de importâncias de tópicos, destaca-se a relação entre as medidas de grau e frequência, onde foram similares em 73% para o conjunto das 15 principais. Porém, a comprovação da relação direta entre grau e frequência se confirmou através do alto valor de correlação (“0,9572”) considerando todo o conjunto de análise. Diante disso, constata-se que a utilização do grau como medida de importância se mostra uma alternativa interessante à frequência.

Por fim, constatou-se a validade dos resultados do ponto de vista quantitativo, uma vez que foi possível auxiliar no entendimento sobre o desenvolvimento da ciência brasileira; por outro lado, no caso da Plataforma Lattes, que possui dados de periódicos e conferências das mais variadas áreas de pesquisas, com diferentes níveis de qualidade, análises apenas quantitativas para destacar tópicos podem induzir a interpretações equivocadas. Com isso, como próximos passos desta pesquisa, espera-se que sejam utilizadas medidas de importância de tópicos que consideram características qualitativas apropriadas ao contexto da Plataforma Lattes, como, o fator de impacto dos periódicos, e em continuação, realizar análises temporais para destacar a evolução dos tópicos estudados.

## REFERÊNCIAS

Brito, A.G.C., Quoniam, L. & Mena-Chalco, J. P., 2016. Exploração da Plataforma Lattes por assunto: proposta de metodologia. *Transinformação*, vol.28, n.1, pp. 77-86.

Cavacini, A. Recent trends in Middle Eastern scientific production. *Scientometrics*, v. 109, n. 1, p. 423-432, 2016.

Chen, G.; Xiau, L. Selecting publication keywords for domain analysis in bibliometrics: A comparison of three methods. *Journal of Informetrics*, v. 10, n. 1, p. 212-223, 2016.

Choi, J.; et. al. Analysis of keyword networks in MIS research for predicting knowledge evolution. *Information & Management*, v. 48, n. 8, p. 371-381, 2011.

Dias, T. M. R., 2016. *Um estudo da produção científica brasileira a partir de dados da Plataforma Lattes*. (Doutorado). Pós-Graduação em Modelagem Matemática e Computacional, Centro Federal de Educação Tecnológica (CEFET-MG).

Dias, T. M. R.; Moita, G. F. A method for the identification of collaboration in large scientific databases. *Em Questão*. v.21, n.2, 2015.

Digiampietri, L. A. *Análise da rede social acadêmica brasileira*. 2015. 160. Tese (Livre Docência). Escola de Artes, Ciências e Humanidades da Universidade de São Paulo USP, São Paulo, 2015.

Heradio, R.; et. al. A bibliometric analysis of 20 years of research on software product lines. *Information and Software Technology*, v. 72, n. 1, p. 1-15, 2016.

Khan, G. F.; Wood, J. Information technology management domain: emerging themes and keyword analysis. *Scientometrics*, v. 105, n. 2, p. 959-972, 2015.

Lane, J. Let's make science metrics more scientific. *Nature*, v. 464, n. 7288, p. 488-489, 2010.

Madlock-Brown, C. R. *A framework for emerging topic detection in biomedicine*. 2014. 110 (Thesis). University of Iowa, 2014.

Mccloskey, D. N, 1998. *The rhetoric of economics*. University of Wisconsin Press.

Medeiros, C.; Mena-Chalco, J. P. The Dynamics of Multidisciplinary Research Networks - Mining a Public Repository of Scientists CVs. *World Social Science: social transformations and the digital ages*, 2013, Montreal, Canadá.

Pereira, J.C.R.; et. al. Who's who and what's what in Brazilian Public Health Sciences. *Scientometrics*, v. 73, n. 1, p. 37- 52, 2007.

Peter, R.S.; et. al. Epidemiologic research topics in Germany: a keyword network analysis of 2014 DGEpi conference. *European Journal of Epidemiology*, v. 31, n.6, p. 635- 638, 2016.

Ravikumar S.; et. al. Mapping the intellectual structure of scientometrics: a co-word analysis of the journal *Scientometrics* (2005–2010). *Scientometrics*, v. 102, n. 1, p. 929-955, 2015.

Trucolo, C.C. *Análise de tendências em redes sociais acadêmicas*. 2016. (Mestrado). Programa de Pós-Graduação em Sistemas de Informação, São Paulo, Brasil.

Trucolo, C. C.; Digiampietri, L. A. Análise de tendências da produção científica nacional na área de Ciência da Informação: estudo exploratório de mineração de textos. *AtoZ: novas práticas em informação e conhecimento*, v. 3, n. 2, p. 87-94, 2014a.

Trucolo, C. C.; Digiampietri, L. A. Trend Analysis of the Brazilian Scientific Production in Computer Science. *Revista de Sistemas de Informação*. 2014b. p. 2-10.

Saes, S.G., 2005. *Aplicação de métodos bibliométricos e da co-word analysis na avaliação da literatura científica brasileira em ciências da saúde de 1990 a 2002*. (Doutorado). Pós-Graduação em Saúde Pública, USP, São Paulo.

ZHANG K.; et. al. A bibliometric analysis of research on carbon tax from 1989 to 2014. *Renewable and Sustainable Energy Reviews*, v. 58, p. 297-310, 2016.

Yi, S. & Choi, J., 2012. The organization of scientific knowledge: the structural characteristics of keyword networks. *Scientometrics*, vol. 90, n. 3, pp. 1016-1026.