

**Análise da evolução da pesquisa em engenharia de transportes****Analysis of the evolution of research in transport engineering**

DOI:10.34117/bjdv6n2-081

Recebimento dos originais: 30/12/2019

Aceitação para publicação: 07/02/2020

**Ercilia de Stefano**

Doutora em Engenharia de Transportes – UFRJ – COPPE – PET  
Professora Adjunta UFF  
Rua Domingos Silvério, 01 – Petrópolis – RJ  
E-mail:ercilia@ufrj.br

**Fátima Regina Neves Lima**

Mestre em Engenharia Civil – UFRJ – COPPE  
Professora Subtituta UFF  
Rua Domingos Silvério, 01 – Petrópolis – RJ  
E-mail:engfatimalima@gmail.com

**Fernanda Santana de Paulo**

Doutora em Agronomia - UFRRJ  
Professora Subtituta UFF  
Rua Domingos Silvério, 01 – Petrópolis – RJ  
E-mail: nandasibylla@yahoo.com.br

**Gabriela Filgueiras Valadão Borges**

Graduando em Engenharia de Produção - UFF  
Rua Domingos Silvério, 01 – Petrópolis – RJ  
E-mail: gborges.engenharia@gmail.com

**Giovane Hammes Vieira**

Graduando em Engenharia de Produção - UFF  
Rua Domingos Silvério, 01 – Petrópolis – RJ  
E-mail: giovanehammes93@gmail.com

**RESUMO**

Este estudo realiza uma análise da evolução da pesquisa em Engenharia de Transportes. Mais de oito mil artigos publicados por um periódico, ao longo de quase 50 anos, foram analisados sob os aspectos da medição de produção científica, da identificação do conhecimento em publicações; o acompanhamento de tendências e evolução na pesquisa. Para realizar os estudos bibliométricos, informétricos e cienciométricos aqui apresentados, foi desenvolvido um software especialista, que é também capaz de analisar, de forma semelhante, outras publicações científicas. O software utiliza técnicas de inteligência artificial, processamento de linguagem natural (PLN), grafos, dentre outros recursos.

**Palavras-chave:** Cienciometria, Processamento de Linguagem Natural, Bibliometria, Engenharia de Transporte

## ABSTRACT

This study is an analysis of the evolution of the research in Transport Engineering. More than eight thousand articles published in a period of almost 50 years were analyzed to identify metrics of scientific production, creation of new knowledge and the rise and fall of trends in the studied. To carry out the bibliometric, computational and scientometric studies presented in this paper, a specialized software was developed, which is also capable of analyzing other scientific publications. The software uses techniques of artificial intelligence, natural language processing (NLP) and presents the results using graphs and charts.

**Keywords:** Scientometrics, Natural Language Processing, Bibliometrics, Transport Engineering.

## 1 INTRODUÇÃO

A trajetória das pesquisas na área da engenharia de transportes pode ser descrita por uma análise detalhada de publicações ao longo do tempo? Este artigo tem o objetivo de responder essa questão. Apresenta uma síntese do estudo realizado, onde analisou-se os mais de oito mil artigos publicados, sob os aspectos da medição de produção científica, da identificação do conhecimento em publicações e o acompanhamento de tendências e evolução na pesquisa (de STEFANO, 2016). Para realizar os estudos bibliométricos, informétricos e cienciométricos foi desenvolvido um software especialista (de STEFANO, et al., 2016), que é também capaz de analisar, de forma semelhante, outras publicações científicas.

A principal contribuição deste artigo é apresentar uma trajetória das pesquisas científicas na área de Engenharia de Transportes, a partir de um estudo científico estruturado com fundamentos na cienciometria, informetria e bibliometria, bem como descrever os passos para a criação do software e do *data mining* utilizado para realizar os estudos. Para se alcançar os resultados esperados foi selecionado o periódico *Transportation Research* (Transportation Research, 2015) como fonte dos dados, por ser uma das revistas mais tradicionais na área de transportes, e que apresenta uma grande diversidade e abrangência de temas.

## 2 ETAPAS PARA A CRIAÇÃO DO SOFTWARE E DATA MINING

A criação de um software especialista teve como objetivo automatizar o estudo cienciométrico, informétrico e bibliométrico de um grande volume de dados (*big data*), aplicando técnicas de inteligência artificial (RUSSEL, et al., 2004), processamento de

linguagem natural e *data mining*. A seguir apresenta-se as etapas realizadas para a elaboração e construção do *software* e do *data mining*.

Para iniciar a construção da base de dados, primeiramente fez-se necessária a obtenção de todos os artigos da revista *Transportation Research*. Para isso, a primeira etapa realizada foi o download de mais de doze mil documentos. Realizar essa tarefa manualmente seria praticamente impossível, logo, utilizou-se a técnica *Web Scraping* (colheita web ou extração de dados da web), que permite a extração de informações de sites web usando um *bot* ou *web crawler*, técnica essa adotada pela maioria dos robôs de busca. *Web Scraping* é muito eficiente para simular a navegação humana.

Após a realização de uma análise prévia, descartou-se em torno de 28% do total dos arquivos baixados, pois tratavam de informativos diversos, índices e listas de editores. A segunda etapa realizada foi a conversão dos arquivos válidos do formato PDF para o formato TXT, para permitir a análise de conteúdo para a criação do banco de dados inteligente.

Na sequência realizou-se o emprego do processamento de linguagem natural. O processamento de linguagem natural consiste na aplicação de métodos e técnicas que possibilitem ao computador extrair a semântica da linguagem humana, que pode ser expressa em textos, como no presente caso, ou em voz (YUE, et al., 2012). Em síntese, o PLN também pode ser visto como “(...) o desenvolvimento de paradigmas ou algoritmos que requeiram máquinas para realizar tarefas cognitivas, para as quais os humanos são habitualmente melhores” (SAGE, 1990).

O significado das palavras e seus agrupamentos etimológicos estão presentes em todos os níveis de análise e em banco de dados; e o estudo de sua dinâmica de frequência e distribuição é relativamente fácil com a ajuda computacional (BJÖRNEBORN e INGWERSEN, 2004). O WordNet (Princeton University, 2015) foi utilizado para classificar as palavras encontradas nos artigos em substantivos, verbos, adjetivos e advérbios e as agrupou em conjuntos de sinônimos cognitivos, e estes conjuntos são interligados por meio de relações lexicais. O WordNet vai além, e estabelece também relações conceituais-semânticas. A realização desta análise léxica é de fundamental importância para a criação do *data mining*.

Na sequência inicia-se o emprego da técnica de *tokenization*, para “quebrar” fluxos de texto em palavras, frases, símbolos ou outros elementos significativos denominados *tokens* (JURISH e WÜRZNER, 2013). Detectar os limites dos *tokens* e frases é uma etapa de pré-processamento importante. Os principais desafios da tarefa são eliminar as ambiguidades de conjugações verbais, por exemplo, bem como identificar classes gramaticais não relevantes

para a aplicação (JURISH e WÜRZNER, 2013). O *software* MetrIA utiliza os critérios definidos por Leydesdorff e Milojevic (2013), para criação dos *tokens*.

Ainda com emprego do WordNet, outro passo realizado na sequência consiste em categorizar os *tokens* pelo critério *part-of-speech*, ou seja, identificar uma categoria de palavras, ou genericamente, da função gramatical que cada *token* exerce dentro da frase. Palavras que são atribuídas às mesmas *part-of-speech* geralmente apresentam comportamento semelhante em termos de sintaxe, e desempenham funções semelhantes dentro da estrutura gramatical das sentenças e, por vezes, em termos de morfologia, em que sofrem inflexão para palavras semelhantes. Vale destacar que a semelhança semântica, também muito útil para a seleção com qualidade das palavras, é um conceito pelo qual uma métrica é dada aos grupos de termos ou documentos baseados na similitude de seus significados (FURLAN, et al., 2013). A classificação por semelhança semântica realizada pelo *software* Wordnet agrupou palavras semelhantes de grafia e de significado.

A seguir foi realizada a remoção das denominadas *stop words* (palavras muito comuns que servem apenas para apoio gramatical, como por exemplo: e, ou, a, isto, este, etc.) do texto. Cada idioma possui uma série de palavras que podem ser consideradas *stop words* de acordo com o contexto.

Na sequência, o algoritmo de *term frequency–inverse document frequency (tf-idf)* foi utilizado para verificar o quão importante é a palavra em um determinado documento (HONG, et al., 2013). Este algoritmo permite classificar os termos por relevância, dada a consulta de um usuário, e permite identificar os principais temas abordados na publicação (AIZAWA, 2003). O estudo da relevância tem como objetivo verificar se ler apenas o resumo é válido para se conhecer o conteúdo de um artigo. O outro estudo é uma análise inter-publicações, ou seja, objetiva-se localizar e listar as 10 publicações mais semelhantes ao artigo indicado pelo usuário. Essa similaridade é muito útil, pois indica ao pesquisador/leitor quais outros textos são similares por conteúdo, facilitando a pesquisa por artigos em uma base tão volumosa como a que foi criada para esse estudo. Para isso realizou-se o mapeamento da interseção de *tokens* idênticos entre os textos e também da totalidade de *tokens*, através da utilização do conceito da similaridade de Jaccard, que determina a similaridade entre dois conjuntos (LESKOVEC, et al., 2014).

Após essa etapa, inicia-se então a mineração de dados ou *data mining*, técnica essa utilizada para executar a extração de conhecimento utilizável de grandes conjuntos de dados (HONG, et al., 2013), envolvendo métodos de inteligência artificial, aprendizado de máquina,

estatísticas e sistemas de banco de dados (BRINN, et al., 2000). Vale destacar que *data mining* é um importante processo empregado para descobrir informações ocultas e potencialmente útil em grandes bases de dados (AGGARWAL *et al.*, 2012).

A mineração de texto, *text mining*, também utilizada no *software* desenvolvido, é uma extensão da mineração de dados e tem como objetivo encontrar modelos úteis, tendências, padrões ou regras em dados textuais não estruturados (HEA, et al., 2013), como as publicações da *Transportation Research*.

Realizadas todas as etapas, obteve-se o banco de dados. O fluxograma a seguir retrata de modo esquematizado os passos realizados para a obtenção do banco de dados, que foi descrito nesta seção. (Figura 1).

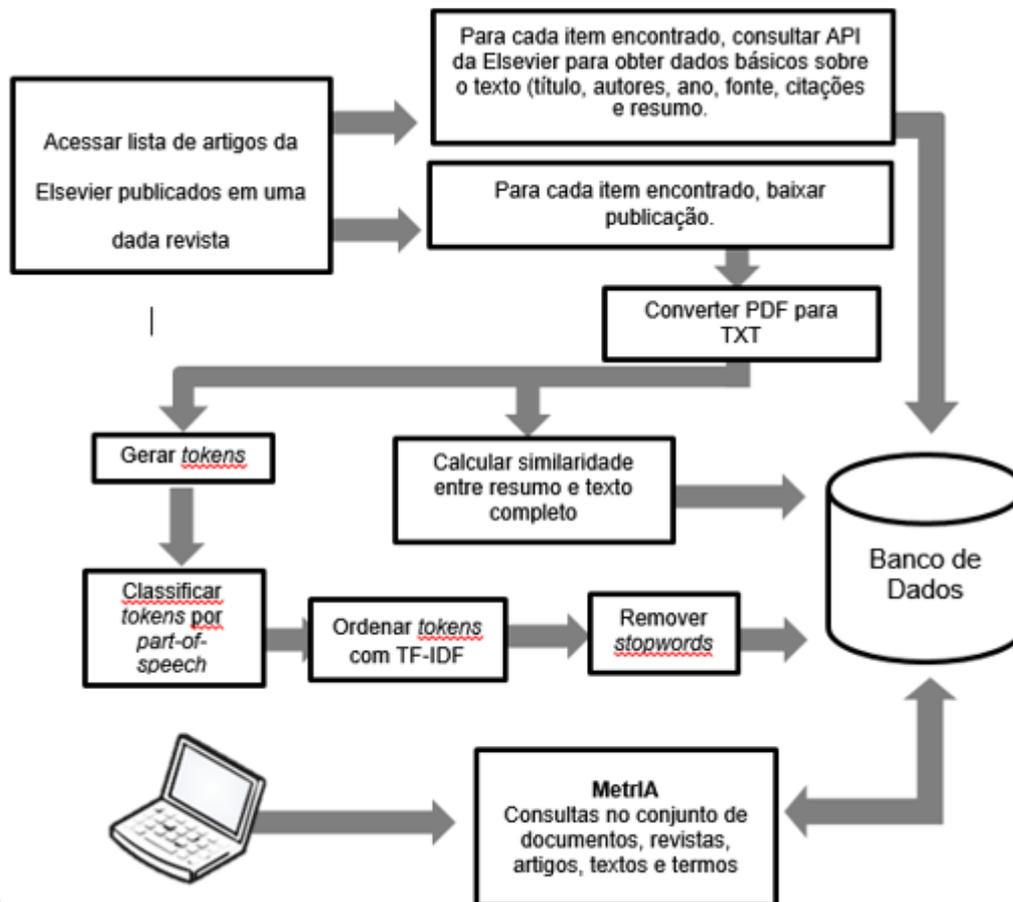


Figura 1 - Fluxograma para obtenção do banco de dados

Com o *data mining* criado, iniciou-se elaboração das telas de consulta do *software*, gráficos, tabela e relatórios conforme as necessidades de apresentação dos estudos. Para esta etapa foram empregadas técnicas convencionais de programação para ambiente Web.

### 3 ANÁLISES REALIZADAS

Na análise por publicações apresenta-se a similaridade, uma variável criada para avaliar quanto o corpo do texto da publicação é similar ao seu próprio resumo. Calculada pelo algoritmo *TF-IDF*, apresenta variação de 0,1, dentro do intervalo entre 0 e 1, sendo 0 nada similar e 1 sendo 100% similar, ou seja, quanto maior o índice apresentado, maior a semelhança entre texto e resumo.

Nesta mesma análise apresenta-se os dez artigos mais similares um artigo selecionado, do ponto de vista de seu conteúdo, indicando referências para o pesquisador. O *software* faz o estudo comparativo de similaridade sem se limitar às revistas do periódico selecionado, aumentando a diversidade de artigos analisados. Exibe-se ainda os 10 termos de maior relevância no *abstract* e os 10 termos de maior relevância no texto completo de cada artigo, exibindo os respectivos TF-IDF. E finalmente, exibe-se as citações do artigo selecionado e o total, conforme visualizado abaixo (Figura 2).

tf-idf		Citações (11)	
Abstract		Texto Completo	
Termo	tf-idf	Termo	tf-idf
transport	1.22	transport	3.62
problem	0.92	urban	2.49
scientific	0.61	theses	2.27
changes	0.61	planning	1.96
paradigm	0.61	research	1.87
research	0.61	modelling	1.56
united	0.61	number	1.32
detected	0.3	problem	1.23
1981	0.3	main	1.23
activity	0.3	subject	1.17

Autor(es)	Título	Fonte
M. Batty	Urban Modelling	Cambridge Univ. Press, Cambridge, MA (1976)
D.A. Hensher	Urban transport planning—The changing emphasis	Socio-Econ. Plan. Sci., 13 (1979), pp. 95–104
B.G. Hutchinson	Urban transport policy and policy analysis methods	Transp. Rev., 1 (1981), pp. 169–186
P.M. Jones	A New Approach to Understanding Travel Behaviour and its Implications for Transportation Planning	Ph.D. thesis, Imperial College, London (1984)
T.S. Kuhn	The Structure of Scientific Revolutions	University of Chicago Press

Figura 2 – Termos mais frequentes no resumo e corpo do texto (tabela 1), e na tabela 2 apresentam-se as citações.

Na análise por autores, o estudo bibliométrico resulta em uma tabela decrescente do nome dos pesquisadores que mais publicaram na revista *Transportation Research*. Na sequência exibe-se os 10 autores que mais publicaram na revista, e o total de publicações (Figura 3).

Total de Publicações	Autor
98	Yang, Hai
97	Hensher, David A.
87	Gudmundsson, Sveinn
78	Daganzo, Carlos F.
70	Bhat, Chandra
65	Sheu, Jih-Biing
65	Cassidy, Michael J.
64	Lam, William H.K.
53	Meng, Qiang
48	Wong, S.C.

Primeira << 1 2 3 4 5 6 7 8 9 10 >> Última

Figura 3 – Autores que mais publicaram na Revista *Transportation Research*.

A análise bibliométrica apresenta o pesquisador com o maior número de publicações, Hai Yang, professor titular do departamento de Engenharia Civil na Universidade de Hong Kong, com 98 artigos publicados. Uma análise em suas publicações demonstra que elas abrangem diversas áreas e aplicações da Engenharia de Transporte, não havendo, portanto, um padrão na temática de seus textos.

Os 10 autores analisados perfazem juntos 0,08% dos autores que publicaram na revista *Transportation Research*, e são responsáveis por publicarem 8,84% dos artigos, ou seja, quase 10% do total. O estudo mostrou que são 12.716 autores distintos no período analisado que publicaram na referida revista.

É possível selecionar um autor e visualizar todas suas publicações. É apresentado o total de artigos do autor, e subtotais anuais, a média de similaridade de seus artigos/resumos, termos mais relevantes de suas publicações, bem como as áreas temáticas de interesse do referido pesquisador. Assim é possível identificar outro(s) autor(es) que publicam temáticas semelhantes. Permite identificar assim outros artigos que possam ser de interesse, e propicia oportunidade de se estabelecer trabalhos de cooperação e parcerias com autores afins.

Outra análise realizada é por termo (palavra), sendo a consulta pode ser de único termo ou de todo o conjunto de termos catalogados pelo *software*. É possível a análise da variação de ocorrência dos termos, tanto por relevância quanto pelo número de ocorrências (frequência). A principal diferença dentre as duas classificações é que a ocorrência por

relevância só registra a ocorrência do termo quando ele é de fato significativo dentro da publicação analisada. Já a análise por frequência é a computada pela simples aparição do termo no trabalho publicado, sendo que neste caso não se avalia a relevância do termo para o artigo. Como exemplo de consulta de relevância, a palavra *emission* apresenta 127 ocorrências significativas, sendo sua primeira ocorrência foi em 1992, sendo constantes desde o ano de 1997. Pode-se inferir que o aumento da ocorrência da palavra *emission* nas publicações retrata a preocupação dos pesquisadores com as temáticas atuais, que visam fomentar ações efetivas para redução de emissão de poluentes. Tratados internacionais de redução de emissão de poluentes, como por exemplo a ECO-92, Protocolo de Kyoto (1997), Rio+10 (2002), bem como políticas internas e ambientalistas buscam soluções para minimizar e reduzir os impactos da poluição, e as publicações acompanham essa tendência.

Os termos de maior relevância da Revista *Transportation Research* no período analisado são: transporte, pesquisa, modelo, viagem, tráfego, veículo, custo e redes. Os dois termos de maior frequência, *transportation* e *research*, são a temática e o título da revista, por isso justifica-se sua maior frequência. A seguir apresenta-se a *tabela 1* com termos de maior frequência:

Tabela 1 – Listagem dos termos de maior frequência.

Frequência do termo	Termo
8391	transportation
8374	research
8343	time
7907	analysis
7729	model
7727	paper
7715	case
7624	university
7471	data
7426	study
7319	set
6927	system
6664	information
6638	point
6555	terms

Grafos foram também utilizados nas análises de dados. Para realizar essas análises subdividiu-se os artigos por períodos de 5 anos, e foram formatados em nós compostos pelas dez palavras mais relevantes de cada texto, de acordo com a classificação do TF-IDF. As arestas entre os nós contabilizam a relação de vizinhança com outras palavras no mesmo texto. O peso de uma aresta foi definido pela quantidade de vezes que uma relação de vizinhança foi identificada. O grau de um nó é a quantidade de arestas relacionadas àquele contabilizadas (BOAVENTURA NETO, 2003).

Neste estudo, a distribuição dos nós no grafo foi realizada com o algoritmo Force Atlas 2. Para facilitar a identificação de agrupamentos em cada grafo foi utilizado cálculo de classe de modularidade. Este cálculo utiliza o algoritmo Louvian que identifica grupos de comunidades. Cada comunidade é identificada por uma cor distinta. Em cada aresta também foi registrada a data da relação, de modo a permitir a filtragem por período histórico e a revista em que a aresta ocorreu.

Observa-se nos grafos que quanto maior é o círculo onde encontra-se o termo, comparado aos demais, maior sua relevância dentro da totalidade de publicações no período determinado. As diferentes cores indicam a modularidade, que é uma medida da estrutura das redes. A modularidade mede a força da divisão de uma rede em módulos (também chamados de grupos, agrupamentos ou comunidades). Redes com alta modularidade têm conexões densas entre os nós dentro de módulos, mas as ligações esparsas entre nós em diferentes módulos (NEWMAN, 2006). Através da modularidade os termos foram agrupados por similaridade.

A análise permitiu acompanhar a evolução das tendências das publicações ao longo do tempo, bem como identificar termos de maior relevância e agrupamentos de termos. Foram elaborados e analisados 10 grafos ao todo (de STEFANO, 2016), e apresenta-se a seguir o décimo grafo (figura 4), que compreende o período de 2010 a 2014, inclusive. Este grafo apresenta grande pulverização entre diversos temas e comunidades, além de apresentar muito mais interconexões entre as arestas, sendo mais eficaz observar termos isolados e seus vizinhos do que se guiar pelas cores dos grupos.

É possível ver a presença expressiva de *data*, associada a termos como *traffic*, *decision*, *congestion* e *design*, o que permite identificar os usos de análises de dados. Estudo de comportamento pode ser visto na parte superior do grafo, representado pela palavra *choice*, associada aos termos *speed*, *highway* e *transit*.

Como destacado anteriormente, os artigos publicados apresentam mais termos significativos, trazendo conteúdos cada vez mais interdisciplinares, apresentando aspectos sociais, de meio ambiente, de pesquisa e de mercado.

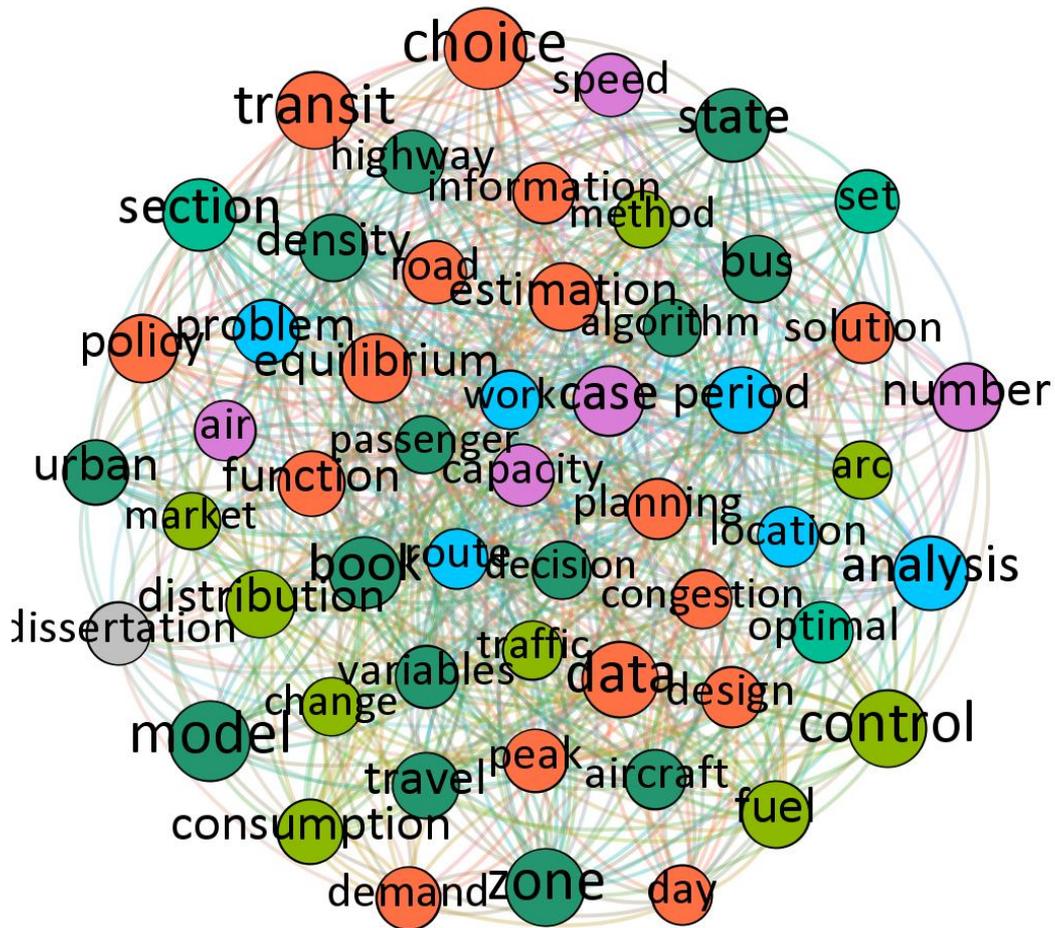


Figura 4 - Grafo relativo ao período de 2010 – 2014 (inclusive).

O software ainda disponibiliza outros estudos bibliométricos e informétricos, como, por exemplo, o número de autores por publicação. Textos com dois autores apresenta o maior percentual de ocorrência, sendo que textos com mais de cinco autores perfaz o menor percentual, como pode ser visualizado na Tabela 2.

Tabela 2 – Numero de autores por publicação.

Textos com 1 autor	2059 (25%)
Textos com 2 autores	2836 (34%)
Textos com 3 autores	1975 (24%)
Textos com 4 autores	889 (10%)
Textos com 5 autores ou mais	443 (5%)

Já o total de artigos agrupados por similaridade mostra que não há ocorrência de similaridade na faixa de 0,7 a 1, ou seja, os artigos mais similares a seus resumos não ultrapassam este patamar (por isso esse intervalo foi agrupado ao serem apresentados na Tabela 3. É interessante notar que mais de 50% dos artigos possuem média entre 0,4 e 0,5 de similaridade. O percentual indicado em *sem similaridade* são os artigos que não possuem resumo (artigos mais antigos publicados na revista).

Tabela 3 – Percentual de similaridade Artigo x Resumo.

## Similaridade Artigo x Resumo

0,7 a 1	0 (0.00%)
0,60 a 0,70	49 (0.60%)
0,50 a 0,60	1495 (18.23%)
0,40 a 0,50	4162 (50.74%)
0,30 a 0,40	1811 (22.08%)
0,20 a 0,30	118 (1.44%)
0,10 a 0,20	7 (0.09%)
0 a 0,10	21 (0.26%)
Sem similaridade	539 (6.57%)

Com objetivo de complementar a análise, apresenta-se a seguir dois gráficos, sendo relativos aos 10 termos mais relevantes do ano inicial da revista, 1967, e o segundo referente ao último ano de análise, 2014. O gráfico da Figura 5 indica que o termo de maior ocorrência no início da revista, *research*, decaiu ao longo dos anos, apresentando um pequeno aumento

apenas no ano de 1992. Verifica-se que outros termos também decresceram ao longo do tempo, como *drivers*, *risk* e *behaviour*.

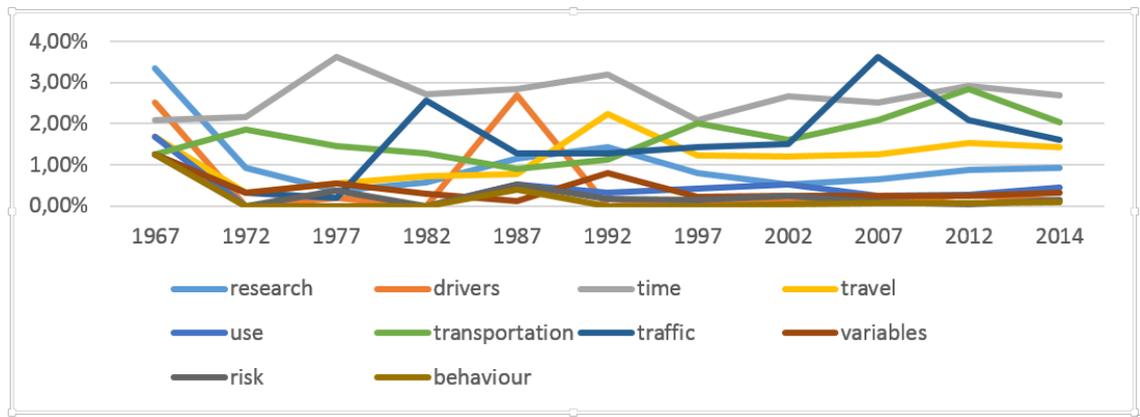


Figura 5 - Gráfico de projeção dos termos mais relevantes em 1967.

Considerando agora os termos mais relevantes no último ano analisado da revista *Transportation Research* (Figura 6), verifica-se que alguns termos se repetem entre os 10 mais significativos. São eles *research*, *time*, *travel*, *transportation* e *traffic*. Logo, 50% dos termos iniciais mais relevantes permanecem na lista dos 10 mais significativos, quando analisado o ano de 2014. Os termos *transportation* e *research* fazem parte da temática da revista, e nada mais correto do que obter essa classificação. Os outros três termos, *time*, *travel*, e *traffic* se inter-relacionam por abordarem viagem, tráfego e tempo, fator econômico de extrema relevância. “Tempo é dinheiro”.

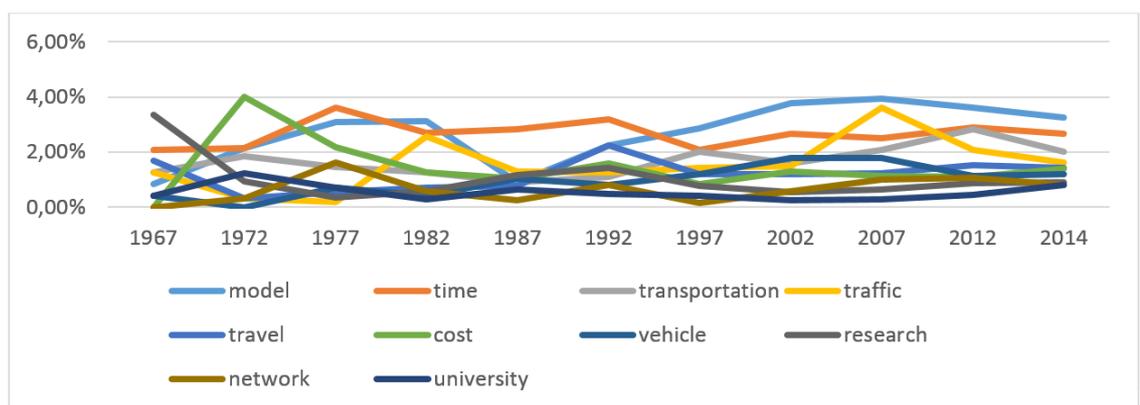


Figura 6 - Gráfico de projeção dos termos mais relevantes em 2014.

Os demais termos presentes entre os 10 mais relevantes de 2014 são *model*, *cost*, *vehicle*, *network* e *university*. Pode-se inferir que modelos criados nos institutos de pesquisa

(universidades) tem como principais focos o custo de veículos e as malhas de transporte público.

Através dos estudos realizados o software identificou 1030 termos significativos nos artigos, ou seja, termos que fazem parte da temática das publicações e que são essenciais para o seu entendimento. No primeiro ano da revista, encontram-se apenas 95 termos da listagem dentre os 1030 termos mais significativos. No ano de 2014 aparecem 878 dos 1030 termos nas publicações. Sem dúvida, houve um aumento na diversidade de temas abordados ao longo dos anos.

Na década de 60 do século passado, verificou-se que os pesquisadores centravam seus trabalhos no transporte rodoviário, onde os termos como risco, tráfego, viagem, comportamento e motoristas se apresentam dentre os temas mais relevantes das publicações. Cabe interpretar como foco principal das pesquisas realizadas há 50 anos os riscos do transporte rodoviário, englobando termos: motoristas, tráfego e risco.

Na outra extremidade, analisando os dados nos últimos anos de estudos publicados na revista *Transportation Research*, o 10º grafo apresenta dezenas de palavras em destaque, quantidade superior ao do início do periódico, com um número muito superior de interconexões entre si, e, além disso, sua proporção de ocorrência (tamanho do círculo), são muito mais proporcionais entre si, quase que idênticos. Isso demonstra que atualmente as pesquisas na área de Engenharia de Transportes abrangem uma diversidade de temáticas muito superior ao seu início, conforme acima mencionado através da análise da ocorrência de termos significativos.

De acordo com os dados apurados, deduz-se que a tendência da Engenharia de Transportes é continuar sua expansão, se interconectando com outras engenharias e também, com as áreas sociais e humanas. Acordos internacionais e a relativização da soberania dos países, alinhados com a globalização, tende a trazer soluções que avancem através das fronteiras nacionais.

O *software* MetrIA apresenta tendências futuras probabilísticas e não determinísticas. A análise resultante dos estudos com o intuito de identificar futuras tendências na área de transporte identificou três como principais grupos encontrados pelo algoritmo:

- Grupo 1 – trata de algoritmos, modelos, variáveis, controle, viagem, passageiros, ônibus, avião e trem. Este grupo criado pelo algoritmo indica uma tendência na criação de modelos para simular situações reais de deslocamento de pessoas.

- Grupo 2 – trata de método, mercado, distribuição, controle, consumo, combustível, ou seja, existe uma tendência de aumento na área de pesquisas de análise do mercado de combustíveis, através do uso de metodologias, com foco na distribuição, controle e consumo.

- Grupo 3 – apresenta termos como trânsito, política, equilíbrio, função, solução, planejamento, dados, design, demanda. O grupo apresenta o foco em pesquisas e soluções públicas para o trânsito, abrangendo design, planejamento, fundamentado em dados, buscando assim soluções equilibradas para atender às demandas.

Desta forma, tem-se três grandes áreas como tendência futura de pesquisa na área de Engenharia de Transportes:

1 – Criação de modelos, algoritmos e simuladores para realizar propostas na área de transporte de pessoas;

2 – Preocupação com combustíveis, seja com seu consumo, controle e mercado (público-alvo);

3 – Planejamento público – soluções para o trânsito abrangendo design e planejamento, buscando assim soluções que atendam equilibradamente à demanda existente.

Verifica-se assim que a área de planejamento está dentre as três principais grandes áreas de tendência da expansão das pesquisas em Engenharia de Transportes, com foco direcionado para a área pública. O investimento no planejamento em transportes é muito evidente nos países desenvolvidos e em cidades com IDH mais elevados. A globalização, os tratados internacionais e a aproximação dos países impulsionam a necessidade de soluções efetivas em transporte para atenderem o bem comum, a coletividade, possibilitando, desta forma, expansões no transporte de pessoas, cargas e bens, tanto interna quanto externamente.

#### **4 CONCLUSÕES FINAIS E RECOMENDAÇÕES**

A análise realizada permitiu identificar uma mudança de paradigma nas pesquisas na área de Engenharia de Transportes. Inicialmente os artigos apresentavam seu foco dirigido ao estudo do transporte rodoviário urbano, e com o passar dos anos expandiram-se as temáticas, abordando atualmente áreas como meio ambiente, saúde, habitação, mobilidade urbana, uso do solo e políticas públicas, dentre tantos outros temas, inclusive, a tecnologia da informação, utilizada neste trabalho. A diversidade na temática das publicações apresentou um aumento de quase 1.000 % de termos relevantes, comparando o primeiro ano e o último da revista.

Como resultado do objetivo proposto identificou-se três grandes áreas mapeadas como tendências futuras em Engenharia de Transportes: Criação de modelos, algoritmos e

simuladores para realizar propostas na área de transporte de pessoas; Preocupação com combustíveis, seja com seu consumo, controle e mercado (público-alvo); e Planejamento público – soluções para o trânsito abrangendo design e planejamento, buscando assim soluções que atendam equilibradamente à demanda existente, e visivelmente são coerentes como a realidade atual.

Conclui-se que é possível realizar um estudo, dentro do espectro temporal, e identificar novas tendências, a partir da análise da série histórica dos termos de maior ocorrência e de maior relevância. É possível observar que ao longo do tempo aumentou exponencialmente o número de temas distintos abordados pelos autores, e como consequência, houve redução no peso (número de ocorrências) das palavras mais relevantes, comparadas às de menor relevância, apresentando um maior equilíbrio entre os temas abordados.

Destaca-se que as três metodologias utilizadas, a cienciometria, a informetria e a bibliometria demonstraram serem adequadas para a análise de um grande volume de publicações científicas, conforme previsto, e que podem ser “traduzidas” em um *software* especialista (de STEFANO *et al.*, 2016), o que tornou possível a análise de um volume tão grande de informações e apresentação dos resultados obtidos.

## REFERÊNCIAS

- AIZAWA, Akiko. 2003.** An information-theoretic perspective of tf-idf measures. Janeiro de 2003, Vol. 39, 1, pp. 45–65.
- BOAVENTURA NETO, Paulo O. 2003.** *Grafos - Teoria, Modelos, Algoritmos. 2ª.* São Paulo : Edgard Blücher, 2003.
- BRINN, T., JONES, M.J. e PENDLEBURY, M. 2000.** Measuring research quality. *Omega.* 2000, pp. 237-239.
- de STEFANO, Ercilia. 2016.** *Análise da evolução da pesquisa em Engenharia de Transportes.* RJ : s.n., 2016. (tese de doutorado apresentada no Programa de Engenharia de Transportes - PET/COPPE/UFRJ).
- de STEFANO, Ercilia, PEIXOTO, Marcio S. e BALASSIANO, Ronaldo. 2016.** Development of a software for metric studies of transportation engineering journals. *Scientometrics.* 109, 2016, Vol. 3, pp. 1579-1591.

- FURLAN, B., BATANOVIC, V. e NIKOLIC, B. 2013.** Semantic similarity of short texts in languages with a deficient natural language processing support. *Decision Support Systems - Elsevier*. 3, 2013, Vol. 55.
- HEA, Wu, ZHAB, Shenghua e LIA, Ling. 2013.** Social media competitive analysis and text mining: A case study in the pizza industry. Junho de 2013, Vol. 33, 3, pp. 464–472.
- HONG, Tzung-Pei, et al. 2013.** Using TF-IDF to hide sensitive itemsets. Junho de 2013, Vol. 38, 4, pp. 502-510.
- LESKOVEC, Jure, RAJARAMAN, Anand e ULLMAN, Jeffrey D. 2014.** *Mining of Massive Datasets*. 5<sup>a</sup>. Stanford : s.n., 2014.
- LEYDESDORFF, Loet e MILOJEVIC, Staša. 2013.** Scientometrics. *School of Informatics and Computing, Indiana University*. 2013.
- NEWMAN, M. E. J. 2006.** Modularity and community structure in networks. *Proc Natl Acad Sci U S A*. 2006.
- Princeton University. 2015.** Wordnet. [Online] 2015. <https://wordnet.princeton.edu/>.
- RUSSEL, S. e NORVIG, P. 2004.** *Inteligência Artificial*. São Paulo : s.n., 2004.
- SAGE, A. P. 1990.** *Concise Encyclopedia of Information Processing in Systems and Organizations*. Nova Iorque : Pergamon, 1990.
- Transportation Research. 2015.** Transportation Research. *Elsevier - ScienceDirect*. [Online] 2015. [Citado em: 25 de Agosto de 2015.] <http://www.sciencedirect.com/science/journal/00411647>.
- YUE, Xiaoguang, et al. 2012.** Analysis of the Combination of Natural Language Processing and Search Engine Technology. *International Workshop on Information and Electronics Engineering (IWIEE)*. 2012.