



## ANÁLISE DA EVOLUÇÃO DA PESQUISA EM ENGENHARIA DE TRANSPORTES

Ercilia de Stefano

Tese de Doutorado apresentada ao Programa de Pós-graduação em Engenharia de Transportes, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia de Transportes.

Orientador: Ronaldo Balassiano

Rio de Janeiro  
Agosto de 2016

ANÁLISE DA EVOLUÇÃO DA PESQUISA EM ENGENHARIA DE TRANSPORTES

Ercilia de Stefano

TESE SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS EM ENGENHARIA DE TRANSPORTES.

Examinada por:

---

Prof. Ronaldo Balassiano, Ph.D.

---

Prof. Marcio Peixoto de Sequeira Santos, Ph.D.

---

Prof<sup>a</sup>. Marilita Gnecco de Camargo Braga, Ph.D.

---

Prof. Elton Fernandes, Ph.D.

---

Prof<sup>a</sup>. Nissia Carvalho Rosa Bergiante, D.Sc.

---

Prof. Ricardo Rodrigues Pacheco, D.Sc.

RIO DE JANEIRO, RJ - BRASIL

AGOSTO DE 2016

Stefano, Ercilia de

Análise da evolução da pesquisa em engenharia de transportes/ Ercilia de Stefano. – Rio de Janeiro: UFRJ/COPPE, 2016.

VIII, 142 p.: il.; 29,7 cm.

Orientador: Ronaldo Balassiano

Tese – UFRJ/ COPPE/ Programa de Engenharia de Transportes, 2016.

Referências Bibliográficas: p. 131-142.

1. Engenharia de transportes. 2. Métricas. 3. Data mining. 4. Processamento de linguagem natural. I. Balassiano, Ronaldo. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Transportes. III. Título.

## **Agradecimentos**

Com muita admiração vai meu primeiro agradecimento, aos mestres, e meus grandes amigos Ronaldo e Marcio, que sempre me apoiaram em todos os sentidos.

Agradeço em especial aos professores que me honram terem aceitado nosso convite para integrar a banca da defesa desta tese.

Agradeço imensamente ao Prof. Watanabe que muito me apoiou e acreditou em mim.

Agradeço à Jane e a Helena pela atenção, profissionalismo e carinho com que sempre desempenharam suas funções.

Agradecimento anônimo a todas e todos mestres que contribuíram com minha formação.

A amiga e dinda da minha filha Zuleide, que está sempre presente em nossa vida, e por isso participou também desta tese.

Apoio de família é sempre fundamental, por isso quero agradecer à minha mãe Elia, sempre presente e faz parte da realização deste sonho.

Com muito amor agradeço ao meu marido Leonardo por toda ajuda.

E por último agradeço à minha princesinha Laura que é a razão e a inspiração da vida da mamãe.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.).

## ANÁLISE DA EVOLUÇÃO DA PESQUISA EM ENGENHARIA DE TRANSPORTES

Ercilia de Stefano

Agosto/2016

Orientador: Ronaldo Balassiano

Programa: Engenharia de Transportes

Este estudo realiza uma análise da evolução da pesquisa em Engenharia de Transportes. A metodologia utilizada analisou toda a série de artigos publicados pelo periódico *Transportation Research*, da editora *Elsevier*, entre 1967 e 2014. São abordados aspectos da medição de produção científica, da identificação do conhecimento em publicações e o acompanhamento de tendências e evolução na pesquisa. Para permitir a análise dos mais de 12 mil itens da publicação foi desenvolvido um software – “MetrlA” - para atender às especificações identificadas, que é também capaz de analisar, de forma semelhante, outras publicações científicas. O software permitiu a realização de análises que mostram a evolução dos temas tratados dentro da *Transportation Research*, o que demonstra sua capacidade para analisar este e outros periódicos.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

ANALYSIS OF THE RESEARCH EVOLUTION IN TRANSPORTATION  
ENGINEERING

Ercilia de Stefano

August/2016

Advisor: Ronaldo Balassiano

Department: Transportation Engineering

This study performs an analysis of the evolution of research in Transportation Engineering. The employed methodology analyzed all the articles published by the journal Transportation Research, from Elsevier, between 1967 and 2014. The study approaches aspects related to the measurement of academic production, knowledge identification in publications and the monitoring of trends and evolution in research. To make the analysis of 12 thousand items possible the researcher developed a software – named “MetRIA” – that fulfills the requirements for this study, which is also capable of analyzing similar journals. The software allowed analysis that show the evolution of different themes in the journal Transportation Research, which demonstrates its capability of analyzing this and other publications.

## Sumário

<b>CAPÍTULO 1 - INTRODUÇÃO .....</b>	<b>1</b>
1.1. Contexto da Pesquisa .....	2
1.2. Objetivos Gerais do Trabalho .....	5
1.3. Objetivos Específicos do Trabalho .....	5
1.4. Hipóteses .....	6
1.5. Importância, Ineditismo e Justificativa da Pesquisa .....	7
1.6. Resultados Esperados/Contribuição da Tese .....	8
1.7. Procedimentos Metodológicos.....	8
1.8. A Organização do Estudo.....	9
<b>CAPITULO 2 – REVISÃO BIBLIOGRÁFICA.....</b>	<b>11</b>
2.1. Visão geral da Engenharia de Transportes .....	11
2.1.1. Origem e evolução da Engenharia de Transportes .....	13
2.1.2. A Engenharia de Transportes e o Brasil.....	16
2.1.3. A importância do Planejamento de Transporte .....	22
2.2. Perfil do leitor-pesquisador .....	27
2.3. Fundamentos da Ciência da Computação a serem empregados no desenvolvimento do <i>software</i> e banco de dados .....	30
2.3.1. Inteligência Artificial .....	31
2.3.2. Processos de Aprendizagem – redes neurais.....	32
2.3.3. Processamento de Linguagem Natural (PLN).....	33
2.3.4. Data mining .....	34
2.3.5. Big data .....	36
2.3.6. Text mining .....	37
2.3.7. Web scraping.....	39
2.3.8. Tokenization .....	39
2.3.9. Separação de <i>part-of-speech</i> e Semelhança Semântica .....	41
2.3.10. Computação Paralela .....	42
2.3.11. TF-IDF .....	43
2.3.12. Similaridade de Jaccard.....	45
2.3.13. Node.js .....	45
2.3.14. WordNet .....	46
2.3.15. Grafos.....	47
<b>CAPÍTULO 3 – METODOLOGIA E PROCEDIMENTOS .....</b>	<b>49</b>
3.1. Metodologia .....	49

3.2. Cienciometria, Informetria e Bibliometria.....	51
3.2.1. Cienciometria.....	52
3.2.2. Bibliometria.....	56
3.2.3. Informetria.....	59
3.3. A escolha do periódico <i>Transportation Research</i> .....	62
3.3.1. Métricas disponibilizadas no site da <i>Transportation Research</i> .....	67
<b>CAPÍTULO 4 – O SOFTWARE MetrIA.....</b>	<b>70</b>
4.1. Etapas de criação do <i>software</i> e do banco de dados.....	71
4.2. Tecnologias empregadas.....	77
<b>CAPÍTULO 5 – ANÁLISE DAS PESQUISAS EM TRANSPORTES.....</b>	<b>79</b>
5.1. Análise por Publicações.....	79
5.2. Análise por Autores.....	84
5.2.1. Seleção de um único autor.....	87
5.3. Análise por Termos.....	90
5.3.1. Termo <i>bicycle</i> .....	90
5.3.2. Termo <i>environment</i> .....	91
5.3.3. Termo <i>emission</i> .....	92
5.3.4. Pesquisa por termos mais frequentes.....	94
5.3.5. Pesquisa por termos de maior relevância.....	95
5.4. Grafos – termos de maior relevância.....	97
5.5.1. Primeiro grafo - intervalo – 1967-1969.....	99
5.5.2. Segundo grafo - intervalo – 1970 a 1974.....	102
5.5.3. Terceiro grafo - intervalo – 1975 a 1979.....	103
5.5.4. Quarto grafo - intervalo – 1980 a 1984.....	104
5.5.5. Quinto grafo - intervalo – 1985 a 1989.....	106
5.5.6. Sexto grafo - intervalo – 1990 a 1994.....	107
5.5.7. Sétimo grafo - intervalo – 1995 a 1999.....	108
5.5.8. Oitavo grafo - intervalo – 2000 a 2004.....	110
5.5.9. Nono grafo - intervalo – 2005 a 2009.....	111
5.5.10. Décimo grafo - intervalo – 2010 a 2014.....	113
5.5. Totais analisados.....	114
5.6. Estudo complementar.....	117
5.7. Conclusões a partir do Estudo de Caso.....	119
5.8. Análise realizada: Humano x Máquina.....	123
<b>CAPÍTULO 6 – CONCLUSÕES FINAIS E RECOMENDAÇÕES.....</b>	<b>127</b>
<b>Referências Bibliográficas.....</b>	<b>131</b>



## **CAPÍTULO 1 - INTRODUÇÃO**

Nos primórdios da humanidade, a cultura, os costumes e os conhecimentos eram repassados aos descendentes através da oralidade. Os primeiros registros físicos foram figuras, hieróglifos até se chegar à escrita utilizada atualmente. Os registros possibilitam a realização de pesquisas temporais pela história da humanidade, onde verifica-se as descobertas, os desafios, as tentativas, acertos e erros na evolução científica ao longo dos anos. Sem dúvida, o fato mais marcante para a universalização do conhecimento foi a utilização da prensa por Gutenberg, por volta dos anos 1439, prensa essa criada pelo chinês Bi Sheng no ano de 1040 (REED, 2011). Esse marco histórico foi um grande facilitador da difusão, replicação e perpetuação do conhecimento.

Na atualidade, ainda são os livros e periódicos os meios mais eficientes para a aquisição e atualização do conhecimento. Os periódicos científicos possuem papel importantíssimo, propiciando visibilidade à produção intelectual, e atuam também como um crivo para a seleção do que é efetivamente publicado. Os artigos são fontes para métricas de aceitação, impacto e difusão do conhecimento.

Um estudo realizado entre os anos de 1980 a 2012 contabilizou a publicação de 38 milhões de artigos neste período, e estima-se que a cada nove anos duplique o número de publicações (VAN NOORDEN, 2014).

Os números retratam o grande volume de conhecimento que é produzido e publicado, e também trazem consigo outras informações importantes. O aumento significativo do número de publicações decorre do aumento do número de cientistas, da expansão de áreas de pesquisas, da valorização de se publicar ciência, no aumento do número de periódicos para se publicar, dentre muitos outros fatores.

A grande maioria destas publicações está disponível para a aquisição imediata e leitura on-line. Toda esta facilidade traz um novo desafio: como acessar, assimilar e validar o conteúdo de tantas fontes distintas?

Neste cenário, delimitando o estudo na área de Engenharia de Transportes, o principal objetivo desta tese é realizar uma análise temporal de publicações nesta área, a fim de identificar sua evolução e tendências futuras.

### **1.1. Contexto da Pesquisa**

A presente tese pretende dar novos passos além da pesquisa iniciada por Santos e Braga (1986), que publicaram o artigo intitulado “*Research Trends in Urban Transport Science: some empirical evidence from Academic Research*”. O artigo aborda o impacto que o desenvolvimento de novas tecnologias traz para o ambiente da pesquisa acadêmica relacionado à área de transportes. Como resultado, concluiu-se que as primeiras formulações formais no processo de planejamento de transporte urbano foram delimitadas nos EUA em meados do século XX.

Santos e Braga (1986) afirmam que os problemas que envolvem a engenharia de transporte são multidimensionais, não podendo ser resolvidos apenas por abordagens sintéticas. Por isto foi possível identificar o desenvolvimento das abordagens que levam em conta fatores como comportamento e psicologia. Esta pesquisa histórica identificou uma mudança de paradigma, onde a metodologia adotada abandonou a visão no qual se preocupava em prover capacidade suficiente para o fluxo de veículos, para um formato preocupado com a função de uso do solo.

A própria definição de “problema a ser solucionado” mudou de acordo com o passar dos anos. O estudo identificou uma redução na produção científica da área de engenharia de transportes nos períodos anteriores à mudança de paradigma. Nos anos de 1972/73 houve uma mudança para um formato centrado em planejamento e nos anos de 1979/80 houve uma nova mudança, para abordagens comportamentais. A mudança no ritmo das pesquisas, segundo o artigo, pode ser atribuída a resultados não satisfatórios de metodologias que se demonstram inadequadas, o que leva à necessidade do desenvolvimento de novas metodologias para se compreender os problemas propostos. A preocupação com questões ambientais foi identificada na pesquisa, notadamente a partir dos anos 70, com o aumento no número de trabalhos sobre poluição sonora e meio ambiente.

Outra pesquisa realizada que serviu de motivação para o presente trabalho ocorreu no período de 2010/2012, com a colaboração de alunos de mestrado e doutorado do Programa de Engenharia de Transportes (PET) da COPPE – UFRJ, onde um novo estudo foi realizado. Foram analisados os resumos de 87 teses de doutorado e 457 dissertações de mestrado, totalizando 544 trabalhos, sendo estes trabalhos referentes aos primeiros trinta anos de existência do programa de Engenharia de

Transportes da COPPE. Os dados bibliométricos e informétricos foram coletados e sintetizados em planilhas, trazendo importantes informações cienciométricas. Assim como o trabalho de Santos e Braga (1986), o estudo realizado também identificou novas áreas de pesquisa, como meio ambiente e turismo.

A proposta desta tese é ampliar significativamente a análise de publicações, e para isso foi selecionado como fonte de dados o periódico *Transportation Research* publicado pela *Elsevier*. A escolha da *Transportation Research* fundamenta-se no reconhecimento internacional da revista, considerada uma das mais relevantes na área, quer seja por seu conteúdo, autores ou pela diversidade de temas pesquisados. Os fatores de impacto das 6 revistas que compõe a *Transportation Research* variam entre 1.473 (parte F) à 2.952 (parte B), sendo todas classificadas como Qualis A1 pela CAPES. O estudo abrange mais de **oito mil artigos** do periódico ao longo de quase 50 anos de existência da revista.

Para realizar com êxito o estudo de milhares de artigos, foi desenvolvido um *software* especialista para automatizar as atividades de leitura, pesquisa, seleção e interpretação das publicações, sem a interferência humana. Isso é possível com o emprego das técnicas de inteligência artificial, *data mining* e processamento de linguagem natural (PLN).

## **1.2. Objetivos Gerais do Trabalho**

O objetivo principal deste trabalho é realizar uma análise temporal da evolução das pesquisas em transporte nas últimas décadas, a partir de um aprimorado estudo cienciométrico, informétrico e bibliométrico de milhares de artigos, visando identificar tendências na área de Engenharia de Transporte.

Para alcançar esse objetivo um *software* foi desenvolvido para realizar de forma autônoma este estudo. Pretende-se obter a partir da análise temporal uma compreensão histórica das mudanças e avanços nas pesquisas na área da engenharia de transportes como um todo, e como consequência identificar tendências. A seguir serão apresentados os Objetivos Específicos que norteiam a presente tese.

## **1.3. Objetivos Específicos do Trabalho**

O primeiro objetivo específico deste trabalho é identificar os critérios para a realização dos estudos cienciométricos, informétricos e bibliométricos dos dados. Para isso, far-se-á necessário determinar quais os algoritmos mais apropriados para atingir o objetivo esperado.

O segundo objetivo específico é a criação de um software especialista para realizar de forma automatizadas os estudos determinados.

O terceiro objetivo específico é realizar classificações nos dados através do uso do software desenvolvido, e identificar os termos mais frequentes e os termos de maior relevância contidos nas publicações, para que se permita verificar a evolução das pesquisas em transportes. A distinção entre termos mais frequentes e mais relevantes será descrita no estudo de caso apresentado no capítulo específico.

#### **1.4. Hipóteses**

Algumas hipóteses norteadoras foram elaboradas fundamentadas nos objetivos acima descritos:

1 - A cienciometria, a informetria e a bibliometria são os métodos mais adequados para realizar a análise de um grande volume de publicações?

2 – É imprescindível a utilização de *software* especialista para realizar a análise de quase dez mil artigos, e apresentar o resultado esperado?

3 – É possível realizar um estudo destes termos dentro do espectro temporal, identificar novas tendências, a partir da análise da série histórica de ocorrência das palavras de maior ocorrência e mais relevantes?

### **1.5. Importância, Ineditismo e Justificativa da Pesquisa**

O progresso científico e a trajetória das pesquisas na área da engenharia de transportes podem ser descritos por uma análise detalhada de publicações ao longo do tempo. A realização desta análise crítica dar-se-á através dos fundamentos da cienciometria, informetria e bibliometria, que busca apresentar estudos objetivos e no caso da presente tese onde a análise será realizada por um *software* especialista a ser desenvolvido, também imparcial pela não interferência subjetiva da análise humana.

Os dados obtidos neste estudo visam permitir traçar uma trajetória dos artigos na área de transportes, fornecendo aos pesquisadores e formuladores de políticas públicas subsídios para a definição de novos estudos e projetos. O *software* a ser desenvolvido para auxiliar a realização destes estudos, traz o ineditismo da automatização de atividades realizadas integralmente ou parcialmente pelo ser humano, e que demandavam anos de leitura, anotações e análises dos dados obtidos, sob o risco de ocorrer erros e inconsistências.

O *software* analisa milhares de publicações e apresenta resultados quase que instantaneamente. Em um notebook core i5 com 4 Gb de memória RAM o resultado das consultas e pesquisas variam de milésimos de segundo até 10 segundos, sendo a consulta mais demorada de termos relevantes. Vale destacar que este *software* pode ser aplicado não somente à área de planejamento em transportes, bem como, a qualquer outra área do conhecimento.

## **1.6. Resultados Esperados/Contribuição da Tese**

A principal contribuição da tese 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. A análise temporal de milhares de publicações pretende apresentar como resultado as novas tendências da área de transportes.

Para se alcançar os resultados esperados foi selecionado o periódico *Transportation Research* como fonte dos dados, por ser uma das revistas mais tradicionais na área de transportes, que apresenta uma grande diversidade e abrangência de temas. O periódico subdivide-se em seis grandes áreas temáticas da Engenharia de Transporte, como melhor descrito no item 3.3 da presente tese.

## **1.7. Procedimentos Metodológicos**

Os procedimentos metodológicos aplicados nesta tese serão brevemente apresentados a seguir, e melhor detalhados no capítulo 3.

1. Primeiramente foi realizado um estudo bibliográfico. Inicia-se por uma revisão sobre a Engenharia de Transportes, na sequência uma pesquisa sobre o perfil do pesquisador-leitor, e encerra-se com um estudo das tecnologias da engenharia da



computação a serem aplicadas no desenvolvimento do *software* de apoio à análise de dados.

2. A partir do método científico hipotético-dedutivo, que parte de um conhecimento prévio da literatura especializada, a próxima etapa aplica as metodologias, cienciometria, informetria e bibliometria, através do *software* desenvolvido, para a obtenção de dados e informações para análise.

3. Uma Pesquisa Exploratória, através da análise das informações apresentadas pelo *software*, dentro das delimitações estabelecidas para alcançar as conclusões previstas.

## **1.8. A Organização do Estudo**

Esta tese está constituída de seis capítulos, conforme a estrutura descrita a seguir:

1 - Introdução – apresenta a motivação para a realização do presente trabalho, os objetivos gerais e específicos, hipóteses, contribuições ao conhecimento e a estrutura da tese a ser apresentada.

2 – Revisão Bibliográfica – apresenta as revisões bibliográficas que norteiam este trabalho, subdividida em três partes distintas: inicialmente apresenta-se uma visão geral na área de transportes; na sequência retrata-se uma pesquisa que inspirou a

criação de um *software* para esta tese; e a terceira grande área apresenta fundamentos da engenharia e ciência da computação para o desenvolvimento do referido *software*.

3 – Metodologia e Procedimentos – descreve a metodologia a ser utilizada na tese, e em um segundo momento apresenta os fundamentos e procedimentos das três metodologias a serem utilizadas para realizar os estudos da presente tese: a cienciometria, a informetria e a bibliometria, e encerra-se com a justificativa da escolha da revista *Transportation Research* para o estudo a ser realizado na presente tese.

4 - *Software Metria* – apresenta uma descrição detalhada das escolhas e soluções, e todo o procedimento adotado para a criação do banco de dados – *data mining*, bem como uma descrição da elaboração do *software*.

5 – Análise das Pesquisas em Transportes – discute e analisa o *data mining* criado tendo como fonte de dados a revista *Transportation Research*. Apresenta-se diversas telas, pesquisas, gráficos e grafos, com o objetivo de realizar um profundo estudo cienciométrico, informétrico e bibliométrico dos dados.

6 – Conclusões e Recomendações – apresenta-se aqui as considerações finais, recomendações e estudos futuros.

Inicia-se a seguir o capítulo 2, que traz os fundamentos teóricos que serão o ponto de partida, fonte de informações e inspiração para a construção do trabalho proposto.

## **CAPITULO 2 – REVISÃO BIBLIOGRÁFICA**

Este capítulo traz uma revisão bibliográfica das temáticas a serem abordadas nesta tese. A análise da evolução da pesquisa em Engenharia de Transportes é tema principal deste trabalho, sendo assim, será apresentada uma visão geral da engenharia de transportes; na sequência, apresenta-se detalhadamente um importante estudo que inspirou a criação de um *software* para auxiliar os estudos desta tese, estudo esse que descreve o perfil do pesquisador-leitor; e na terceira parte apresenta-se o referencial teórico das tecnologias que serão utilizadas no desenvolvimento do *software* especialista.

### **2.1. Visão geral da Engenharia de Transportes**

A Engenharia de Transportes surgiu da necessidade do deslocamento planejado de pessoas e mercadorias, em ambientes de constantes mudanças, e que exigem cada vez mais rapidez, eficiência e eficácia (KUMARES, et al., 2002). Aplicam-se na Engenharia de Transportes os princípios científicos e tecnológicos para planejar, projetar, construir, operar e manter toda e qualquer modalidade de transporte e sistemas multimodais, (CHAKROBORTY e DAS, 2012) com o objetivo de alcançar um transporte seguro, eficiente, rápido, confortável, prático, econômico e compatível com as necessidades locais, com o meio ambiente e com a sociedade.

Mais da metade da população mundial atualmente vive em áreas urbanas, e enfrenta desafios de transporte e mobilidade sem precedentes. Com rápido crescimento populacional e econômico, as exigências de mobilidade urbana estão em constante aumento (CERVERO, 2013). Ainda para o autor, a área de transportes tem papel fundamental na definição do futuro econômico das cidades, por interligar os trabalhadores aos postos de trabalho, matérias-primas à agricultura e produção de mercadorias; e também o deslocamento humano para atividades de lazer.

Importante destacar que a qualidade do sistema de transportes é um dos índices mais importantes que medem o desenvolvimento econômico de um país, de acordo com o *World Economic Forum* (2015). Para Novaes (2007) o transporte “agrega valor de lugar, de tempo, de qualidade e de informação à cadeia produtiva”.

As atividades que o Engenheiro de Transportes desenvolve abrangem uma diversa gama de áreas, indo além do planejar e projetar estruturas. Contempla áreas como planejamento de operações, logística, análise de redes, financiamento e políticas públicas, meio ambiente, estudos de impactos socioambientais, cada vez mais necessários nos grandes centros urbanos (HOEL, et al., 2011). É fato que o planejamento em engenharia de transportes é essencial no mundo atual, e bem empregado, propicia realizar, com efetividade, o deslocamento de pessoas, bens e produtos. Também é fato e notório a importância desta engenharia e sua relação direta com o desenvolvimento econômico e social das diferentes nações.

O homem e a sua capacidade de produzir não se limita mais ao meio onde vive, como no passado. Antigamente o transporte era considerado apenas uma atividade de apoio ao negócio, hoje o transporte é essencial à existência humana. BALLOU (2006)

destaca que o consumo e a produção puderam experimentar uma separação geográfica a partir do aperfeiçoamento dos sistemas logísticos. “A produção excedente poderia então ser enviada, com vantagem econômica, a outras áreas produtoras (ou consumidoras), e os artigos necessários de escassa ou inexistente produção local seriam importados/exportados”.

A grande importância econômica, social e política faz da Engenharia de Transportes uma das áreas de pesquisa mais estratégicas. A revisão bibliográfica referente à Engenharia de Transportes apresenta a seguir um histórico de sua evolução, um retrato atual no Brasil, e ressalta a importância do planejamento dentro da Engenharia de Transportes.

### **2.1.1. Origem e evolução da Engenharia de Transportes**

Apresenta-se a seguir um breve histórico da participação da Engenharia de Transportes no progresso social e tecnológico. Faz-se destacar os principais marcos que revolucionaram esta área da Engenharia, e sua interferência direta nas mudanças no rumo na história da humanidade.

Na época das cavernas, os seres humanos deslocavam-se andando ou nadando. Com os adventos da domesticação de animais, o invento de pequenas canoas, e na sequência a invenção da roda, surgiu a oportunidade de transportar maiores volumes, cargas mais pesadas e à maiores distâncias. O início da utilização de

artefatos no transporte trouxe maior velocidade de deslocamento e possibilitou também o aumento na duração do deslocamento (RODRIGUE, et al., 2013).

O transporte na água foi, e ainda é, uma das formas mais eficientes de deslocamento de grandes quantidades para grandes distâncias. Em decorrência desta importância, a maioria das cidades na antiguidade nasceram às margens de rios ou a beira-mar, e sempre que possível na intersecção de ambos.

Em terra, o transporte rodoviário era tracionado por cavalos, bois ou mesmo por seres humanos, e transportavam mercadorias por estradas rudimentares, raramente pavimentadas. Os persas e os romanos foram os primeiros a construir ruas pavimentadas com pedras, com o objetivo de aumentar a velocidade de deslocamento de seus exércitos, suprimentos e comércio. Já na época medieval surgiram as primeiras estradas pavimentadas com alcatrão no oriente médio.

Até o advento da Revolução Industrial, o transporte permaneceu lento e caro, e por isso a produção e o consumo eram localizados e tão próximos uns dos outros quanto possível. O início do transporte ferroviário data do Século XVI com o desenvolvimento do motor a vapor, este que foi considerado um dos maiores inventos naquela época, juntamente com a máquina de tear movida a vapor (BOGART, 2012).

Por sua estreita ligação com a Revolução Industrial, o transporte ferroviário se desenvolveu com maior intensidade na Inglaterra (BOGART, 2013). As locomotivas eram movidas a vapor, gerado a partir da queima de carvão mineral. Após o surgimento deste inovador transporte, rapidamente sua tecnologia se alastrou para outros pontos da Europa e do mundo. O motor a vapor também foi utilizado em embarcações com grande êxito.

Vale destacar que foi no século XIX que a Engenharia de Transportes iniciou sua cisão com a Engenharia Civil, ganhando autonomia. Com o desenvolvimento do motor à combustão e do automóvel na virada para o século XX, o transporte rodoviário tornou-se mais viável, surgindo assim as primeiras transportadoras privadas. As primeiras rodovias foram construídas no século XIX (HOEL, et al., 2011).

No início do século XX ocorreu um dos maiores inventos na área de transportes: o avião. Após a Primeira Grande Guerra o avião tornou-se comercial e passou a ser o mais rápido meio de transporte de pessoas e bens a longas distâncias.

Após a Segunda Guerra Mundial, os veículos automotivos e os aviões ocuparam grande parte do transporte, reduzindo o fluxo de transporte de mercadorias e de passageiros nas modalidades ferroviário e aquaviário para curtas distâncias.

A década de cinquenta trouxe outra grande inovação no transporte de cargas, a introdução de *containers* de tamanho padronizado para o transporte de grandes volumes, o que efetivamente trouxe enormes ganhos no transporte de mercadorias. Foi um dos precursores da globalização do comércio (RODRIGUE, et al., 2013).

As viagens aéreas internacionais tornaram-se muito mais acessíveis nos anos 60, graças ao motor a jato. Na mesma época surgiram os primeiros trens de alta velocidade na Ásia e na Europa, concorrendo assim com os aviões em rotas de longa distância.

Atualmente temos motores mais potentes e velozes, que aumentaram em muito a velocidade de deslocamento. Novas tecnologias, como por exemplo, o trem de levitação magnética tem apresentando grande êxito (GARBER e HOEL, 2002), e

recentemente o Japão alcançou a velocidade de 602km/h com esta tecnologia. Já na área automobilística surgem os carros movido a hidrogênio, água e elétricos.

Na breve análise acima descrita, verifica-se que contar a história da engenharia de transportes é contar a história da própria humanidade. A presença inquestionável da engenharia de transportes está no deslocamento diário, no abastecimento do comércio como um todo, no lazer, na indústria, enfim, é imprescindível a vida moderna.

### **2.1.2. A Engenharia de Transportes e o Brasil**

O setor de transportes é parte integrante no desenvolvimento da sociedade moderna. O bem-estar de uma localidade está diretamente associado à qualidade de seu transporte, indispensável este para a mobilidade de pessoas e bens.

No Brasil, segundo o IBGE (2013), tem-se 2,8 milhões de decolagens e pousos de aeronaves anualmente; quase 180 milhões de embarques e desembarques de passageiros; 1,5 milhão de toneladas de carga e descarga aérea movimentada; frota nacional de 76,14 milhões de veículos; 1.560 mil km de malha rodoviária, 30 mil km de malha ferroviária e 29 mil km de malha hidroviária. Apesar de números tão expressivos, o próprio IBGE (2013) considerada inadequada a matriz de transportes no Brasil - de cargas e passageiros, quando comparada com outros países de dimensões semelhantes.



O déficit no transporte reflete negativamente no PIB (Produto Interno Bruto) e na balança comercial. Com a competitividade intensificada pelos efeitos da globalização, o Brasil vem ano-a-ano perdendo espaço no comércio internacional. Uma rápida análise no PIB fundamenta essa afirmação. No período de 2005-2010 o PIB era de 4,3% ao ano, passando no 2011-2013 para 2%. No mesmo período os investimentos caíram de 9,2% para 2,3% ao ano. Segundo o *World Economic Forum* (2015) “(...) o Brasil está em 114.º lugar entre 148 países no quesito infraestrutura.

A figura 1 publicada pelo *World Economic Forum* (2015) refere-se ao biênio de 2014-2015. Ela vem corroborar com as afirmativas acima, onde visualiza-se como terceiro maior problema brasileiro “*inadequate supply of infrastructure*”, considerando os negócios internacionais.

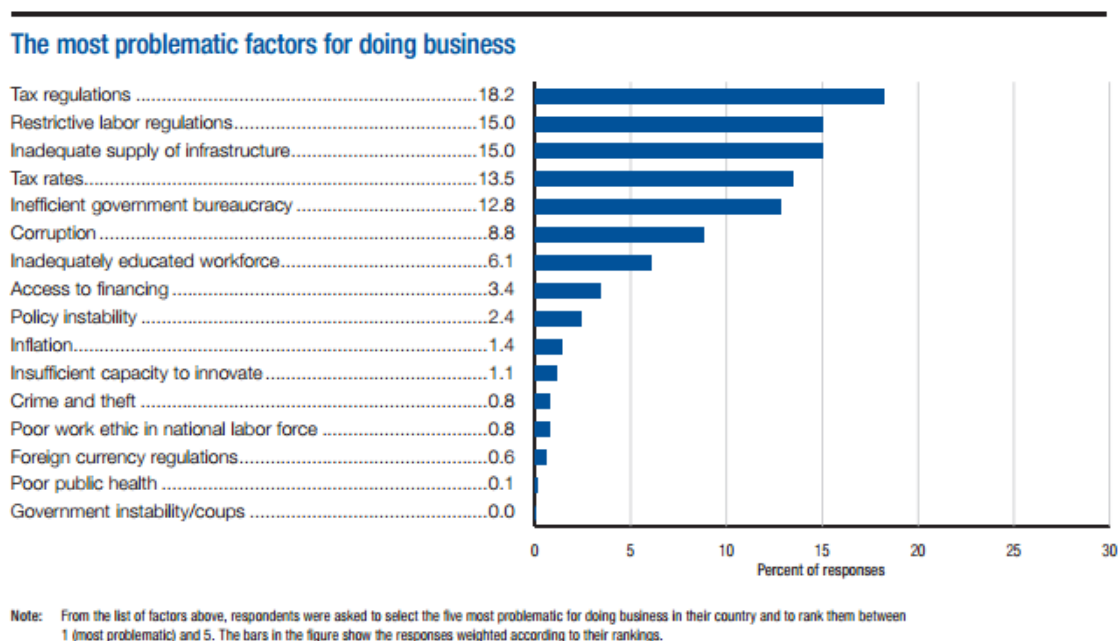


Figura 1 - Gráfico dos problemas brasileiros.

Vale exemplificar que atualmente o custo do frete interno para a exportação de soja representa aproximadamente 25% do valor do produto e no caso do milho cerca 50%, sendo que este está próximo de se tornar inviável sua comercialização (RESENDE, 2013). O prejuízo anual do Brasil com custos logísticos é em torno de 5,6% do Produto Interno Bruto (PIB), o equivalente a US\$ 83,2 bilhões, que abrangem a burocracia desmedida; limitada infraestrutura de estradas, ferrovias, portos e aeroportos. Resende (2013), coordenador do Núcleo CCR de Infraestrutura e Logística da Fundação Dom Cabral afirma em reportagem publicada: “O Brasil não investe o suficiente para recuperar nem o que perde”.

Para Corrêa (2014), um grave entrave nas nossas exportações é o chamado “custo Brasil”, que são desvantagens competitivas que o país tem nos campos estrutural, burocrático e econômico, como por exemplo “a precariedade da infraestrutura de transportes, a insegurança jurídica, a burocracia excessiva para se criar e manter uma empresa, os altos custos trabalhistas, a alta carga tributária, entre outros”.

Chagas et. al. (2015) afirma que o Brasil é conhecido no cenário do comércio exterior por seus “famosos gargalos logísticos”, tendo como principais entraves “a péssima matriz de transporte brasileira na qual prevalece o transporte rodoviário; os muitos problemas de infraestrutura logística; o excesso de burocracia; a alta carga tributária, entre outros gargalos que “estrangulam” quaisquer tipos de planejamento logístico”.

Para Cano (2012), o investimento em transporte vem em decorrência do desenvolvimento do país:

“O resultado de um longo processo de crescimento econômico, com elevado aumento da produtividade média, sem o qual o excedente não cresce o bastante para acelerar a taxa de investimento e diversificar a estrutura produtiva e do emprego. Esse processo intensifica a industrialização e urbanização para transformar de maneira progressista as estruturas sociais e políticas do país”.

Ainda para o autor, o crescimento econômico provoca uma transformação que ocorre e se intensifica a urbanização, induzindo também o crescimento do comércio, transportes, finanças, saúde, educação e outros. Já Macbry (2007) destaca que os investimentos realizados em infraestrutura de transportes, de acordo com dados do Banco Mundial, geram 15 mil empregos para cada US\$ 1 bilhão investidos.

Mesmo sendo o modal marítimo o meio de transporte mais utilizado internacionalmente para o transporte de mercadorias, representando 95% da carga movimentada no comércio exterior, não tem como esquecer que no Brasil o modal rodoviário é o mais significativo meio de transporte interno nacional, representando cerca de 60% de toda carga interna movimentada, atingindo, no Estado de São Paulo, a cifra de 90% do volume transportado. Ele é determinante para que as cargas cheguem aos portos.

Com relação ao transporte público, Balassiano (2007) ressalta as principais estratégias e políticas a serem priorizadas, principalmente, por países emergentes. Balassiano (2007) apresentou um estudo resultante da análise dos países integrantes do BRICs, comparados a países denominados desenvolvidos, onde classificou como sete as principais estratégias e políticas, apresentadas a seguir:

a) O planejamento integrado – o planejamento dos sistemas de transportes não pode ser tratado de forma independente, é importante valer a “correlação existente entre transportes e uso do espaço urbano”. Gestores públicos que não atuam com

políticas públicas que perpassam o âmbito de secretarias, “engessam” a máquina pública e não obtém o resultado efetivo para as soluções de transporte.

b) Integração modal – “mostra que o transporte público pode operar de forma coordenada e eficiente”, e apresenta como resultado o aumento da mobilidade e acessibilidade aos usuários. Cidades que integram eficientemente seus modais, por facilitarem o acesso aos meios de transporte coletivo aumentam o uso dos meios de transporte coletivos, reduzindo principalmente o número de veículos particulares nas vias. Cervero et al. (2004) apresenta o conceito do Desenvolvimento Orientado pelo Transporte (DOT), que ganhou popularidade como uma estratégia de crescimento inteligente. Uma pesquisa nos EUA identificou recentemente mais de 100 DOTs (Cervero et al., 2004). DOT é sem dúvida a forma mais concreta de crescimento inteligente onde o desenvolvimento resulta do uso misto do solo e ocorre em torno de estações de embarque e desembarque, apresentando como resultado positivo o aumento do número de passageiros. DOT reduz o congestionamento do tráfego, melhora a qualidade do ar pela redução das emissões de gases nocivos, e aumenta a segurança dos pedestres nos bairros atendidos. Os estudos concluem que os moradores que vivem perto de estações ferroviárias dos EUA apresentam a probabilidade cinco a seis vezes maior de trocar o transporte individual pelo coletivo.

c) Implantação de sistemas de ônibus expressos operando com prioridade nas vias – recentemente denominado como *BRT (Bus Rapid Transit)* tem a cidade de Curitiba como pioneira, e que melhor caracteriza as vantagens de operação desse sistema. Há décadas funcionando, Curitiba inspirou centenas de cidades no mundo a adotarem o modelo.

d) Áreas de uso misto – permite “a redução do número de deslocamentos feitos por veículos privados, incentiva a racionalização do sistema de transportes públicos”, além de reduzir a poluição ambiental. Para isso faz-se necessária a criação

de grandes estacionamentos e bicicletários próximos a grandes estações de trem ou metrô para permitir essa utilização.

e) Coordenação nas diferentes esferas da administração – “sem uma perfeita integração entre as esferas administrativas e políticas, o risco de desperdício de recursos estará sempre presente (...)”. Há a necessidade de uma profunda revisão, no caso brasileiro, do estatuto das cidades, para que se identifique que políticas públicas de transporte perpassam os limites dos municípios, muitas vezes de um estado, inclusive.

f) Fiscalização e monitoração – “políticas que promovam de forma adequada a fiscalização da operação de serviços de transporte público, deverão estar contempladas no planejamento estratégico urbano (...)”. A falta de fiscalização e controle da efetividade dos meios de transporte e da implementação das políticas de planejamento podem acarretar um caos no sistema como um todo.

g) Informação e marketing dos serviços de transportes – a falta de informação sobre a estrutura pública de transporte disponível reduz o número de usuários. Sem dúvida, a falta de informação por parte do usuário do transporte oferecido e a falta de informação por falta do provedor de soluções da real necessidade dos usuários levam à redução da utilização dos transportes disponíveis.

Ainda de acordo com Balassiano (2007), os países emergentes devem não apenas seguir os passos dos países ditos industrializados, mas devem trocar entre si experiências que podem ser mais criativas, uma característica forte do Brasil, bem como economicamente mais viáveis.

Vale ressaltar que a oferta de transporte coletivo é apenas parte da solução aos problemas urbanos de mobilidade. O transporte deve também atender o usuário no

horário que necessita, deve ser seguro e oferecido com qualidade. O sucateamento da frota é uma realidade na maioria dos municípios brasileiros, o que em alguns casos, acarreta o aumento do uso do transporte individual.

O cenário atual da infraestrutura de transportes de cargas e de pessoas tem se mostrado um grande bloqueio à expansão e ao crescimento econômico do país. A falta de infraestrutura não reflete apenas no comércio exterior, mas sua precariedade eleva o valor dos produtos no mercado interno e, por consequência, torna nossos produtos menos competitivos no comércio internacional. Analisando os autores, conclui-se que políticas públicas adequadas, planejamento e investimentos na área de transportes são essenciais para a retomada do crescimento brasileiro.

### **2.1.3. A importância do Planejamento de Transporte**

O planejamento de transporte é a principal área da engenharia de transportes, sendo essencial para atender as altas demandas de deslocamento, com o menor impacto socioambiental (LEVISON, et al., 2015). A área vem ganhando ainda importância, principalmente por sua obrigatoriedade para execução de obras públicas, para fins de liberação, fiscalização e também para a obtenção de recursos, bem como no sentido de obter maior efetividade para atingir o objetivo pretendido. Seus aspectos se relacionam principalmente com o planejamento urbano e sua execução envolve decisões não somente técnicas, pois geralmente sofre interferência direta de fatores políticos (CASSETTA, 2013).

Para Ferrari (1979), o planejamento de transporte é um método contínuo destinado à solução de demandas que afetam diretamente a sociedade, demandas essas, que bem atendidas, antecipam a solução de problemas futuros. Para obter bons resultados, o planejamento de transporte ideal deve objetivar o alcance do equilíbrio entre oferta-demanda e custo-benefício, e ter como consequência o aumento da sustentabilidade econômica e ambiental.

A ANTP (1997) define o planejamento de transporte como “a atividade que define a infraestrutura necessária para assegurar a circulação de pessoas e mercadorias e que organiza os sistemas de transporte que estão sujeitos à regulamentação pública, inclusive a tecnologia e nível de serviço a ser ofertado”. Tem como objeto o ambiente de circulação como um todo.

Vasconcellos (2000) tem uma visão mais ampla do planejamento de transporte, sendo este um elemento de apoio ao desenvolvimento urbano, regional e nacional, e não apenas um fim em si mesmo. A partir desta visão holística, Vasconcellos (2001) define três diferentes técnicas baseadas no controle sobre o espaço:

- a) Planejamento urbano – define padrões de ocupação e uso do solo;
- b) Planejamento de transporte – define a estrutura e a oferta física e operacional da circulação, formas de acesso e interconexões dos diferentes meios de transporte, a médio e longo prazo;
- c) Planejamento de circulação – define a distribuição do espaço disponível para a circulação entre os diversos usuários.

Como principal objeto do planejamento de transporte tem-se o sistema de transporte, que é “um conjunto de elementos organizados de forma a dirigir a ação do

sistema de *inputs* em direção a metas e objetivos específicos” (HUTCHINSON, 1979). Tedesco (2008) considera o sistema de transporte como “um conjunto de elementos que se encontram em interação para um dado objetivo”.

O sistema de transporte é composto, segundo Ortúzar e Willumsen (1997) de: infraestrutura, sistema de gestão e por um conjunto de modais de transporte e de seus operadores. Existem diferentes formas de classificação dos componentes, dependentes do interesse e da empregabilidade. Outro exemplo de classificação apresentada por Papacostas e Prevedouros (1993), um sistema de transporte consiste das instalações fixas, as entidades de fluxo e dos sistemas de controle. Manheim (1979) apresenta um estudo detalhado, onde apresenta os principais componentes e os subsistemas do sistema de transporte, como pode ser visualizado na Tabela 1.

Tabela 1 - Componentes dos sistemas de transporte - Manheim (1979)

componentes principais	subsistemas
sistema de carregamento (veículo, esteira, duto)	sistema de compartimento
	sistema de suporte (transmite o peso do veículo para o sistema viário ou outro meio de suporte)
	sistema de força e propulsão
	sistema de controle e direção
	suporte à tripulação
	serviços de suporte ao passageiro ou carga
	sistema de embarque/desembarque
sistema viário	sistema de suporte (transmite o peso para um meio de suporte)
	sistema de propulsão e força
	sistema de controle e direção
instalação de transferência	sistema de controle e direção
	sistema de embarque/desembarque
	sistemas de serviço para o veículo
	sistemas de estocagem
	sistemas de suporte ao passageiro ou carga
sistema de manutenção	sistema de manutenção de veículo
	sistema de manutenção de via
	sistema de manutenção de instalações de transferência
sistema de gerenciamento	serviços de suporte ao embarque
	sistemas operacionais
	sistema de mercado
	sistema de controle e comunicação
	sistema de pessoal
	sistema financeiro
	sistema de análise e planejamento
estrutura organizacional.	



Novas tecnologias apoiam a modelagem dos sistemas de transporte, como por exemplo, a criação e manutenção de sistemas inteligentes, análises estatísticas, inteligência artificial, agentes inteligentes, dentre outros. Usando modelos e simuladores é possível antever os padrões de transporte futuros, se antecipar e aplicar soluções para os problemas previstos (BOWERSOX e CLOSS, 2006).

O planejamento do transporte passa a ser um determinante para o crescimento sustentável das cidades. Para Cervero (2013), as cidades do futuro devem tornar-se mais sustentáveis e o planejamento de transporte tem um papel importante a desempenhar a este respeito. A ideia de uma mudança de paradigma no transporte urbano está ganhando força em muitas partes do mundo. Essa mudança não está ligada apenas à poluição ambiental causada pelos combustíveis fósseis, mas tem como objetivo criar cidades economicamente viáveis e socialmente mais justas, tendo a cidade de Dubai como exemplo. Em particular, uma mudança para a concepção de cidades mais compactas. A cidade do futuro será aquela que permite que as pessoas facilmente se locomovam a pé, em veículos de duas rodas, ônibus e trens.

Ainda de acordo com Cervero (2013), vale destacar que os sistemas de transporte urbano precisam ser inclusivos, proporcionando oportunidades de mobilidade para todos. Em uma cidade dependente do carro, aqueles sem acesso a um veículo particular - muitas vezes os economicamente menos favorecidos, deficientes físicos, jovens, idosos - são incapazes de ter acesso a oportunidades e serviços.

A seguir, apresenta-se duas importantes pesquisas realizadas por Tenopir et. al (2013) que descrevem o perfil do pesquisador-leitor em cinco Universidades nos

Estados Unidos e em duas na Austrália. Essa pesquisa contribuirá para a especificação das funcionalidades apresentadas no *software* especialista a ser desenvolvido.

## 2.2. Perfil do leitor-pesquisador

A pesquisa científica utiliza como ponto de partida para sua evolução conhecimentos pré-existentes, acumulados ao longo da história da ciência, até os estudos mais recentes (SciELO, 2014).

Probst et. al. (2002) afirmam que “para quantificar o conhecimento, precisamos externalizá-lo, (...) para ser medido de forma exata”. Para isso se faz necessário que determine corretamente os parâmetros, o que é mensurável e qual resultado se pretende atingir com as medições.

Uma pesquisa muito relevante para a presente tese foi realizada em cinco Universidades nos Estados Unidos e em duas na Austrália, conduzida por Carol Tenopir e seus colaboradores do *Center for Information and Communication Studies, University of Tennessee*, em *Knoxville*, Estados Unidos, no ano de 2012. A pesquisa busca traçar um perfil do pesquisador-leitor (TENOPIR, et al., 2013).

Os pesquisadores-leitores deparam-se com o desafio constante de selecionar o que é relevante ler, e acompanhar os recentes desenvolvimentos em uma determinada área, através da leitura de material impresso e on-line. Atualmente existem à disposição dos acadêmicos centenas de assinaturas e licenças para acessar artigos científicos, livros e outros materiais, no formato online ou impressos.

A pesquisa realizada por Tenopir (2013) aponta resultados interessantes sobre os hábitos de leitura dos pesquisadores nos dois países.

Sessenta e nove por cento dos artigos lidos por acadêmicos nas universidades da Austrália que foram alvo da pesquisa e 54% dos pesquisadores nas universidades norte-americanas acessam conteúdo que provém de assinaturas de bibliotecas e a quase totalidade (97% na Austrália, e 93% nos EUA) são em formato eletrônico. Isso não altera a preferência da leitura pelo material impresso, onde 60% dos pesquisadores na Austrália e 51% nos EUA realizam a leitura na tela de computadores. Outros dispositivos de leitura tem a preferência de 48% dos pesquisadores dos EUA e 42% da Austrália.

Nos dois países, o índice de leitura de livros eletrônicos (*ebooks*) é baixo (7% na Austrália e 8% nos EUA). Os livros são obtidos com maior frequência (37% no caso da Austrália e 39% caso dos EUA) de bibliotecas do que através de compras pessoais (29% no caso da Austrália e 22% no caso dos EUA).

O número de downloads de artigos na Universidade do Tennessee mostrou uma leve queda no ano acadêmico 2010-2011 (4,3 milhões de downloads), em comparação com as taxas sempre crescentes nos anos anteriores (4,5 milhões em 2009-2010 e 2,6 milhões em 2008-2009). Paralelamente, o número médio individual de artigos que os acadêmicos declararam ler em 2011-2012 foi de 280 contra 288 em 2005, um aumento muito pequeno.

A leitura de artigos científicos é essencial para seu trabalho acadêmico, afirmaram os pesquisadores que participaram da pesquisa e que 74% dos artigos, 58% dos livros e 45% de outros materiais são lidos principalmente para conduzir pesquisas e redigir artigos.

O tempo empregado nesta leitura é considerável, em média 49 minutos por artigo, 1 hora e 46 minutos por livro e 42 minutos para outras publicações. Os pesquisadores leem em média 22 artigos, 7 livros e 10 publicações de outros tipos por mês. Isso soma 216 horas ao ano para ler artigos, 148 horas para livros e 84 horas para outros tipos de publicação, totalizando 448 horas empregadas apenas em leitura.

Acadêmicos que receberam prêmios por seus trabalhos leem mais do que a média mensal: 30 artigos, 9 livros e 14 publicações de outro tipo. Pesquisadores que publicam mais também leem mais, segundo Tenopir et al. (2013), atingindo aproximadamente as marcas dos pesquisadores premiados.

As pesquisas apontam que os pesquisadores leem e escrevem comentários em mídias sociais com maior frequência do que publicam artigos, acompanhando assim a crescente importância das redes sociais. Vale ressaltar que as redes sociais apresentam alcance efetivo como novas fontes de influência e impacto científico. Os cientistas reconhecem o valor das mídias sociais como fonte de inspiração e novas ideias, apesar de serem de opinião de que estas não substituem os artigos tradicionais.

Presume-se que na época da publicação em papel, os pesquisadores liam o artigo por inteiro. No formato digital, porém, na maior parte do tempo os pesquisadores navegam pelo texto em busca de trechos específicos de informação. De fato, ao examinar os rastros digitais de acesso a artigos deixados pelos pesquisadores, Tenopir et al. (2013) verificaram que os pesquisadores mantêm quatro ou cinco janelas do navegador abertas ao mesmo tempo em diferentes artigos, e que os artigos com maior probabilidade de serem lidos são os mais curtos.

Inspirado nas pesquisas realizadas por Tenopir e seus colaboradores, o *software* desenvolvido para realizar as pesquisas da presente tese tem também como objetivo coadjuvante facilitar a leitura do pesquisador, apresentando a ele artigos de fato relevantes a temática que pretende pesquisar, e assim, como resultado, aumentar a leitura de artigos completos.

A proposta de um “selecionador” inteligente de artigos torna-se viável com a utilização de recursos da área de inteligência artificial, processamento de linguagem natural (PLN) e algoritmos para tratamento de dados. É possível criar grandes bancos de dados de artigos – *Big Data* e, a partir deste banco, apresentar ao pesquisador as publicações que são de maior relevância dentro de seu interesse e necessidade de pesquisa, comparar similaridade entre artigos, similaridade de artigo com seu próprio resumo, similaridade entre autores, busca por palavras-chave, interseção de palavras-chave na busca, dentre outros aspectos abordados.

A seguir apresentar-se-á conceitos, técnicas e fundamentos da tecnologia da informação e algoritmos que serão utilizados na criação do *software* especialista para realização do estudo em foco.

### **2.3. Fundamentos da Ciência da Computação a serem empregados no desenvolvimento do *software* e banco de dados**

A análise proposta por essa tese é desafiadora, principalmente pelo volume de artigos a serem analisados - mais de oito mil. Por isso a decisão de desenvolver um

*software* especialista para automatizar todo o processo de leitura dos artigos, interpretação e criação do banco de dados – *data mining*. Não foi encontrado qualquer outro *software* que realizasse os estudos métricos desejados de dados não estruturados, como são os artigos, e a posteriori realizasse os estudos cienciométricos, informétricos e bibliométricos desejados.

A seguir, apresentam-se conceitos de tecnologias e algoritmos a serem empregados na concepção do *software* especialista a ser desenvolvido.

### **2.3.1. Inteligência Artificial**

A inteligência artificial (IA) é o cerne do *software* a ser desenvolvido. Sua aplicação neste contexto tem o objetivo de substituir o leitor e analisar o conteúdo verdadeiramente relevante das publicações. A inteligência artificial é responsável por realizar o processamento de linguagem natural e transformar os dados em informações para permitir a realização dos estudos métricos.

Para os autores Russell e Norvig (2004), a inteligência artificial é “o estudo de agentes que recebem percepções do ambiente e executam ações. Cada agente implementa uma função que mapeia sequências de percepções em ações (...)”. Os agentes no caso serão os algoritmos de processamento de linguagem natural que extrairão dos artigos as informações necessárias para a criação do *data mining*.

Para Sage (1990), a inteligência artificial “é o desenvolvimento de paradigmas ou algoritmos que requeiram máquinas para realizar tarefas cognitivas, para as quais os humanos são atualmente melhores”. O autor complementa ainda que um sistema de IA deve “ser capaz de fazer três coisas: armazenar conhecimento, aplicar conhecimento armazenado para resolver problemas e adquirir novo conhecimento através da experiência”.

### **2.3.2. Processos de Aprendizagem – redes neurais**

Para Mendel e McLaren (1970) :

“Aprendizagem é um processo pelo qual os parâmetros livres de uma rede neural são adaptados através de um processo de estimulação pelo ambiente no qual a rede está inserida. O tipo de aprendizagem é determinado pela maneira pela qual a modificação dos parâmetros ocorre.”

Para que ocorra adequadamente a aprendizagem, se faz necessário a solução do problema, através do que é chamado de algoritmo de aprendizagem.

Ele é desenvolvido com um conjunto de ferramentas, que são uma variedade de algoritmos de aprendizagem. Escolhe-se o algoritmo de acordo com o problema a ser solucionado.

Para Haykin (2001), a função primordial de uma rede neural é a habilidade de aprender e de aprimorar seu desempenho ao longo do processo de aprendizagem.



“Idealmente, a rede se torna mais instruída sobre seu ambiente após cada iteração do processo de aprendizagem”.

### **2.3.3. Processamento de Linguagem Natural (PLN)**

O processamento de linguagem natural (PLN) consiste na aplicação de métodos e técnicas que possibilitem ao computador extrair a semântica da linguagem humana expressa em textos e voz. É uma estratégia de modelagem cognitiva. Suas principais aplicações são na tradução automática de textos, sistemas de diálogo, aprendizagem de máquina, sistemas especialistas, mineração de dados, robôs e recuperação de informações (YUE, et al., 2012).

O processamento de linguagem natural faz parte da classificação dos sistemas inteligentes, pois utiliza técnicas e metodologias específicas, tais como: aquisição do conhecimento, aprendizado de máquina, lógica *fuzzy*, mineração de dados e de textos. Rezende et. al. (2005) define que os sistemas para serem considerados inteligentes devem ter “habilidade para usar conhecimento para desempenhar tarefas ou resolver problemas” e ter “a capacidade para aproveitar associações e inferência para trabalhar com problemas complexos que se assemelham a problemas reais”.

No desenvolvimento do *software* desta tese o processamento de linguagem natural será um dos procedimentos mais importantes, pois será responsável pela “leitura” dos artigos, interpretação e extração de informações para a construção do

banco de dados. Em síntese, o *software* desenvolvido para a tese tem a habilidade de compreender a linguagem humana e conectar informações de forma associativa.

Como curiosidade vale lembrar que o PLN tem evoluído muito desde a época de cartões perfurados processados em lote, em que a análise de uma frase levava até 7 minutos, e na atualidade, na era do Google, milhões de páginas Web podem ser processadas em menos de um segundo (CAMBRIA e WHITE, 2014).

#### **2.3.4. Data mining**

*Data mining* ou mineração de dados é um subcampo interdisciplinar da ciência da computação. A tecnologia de mineração de dados ajuda a extrair conhecimento utilizável a partir de grandes conjuntos de dados, (HONG, et al., 2013) envolvendo métodos na intersecção da inteligência artificial, aprendizado de máquina, estatística e sistemas de banco de dados.

A mineração de dados é uma das áreas que vem ganhando muito espaço e cresce em ritmo acelerado, com novos métodos, metodologias e apresenta resultados em várias áreas do conhecimento como medicina, ciência da computação, bioinformática, previsão do mercado de ações, previsão do tempo, texto, áudio e processamento de vídeo, dentre outros (PHRIDVIRAJA e GURURAOB, 2014).

De acordo com Carvalho (2001), “a classificação é uma das técnicas mais utilizadas do data mining simplesmente porque é uma das tarefas cognitivas humanas mais realizadas no auxílio à compreensão do ambiente em que vivemos”.

O objetivo global do processo de mineração de dados é extrair informações a partir de um conjunto de dados e transformá-lo em uma estrutura compreensível para uso a posteriori. O processo inicia-se na etapa de análise crua, que envolve bancos de dados, gerenciamento dos aspectos dos dados, pré-processamento de dados, modelos e inferências, métricas, estruturas complexas e vai até o pós-processamento das estruturas descobertas, visualização e atualização.

A tarefa real de mineração de dados é a análise automática ou semiautomática de grandes quantidades de dados, anteriormente desconhecidos, para extrair importantes grupos de registros de dados (análise de cluster), registros incomuns (detecção de anomalias) e dependências. Pode ser resumido pela definição de Aggarwal et. al. (2012), que *data mining* é o processo de descobrir informações ocultas e potencialmente úteis em grandes bases de dados

*Data mining* vai ser empregado na análise dos dados do *software* desenvolvido, com o objetivo de encontrar tendências e padrões dentro do banco de dados criado. A ferramenta a ser criada para análise de similaridade entre resumos e textos, e textos-textos valer-se-á desta tecnologia.

### 2.3.5. Big data

*Big data* é um termo amplo utilizado para conjuntos de dados muito grandes ou complexos, onde as aplicações de processamento de dados tradicionais são insuficientes. Os desafios atendidos pelo *big data* incluem análise, captura, tratamento de dados, pesquisa, compartilhamento, armazenamento, transferência, visualização e garantia de privacidade.

Para Cumbley e Church (2013) *big data* também se caracteriza por uma mudança na natureza dos dados a serem armazenados. Tradicionalmente, os dados seriam armazenados num formato estruturado para maximizar o seu conteúdo informativo. Por exemplo, uma base de dados relacional tem um determinado número de campos, cada um dos quais vai conter um tipo específico de dados num formato específico. A estrutura tradicional faz com que seja fácil de manipular e processar através da aplicação de regras determinísticas simples para que os dados sejam localizados.

Ainda de acordo com Cumbley e Church (2013), uma grande quantidade de dados não fornece necessariamente uma série de informações úteis. Por isso é de grande importância selecionar o que de verdade é relevante para constituir o banco de dados.

A análise de conjuntos de dados pode encontrar novas correlações, como por exemplo, tendências de negócios, prevenir doenças, combate ao crime e assim por

diante. Na presente tese busca-se padrões e tendências em publicações na área da Engenharia de Transportes.

Como os conjuntos de dados em sua grande maioria, inevitavelmente, em tamanho e espaço de armazenamento, o *software* a ser desenvolvido irá prever essa expansão e seu banco de dados foi criado dentro dos princípios do *big data*, suportando facilmente a inserção de novos artigos e publicações.

### **2.3.6. Text mining**

*Text mining* ou mineração de texto é aproximadamente equivalente à análise de texto e refere-se ao processo de extrair informações significativas a partir de dados textuais não estruturados (HEA, et al., 2013). Informação de alta qualidade é normalmente obtida através da identificação de padrões e tendências através de meios como o padrão estatístico de aprendizagem.

A mineração de texto é uma extensão da mineração de dados para dados textuais. Mineração de texto tem seu foco em encontrar modelos úteis, tendências, padrões ou regras de dados textuais não estruturados, como arquivos de texto, arquivos HTML, mensagens e e-mails (HEA, et al., 2013), sendo muito útil no presente trabalho para analisar as publicações, pois o conteúdo dos artigos não é padronizado e também não estruturado.

A mineração de texto normalmente envolve o processo de estruturação do texto de entrada, que compreende em analisar, conjuntamente com a inserção de alguns recursos linguísticos e a remoção de outros, e subsequente armazenamento em um banco de dados.

Na mineração de texto, inicialmente, os documentos de texto não estruturados são processados utilizando técnicas de processamento de linguagem natural para extrair palavras-chaves, rotulando os itens de texto. Na sequência, técnicas de mineração de dados clássicos são aplicadas sobre os dados extraídos (palavras-chaves) para descobrir padrões relevantes, repetindo até que a informação seja extraída (BAGHELA e TRIPATHI, 2012). Em seguida, realiza-se a fase de análise de texto.

A análise de texto envolve a recuperação da informação, análise léxica para estudar distribuições de frequência de palavras, reconhecimento de padrões, marcação/anotação, extração de informações, técnicas de mineração de dados, incluindo ligação e análise de associação, visualização e análise preditiva. Concluindo, é essencialmente transformar textos em dados para análise, através da aplicação de processamento de linguagem natural (PLN) e métodos analíticos.

A função da mineração de dados na construção do futuro banco de dados possibilitará armazenar no banco informações relevantes para a aplicação do PLN, retirando termos sem valor semântico ou com baixa relevância para a análise.

### **2.3.7. Web scraping**

*Web scraping* (colheita *web* ou extração de dados da *web*) é uma técnica de extração de informações de sites de *software* de computador. Normalmente, os *softwares* de *web scraping* estão intimamente relacionados com a indexação da *web*, que indexa informações utilizando um *bot* ou *web crawler*, uma técnica universal adotada pela maioria dos robôs de busca.

No presente trabalho a técnica *web scraping* será utilizada para simular a navegação humana, através de um *software* desenvolvido para esse fim específico. O *software* permitirá realizar o download dos mais de 10 mil artigos do site da revista a ser analisada, de forma totalmente automatizada, sem a interferência humana.

A decisão de utilizar um robô para realizar os downloads foi em função do site da *Elsevier* não possuir opção para download conjunto de todas as publicações. Os downloads serão realizados individualmente, onde o *software* simulará a digitação e as seleções realizadas com mouse pelo usuário.

### **2.3.8. Tokenization**

A análise léxica *tokenization* é o processo de quebrar um fluxo de texto em palavras, frases, símbolos ou outros elementos significativos chamados *tokens*. A lista de *tokens* tornam-se entrada para outro processamento adicional, como análise ou de

mineração de texto. *Tokenization* é útil tanto em linguística (onde é uma forma de análise morfológica), como em ciência da computação, onde forma parte da análise lexical.

Detectar limites dos *tokens* e frases é uma etapa de pré-processamento importante em aplicações de processamento de linguagem natural. Os principais desafios desta tarefa são eliminar as ambiguidades de conjugações verbais, por exemplo, bem como identificar classes gramaticais não relevantes para a aplicação (JURISH e WÜRZNER, 2013).

Normalmente a “tokenização” é realizada por palavras. Em idiomas que usam espaços entre palavras (como a maioria dos que usam o alfabeto latino, e a maioria das linguagens de programação), esta abordagem é bastante simples. No entanto há muitas exceções tais como contrações, palavras com hífen, *emoticons* e construções maiores, como URLs, por exemplo.

O *software* a ser desenvolvido utilizará como critérios de criação dos *tokens* os estabelecidos pelo autor Leydesdorff (2001). O autor sugere que com exceção de advérbios derivados diretamente de adjetivos, todos os advérbios, números, pronomes, conjugações verbais do ‘to be’, ‘to have’, ‘may’, ‘will’, ‘shall’, ‘can’, etc., (vale salientar que a base de dados no *software* é em inglês) sejam excluídos. As formas singular e plural são consideradas como a mesma palavra, e comparativos e superlativos são considerados como a forma básica do adjetivo. Conjugações verbais são reduzidas à forma do presente do indicativo para eliminar ambiguidades. Todas as abreviaturas de uma letra foram excluídas da análise.

Para a contagem de ocorrência de *tokens* não será estipulado limites de palavras por artigo, ao contrário do que o autor empregou em sua análise. O *software*



irá restringir o limite de dez palavras por artigo e dez palavras por resumo somente para o cálculo da similaridade de artigo/resumo e artigos/artigos.

### **2.3.9. Separação de *part-of-speech* e Semelhança Semântica**

*Part-of-speech* é uma categoria de palavras, ou genericamente, dos itens lexicais, que têm semelhantes propriedades gramaticais. Palavras que são atribuídas às mesmas *part-of-speech* geralmente apresentam comportamento semelhante em termos de sintaxe que desempenham funções semelhantes dentro da estrutura gramatical das sentenças e, por vezes, em termos de morfologia, em que sofrem inflexão para palavras semelhantes. A classificação mais comumente identificada dos elementos de linguagem são: substantivos, verbos, adjetivos, advérbios, pronomes, preposições, conjunções, interjeições e com menor frequência: numerais, artigos e determinantes.

A semelhança semântica é o princípio no qual uma métrica é dada aos grupos de termos ou documentos baseados na similitude de seus significados. Este é um conceito importante na compreensão da linguagem natural, pois nos permite fazer comparações, análises e tirar conclusões.

Desempenha um papel muito importante em diversos campos da inteligência artificial, como a categorização automática e sumarização, tradução automática, recuperação de informação (FURLAN, et al., 2013). A semelhança semântica será

empregada no banco de dados do *software* a ser desenvolvido para agrupar palavras semelhantes de grafia e de significado.

### **2.3.10. Computação Paralela**

A computação paralela é um tipo de computação onde vários processos são realizados simultaneamente, por considerar que por muitas vezes grandes problemas podem ser divididos em partes menores, e que então são resolvidos ao mesmo tempo.

A computação paralela encontra-se em crescimento contínuo em sua aplicação em função do aumento no desempenho de aplicativos. Para que o desempenho de um aplicativo atinja seu ápice de velocidade de processamento, uma parcela significativa de seus cálculos devem ser executados em paralelo (HWU, 2014).

Na presente tese os algoritmos do *software* a ser desenvolvido irão “rodar” em paralelo, ou seja, muitas das funções serão executadas simultaneamente. A computação paralela trará expressivo ganho de tempo e de performance para a obtenção das informações, processamento dos dados e geração de pesquisas e gráficos.

### 2.3.11. TF-IDF

O *tf-idf*, abreviação do inglês *term frequency–inverse document frequency*, ou seja, frequência do termo–inverso da frequência nos documentos, é utilizado para verificar o quão importante é uma palavra para um determinado documento (HONG, et al., 2013). O *tf-idf* é frequentemente utilizado como fator de ponderação em recuperação de informação e em mineração de texto. O valor de *tf-idf* aumenta proporcionalmente ao número de vezes que uma palavra aparece no documento, mas é compensado pela frequência da palavra no total analisado (AIZAWA, 2003).

No caso do *software* a ser desenvolvido, será verificada a frequência de um termo em um documento comparada ao total de palavras válidas, no caso de estar utilizado em conjunto com o PLN, ou seja, o número de vezes em que o termo *t* ocorre no documento *d*. O peso de um termo em um documento é diretamente proporcional à sua frequência no mesmo.

O primeiro passo no cálculo é identificar a frequência dos termos, que é a frequência bruta da quantidade de vezes que o termo *t* ocorre no documento *d*, como descrito na fórmula abaixo:

$$tf(t, d) = f_{t,d}$$

A segunda parte do cálculo é obter a frequência invertida no documento, que indica quanta informação a palavra fornece. Este dado aponta se uma palavra é

rara ou comum em um conjunto de documentos. É o resultado de uma fração multiplicada por um escalar logarítmico.

$$\text{idf}(t, D) = \log \frac{N}{|\{d \in D : t \in d\}|}$$

- Onde  $N$  é o total de documentos no corpus a ser analisado;
- $|\{d \in D : t \in d\}|$  é o número de documentos onde o termo  $t$  ocorre.

A última etapa do cálculo consiste em obter o produto destes dois índices:

$$\text{tfidf}(t, d, D) = \text{tf}(t, d) \cdot \text{idf}(t, D)$$

Quanto maior o índice obtido, mais importante é a palavra. A vantagem desta abordagem é que palavras que ocorrem mais obtêm índices baixos, pois são consideradas de apoio gramatical. Palavras que ocorrem menos possuem mais chance de terem peso semântico maior e por isso obtêm índices mais elevados

### **2.3.12. Similaridade de Jaccard**

A semelhança entre artigos, ou entre artigos e resumos, no caso do *software* a ser desenvolvido, define-se pela interseção de *tokens* idênticos entre os textos. Esta semelhança é definida pela similaridade de Jaccard (LESKOVEC, et al., 2014).

A similaridade de Jaccard de dois conjuntos é o total de elementos da interseção dividido pelo total de elementos do conjunto união formado entre ambos.

### **2.3.13. Node.js**

*Node.js* (2015) é um interpretador de código *JavaScript* que é executado no servidor. Seu objetivo é possibilitar a criação de aplicações de alta escalabilidade, com códigos capazes de manipular dezenas de milhares de conexões simultâneas, em uma única máquina física. O *Node.js* é baseado no interpretador V8, criado pelo Google.

*Node.js* possui arquitetura orientada a eventos, e sua *API* (*Application Programming Interface*) de entrada e saída não possui bloqueios por ser projetada para otimizar a saída, apresentação dos resultados e escalabilidade de um aplicativo para aplicações web em tempo real. O Google tem uma grande porcentagem dos módulos básicos escritos em *JavaScript*. A abordagem permite que seja utilizada a mesma linguagem tanto no cliente quanto no servidor, o que aumenta as possibilidades de reaproveitamento de código.

A escolha do *Node.js* para o desenvolver o *software* é visando obter velocidade às consultas e saída de dados, permitir o *software* ser multiplataforma, ou seja, por rodar via web, independentemente do sistema operacional que o usuário esteja utilizando em sua máquina, podendo inclusive ser acessado por *smartphones*.

#### **2.3.14. WordNet**

*WordNet* (Princeton University, 2015) é um grande banco de dados léxico do idioma inglês. Léxico é um conjunto das unidades significativas de uma dada língua, num determinado momento da sua história. Substantivos, verbos, adjetivos e advérbios são agrupados em conjuntos de sinônimos cognitivos e são interligados por meio de relações lexicais, no caso do *WordNet* também através de relações conceituais-semânticas.

O *WordNet* se assemelha a uma enciclopédia, na medida em que os grupos de palavras são criados com base em seus significados. Apesar da semelhança, no entanto, há algumas diferenças importantes. As palavras que se encontram em estreita proximidade uma da outra na rede são semanticamente desambiguadas e os agrupamentos ocorrem por similaridade.

A principal relação entre palavras no *WordNet* é a sinonímia, ou seja, palavras que denotam o mesmo conceito e que são permutáveis em muitos contextos, como por exemplo, as palavras *carro* e *automóvel*. Os sinônimos são agrupados em conjuntos não ordenados no *WordNet*.

### 2.3.15. Grafos

A teoria dos grafos estuda objetos combinatórios — os grafos — que consiste em um modelo para a solução de muitos problemas em vários ramos da matemática, da informática, da engenharia e da indústria. Muitos problemas envolvendo grafos tornaram-se célebres por serem um grande desafio intelectual, também abrangem importantes aplicações práticas, como por exemplo o famoso dilema do caixeiro viajante.

Os grafos permitem a abstração necessária para codificar relacionamentos entre pares de objetos. Para o autor Boaventura Netto (2003), “um grafo é uma estrutura  $G = (X,U)$  onde  $X$  é um conjunto discreto e  $U$  é uma família cujos elementos  $u$  (não vazios) são definidos em função dos elementos  $x$  de  $X$ ”.

Grafos são modelos matemáticos formados por estruturas simples que consistem de um conjunto de vértices e um conjunto de arestas. Os vértices em geral representam objetos concretos ou abstratos, em diversas áreas do conhecimento humano, e as arestas indicam relações entre esses objetos. De acordo com Nascimento e Ferreira (2005), define-se um grafo  $G$  (não-direcionado) como um par  $(V, E)$ , onde  $V$  é um conjunto de vértices e  $E$  é um conjunto de arestas.

Exemplificando, um grafo  $G = (V,E)$  é um conjunto não-vazio  $V$ , cujos elementos são chamados vértices, e um conjunto  $E$  de arestas. Uma aresta é um par não-ordenado  $(v_i,v_j)$ , onde  $v_i$  e  $v_j$  são elementos de  $V$ . Normalmente, utiliza-se uma

representação gráfica de um grafo. A seguir um exemplo de grafo, com a sua representação gráfica:

$$V = \{v1, v2, v3, v4, v5\}$$

$$E = \{a1, a2, a3, a4, a5, a6, a7, a8\}$$

onde:

$$a1 = (v1,v2), a2 = (v1,v3), a3 = (v1,v3), a4 = (v2,v3), a5 = (v2,v5), a6 = (v5,v5), \\ a7 = (v3,v5), a8 = (v3,v4).$$

No presente trabalho os grafos serão empregados para determinar relevâncias de termos, permitindo assim avaliar quais temas tiveram mais destaque nas publicações ao longo do tempo, além de criar grandes grupos com os termos de maior relevância, grupos esses que trazem grande significância à interpretação pretendida.

O próximo capítulo traz uma apresentação da metodologia a ser utilizada, que se valerá dos fundamentos da revisão bibliográfica, para solucionar as questões propostas nesta tese, com o auxílio das métricas que serão também apresentadas no próximo capítulo.



## CAPÍTULO 3 – METODOLOGIA E PROCEDIMENTOS

“O conhecimento científico surge da necessidade de o homem não assumir uma posição meramente passiva, de testemunha dos fenômenos, sem poder de ação ou controle dos mesmos. Cabe ao homem, otimizando o uso da sua racionalidade, propor uma forma sistemática, metódica e crítica da sua função de desvelar o mundo, compreendê-lo, explicá-lo e dominá-lo” (KÖCHE, 2011).

“De pé sobre o ombro de gigantes”. (Isaac Newton)

O presente capítulo tem como objetivo abordar a metodologia científica a ser adotada, bem como apresentar a cienciometria, informetria e bibliometria, metodologias essas que fundamentam as análises das pesquisas em transportes a serem realizadas pelo *software* a ser desenvolvido. A motivação para a escolha da revista *Transportation Research* encerra o presente capítulo desta tese.

### 3.1. Metodologia

O método científico é um conjunto de técnicas para investigar fenômenos, seja para a aquisição de novos conhecimentos, corrigir ou integrar conhecimentos prévios.

Traz também procedimentos lógicos com o objetivo de produzir conhecimento científico. Esses procedimentos variam de uma área para outra, sem nunca perder seu foco na produção científica. O método científico é construído de forma que a ciência e suas teorias evoluam ao longo do tempo (MARCONI e LAKATOS, 2003).

A metodologia científica tem sua origem em Aristóteles, passa por Descartes, que propôs decompor o problema em pequenas partes (pequenas partes compõem o sistema como um todo) e a posteriori foi desenvolvido empiricamente por Isaac Newton. Karl Popper trouxe como colaboração que hipóteses devem ser submetidas também a testes de falseamento e não apenas a comprovação de veracidade. Ainda para Popper, a ciência é um conhecimento provisório e que nunca se prova uma teoria científica em definitivo. Já Thomas Kuhn traz como contribuição para o método científico os paradigmas como elementos essenciais, denominando as mudanças de paradigma de revoluções científicas (KÖCHE, 2011).

As bases lógicas do conhecimento científico têm como principais vertentes os métodos: indutivo, dedutivo, hipotético-dedutivo, dialético e fenomenológico. O método científico requer não apenas seguir uma metodologia, mas também um sujeito ativo, inteligente, imaginativo e criativo. A metodologia empregada nesta tese é a hipotético-dedutiva, onde a investigação científica passa pela formulação de uma hipótese de uma forma que poderia concebivelmente ser falseada por um teste em dados observáveis.

A seguir as etapas do método hipotético-dedutivo:

- Experiência pessoal: reconhecer o problema e tentar dar sentido a ele. Coletar dados e procurar explicações anteriores.
- Formar uma conjectura: propor uma explicação para o problema.

- Deduzir previsões da hipótese: se você assumir que algo é verdadeiro, que caminhos seguir?
- Teste: procurar evidências que entrem em conflito com as possíveis soluções, a fim de refutar. Voltar ao passo dois até a conclusão.

A partir desta perspectiva, a presente tese irá se valer de métricas para obter os resultados para as análises necessárias para atingir os objetivos desta pesquisa. As métricas serão transformadas em algoritmos computacionais para automatizar todo o processo humano de leitura e interpretação, através da utilização do processamento de linguagem natural, inteligência artificial e grafos, representando os resultados através de tabelas e gráficos.

A seguir apresenta-se uma descrição das três metodologias a serem empregadas nos algoritmos do *software* que será desenvolvido, sua relevância e importância para o estudo da relevância das publicações ao longo do tempo.

### **3.2. Cienciometria, Informetria e Bibliometria**

Medir e analisar a produção científica e tecnológica é uma atividade complexa, mas essencial como instrumento de impacto da produção científica. Busca-se com uso de métricas reduzir avaliações subjetivas e assim democratizar o processo de valorização do progresso acadêmico e científico. São as métricas, por exemplo, que calculam os rankings dos periódicos e revistas em cada área do conhecimento (ZAVADSKAS, et al., 2014), bem como classificam pesquisadores e suas publicações.

As métricas atualmente difundidas e aceitas para medir o conhecimento científico nasceram em meados do século XX. São elas a cienciometria, a informetria e a bibliometria, métricas essas que são o foco de desta seção do capítulo 3.

Para o autor Macias-Chapula (1998):

“As abordagens informétricas, bibliométricas e cienciométricas, pelas quais a ciência pode ser retratada através dos resultados que alcançam, são baseadas na noção de que a essência da pesquisa científica é a produção de conhecimento e que a literatura científica é um componente desse conhecimento”.

Os fundamentos da cienciometria, da informetria e da bibliometria serão detalhados a seguir, com destaque para suas características, aplicações e resultados alcançados.

### **3.2.1. Cienciometria**

A cienciometria surgiu na antiga URSS e Europa Oriental e foi empregada especialmente na Hungria em seus primórdios, nos anos 60, sendo a publicação *Scientometrics*, dos russos Nalimov e Mulchenko, considerada como a precursora do tema. As primeiras definições consideravam a cienciometria como "a medição do processo informático", onde o termo "informático" significava "a disciplina do conhecimento que estuda a estrutura e as propriedades da informação científica e as leis do processo de comunicação", por Mikhilov et al. (apud (SPINAK, 1996).

A cienciometria atualmente vem a ser conceituada como uma avaliação quantitativa da produção científica e tem como objetivo analisar, identificar e medir as diferentes atividades de pesquisa que abrangem as áreas do conhecimento, instituições e países. É uma grande aliada para projetos de inovação tecnológica (LEYDESDORFF e MILOJEVIC, 2013). Vale destacar que a cienciometria observa diretamente o processo de produção do conhecimento sem avaliar diretamente os resultados produzidos nessas publicações.

A cienciometria utiliza-se de indicadores científicos – avalia o marco temporal, quantificando o passado e presente com vistas a uma proposta de intervenção no futuro. Por isso a cienciometria também é objeto e fonte de pesquisa para historiadores, sociólogos e demais interessados na evolução científica.

A cienciometria tem infinitas áreas de aplicação, sendo também relevante ferramenta para governos e instituições de pesquisas avaliarem o progresso científico, reavaliarem áreas de foco de investimento em pesquisa, elaborarem políticas de governo e estratégias de gestão universitária e, com isso, apoiar o desenvolvimento científico e tecnológico dentro de suas necessidades específicas (SILVA e BIANCHI, 2001).

Já para os pesquisadores a cienciometria é uma poderosa ferramenta para a análise das citações, estudo da evolução de pesquisas em áreas temáticas determinadas, surgimento de novas áreas de estudo e importância na atualidade, bem como sua evolução em uma janela temporal.

A análise quantitativa traz um grande diferencial em avaliar a ciência da perspectiva métrica, sendo assim imparcial, diferentemente das opiniões e dos julgamentos realizados por colegas especialistas, pares, bancas, etc.

Vale ressaltar um importante marco da cienciometria. Ela foi a ferramenta que o pesquisador de linguística, o norte-americano Eugene Garfield, utilizou para propor a criação de um índice para citação científica *Science Citation Index*, (2014). O objetivo do SCI era medir e demonstrar como se dá a propagação das ideias na área científica. A utilização do *Science Citation Index* por governos de países como Estados Unidos, Canadá e Austrália como indicadores de Ciência e Tecnologia abriu espaço para estudos sobre políticas para a produção científica.

A aplicação da ciência da computação à cienciometria, após anos os trabalhos de Garfield (1996, 1998 e 2000), tem como principal resultado servir de inspiração para a criação do *Pagerank*, o algoritmo responsável pela ordenação das informações pelo Google em seus primeiros anos. O algoritmo usa como base os links entre as páginas e permitiu ao Google dominar o mercado de mecanismos de busca (BRIN e PAGE, 2001).

O segundo grande marco científico da cienciometria foi uma pesquisa realizada por Small e Griffith (1974). O objetivo foi avaliar a importância de autores em seu ambiente acadêmico, através de identificação das co-citações. O produto da referida pesquisa foi o resultado de uma combinação de dados bibliométricos com dados cienciométricos, de modo a explicar o desenvolvimento das ciências.

Até os anos 70, questões sobre o crescimento e a dinâmica do conhecimento científico pertenciam ainda ao campo da filosofia. O principal ponto na filosofia da

ciência era validar o conhecimento de forma qualitativa. Era notório que a reconstrução filosófica era analiticamente distinta de questões sobre como este conhecimento estava sendo descoberto (RUIZ e JORGE, 2002).

Já nos anos 80 houve uma importante cisão na cienciometria, que ficou com o foco exclusivo nas medidas quantitativas das produções acadêmicas, enquanto o foco qualitativo foi direcionado para o estudo sistemático da “sociologia do conhecimento científico”.

O crescimento contínuo do número de publicações estreitou a ligação da cienciometria com área de computação científica, sendo esta uma grande facilitadora para realizar análises de grandes conjuntos de dados e informações. O avanço da capacidade de processamento computacional inclusive viabilizou estudos anteriormente inviáveis.

De uma perspectiva metodológica, a escolha de artigos científicos como unidade de análise de estudos cienciométricos são os que mais contribuem, pois descrevem o desenvolvimento científico no nível de crescimento do conhecimento e também contém a maior parte das atribuições relevantes organizadas como autores, endereços de instituições, publicações, etc.

A cienciometria possui estreita interface com outras duas métricas, bibliometria e a infometria (LEYDESDORFF e MILOJEVIC, 2013), a serem apresentadas a seguir.

### **3.2.2. Bibliometria**

A bibliometria é o estudo dos aspectos quantitativos da produção, disseminação e uso da informação registrada (TAGUE-SUTCLIFFE, 1992). A bibliometria, portanto, desenvolve padrões matemáticos para medir esses processos. Os principais objetos observados pela bibliometria são livros, documentos, artigos e revistas, entre outros.

A análise bibliométrica permite medir a informação científica e avaliar sua eficiência e produtividade. Importante ressaltar que as relações entre as publicações podem fornecer modelos em rede do desenvolvimento científico (MARKSCHEFFEL, 2011).

Dentro da perspectiva humanista, a publicação de novos descobrimentos é uma etapa essencial do processo de investigação, permitindo ao cientista obter o reconhecimento de seu próprio trabalho.

É notório que a publicação científica é um elemento essencial e sólido da pesquisa, sendo subsídio para os indicadores bibliométricos, que os validam como uma medida indireta da atividade científica. Também é uma condição reconhecida que um novo conhecimento somente adquire o seu valor quando ele é difundido dentro da comunidade científica, aceito, replicado, aprimorado, e sua expansão poderá contribuir para o avanço científico (RUIZ e JORGE, 2002).

Dois são os principais indicadores bibliométricos:



- Os indicadores quantitativos da atividade científica refletida no número de publicações por autor, por periódico, por ano, etc.

- Os indicadores de impacto, que baseados nos indicadores quantitativos caracterizam a importância do tema, do pesquisador, do periódico, etc.

O número de publicações permite comparar o grau de especialização de uma instituição, centro ou uma região em um determinado tema ou assunto. Assim é possível identificar o quanto uma unidade científica se dedica para o estudo de uma área específica, comparado a outros centros, e ainda o impacto na produção científica nacional.

Estudos bibliométricos apontam o Brasil como detentor de 1,2% da produção científica mundial, de acordo com as publicações indexadas no *SCI - Science Citation Index* (2014), e ocupa a 17ª posição entre os países ranqueados.

Vale lembrar que o número de publicações é um indicador meramente quantitativo, que não leva em conta a qualidade ou a importância do conteúdo publicado, além de que cada publicação desperta diferentes interesses no leitor, e podem ou não colaborar para o progresso científico da área.

Na busca da valorização e mensuração da qualidade dos trabalhos publicados foram criados dois outros indicadores: o número de citações que um artigo recebe na literatura e o fator de impacto da revista ou periódico em que a publicação é feita.

O número de citações recebidas por um artigo é um indicador da influência e contribuição que o seu conteúdo trouxe para a comunidade científica. Descobertas e

avanços que fundamentam novas linhas de pesquisa, teóricas ou práticas, certamente despertam interesse e serão frequentemente citados em trabalhos posteriormente publicados.

O fator de impacto foi concebido por Garfield (1998, 2000), Diretor Emérito do *Institute for Scientific Information* (ISI), em 1955. Atualmente este indicador é gerado a partir do *Science Citation Index* (SCI) que é uma base de dados multidisciplinar de ciência e tecnologia. O SCI (2014) engloba aproximadamente 8.000 revistas, selecionadas através de critérios de qualidade científica formal e também considerando o reconhecimento da comunidade científica, através do número de citações recebidas. Ele avalia a influência de um determinado artigo sobre as pesquisas afins em uma determinada janela temporal. Para Garfield (1996) o impacto de um trabalho sobre os pesquisadores da área seria um indicador indireto de qualidade.

Não há unanimidade quanto à validade deste indicador. Para Brinn et al. (2000) o número de citações de um artigo científico não depende somente de sua qualidade, mas também de fatores como o reconhecimento do(s) autor(es) como especialistas da área, instituição de pesquisa, contemporaneidade do tema abordado, o idioma de publicação e também a revista ou periódico em que foi publicado. Para eles, esse indicador deve ser considerado como um indicador parcial de qualidade.

Ter um artigo publicado em periódicos de alto fator de impacto, como a *Nature* ou a *Science*, ou a própria *Transportation Research* da editora *Elsevier* é considerado, pela maioria dos pesquisadores, um critério que determina a qualidade do trabalho, pois, em geral, esses periódicos apresentam critérios rigorosos de análise e aprovação.

Além do fator de impacto, é importante considerar também dois outros indicadores bibliométricos: o índice de imediatismo ou de rapidez e o de vida média de um determinado artigo. O índice de imediatismo é determinado pela rapidez de citação dos trabalhos de uma revista ou periódico. Quanto menor for este índice tanto maior será o valor da revista, pois suas publicações são rapidamente utilizadas por outros pesquisadores para transformá-las em novos conhecimentos. O índice de vida média (ou meia-vida) permite conhecer o grau de envelhecimento ou obsolescência de um artigo. Ele é definido como o número de anos retrospectivamente necessários para alcançar 50% das citações dos artigos desta revista.

Finalmente, devemos destacar que o ritmo decadencial das bibliografias referenciadas nas diferentes áreas do saber varia em função da velocidade do progresso científico de cada uma delas, o que determina a rapidez com que os trabalhos são citados. Em áreas de rápido crescimento os pesquisadores tendem a citar os trabalhos mais recentes, enquanto nas de evolução mais lenta citam-se trabalhos mais antigos. Portanto, diferenças na contagem do número de citações de artigos publicados por pesquisadores de áreas distintas não implicam necessariamente numa diferença no impacto dos artigos e nem na qualidade das pesquisas realizadas (GARFIELD, 2000).

### **3.2.3. Informetria**

A Informetria, conceituada por Tague-Sutcliffe (1992), tem seu foco na ocorrência de palavras, documentos, bases de dados etc. O principal objetivo da

Informetria é melhorar a eficiência na recuperação dos documentos. Diferentemente de citações, as palavras podem também ser encontradas em outros tipos de documentos, como registros de conferências e arquivos governamentais (MARKSCHEFFEL, 2011).

“A informação é um dado analisado e contextualizado”, de acordo com Rezende (2005). O dado é um elemento puro, que pode ser quantificado em um determinado evento, já o conhecimento é “a habilidade de criar um modelo mental que descreva o objeto e indique as ações a implementar”.

É importante observar que para a construção do conhecimento o indivíduo realiza comparações e combinações anteriormente apreendidas (hiperconexões) que são dependentes “de nossos valores e nossa experiência e sujeito às leis universalmente aceitas” (REZENDE, 2005). A natureza do conhecimento humano utiliza a compreensão, análise e síntese de informações para a tomada de decisões inteligentes.

O uso de palavras para reconstruir as ciências (ex: em termos de mapeamento cienciométrico) aparentemente soa mais atrativo de um ponto de vista de ciência da informação. Palavras têm significados, estão presentes em todos os níveis de análise e em banco de dados o estudo da dinâmica de sua frequência e distribuição é relativamente fácil com a ajuda do computador (BJÖRNEBORN e INGWERSEN, 2004).

A informetria engloba a cienciométrica e a bibliometria por abranger todas as palavras, criando a partir desta base de dados uma estrutura para diversas análises. O Processamento de Linguagem Natural (PLN) aplicado à informetria possibilita extrair diversos relatórios, gerar gráficos e antever tendências. Para Silva e Bianchi (2001),

essas análises possibilitam identificar importantes atividades em áreas específicas e extrair desses dados indicadores.

Para encerrar as três metodologias, apresenta-se a seguir na Tabela 2 um quadro comparativo que apresenta uma síntese dos conceitos descritos anteriormente, classificado por objetos de estudo, variáveis, métodos e objetivos de bibliometria, cienciometria e informetria, de acordo com o pesquisador McGrath (1989).

Antes do encerramento do presente capítulo, será apresentada a revista *Transportation Research*, a motivação de sua escolha para a construção do banco de dados para realizar os estudos da presente tese e quais métricas considera para avaliar o conteúdo publicado.

Tabela 2 - Quadro comparativo entre as três metodologias

<b>Tipologia</b>	<b>Bibliometria</b>	<b>Cienciometria</b>	<b>Informetria</b>
<b>Objetos de Estudo</b>	Livros, documentos, revistas, artigos, autores, usuários	Disciplinas, assunto, áreas, campos	Palavras, documentos, bases de dados
<b>Variáveis</b>	Número de empréstimos (circulação) e de citações, frequência de extensão de frases, etc.	Fatores que diferenciam as sub-disciplinas. Revistas, autores, documentos. Como os cientistas se comunicam.	Difere da cienciometria no propósito das variáveis; por exemplo medir a recuperação, a relevância, a revogação, etc.
<b>Métodos</b>	Ranking, frequência e distribuição.	Análise de conjunto e de correspondência.	Modelo vetor-espaço, modelos booleanos de recuperação, modelos probabilísticos; linguagem de processamento, abordagens baseadas no conhecimento, tesouros.

<b>Objetivos</b>	Alocar recursos: tempo, dinheiro, etc.	Identificar domínios de interesse. Onde os assuntos estão concentrados. Compreender como e com qual frequência os cientistas se comunicam.	Melhorar a eficiência da comunicação
------------------	--	--	--------------------------------------

### 3.3. A escolha do periódico *Transportation Research*

A escolha do periódico *Transportation Research* (2015) publicado pela *Elsevier* foi decorrente do reconhecimento internacional da revista, considerada uma das mais relevantes na área de transportes, tanto pela qualidade de seu conteúdo, como pelo reconhecimento internacional dos autores que nela publicam, considerando também o fator de impacto dito anteriormente.

A *Elsevier* existe há mais de 135 anos e é uma empresa de publicação de revistas, jornais e periódicos nas áreas científica, médica e técnica. Atualmente seu conteúdo é disponibilizado por meio digitalizado em bases de dados proprietária, onde é possível acessar seu conteúdo através do pagamento de uma assinatura ou compra específica do material de interesse.

A *Transportation Research* abrange todos os modos de transporte de passageiros e mercadorias, bem como análise, formulação e avaliação de políticas; planejamento; interação com o ambiente político, socioeconômico e físico; design, gestão e avaliação de sistemas de transporte. Atualmente é subdividida em 6 revistas,

de A até F, para facilitar localização temática das publicações. As 6 se complementam e formam um conjunto coeso e abrangente de pesquisa na área da ciência do transporte.

Propostas de inovação tecnológica dentro de uma estrutura social ou política também são aprovados para publicação nesta revista. A *Transportation Reserch* é uma revista internacional, aspecto esse observado pela exposição de problemas das regiões industrializadas e não industrializadas.

A primeira publicação da revista *Transportation Research* foi em 1967, e nesta época a revista publicava um número por ano, subdividido em 4 trimestres. Em 1979, a revista subdividiu-se pela primeira vez em duas partes, sendo que a *Transportation Research A* ficou com a parte geral e a B ficou com a metodológica.

A revista *Transportation Research B* publica artigos sobre os diversos aspectos metodológicos, particularmente aqueles que necessitam de análise matemática. O tema central da revista é o desenvolvimento de soluções de problemas relacionados aos aspectos do design e/ou análise dos sistemas de transporte. Contempla as áreas de fluxo de tráfego; projeto e análise de redes de transporte; controle e programação; otimização; teoria das filas; logística; redes de fornecimento; desenvolvimento e aplicação de modelos estatísticos, econométricos e matemáticos para resolver problemas de transporte; modelos de custos, preços e/ou de investimentos; metodologias de custo-benefício.

Em 1992 a *Transportation Research A* passou a ser denominada de Política e Prática, mantendo seu histórico separado da *Transportation Reserch A* inicial.

No ano de 1993 surgiu a *Transportation Research* parte C, que contempla os temas ligados a Tecnologias Emergentes. Tem como principais focos pesquisas acadêmicas que tratam de desenvolvimento, aplicações e implicações, no domínio dos transportes, das tecnologias de áreas como a ciência da computação, eletrônica, sistemas de controle, inteligência artificial e telecomunicações, dentre outros. O interesse não é nas tecnologias ou metodologias individuais, por si só, mas em suas implicações finais para o planejamento, projeto, operação, controle, gestão, manutenção e reabilitação de sistemas de transporte, serviços e componentes. De particular interesse são os impactos das tecnologias emergentes sobre o desempenho do sistema de transporte, em termos de nível de serviço, capacidade, segurança, confiabilidade, consumo de recursos e do meio ambiente, economia e finanças, privacidade, padrões e responsabilidade.

Já no ano de 1996 foi criada a *Transportation Research* parte D, com o objetivo de apresentar pesquisas nas áreas de Transporte e Meio Ambiente. Seu foco são publicações de pesquisas originais sobre os impactos ambientais do transporte, as respostas políticas para esses impactos e suas implicações para a concepção, planejamento e gestão dos sistemas de transporte. Também aborda os efeitos locais e imediatos das redes de transporte sobre os ambientes de áreas geográficas específicas, para as mais amplas implicações globais do esgotamento dos recursos naturais e da poluição atmosférica.

No ano subsequente, uma nova subdivisão foi lançada, a *Transportation Research* parte E, com o foco na Logística e transporte comentado, que publica artigos informativos vindos de todo o espectro da logística e da pesquisa em transporte. Destacam-se como principais temas abordados a Economia no Transporte, incluindo



funções de custo e produção, capacidade, demanda, preços, externalidades, estudos modais; infraestrutura de transporte e avaliação de investimentos; avaliação de políticas públicas; estudos empíricos sobre práticas e desempenho de gestão; logística, operações e modelos; logística e tópicos de gerenciamento de cadeia de suprimentos.

Em 1998, foi lançada a última subdivisão, F, que aborda publicações sobre Psicologia e Comportamento no Trânsito, ou seja, os aspectos comportamentais e psicológicos do tráfego e transporte. O objetivo da *Transportation Research F* é aprimorar a qualidade dos estudos empíricos e estimular a aplicação dos resultados da investigação na prática. A TR F oferece um fórum para pesquisadores de transporte, psicólogos, ergonomistas, engenheiros e atores políticos que tenham interesse na psicologia do trânsito e do transporte.

Na Figura 3 visualiza-se o gráfico dos totais de publicações de periódicos ao longo dos anos de existência da *Transportation Research*. O gráfico exibido foi gerado a partir do *software* MetrIA, desenvolvido para atender aos estudos apresentados nesta tese.

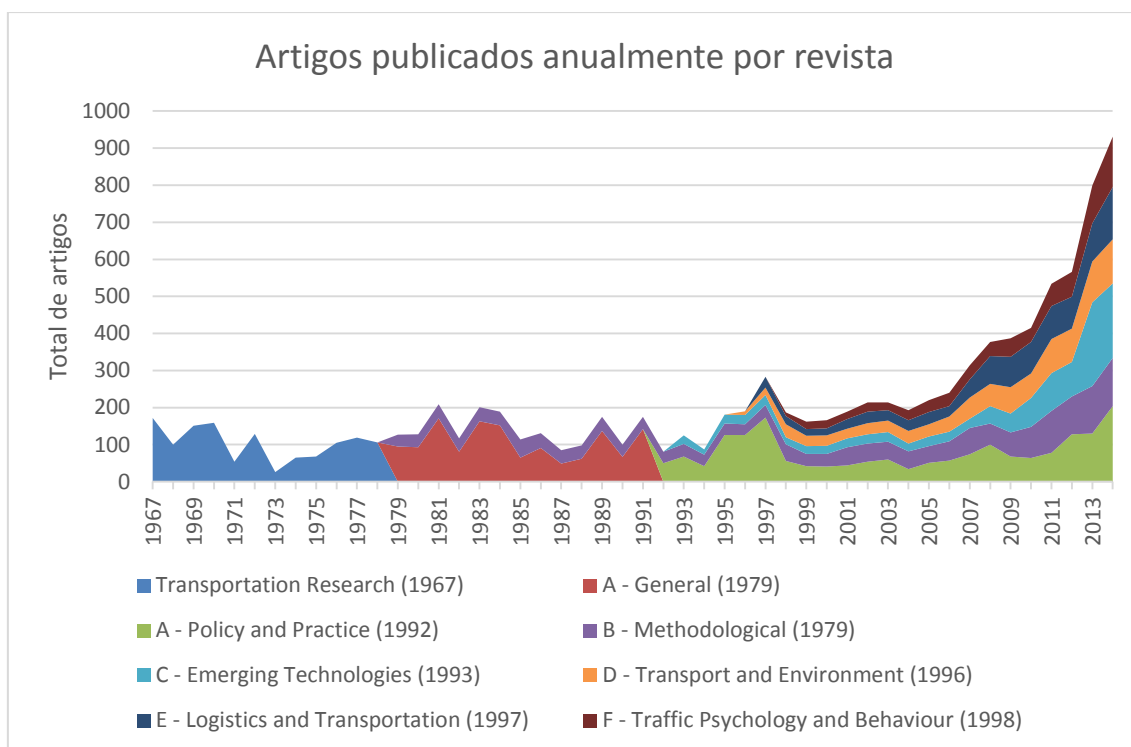


Figura 2 - Totais de publicações por revista ao longo do tempo de existência da revista.

A partir de estudos bibliométricos verifica-se um aumento significativo no número de artigos publicados pela revista *Transportation Research* ao longo dos anos.

Vale ressaltar que a consulta ilimitada realizada no portal *Elsevier* foi possível através de *proxy* da UFRJ, que em parceria com a CAPES disponibiliza para seu corpo discente e docente acesso gratuito, mediante autorização emitida pela biblioteca do campus. Logo, todos os artigos foram baixados de forma oficial.

### 3.3.1. Métricas disponibilizadas no site da *Transportation Research*

Os principais indicadores de qualidade de uma publicação, na perspectiva da revista *Elsevier* e internacionalmente aceitas são SNIP, SJR e IF e IF 5 anos.

O SNIP e o SJR ([www.journalmetrics.com](http://www.journalmetrics.com)) são índices bibliométricos projetados para auxiliar estudantes, pesquisadores, editores e bibliotecários como estão as citações nas áreas de conhecimento, medindo-os e verificando seu fator de impacto e relevância.

O *Source Normalized Impact per Paper* - SNIP (<http://www.journalmetrics.com/snip.php>) foi criado pelo Professor Henk Moed do *Centre for Science and Technology Studies* (CWTS) da Universidade de Leiden, na Holanda. Ele mede o impacto da citação contextual, ponderando citações com base no número total de citações em uma área. O impacto de uma única citação recebe maior valor em áreas em que as citações são menos prováveis, e vice-versa.

O SNIP é definido como a relação entre a contagem da citação de um artigo em uma revista e do potencial de citação em sua área. Destina-se a permitir a comparação direta de fontes, em diferentes campos.

Todos os resultados empíricos são obtidos a partir da base de dados de resumos e indexação *Scopus*. Os valores SNIP são atualizados uma vez por ano, proporcionando uma visão atualizada do panorama da investigação. Ele também pode ajudar os acadêmicos a identificar que revistas estão melhor ranqueadas dentro de sua área.

O SJR (*SCImago Journal Rank*) é um índice bibliométrico elaborado pelo pesquisador Felix de Moya, da Universidade de Granada na Espanha. SJR utiliza algoritmo semelhante ao *Google Pagerank*, atribuindo pesos diferentes às citações dependendo da fonte que citou o documento. Mede o impacto direto da área de pesquisa, da qualidade e da reputação do periódico sobre o valor de uma citação. No site <http://www.scimagojr.com/> podemos obter oito pesquisas referente aos periódicos. Ele utiliza os 3 últimos anos para a análise.

Já o *Impact Factor* – IF, ou fator de impacto FI, reflete o número médio de citações de artigos recentes publicados em uma determinada revista. É frequentemente usado para determinar a importância relativa de uma revista dentro de sua área, onde revistas com maiores fatores de impacto são consideradas mais importantes do que aquelas com menores impactos. O fator de impacto foi idealizado por Eugene Garfield, fundador do *Institute for Scientific Information*. Fatores de impacto são calculados anualmente a partir de 1975 para as revistas que são indexadas no *Journal Citation Reports*, e pertencem hoje a Thomson Reuters.

Na Tabela 3 é possível visualizar os 4 fatores de impacto que a própria *Transportation Research* publica em seu site:

Tabela 3 - Tabela de indicadores disponibilizados no site da revista

<b>Revista</b>	<b>SNIP</b>	<b>SJR 2014</b>	<b>IF</b>	<b>IF 5 anos</b>
Transportation Research A	2.221	2.256	2.789	3.563
Transportation Research B	2.642	2.819	2.952	4.116
Transportation Research C	2.917	2.040	2.818	3.402
Transportation Research D	1.714	1.405	1.937	2.603
Transportation Research E	2.206	2.293	2.676	3.513
Transportation Research F	1.434	0.976	1.473	2.245

O próximo capítulo traz a descrição do *software* MetrlA. O nome escolhido para o *software* é resultante da concatenação do sufixo *metria* encontrado nas três metodologias acima descritas e das iniciais das palavras inteligência artificial (IA).

## CAPÍTULO 4 – O SOFTWARE MetrIA

Este capítulo tem por objetivo detalhar todas as etapas da criação do *software* MetrIA e de seu banco de dados. Como dito anteriormente, o *software* MetrIA tem como função automatizar o estudo cienciométrico, informétrico, bibliométrico de um grande volume de dados, aplicando técnicas de inteligência artificial, processamento de linguagem natural e *data mining*.

Pergunta-se: utilizar um *software* já existente ou desenvolver de um novo *software*? Em uma rápida pesquisa realizada em publicações científicas, não foi encontrado um *software* com todas as funcionalidades esperadas. Com isso, identificou-se um nicho de mercado a ser explorado e foi crucial para confirmar a decisão em desenvolver uma ferramenta própria, visando futuramente atender pesquisadores da UFRJ e de outras instituições interessados nos resultados obtidos com o MetrIA.

A seguir apresenta-se a descrição de todas etapas realizadas para a elaboração e construção do *software* e do banco de dados. A descrição inicia-se desde a forma de obtenção dos mais de 8 mil artigos, quanto à aplicação de técnicas e conceitos apresentados no referencial teórico, tanto para a criação do *data mining* quanto para a criação e desenvolvimento do *software*.

#### 4.1. Etapas de criação do *software* e do banco de dados

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 a ser realizada é o download de cada artigo. Realizar essa tarefa manualmente seria praticamente impossível, logo, para isso 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.

Com a criação de um *software* específico que simula a interação humana com um site, economizou-se milhares de cliques e acessos às páginas da revista analisada. Ainda assim foram necessárias mais de 160h de processamento em um notebook core i5 com 4gb de memória RAM, para que todos os arquivos fossem baixados do site e pré-processados. Foi cogitada a utilização dos supercomputadores da COPPE, mas, acabou-se optando pelo desenvolvimento *home office*, em função da utilização do robô não ser percebida desta forma pelo portal de acesso UFRJ/CAPES, fato que poderia ter inviabilizado a realização dos downloads imprescindíveis para esta tese.

Após a realização de uma análise prévia, descartou-se em torno de 20% do total dos arquivos baixados, pois se tratavam de informativos diversos, como índices, listas de editores e informativos. A partir da obtenção dos arquivos, que se encontram no formato PDF, a segunda etapa foi utilizar um *software* de conversão dos arquivos válidos para o formato TXT, para que fosse iniciada toda a análise de conteúdo para a

criação do banco de dados inteligente. Para realizar essa tarefa foi utilizado o software de código aberto e de linha de comando denominado PDFtoTXT.

Não é possível realizar análises em documentos no formato PDF, pois este formato é essencialmente para leitura e impressão, e não permite, a priori, edição do texto, por exemplo. Já o texto no formato TXT permite acesso ilimitado ao seu conteúdo, permitindo inclusive que o *software* criado acesse seu conteúdo para a criação do banco de dados e *data mining*. A versão PDF dos artigos também foi mantida na base de dados, pois o *software* MetrIA a utiliza para exibição ao usuário, na modalidade leitura e impressão.

Com os textos já no formato TXT, e apresentados de forma não estruturada, inicia-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).

É neste momento que o WordNet participa na criação do *data mining*. Vale lembrar que 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 classificou 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 estabelecendo também relações conceituais-semânticas.

A realização desta análise léxica é de fundamental importância. Na sequência inicia-se o emprego da técnica de *tokenization*, utilizada 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*, como descrito no capítulo 2.

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).

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. 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 3)

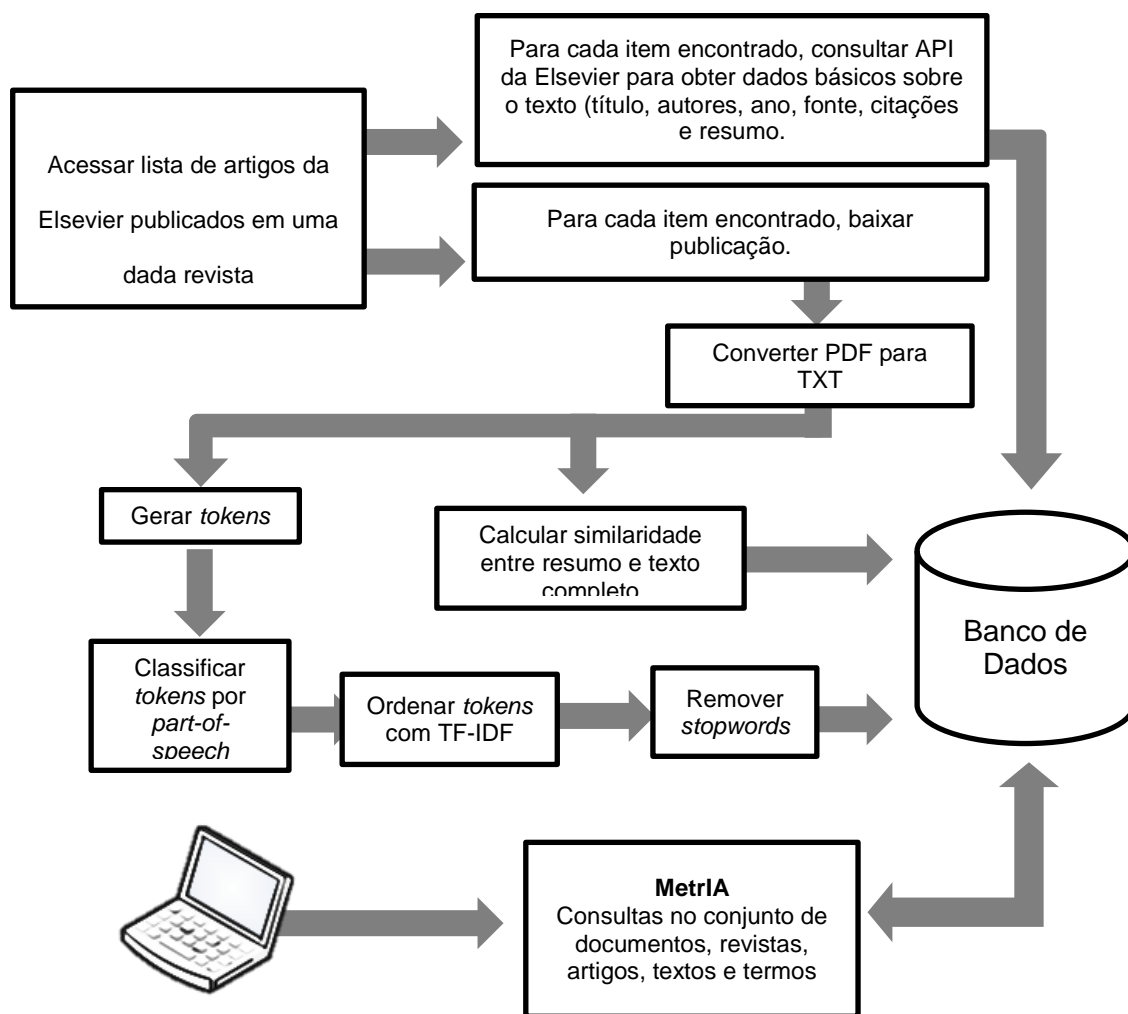


Figura 3 - Fluxograma para obtenção do banco de dados

Com o data mining criado, a etapa seguinte foi a construção 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, por não apresentarem um diferencial como nas etapas anteriormente descritas, não serão detalhadas como as

etapas anteriores, que demandaram técnicas específicas de processamento de linguagem natural, inteligência artificial e algoritmos.

Além dos estudos métricos já anteriormente descritos, visa-se determinar outros dois estudos inéditos. Primeiramente determinar a similaridade interna de um artigo, comparando seu resumo e conteúdo de sua publicação, ou seja, determinar o quanto uma publicação é similar ao seu próprio resumo. A relevância é 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).

A escolha dos resultados exibidos dos estudos cienciométricos, informétricos e bibliométricos realizado pelo MetrIA foram os resultados priorizados para a versão do *software* apresentado nesta tese, sendo inúmeras outras consultas possíveis, que pretendem-se implementar em uma próxima versão do *software*.

## 4.2. Tecnologias empregadas

Diferentes tecnologias foram utilizadas para o desenvolvimento do *software*, de acordo com a especialidade necessária para o melhor resultado e desempenho de cada componente elaborado, conforme descrito a seguir:

- Interface – é o que o usuário vê e que permite sua interação com o *software*: HTML5, *AngularJs* e *JavaScript*.

- Estrutura de processamento de dados – parte do *software* que permite que os dados sejam acessados e manipulados: *Node.js*.

- Banco de dados – estrutura que armazena fisicamente os dados acessados pelo *software*: MySQL.

- Processamento de linguagem natural – validação das palavras relevantes e criação de agrupamentos entre elas: *Wordnet*.

- Grafos – leitura do banco de dados, processamento e exibição dos grafos: Gephi.

Os *softwares* mencionados são de uso livre. Vale destacar que o *software* MetrIA foi desenvolvido para rodar em plataforma Web. Com isso ganha flexibilidade por ser independente do sistema operacional utilizado, pois roda dentro de um browser (aplicativo para acessar páginas de internet).

O próximo capítulo traz o estudo de caso realizado pelo *software* MetrIA com os mais de oito mil artigos da revista *Transportation Research*, apresentando os resultados obtidos pelas análises do *data mining*. O emprego da cienciometria, informetria e bibliometria é apresentado nos gráficos, grafos e tabelas, com o objetivo principal de verificar se o estudo de uma série histórica de publicações permite antever pesquisas futuras.

## CAPÍTULO 5 – ANÁLISE DAS PESQUISAS EM TRANSPORTES

O estudo a ser apresentado trata de uma análise cientométrica, informétrica e bibliométrica realizada com mais de oito mil publicações da revista *Transportation Research*, abrangendo seus 48 anos de existência.

As principais funcionalidades desenvolvidas para o *software* MetrIA serão apresentadas. Vale salientar que o *software* apresenta grande facilidade para a criação de novas consultas e relacionamentos de dados, bem como de extração de outras informações do *data mining*.

A utilização da tecnologia da informação nos transportes propicia criar simuladores, controladores de tráfego, *softwares* logísticos e toda diversidade de ferramentas para dar suporte a essa engenharia.

Este capítulo irá apresentar uma nova aplicação da TI à engenharia de transportes. Serão apresentadas as funcionalidades práticas do *software* MetrIA, criado especificamente para a presente tese, apresentando os diversos resultados obtidos através de tabelas, gráficos e grafos, e ao final, apresenta-se uma breve conclusão da análise do estudo de caso realizado.

### 5.1. Análise por Publicações

A primeira análise disponibilizada é por publicações. A partir das escolhas dos filtros *publicação* e *ano*, é possível visualizar, como na figura a seguir, todos os artigos,

classificados em ordem crescente de ano de publicação. A coluna similaridade apresentada é uma variável criada pelo MetrIA, 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.

O título da publicação, como pode ser observado na Figura 4, é um hiperlink que possibilita acessar em uma nova tela, o texto original e outros estudos específicos acerca do artigo selecionado.

Ano	Similaridade	Título
1967	0.49	<a href="#">Systems design of electric automobiles</a>
1967	0.54	<a href="#">Car following models and the fundamental diagram of road traffic</a>
1967	0.45	<a href="#">Models for traffic assignment</a>
1967	0.56	<a href="#">Toward a better understanding of the intervening opportunities model</a>
1967	0.62	<a href="#">A simple renewal model of throughput at an oversaturated signalized intersection</a>
1967	0.60	<a href="#">Fourier analysis of the Weiss-Herman traffic equation</a>
1967	0.65	<a href="#">Simulation of two-way traffic on an isolated two-lane road</a>
1967	0.59	<a href="#">Queues with state-dependent service times</a>
1967	0.47	<a href="#">Road traffic counting distributions</a>
1967	0.53	<a href="#">A deterministic queueing model</a>

Figura 4 - Tela de listagem dos artigos classificados em ordem crescente de ano.



A nova tela que pode ser acessada ao selecionar o título de uma publicação exibe cinco tabelas distintas, conforme descrito a seguir. Na primeira tabela visualiza-se as principais informações da publicação como: título; ano de publicação; link para acesso direto ao site da *Science Direct* (Elsevier); o nome de autores no formato hiperlink (que permite o acesso a uma nova tela - a página do autor); palavras-chave; a revista no qual foi publicado o artigo no formato hiperlink (permite abrir uma nova tela com todos os periódicos da referida revista selecionada); e finalmente, o percentual de similaridade resumo/texto, conforme detalhado anteriormente. A segunda tabela exibe a publicação no formato original disponibilizado no site da Elsevier, onde pode-se ler o arquivo no formato PDF, realizar o download ou impressão do mesmo.

A terceira tabela apresenta os dez artigos mais similares ao artigo em foco, do ponto de vista de seu conteúdo, indicando referências relevantes para futuras consultas. O *software* faz o estudo comparativo de similaridade sem se limitar à revista do periódico selecionado. A decisão pela análise inter-revistas é identificar se a temática do artigo corresponde à revista que o publicou. Por este motivo decidiu-se também exibir o resultado do cálculo percentual de quantos artigos pertencem à mesma revista do artigo principal. A similaridade é calculada pelo algoritmo *TF-IDF* a partir dos *tokens* das publicações. Exibe-se ao lado do nome dos artigos similares a revista na qual o artigo foi publicado (Figura 5).

Informações do Artigo	
Título	Research trends in urban transport science: Some empirical evidence from academic research
Ano	1988
Link	<a href="http://www.sciencedirect.com/science/article/pii/0191260788900635">http://www.sciencedirect.com/science/article/pii/0191260788900635</a>
Autor(es)	Santos, M.P.S.; de C. Braga, M.G.
Palavras-chave	
Publicação	Transportation Research Part A: General
Similaridade resumo/texto	40%

Artigos Similares (6 na mesma revista)		
Similaridade	Título	Revista
30%	Consumer mobility in the Scottish Isles: The impact of internet adoption upon retail travel patterns	Transportation Research Part A: Policy

Figura 5 – Tabelas 1, 2 e 3 da pesquisa por publicações.

Já a quarta tabela exhibe os 10 termos de maior relevância no *abstract* e os 10 termos de maior relevância no texto completo, exibindo os respectivos TF-IDF. O TF-IDF do texto completo utilizado para calcular e localizar os 10 artigos mais similares, como citado anteriormente, e explicitado anteriormente na Figura 5. Finalmente a quinta tabela exhibe as citações do artigo selecionado e o total, conforme visualizado abaixo (Figura 6).

tf-idf		Citações (11)	
<b>Abstract</b>		<b>Texto Completo</b>	
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 6 –Termos mais frequentes no resumo e corpo do texto (tabela 4), e na tabela 5 apresenta-se as citações.

Destaca-se, nas consultas por publicação, a versatilidade de busca de publicações por ano, por revista, e, ao localizar uma publicação de interesse, ao clicar em seu título, é possível ter acesso completo ao seu conteúdo, autores, palavras-chave, termos relevantes, similaridade e referências bibliográficas. Também é possível acessar outros 10 artigos semelhantes, através do cálculo de similaridade entre textos. A consulta também permite acessar diretamente todas as publicações de um determinado autor da publicação selecionada.

O objetivo da comparação entre resumo e artigo é a verificar uma questão: é possível determinar o conteúdo de um artigo lendo apenas o seu resumo (abstract)?

## 5.2. Análise por Autores

A segunda opção de análise criada é por autores, onde visualiza-se os nomes dos autores no formato hiperlink, permitindo assim acessar em uma nova janela a página do autor, como veremos mais adiante na Figura 15. Os autores aparecem nesta tela classificados em ordem decrescente de número de publicações (igura 7). Vale destacar que, em todos os mapeamentos, há opções de filtros por ano e por revista.

O estudo bibliométrico de autores apresentado a seguir ilustra os pesquisadores que mais publicaram na revista *Transportation Research*, onde os filtros aplicados foram: todas as revistas e em todos os anos. Em virtude do grande número de pesquisadores que publicaram na revista, 12.716 ao todo, dar-se-á na sequência destaque aos 10 autores que mais publicaram na revista (Figura 7).

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

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

Figura 7 – Autores que mais publicaram na Revista *Transportation Research*.

Através desta análise bibliométrica conclui-se que o pesquisador com o maior número de publicações é Hai Yang, professor titular do departamento de Engenharia Civil na Universidade de Hong Kong, com 98 artigos publicados. Realizando uma análise em suas publicações nota-se que abrangem diversas áreas e aplicações da Engenharia de Transporte, não havendo, portanto, um padrão na temática de seus textos. O demais pesquisadores elencados entre os 10 maiores escritores apresentam também perfil diversificado. Por serem pesquisadores-professores, e muitos ainda acumulam a função de coordenação, supõe-se que certamente têm a oportunidade de participar de diferentes pesquisas, abrindo assim o horizonte temático.

O pesquisador que possui o segundo maior número de publicações na revista *Transportation Research* é David A. Hensher, com 97 artigos. David é Professor de Gestão e Diretor do Instituto de Estudos de Transporte e Logística na Universidade de Sydney, na Austrália. O terceiro pesquisador em destaque é Sveinn Gudmundsson com 87 publicações. Ele é professor de Gerenciamento Estratégico no *Toulouse Business School*, na França.

Carlos F. Daganzo com 78 publicações é o quarto pesquisador da lista. Ele é professor do departamento de Engenharia de Transportes na Universidade de Berkeley na Califórnia. Já o Dr. Chandra R. Bhat é diretor e professor no Centro de Pesquisa em Transporte na Universidade do Texas em Austin. É o quinto da lista com 70 publicações.

Diretor e Professor do Instituto de Transportes e Tráfego da Universidade Nacional Chiao Tung de Taiwan, Jiu-Bing Sheu publicou 65 artigos na referida revista, sendo o sexto no ranking, sendo seguido pelo professor do departamento de Engenharia

de Transportes Michel J. Cassidy da Universidade da Califórnia, em Berkeley. Assim como Jiu-Biing Sheu publicou 65 artigos na *Transportation Reserch*.

O Diretor e Professor do Departamento de Engenharia Civil e de Transportes William Lam Hing-keung, da Universidade Politécnica de Hong Kong aparece em oitavo lugar com 64 publicações. O professor Dr. Qiang Meng aparece em nono lugar com 53 publicações. Ele é professor no Departamento de Engenharia Ambiental e Civil da Universidade Nacional de Singapura. E concluindo a lista dos 10 principais publicadores, o Professor da área de Engenharia de Transportes S.C. Wong da Universidade de Hong Kong, na China, per fez um total de 48 publicações.

Para sintetizar essas informações, realizou-se uma análise informétrica dos 10 autores com o maior número de publicações, agrupando-os por país, e também apresentando o percentual de publicação dentre os 10 mais autores, conforme apresentado na Tabela 4:

Tabela 4 - Universidades e autores mais ativos e suas publicações (dentre os 10+).

<b>Universidade/País</b>	<b>Total de autores</b>	<b>%</b>	<b>Publicações</b>	<b>%</b>
Hong Kong/China	3	30	210	28,97%
Berkeley/EUA	2	20	143	19,72%
Sidney/Austrália	1	10	97	13,38%
Toulouse/França	1	10	87	12,00%
Texas/EUA	1	10	70	9,66%
Taiwan/Taiwan	1	10	65	8,97%
Singapura/Singapura	1	10	53	7,31%
<b>TOTAL</b>	<b>10</b>	<b>100</b>	<b>725</b>	<b>100</b>

Ao analisar a tabela acima verifica-se que metade dos autores são de origem asiática, e foram responsáveis pela publicação de mais de 45% dos artigos dos 10 principais autores. Dos 725 artigos totais publicados pelos 10 pesquisadores, 28,97% foram publicados pela Universidade de Hong Kong.

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. Outro dado de destaque é que dos 10 pesquisadores, 4 acumulam a função de professor e diretor do departamento que lecionam.

A grande variedade de pesquisadores que publicaram na revista, 12.716 autores distintos, ilustra o quão enriquecedor é a diversidade de universidades e países que compartilham seu conhecimento através da *Transportation Research*, o que, sem dúvida também fundamenta o destaque e o foco desta revista como um todo.

#### **5.2.1. Seleção de um único autor**

Ao selecionar um dos autores conforme apresentado na Figura 7, uma nova tela é exibida. Nesta nova tela é possível visualizar todas as publicações do autor selecionado, sendo que os títulos das publicações são apresentados no formato de hiperlink (permitindo o acesso direto), e apresentam-se listados em ordem crescente de ano de publicação. O índice de similaridade de cada artigo precede o nome das publicações, conforme visualiza-se na Figura 8.

Artigos de sua autoria		
Ano	Similaridade	Título
1997	0.38	<a href="#">Substitution and complementarity in aviation: airports vs airlines</a>
2002	0.46	<a href="#">How do carriers price connecting flights? Evidence from intercontinental flights from Europe</a>
2003	0.43	<a href="#">Inefficiencies and scale economies of European airport operations</a>
2009	0.45	<a href="#">Modelling the joint access mode and railway station choice</a>
1996	0.48	<a href="#">Speed behaviour of car drivers: A statistical analysis of acceptance of changes in speed policies in the Netherlands</a>
1999	0.44	<a href="#">The social support for policy measures in passenger transport. A statistical analysis for the Netherlands</a>
2000	0.47	<a href="#">Non-motorised modes in transport systems: a multimodal chain perspective for The Netherlands</a>
2000	0.45	<a href="#">The accessibility of railway stations: the role of the bicycle in The Netherlands</a>
2002	0.43	<a href="#">Environmental effects of a kilometre charge in road transport: an investigation for the Netherlands</a>
2002	0.54	<a href="#">Why railway passengers are more polluting in the peak than in the off-peak: environmental effects of capacity management by railway</a>

Figura 8 – Publicações do autor selecionado, como títulos no formato de hiperlink.

Ao final da listagem de publicações é apresentado o total de artigos do autor, e a média de similaridade de seus artigos/resumos. Na sequência, exibe-se dois gráficos. O primeiro apresenta o número de publicações do autor por ano e o segundo gráfico apresenta as vinte palavras mais significativas das publicações do referido autor, totalizando o número de ocorrências, como apresentado na Figura 9.



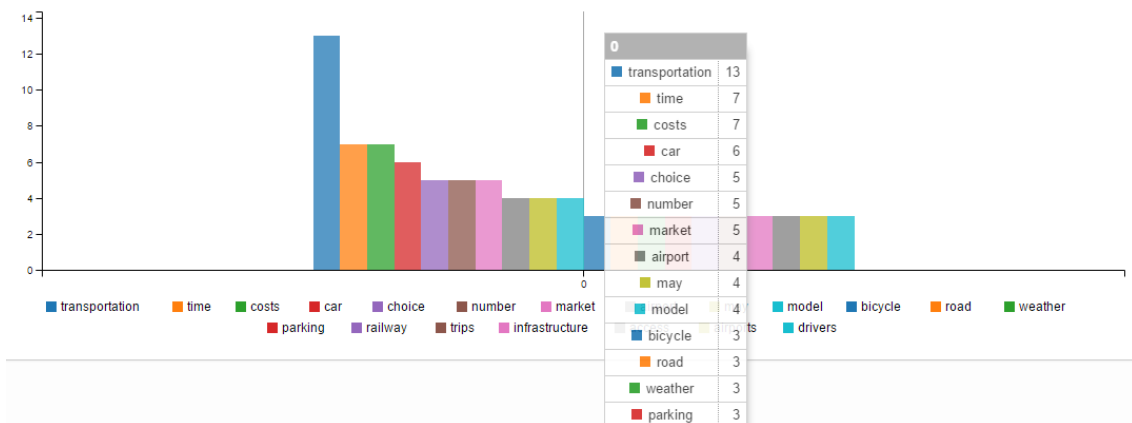


Figura 9 –Vinte termos de maior relevância nas publicações do autor selecionado.

A vantagem desta consulta é ter acesso a todas as publicações de um determinado autor (estudo bibliométrico), contabilizá-las, identificar seu ano de publicação, além de conhecer quais são os termos mais relevantes de suas publicações, verificando assim as temáticas de interesse do referido pesquisador (estudo informétrico).

O objetivo de determinar quais são os termos mais relevantes de um autor é poder identificar outro(s) autor(es) que publicam temáticas semelhantes. Essa informação é importante para o leitor conhecer outros autores que publicam dentro da mesma temática, identificando outros artigos que possam ser de interesse, e para o pesquisador essa informação é importante para poder estabelecer trabalhos de cooperação e parcerias com autores afins de suas áreas de publicação.

### 5.3. Análise por Termos

O MetrIA permite a análise por termo, sendo a consulta de um único termo ou a consulta de todo o conjunto de termos catalogados pelo *software*. O último tipo de pesquisa mencionado exibe todos os termos classificados em ordem decrescente de número de ocorrências, possibilitando ainda o uso dos filtros por revista e por ano. O critério utilizado para o estudo dos termos é a informetria.

A seguir apresenta-se três consultas de termos individualizados realizadas dentro do *software* MetrIA, que exemplificam as informações que podem ser obtidas nesta opção de consulta. As referidas consultas utilizaram como parâmetro de filtros todas as revistas e todos os anos.

#### 5.3.1. Termo *bicycle*

A consulta exibida na Figura 10 é referente ao termo ***bicycle***. O termo ocorre de forma significativa em 45 artigos, e o *software* MetrIA exibe-os listados, em tela onde os títulos são apresentados no formato hiperlink, permitindo consulta direta. Verifica-se que a primeira ocorrência relevante do termo ***bicycle*** foi no ano de 1976, e a seguir nos anos de 1981 e 1984. A partir do ano de 1999 seu número de ocorrências aumentou significativamente (Figura 10).

De fato, nos últimos anos vem ocorrendo a valorização e o incentivo ao uso de meios de transporte não poluentes, indicando a sintonia dos pesquisadores à temática contemporânea e às políticas públicas que vêm sendo implantadas, principalmente nos países desenvolvidos.

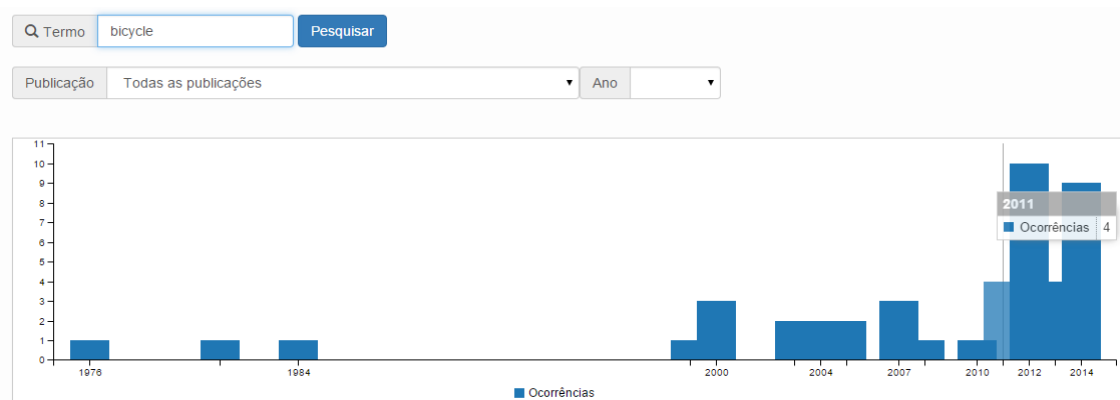


Figura 10 – Consulta por termo da palavra *bicycle*.

### 5.3.2. Termo *environment*

A segunda consulta realizada foi com o termo ***environment***. A consulta identificou, através dos princípios da informetria, 38 artigos, conforme gráfico exibido na Figura 11. Verifica-se que sua primeira ocorrência relevante foi no ano de 1968, e houve uma segunda ocorrência em 1976. A partir de 1994 as ocorrências foram quase contínuas, e, a partir de 2003, sua ocorrência passou a ser anual, aumentando expressivamente.

O crescimento do número de artigos referente ao termo *environment* retrata o cenário mais recente onde as políticas públicas locais e internacionais relacionadas ao tema, incluindo diversos encontros e tratados, como por exemplo a ECO-92, Protocolo de Kyoto (97), Rio+10 (2002). A Figura 11 ilustra o significativo aumento da ocorrência deste termo.

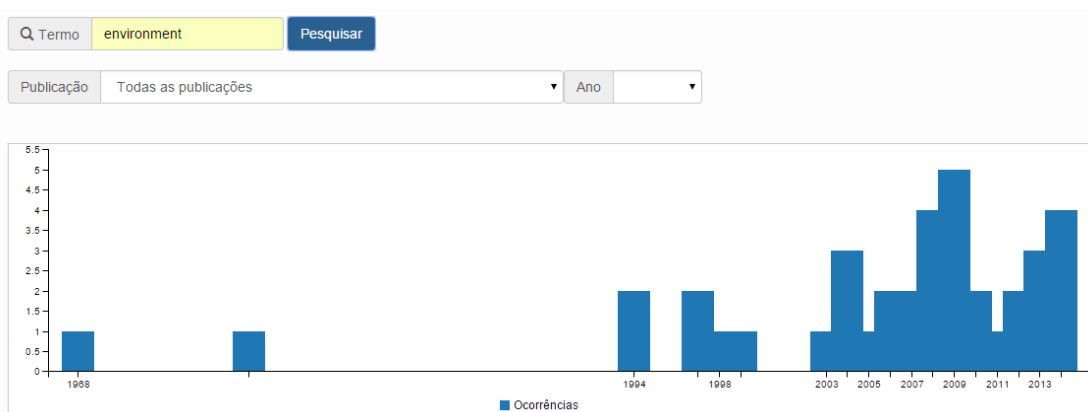


Figura 11 – Consulta por termo da palavra *environment*.

### 5.3.3. Termo *emission*

A palavra *emission* apresenta 127 ocorrências significativas, sendo sua primeira ocorrência bem mais recente que as duas primeiras apresentadas foi em 1992, sendo constantes suas publicações desde o ano de 1997. 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 (97), Rio+10 (2002), bem como políticas internas, empresas-cidadãs e ambientalistas buscam soluções para minimizar e reduzir os impactos da poluição, e fica evidente que as publicações acompanham essa tendência, conforme Figura 12.

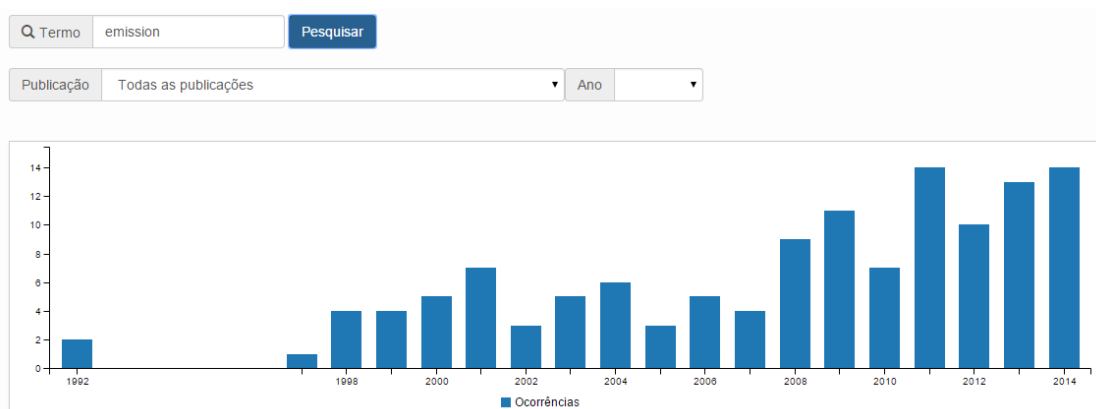


Figura 82 – Consulta por termo da palavra *emission*.

Os exemplos escolhidos apresentam a ocorrência relevante dos termos selecionados ao longo do tempo. É possível a análise da variação de ocorrência dos termos, tanto por relevância quanto pelo número de ocorrências (frequência), como será exibido a seguir. 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.

#### 5.3.4. Pesquisa por termos mais frequentes

Com base na cienciometria e informetria, elaborou-se a pesquisa por termos mais frequentes. A Tabela 5 a seguir exibe os 15 termos que ocorreram com maior frequência em todas as revistas ao longo de todos os anos. Há a possibilidade de realizar a mesma consulta selecionando apenas uma única publicação e/ou por ano selecionado. Os termos encontram-se no formato de hiperlink, e, ao selecionar uma palavra, uma nova tela exibe a listagem com os artigos que contém a palavra selecionada.

Os dois termos de maior frequência, *transportation* e *research*, são a temática e o título da revista, por isso justifica-se com clareza a maior frequência. Verifica-se, consultando os gráficos de cada uma das palavras, que o número de suas ocorrências acompanha o número de publicações ao longo dos anos. A variação crescente justifica-se pelo aumento gradual de publicações ao longo dos 48 anos analisados, quer seja pela criação das novas subdivisões da revista, quer seja pelo aumento no número de aceite de artigos pelas revistas ao longo dos anos.

Na Tabela 5 possibilita selecionar uma palavra-chave e a partir desta seleção, uma outra tela é exibida, que apresenta a listagem, também selecionável, com todos os artigos onde o termo escolhido ocorre.

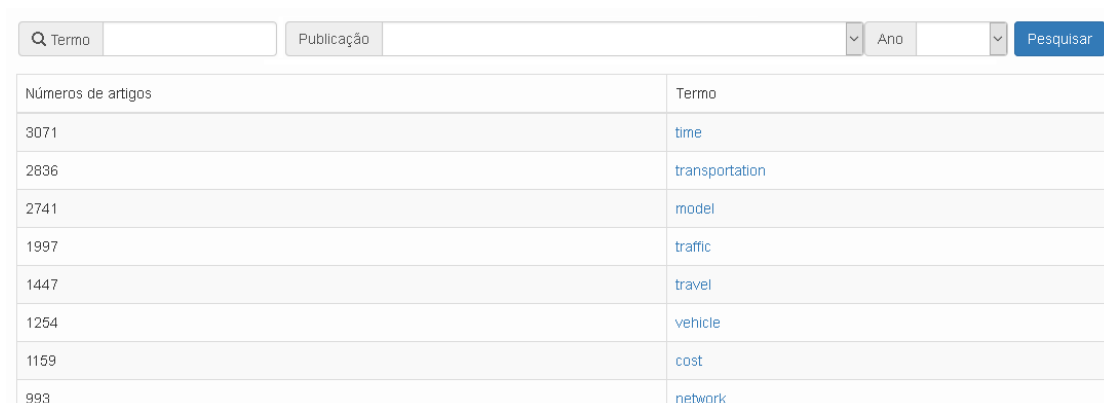
Tabela 5 – Listagem dos termos de maior frequência.

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

### 5.3.5. Pesquisa por termos de maior relevância

A pesquisa por termos mais relevantes, diferentemente dos termos de maior frequência, apresenta os *tokens* que são relevantes na publicação, ou seja, os que fazem parte da temática principal da publicação. Assim como na listagem de termos de maior frequência, visualiza-se os *tokens* em ordem decrescente de relevância.

A Figura 13 exibe as palavras de maior relevância de todas as revistas ao longo de todos os anos. Há a possibilidade de realizar a mesma consulta filtrando por uma única publicação apenas, e filtrando pelo ano selecionado.



The screenshot shows a search interface with a search bar containing 'Termo', a filter for 'Publicação', and a filter for 'Ano'. A blue 'Pesquisar' button is on the right. Below the search bar is a table with two columns: 'Números de artigos' and 'Termo'. The table lists the following terms and their corresponding article counts:

Números de artigos	Termo
3071	time
2836	transportation
2741	model
1997	traffic
1447	travel
1254	vehicle
1159	cost
993	network

Figura 13 - Listagem de termos de maior relevância.

Verifica-se que os totais são menores na consulta de relevância do que na consulta de termos de maior frequência, pois a segunda consulta apresenta apenas os artigos onde de fato o termo seja relevante em comparação ao conteúdo do artigo.

Comparando-se as consultas dos termos mais relevantes com os termos de maior frequência, observa-se que a consulta de maior frequência não nos traz resultados muito significativos quanto aos artigos analisados.

Na consulta por relevância vemos que ao longo dos 48 anos analisados da revista *Transportation Research* que o principal tema abordado foi o tempo. A seguir temos como mais relevantes os termos: modelo, viagem, tráfego, veículo, custo e redes. Em uma análise preliminar pode-se afirmar que os modelos são muito importantes para se planejar os transportes, pois permite realizar testes para a decisão de se implantar



ou não uma determinada solução. Os termos viagem, tráfego, custo, veículo e redes abrangem tanto o transporte de bens e passageiros, bem como traz como informação que os artigos em sua maioria apresentam a temática mais prática e objetiva, atendendo à temática do transporte em si, não estando associados a pesquisa, análises, políticas públicas e conjecturas.

Como os termos analisados pertencem ao conjunto de todos os artigos, de todo o período temporal, realizou-se um estudo mais aprofundado, em intervalos de cinco anos, através da utilização de grafos, conforme apresentados a seguir, com o objetivo de verificar se houveram variações temáticas ao longo dos anos, ou pode-se considerar que as pesquisas realizadas por frequência e relevância agrupadas retratam todas as janelas temporais.

O MetrIA além de realizar os estudos propostos também se apresenta como um facilitador de leitura para o pesquisador/usuário. Apresenta uma navegação dinâmica, permitindo acessar artigos, autores e consultas através de diferentes modos, facilita o leitor localizar publicações de seu interesse, reduzindo seu tempo de procura e acesso, como visto acima.

#### **5.4. Grafos – termos de maior relevância**

A base de dados do MetrIA foi organizada de modo a permitir a exportação dos dados, o que possibilitou a construção de grafos através do *software* Gephi (BASTIAN, et al., 2009). A primeira versão do MetrIA ainda não possui seu modo próprio de geração

de grafos e o *software* Gephi permite a construção de grafos direcionados e não-direcionados, que atendeu às análises desejadas.

Os dados processados pelo MetrIA 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. Por exemplo, se entre as principais palavras de um texto estão os termos *road*, *bus* e *time*, estes três nós possuem ligações entre si. 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.

Neste estudo a distribuição dos nós no grafo foi realizada com o algoritmo Force Atlas 2, implementado pelo *software* Gephi. 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. De posse destas informações apresenta-se a seguir os seguintes grafos, agrupando as revistas, nos referidos períodos visando uma análise da abordagem temática ao longo do tempo.

Períodos analisados:

- 1967-1969;
- 1970-1974;
- 1975-1979;
- 1980-1984;

- 1985-1989;
- 1990-1994;
- 1995-2000;
- 2001-2004;
- 2005-2009;
- 2010-2014.

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. Os grafos apresentados a seguir apresentam agrupamento de termos por cor, agrupamentos esses criados pelo *software* Gephi, com base nas interconexões dos termos, como anteriormente descrito.

#### **5.5.1. Primeiro grafo - intervalo – 1967-1969**

O primeiro grafo retrata o período de 1967-1969 (inclusive), conforme ilustra a Figura 14. Este grafo abrange um intervalo de tempo inferior a 5 anos, o que não acontece nos demais. Foi necessária a criação de um grupo menor em função do espaço temporal ser de 48 anos.

Neste grafo é possível identificar cinco grupos, sendo o rosa e o laranja os de maior significância. O grupo de cor rosa possui como principais termos *performance*, *participants*, *speed*, *study*, *task*, *safety* e *data*, que representa as abordagens teóricas e estudos com grupos controlados.

O grupo de cor laranja possui como destaque as palavras *road*, *risk*, *traffic*, *model*, *design*, *simulation* e *urban*, o que mostra a realização de trabalhos envolvendo o desenvolvimento de modelos de simulação de tráfego urbano.

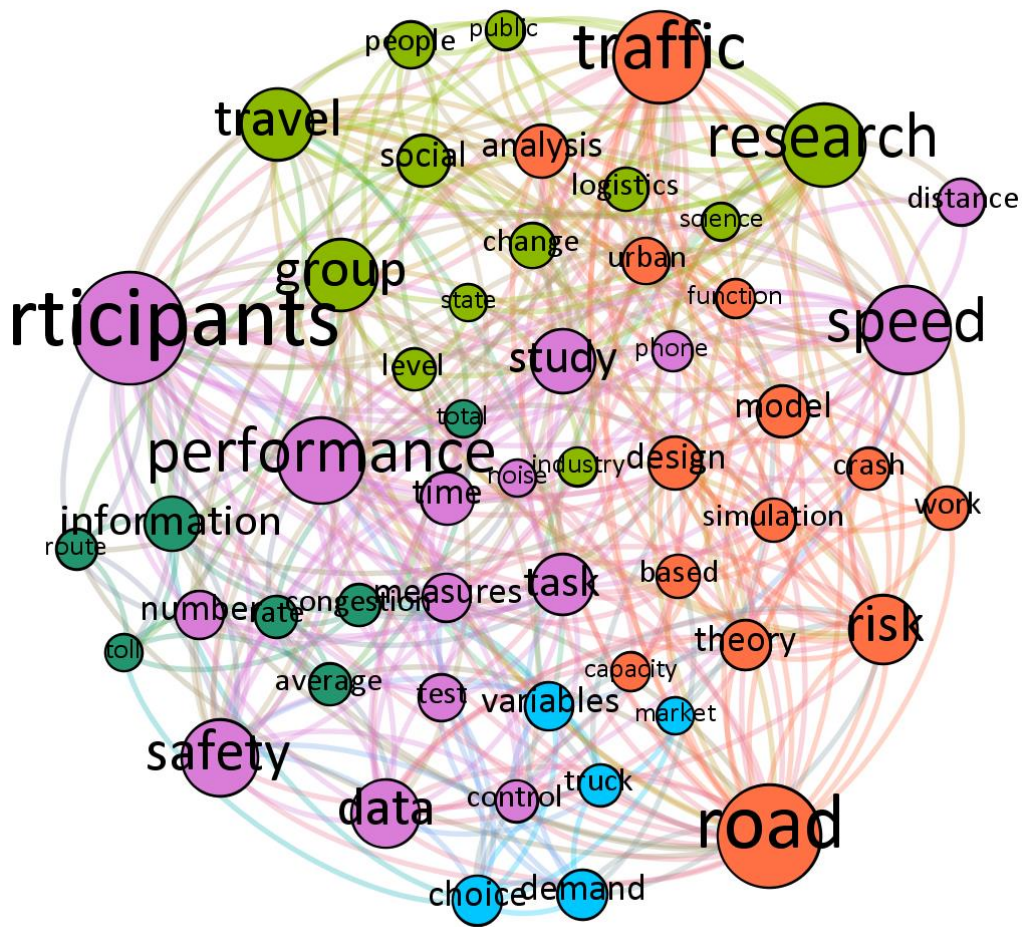


Figura 14 - Grafo relativo ao período de 1967-1969 (inclusive).

Vale ressaltar que as cores se referem aos grupos criados pelo algoritmo em função da frequência apresentada nas conexões e entre os termos. Interessante ressaltar a coerência dos termos apresentados nos distintos grupos de cores criados.

### 5.5.2. Segundo grafo - intervalo – 1970 a 1974

O segundo grafo que compreende o período de 1970 a 1974, inclusive, apresentado na Figura 15, onde verifica-se que um maior número de palavras compartilha os maiores destaques. Neste grafo a cor laranja mostra um conjunto de termos relacionados à construção de modelos de logística e manutenção de cadeia de suprimentos: *demand, time, chain, traffic, model, logistics* e *management*.

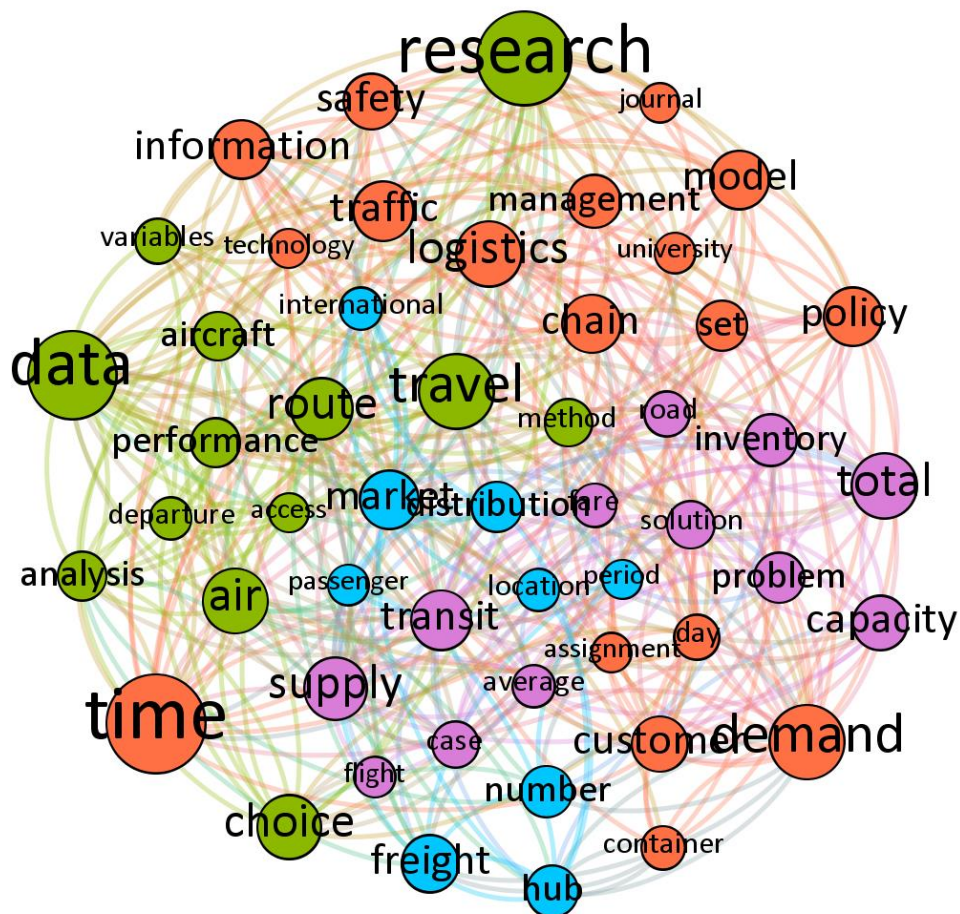


Figura 15 - Grafo relativo ao período de 1970-1974 (inclusive).

O grupo rosa também traz termos relacionados a esta área, mas demonstra uma abordagem mais prática: *capacity, inventory, average, solution* e *problem*. Vale destacar que houve um aumento no número de termos relevantes, ou seja, houve um aumento na diversidade de temas tratados pelos artigos. O grupo azul apresentou um aumento no número de termos significativos, incluindo os termos *passenger, period, location* e *distribution*, o que nos leva a concluir a valorização do transporte de pessoas, demanda das localidades e frequência, além de sua distribuição e alcance físico.

### **5.5.3. Terceiro grafo - intervalo – 1975 a 1979**

O terceiro grafo que compreende o período de 1975 a 1979, inclusive, e está representado na Figura 16. Neste grafo foram identificados pelo Gelphi três grandes grupos, sendo o laranja associado à área de logística de cadeia de suprimentos, o azul com foco no operacional.

A cor verde apresenta os nós que demonstram a utilização comercial de transporte de passageiros: *demand, capacity, route, congestion, passengers, public* e *customer*, ou seja, a preocupação com a efetividade de transporte coletivo começa a ser fortemente abordada nos artigos.

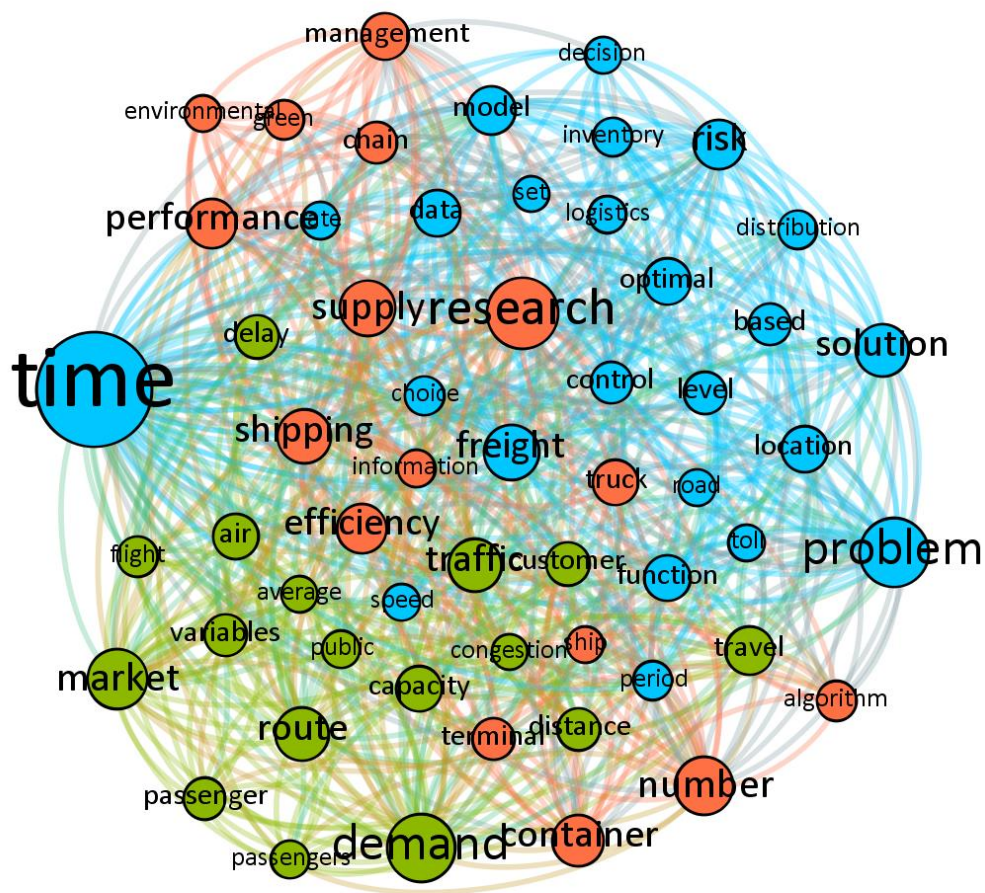


Figura 16 - Grafo relativo ao período de 1975-1979 (inclusive).

#### 5.5.4. Quarto grafo - intervalo – 1980 a 1984

O quarto grafo que compreende o período de 1980 a 1984, inclusive, e está representado na figura 17. Neste grafo é possível observar um aumento na diversidade de temas abordados, totalizando 6 principais grupos.



O grupo laranja mostra pesquisas nas áreas de políticas públicas e planejamento de transporte, apresentando termos como *highway*, *planning*, *simulation*, *parameters*, *book* e *policy*.

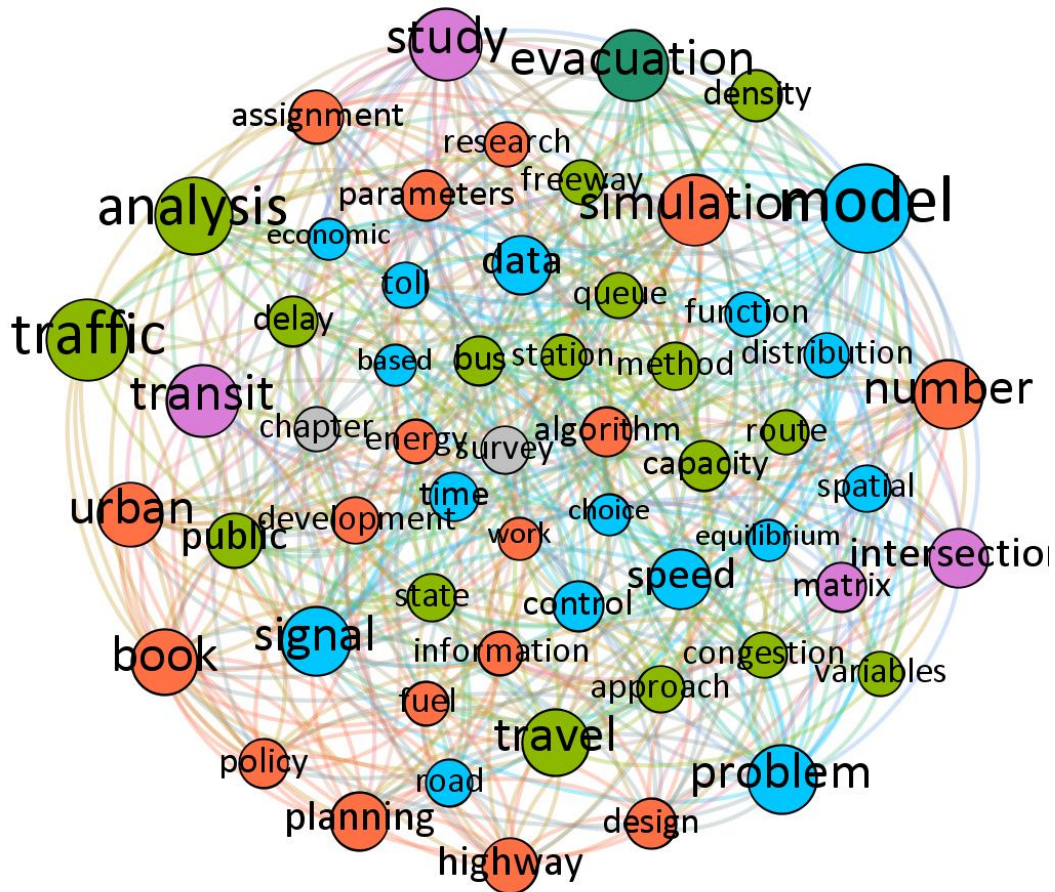


Figura 17 - Grafo relativo ao período de 1980-1984 (inclusive).

Modelos teóricos podem ser vistos na cor azul clara, com *model*, *function*, *problem*, *time*, *speed*, *choice*, *data* e *spatial*, e a cor verde claro um importante grupo

com foco no transporte público, verificado nos termos *bus*, *station*, *travel*, *traffic*, *public*, *congestion*, *route*, *capacity*, dentre outras de destaque.

### 5.5.5. Quinto grafo - intervalo – 1985 a 1989

O quinto grafo que compreende o período de 1985 a 1989, inclusive, e está representado na Figura 18. Neste grafo se destacam cinco grandes grupos, onde o grupo rosa apresenta como destaque os termos *urban*, *public*, *road*, *work*, *department* e *safety*, denotando trabalhos sobre políticas públicas de transporte.

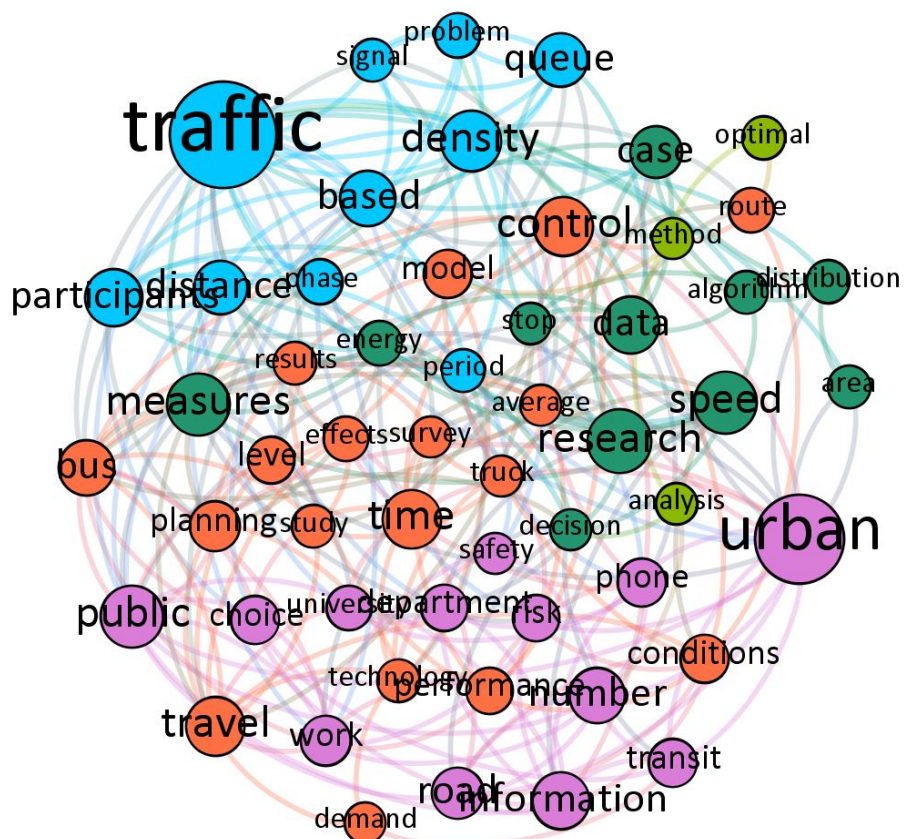


Figura 18 - Grafo relativo ao período de 1985-1989 (inclusive).

Já o agrupamento de cor azul clara apresenta termos relacionados ao estudo de modelos de fluxo de veículos, com termos como *traffic, density, queue, distance* e *problem*. O agrupamento laranja aborda *time, control, planning, technology*, ou seja, este grupo apresenta o planejamento como característica mais determinante.

#### **5.5.6. Sexto grafo - intervalo – 1990 a 1994**

O sexto grafo que compreende o período de 1990 a 1994, inclusive, e está representado na figura 19. Neste grafo se observa uma distribuição ampla na frequência dos termos e seis comunidades com divisão pouco precisa.

É possível identificar uma concentração no verde mostarda com *speed, traffic, level, urban* e *data*. No verde escuro se observa os termos *information, variables, zone* e *passanger*. Este grupo se destaca por estar permeado nos demais grupos, sendo o grupo dominante nesta consulta. É possível interpretar nestes grupos a presença de modelos computadorizados de análise, em função da difusão dos microcomputadores e redução de custo dos equipamentos.





Outro aspecto relevante a ser observado no sétimo grafo é a proporção do tamanho das bolas coloridas. Isso reflete a divisão do peso dos termos, ou seja, aumentou a diversidade de temas abordados, e esse aumento foi de forma homogênea, ou seja, não houve um grande destaque para um determinado tema. As publicações neste período foram muito equilibradas, avaliando-as de acordo com a pertinência temática dos mesmos.

#### **5.5.8. Oitavo grafo - intervalo – 2000 a 2004**

O oitavo grafo que compreende o período de 2000 a 2004, inclusive, e está representado na figura 21. Neste grafo também se identifica uma pulverização nos temas tratados e dificuldade em se estabelecer grupos claros. As cores se misturam, ou seja, cada publicação abrange diversos temas relevantes, não ocorrendo uma separação física das cores, ou seja, dos grupos. As publicações são multitemáticas. Esta tendência ilustra claramente a ampliação do escopo dos artigos.

Os grupos destacam abordagem sobre questões de eficiência energética, como CO<sub>2</sub>, *transit*, *energy*, *efficiency* e *rate*. O quesito meio ambiente também tem presença significativa, com termos como *noise*, *environmental*, *urban*, *air* e *effects*.



ocorrência do termo *energy*, e sua interação com os diferentes grupos. De um lado, o termo está relacionado com questões de planejamento urbano, presentes nos grupos laranja e rosa, e também com o lado ambiental, tendo relação com *noise*.

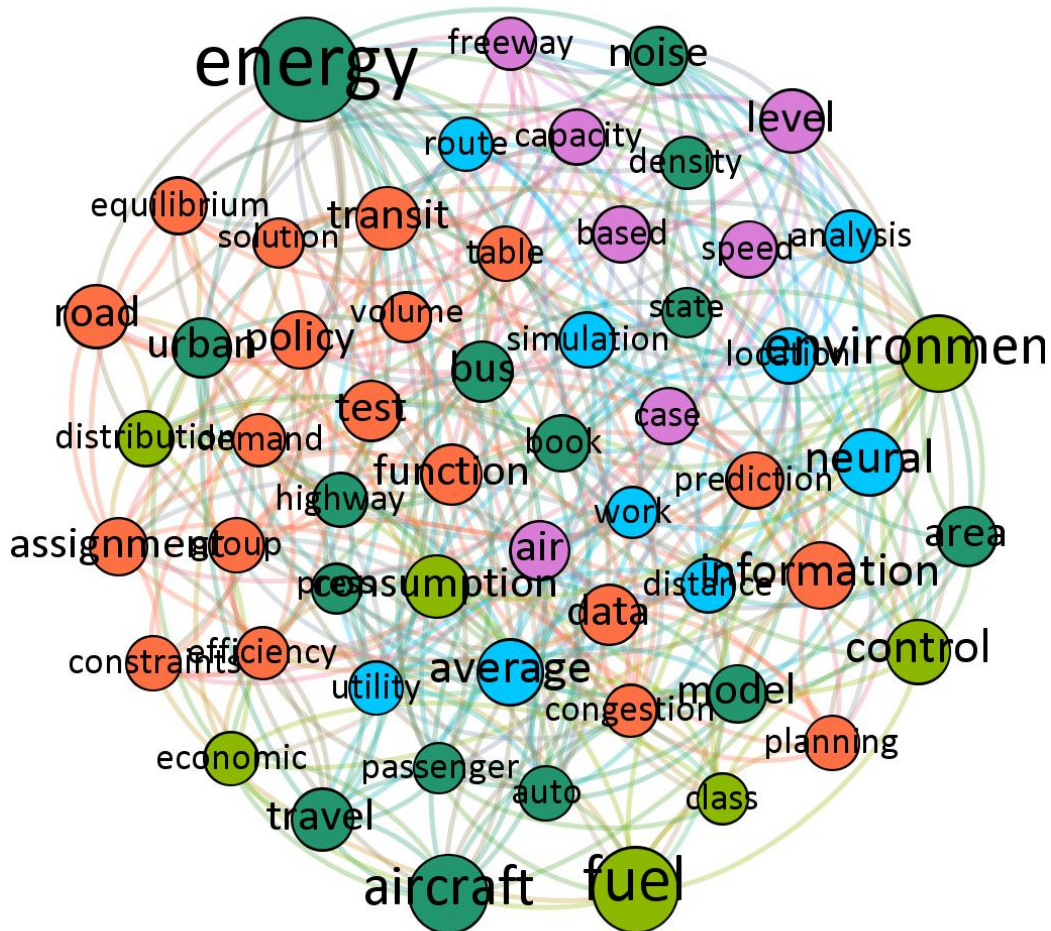


Figura 22 - Grafo relativo ao período de 2005 – 2004 (inclusive).

O grupo verde claro relaciona combustível com meio ambiente, controle e distribuição, localizados na periferia do grafo, ou seja, são questões que inter-relacionam com todas as demais temáticas apresentadas.



#### 5.5.10. Décimo grafo - intervalo – 2010 a 2014

O décimo grafo que compreende o período de 2010 a 2014, inclusive, e está representado na Figura 23. 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.

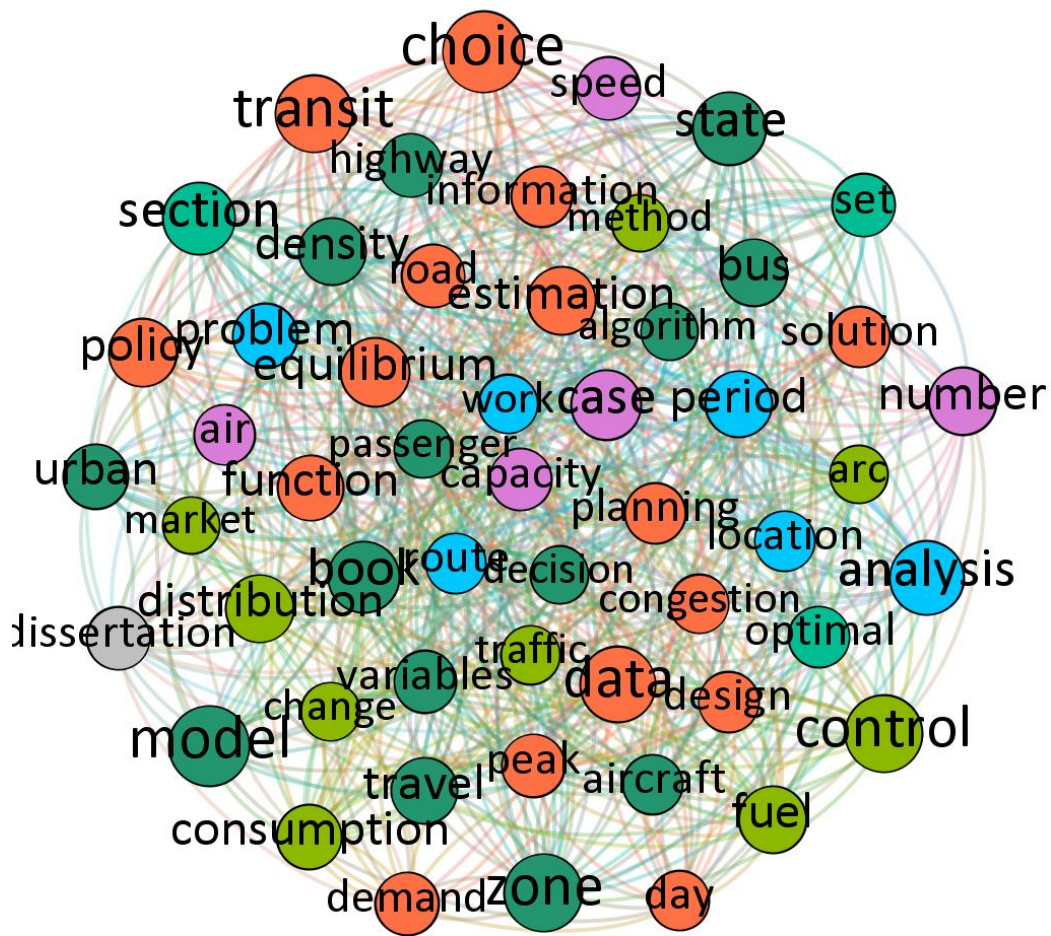


Figura 23 - Grafo relativo ao período de 2010 – 2014 (inclusive).

### 5.5. Totais analisados

A consulta por totais exibe estatísticas gerais e relevantes de todos os artigos. Os totais cienciométricos mostram o volume de dados utilizados para essa pesquisa.

A primeira tabela apresentada em estatísticas gerais exibe o total de publicações, 8202, o total de autores que publicaram na *Transportation Reserch*, 12716 e o total de *tokens* mapeados pelo MetrIA, que totalizam 7925, de acordo com a Tabela 6:

Tabela 6 - Totais de textos, autores e tokens.

<b>Total de textos</b>	8202
<b>Total de autores</b>	12716
<b>Total de tokens</b>	7925

A segunda tabela apresenta o Artigos por Revista. Ela exibe o total de artigos publicados em cada parte da revista *Transportation Research*. Manteve-se aqui o histórico, momento anterior ao da subdivisão nas atuais seis revistas, conforme Tabela 7.

Tabela 7 – Totais de publicações por revista.

### Artigos Por Revista

<b>Transportation Research</b>	504 (6.14%)
<b>Transportation Research Part A: General</b>	614 (7.49%)
<b>Transportation Research Part A: Policy and Practice</b>	1467 (17.89%)
<b>Transportation Research Part B: Methodological</b>	1806 (22.02%)
<b>Transportation Research Part C: Emerging Technological</b>	1130 (13.78%)
<b>Transportation Research Part D: Transport and Enviroment</b>	972 (11.85%)
<b>Transportation Research Part E: Logistics and Transportation Rev</b>	971 (11.84%)
<b>Transportation Research Part F: Traffic Psychology and Behaviour</b>	738 (9.00%)

Na sequência, disponibilidade dos resumos informa o percentual de publicações que possuem resumo (93,43%). No presente trabalho optou-se por manter os 6,57% de artigos sem resumo, como na versão original do autor, para não intervir no cálculo da similaridade.

Número de autores indica o número total 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 8.

Tabela 8 – Numero de autores por publicação.

### Número de autores

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%)

O total de artigos por faixa de 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 9. 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.

Tabela 9 – Percentual de 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%)

## 5.6. Estudo complementar

Com objetivos 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 da presente tese, 2014. O objetivo de analisarmos as 10 palavras mais relevantes no início do periódico e as 10 palavras mais relevantes do término da análise do periódico (2014), verificar o comportamento dos termos que se destacavam há 50 anos atrás até o presente momento, e também o inverso.

O gráfico da Figura 24 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*.

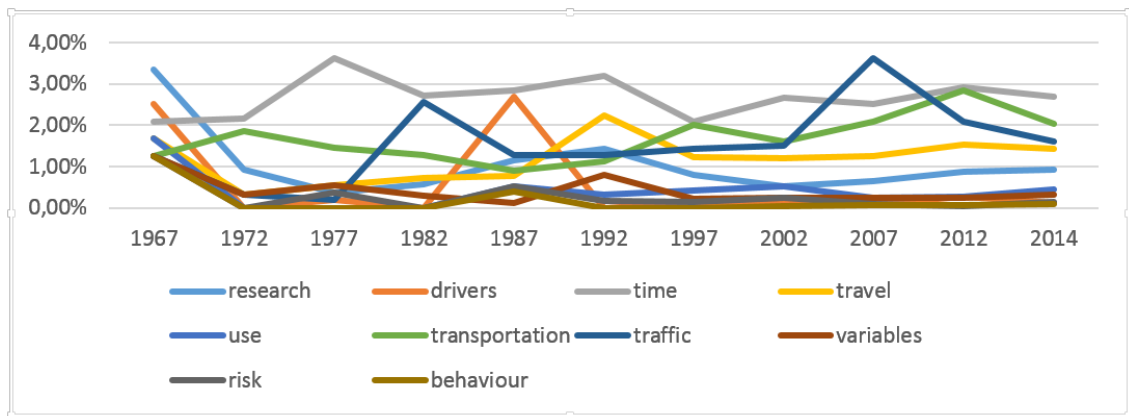


Figura 24 - Gráfico de projeção dos termos mais relevantes em 1967.

Considerando agora os termos mais relevantes no último ano analisado da revista *Transportation Research* (Figura 25), 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”.

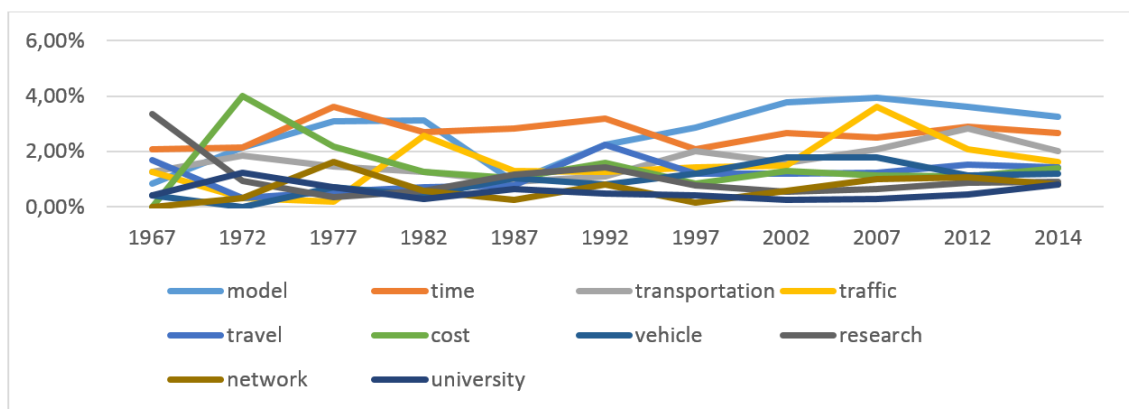


Figura 25 - Gráfico de projeção dos termos mais relevantes em 2014.

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.

### **5.7. Conclusões a partir do Estudo de Caso**

Pretende-se aqui analisar os resultados obtidos, com o fim de traçar uma perspectiva futura das pesquisas nas diversas áreas do transporte e, indiretamente, analisar o resultado do emprego da ciência da computação para a obtenção desses resultados.

O estudo de caso acima apresentado foi realizado com fundamento na cienciometria, informetria e bibliometria baseado em mais de 8 mil artigos da revista *Transportation Research*. Foram realizadas totalização de autores, artigos, artigos por divisão da revista, autores mais relevantes por revista, termos de maior ocorrência, termos de maior relevância, enfim, foi realizado um “raio-x” nas publicações, com o objetivo de responder à principal hipótese da tese: a partir de uma análise temporal é possível determinar novas tendências na área de transporte?

No período inicial analisado, identificou-se como relevantes os termos *road*, *risk*, *traffic*, *model*, *design*, *simulation* e *urban*, termos estes apresentados no grafo da

figura 21. O grafo apresenta a relevância das palavras por suas ocorrências e interdependências, fato esse verificado pelo tamanho do círculo que as representa, mostrando assim, que os artigos iniciais eram mais similares entre si, abordavam principalmente o transporte urbano rodoviário, e não apresentavam muita diversidade em suas temáticas.

Fundamentando com números, o MetrIA 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, é uma prova irrefutável da diversidade de temas que o periódico *Transportation Research* aumentou ao longo dos anos. Além da revista *Transportation Research* ter criado 6 subdivisões e com isso motivar a diversidade.

Um grande diferencial do *software* MetrIA é que ele permite localizar termos de relevância, diferentemente da simples localização de uma palavra como nos sites de busca. Para o MetrIA o termo deve ter relevância para ser ranqueado, isso graças ao processamento de linguagem natural. Ao selecionar uma das publicações que possui o termo procurado, é possível identificar as dez mais similares à mesma, direcionando o leitor/pesquisador objetivamente a conteúdos relevantes à sua busca, reduzindo assim seu tempo de busca, tornando-a objetiva.

Analisando o gráfico das palavras mais relevantes no mesmo período, sem considerar suas interconexões com outras publicações, calculados esses pelo algoritmo do *TF-IDF*. Os termos *research, time, drivers, travel, user, transportation, variable,*



*behavior, risk e traffic* apresentam-se como os mais relevantes. Verifica-se que os gráficos citados são coerentes com os grafos, e os termos também estabelecem relação mais significativa com o transporte rodoviário urbano.

Esta análise contempla os últimos anos da década de 60 do século passado, Conclui-se, desta forma, 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 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 de fronteiras.

O *software* MetrIA apresenta tendências futuras probabilísticas e não determinísticas. O grafo apresentado na Figura 30, que foi calculado a partir da análise dos termos mais frequentes dentre as publicações, orienta essa análise, e apresenta dezenas de termos proporcionalmente relevantes. A análise feita para apresentar as futuras tendências na área de transporte identificou três como principais grupos encontrados pelo algoritmo:

- Grupo verde escuro – 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 verde claro – 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 laranja – 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.

Conclui-se que temos 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.

### **5.8. Análise realizada: Humano x Máquina**

É importante ressaltar algumas das principais diferenças entre as análises realizadas por seres humanos e pelo *software* MetrIA. As pesquisas realizadas por Santos e Braga (1986) e pelos alunos do Programa de Engenharia de Transportes da

UFRJ no início desta década servem de parâmetro para traçarmos algumas das distinções:

1 - O *software* tem como base algoritmos de Inteligência artificial e processamento de linguagem natural, e, sendo assim, não possibilita uma classificação cartesiana, com 100% de acerto, como por exemplo: “o artigo refere-se à zona rural ou zona urbana?”. Como será apresentado nas tabelas, gráficos e, principalmente, nos grafos é possível identificar grupos temáticos, grandes áreas de pesquisa, mas não é possível certas classificações objetivas apresentadas tanto no artigo de Santos e Braga (1986) e pelos alunos do Programa de Engenharia de Transportes da UFRJ, como por exemplo: 100% de certeza que este artigo tem como foco principal o transporte rural.

2 - Outro ponto importante é referente ao tempo gasto em leitura e anotações (humanos) x tempo de processamento de textos (máquina). Para realizar a análise apresentada nesta tese foram processados mais de dez mil artigos originalmente, e foram descartados cerca de 20% de documentos que se encontravam na base de dados da Elsevier e que não eram artigos de fato. Considerando que os alunos do PET realizaram as pesquisas em 2 anos lendo apenas o abstract de teses e dissertações, quantos anos levaria uma única pessoa lendo mais de 8 mil artigos na íntegra (o MetrIA ainda realiza estudo comparativo entre resumo e corpo do texto), além do registro de todas as anotações das análises apuradas durante a leitura?

3 – Não há a possibilidade de realizar manualmente todas as conexões e cálculos de peso dos grafos. Para fins de análise foram apresentados os termos de maior relevância, mas os cálculos foram realizados com TODOS os termos constantes

na base de dados, contabilizando todas suas ocorrências e inter-relações com as demais palavras.

4 – Os cálculos de similaridade, tanto resumo x texto, principalmente a análise de similaridade de cada publicação comparada com mais de 8 mil artigos para determinar as 10 mais similares com relação ao conteúdo seria praticamente impossível de ser realizada manualmente, pois as inter-relações calculadas são da ordem de grandeza de  $10^8$ .

5 – O cálculo de frequência de termos, considerando que a base de dados apresenta mais de mil referências, se realizado manualmente artigo por artigo, resultaria em uma análise na ordem de grandeza de  $10^7$  (só para determinar se a palavra aparece ou não no artigo). Sem contar que para a criação da lista de termos seria necessária uma prévia leitura de todo o material, e a posteriori uma releitura anotando o número de ocorrências de palavra por palavra de cada artigo, para assim definir os termos de maior ocorrência, e também para determinar as palavras mais relevantes de acordo com o contexto. Para o *software* esse processamento ocorre em menos de 1 minuto.

6 – As totalizações de autores é a tarefa mais viável a ser realizada pelo ser humano, sendo que o total dos textos analisados possui mais de 12 mil autores distintos.

7 – Localizar artigos que tenham uma determinada palavra, que não seja apresentada nem no resumo e nem entre as palavras chaves, pelo ser humano, o obrigaria a ler na íntegra todas as publicações a cada nova necessidade de busca.

8 – O Google nem sempre é muito eficiente em localizar o conteúdo de arquivos no formato PDF, formato disponibilizado pela revista Elsevier. Sendo assim utilizar os

algoritmos de busca do MetrIA aumenta a efetividade de busca de termos, tanto por relevância quanto à critério do usuário.

O objetivo de elencar esses tópicos é demonstrar que estudos cienciométricos, informétricos e bibliométricos de um grande volume de dados só se viabiliza quando realizado por um *software* especialista, ou por uma grande equipe de pessoas, as quais correm o risco maior de erros durante as anotações das informações oriundas da leitura do material.

Na sequência apresenta-se o último capítulo que traz as conclusões finais, análise das hipóteses elencadas no capítulo um, trabalhos futuros, bem como recomendações da presente tese.

## CAPÍTULO 6 – CONCLUSÕES FINAIS E RECOMENDAÇÕES

A análise das tabelas, gráficos e grafos apresentados 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 foi verificada pelo *software* MetrlA, que apresentou um aumento de quase 1.000 % de termos relevantes, comparando o primeiro ano e o último da revista.

O transporte deixou de ser visto somente sob a ótica do transporte em si, e passou a ser estudado, pesquisado e projetado a partir da necessidade de se pensar o transporte para atender o deslocamento de pessoas e bens a baixos custos, com qualidade e rapidez, englobando fatores como os aspectos sociais, comportamentais, ambientais, psicológicos e da qualidade de vida dos seres vivos, dentre outros.

A análise realizada pelo *software* MetrlA confirma a hipótese de que é possível antever tendências em Engenharia de Transportes a partir do estudo da série histórica de publicações científicas ao longo dos anos. As três grandes áreas mapeadas pelo *software* apresentadas como tendências futuras em Engenharia de Transportes são: 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. As três grandes áreas apresentadas são áreas que, principalmente nos países desenvolvidos, vêm crescendo e ganhando investimentos em pesquisas e na implantação de políticas públicas. No caso brasileiro, identificar essa demanda futura é importante para se repensar as políticas públicas que hoje vêm sendo executadas em todas as esferas, municipal, estadual e federal.

A última hipótese levantada e confirmada é 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, hipótese confirmada pelo *software* MetrIA e pelos grafos. 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.

O *software* destaca-se por ter demonstrado ser uma nova ferramenta facilitadora na leitura de artigos, por apresentar diversas informações cienciométricas, informétricas e bibliométricas sobre os mesmos, além de inter-relacionar as publicações constantes no banco de dados, quer seja pelo autor, revista, ano de publicação, similaridade, relevância ou ocorrência de termos. O *software* MetrIA demonstrou ainda ser muito versátil pela facilidade na criação de novas consultas e relatórios e também pela possibilidade de ser utilizado em outras áreas do conhecimento e base de dados.

Concluindo, destaca-se que as três metodologias utilizadas na presente tese, a ciencimetria, 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, o que tornou possível a análise de um volume tão grande de informações e apresentação dos resultados obtidos.

Como trabalhos futuros, além de novas consultas, tem-se como objetivo ampliar a base de dados, incluindo publicações de outras revistas especializadas na área de engenharia de transportes, bem como teses e dissertações da área. O aumento da base de dados aumenta a inteligência e acurácia das análises, pois para o processamento de linguagem natural e para as análises realizadas baseadas em inteligência artificial, maior a fonte de informações, mais refinada são as análises. Acredita-se que ter-se-á um grande *big data* de Engenharia de Transportes, tornando-se referência de consulta para pesquisadores de todo o mundo.

Para o *software* MetrIA tem-se como objetivo empregar técnicas de redes neurais na base de dados, visando antever novas tendências para a realização de um estudo comparativo com os resultados obtidos pelos grafos. Pretende-se incorporar novos algoritmos de processamento de linguagem natural, para aumentar ainda mais a inteligência do sistema. Também como projeto futuro se prevê a criação de uma base de dados própria na língua portuguesa, visando o estudo das teses e dissertações produzidas no Brasil. Pretende-se também disponibilizar comparações e conexões entre outras áreas do conhecimento, buscando ampliar as interconexões com áreas afins.

Para encerrar, destaca-se que o maior desafio apresentado por esta tese interdisciplinar foi realizar estudos aprofundados de temas tão distintos como a engenharia de transportes, metodologias de estudo (cienciometria, informetria e bibliometria), inteligência artificial (englobando aqui processamento de linguagem natural e *data mining*) e programação propriamente dita. Na tese se buscou equilibrar a

apresentação de cada temática. Das etapas realizadas, foi o *software* MetrlA que demandou maior tempo de planejamento, concepção e execução.

## REFERÊNCIAS BIBLIOGRÁFICAS

- AGGARWAL, N., KUMAR, A., KHATTER, H., & KHA, H. **Analysis the effect of data mining techniques on database**, 47, Maio 2012. 164–169. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0965997812000038>>. Acesso em: 25 Março 2015.
- AIZAWA, A. **An information-theoretic perspective of tf-idf measures**, 39, n. 1, Janeiro 2003. 45–65. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0306457302000213>>. Acesso em: 25 Março 2015.
- ANTP (Associação Nacional de Transportes Públicos). **Transporte Humano: Cidades com qualidade de vida**. São Paulo: [s.n.], 1997.
- BAGHELA, V. S.; TRIPATHI, S. P. **Text Mining Approaches To Extract Interesting Association Rules from Text Documents**, 9, n. 3, Maio 2012. Disponível em: <<http://www.ijcsi.org/papers/IJCSI-9-3-3-545-552.pdf>>. Acesso em: 25 Março 2015.
- BALASSIANO, R. **Transporte sustentável em países emergentes: Uma análise prospectiva**, 2007.
- BALLOU, R. H. **Gerenciamento da cadeia de suprimentos / logística empresarial**. 5ª. ed. Porto Alegre: Bookman, 2006.
- BASTIAN, M.; HEYMANN, S.; JACOMY, M. **Gephi: An Open Source Software for Exploring and Manipulating Networks**. International AAAI Conference on Web and Social Media, 2009. Disponível em: <<http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154>>. Acesso em: 23 mar. 2016.

BJÖRNEBORN, L.; INGWERSEN, P. **Toward a basic framework for webometrics**, Journal of the American Society for Information Science and Technology. United States, 55, n. 14, Dezembro 2004. 1216–1227. Disponível em: <[http://www.researchgate.net/publication/264289068\\_Toward\\_a\\_basic\\_framework\\_for\\_Webometrics](http://www.researchgate.net/publication/264289068_Toward_a_basic_framework_for_Webometrics)>. Acesso em: 29 Janeiro 2015.

BOAVENTURA NETO, P. O. **Grafos - Teoria, Modelos, Algoritmos**. 2ª. ed. São Paulo: Edgard Blücher, 2003.

BOGART, D. **Inter-Modal Network Externalities and Transport Development: Evidence from Roads, Canals, and Ports During the English Industrial Revolution**, Califórnia, 2012. Disponível em: <[http://www.researchgate.net/publication/5129827\\_Inter-Modal\\_Network\\_Externalities\\_and\\_Transport\\_Development\\_Evidence\\_from\\_Roads\\_Canals\\_and\\_Ports\\_During\\_the\\_English\\_Industrial\\_Revolution](http://www.researchgate.net/publication/5129827_Inter-Modal_Network_Externalities_and_Transport_Development_Evidence_from_Roads_Canals_and_Ports_During_the_English_Industrial_Revolution)>. Acesso em: 29 Janeiro 2015.

BOGART, D. **The Transportation Revolution in Industrializing Britain: A Survey, California**, 2013. Disponível em: <<https://ideas.repec.org/p/irv/wpaper/121306.html>>.

BOWERSOX, D. J.; CLOSS, D. J. **Gestão logística de cadeias de suprimentos**. São Paulo: Bookman, 2006.

BRIN, S.; PAGE, L. **Dynamic data mining: Exploring large rule space by sampling**, Stanford, 2001. Disponível em: <<http://ilpubs.stanford.edu:8090/424/>>.

BRINN, T.; JONES, M. J.; PENDLEBURY, M. **Measuring research quality**. Omega, 2000. 237-239. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0305048399000481>>. Acesso em: 14 Abril 2015.

CAMBRIA, E.; WHITE, B. **Jumping NLP Curves: A Review of Natural Language Processing Research**. Computational Intelligence Magazine, IEEE, 9, n. 2, 2014.

Disponível em: <<http://sentic.net/jumping-nlp-curves.pdf>>. Acesso em: 14 abril 2015.

CANO, W. **A desindustrialização no Brasil**, 21, n. especial, 2012. Acesso em: 15 Julho 2015.

CARVALHO, L. A. V. **Datamining - a mineração de dados no Marketing, Medicina, Economia, Engenharia e Administração**. São Paulo: Érica, 2001.

CASCETTA, E. **Transportation Systems Engineering: Theory and Methods**. Napoli: Springer Science & Business Media, 2013.

CERVERO, R. **Office Development, Rail Transit, and Commuting Choices**. Journal of Public Transportation, 9, 2006.

CERVERO, R. **Transport Infrastructure and the Environment: Sustainable Mobility and Urbanism**. Planocosmo International Conference Bandung Institute of Technology, Berkeley, Outubro 2013. Acesso em: 29 maio 2016.

CHAGAS, H. X.; MOURA, V. A.; AKABANE, G. K. **Logística de Comércio Exterior brasileira: Uma análise do Índice de Desempenho Logístico brasileiro no cenário mundial**, São Paulo, 2015. Disponível em: <[http://www.centropaulasouza.sp.gov.br/pos-graduacao/workshop-de-pos-graduacao-e-pesquisa/010-workshop-2015/workshop/trabalhos/Sistemas\\_Produtivos/Gest\\_de\\_Operac\\_Logist/Logist\\_Comercio\\_exterior\\_brasileira.pdf](http://www.centropaulasouza.sp.gov.br/pos-graduacao/workshop-de-pos-graduacao-e-pesquisa/010-workshop-2015/workshop/trabalhos/Sistemas_Produtivos/Gest_de_Operac_Logist/Logist_Comercio_exterior_brasileira.pdf)>. Acesso em: 31 Outubro 2015.

CHAKROBORTY, P.; DAS, A. **Principles of Transportation Engineering**. Nova Deli: PHI Learning Pvt. Ltd., 2012. 536 p. Disponível em: <[https://books.google.com.br/books?id=UzOixH-gBQcC&dq=transportation+engineering+journals+history+of+transportation&hl=pt-BR&source=gbs\\_navlinks\\_s](https://books.google.com.br/books?id=UzOixH-gBQcC&dq=transportation+engineering+journals+history+of+transportation&hl=pt-BR&source=gbs_navlinks_s)>.

CORRÊA, G. M. **Uma análise do comércio exterior brasileiro de automóveis no contexto da discussão sobre a desindustrialização**. Universidade de Brasília – UnB, Brasília, 2014. Disponível em: <[http://bdm.unb.br/bitstream/10483/8603/1/2014\\_GuilhermeMoriciCorrea.pdf](http://bdm.unb.br/bitstream/10483/8603/1/2014_GuilhermeMoriciCorrea.pdf)>. Acesso em: 24 Agosto 2015. Monografia.

CUMBLEY, R.; CHURCH, P. **Is “Big Data” creepy?**, 29, n. 5, Outubro 2013. 601-609. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0267364913001349>>. Acesso em: 25 Março 2015.

FERRARI, C. **Curso de Planejamento Municipal Integrado**. 2ª. ed. São Paulo: Pioneira, 1979.

FURLAN, B.; BATANOVIC, V.; NIKOLIC, B. **Semantic similarity of short texts in languages with a deficient natural language processing support**. Decision Support Systems - Elsevier, 55, 2013. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0167923613000614>>. Acesso em: 14 abril 2015.

GARBER, N. J.; HOEL, L. A. **Traffic and Highway Engineering**. 3ª. ed. Califórnia: Thomson Learning, 2002.

GARFIELD, E. **How can impact factors be improved?**, 1996. Disponível em: <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2351785/>>. Acesso em: 25 abril 2014. 313, 411-413.

GARFIELD, E. **The use of journal impact factors and citation analysis for evaluation of science**, 1998. Disponível em: <[http://www.garfield.library.upenn.edu/papers/eval\\_of\\_science\\_oslo.html](http://www.garfield.library.upenn.edu/papers/eval_of_science_oslo.html)>. Acesso em: 15 Agosto 2014.

GARFIELD, E. **Use of Journal Citation Reports and Journal Performance Indicators in measuring short and long-term journal impact**, 2000. Disponível em:

<[http://www.researchgate.net/publication/12262737\\_Use\\_of\\_Journal\\_Citation\\_Reports\\_and\\_Journal\\_Performance\\_Indicators\\_in\\_measuring\\_short\\_and\\_long\\_term\\_journal\\_impact](http://www.researchgate.net/publication/12262737_Use_of_Journal_Citation_Reports_and_Journal_Performance_Indicators_in_measuring_short_and_long_term_journal_impact)>. Acesso em: 14 agosto 2014. 41 (4), 368-374.

HAYKIN, S. **Redes Neurais. Bookman**. 2ª. ed. São Paulo: Bookman, 2001.

HEA, W.; ZHAB, S.; LIA, L. **Social media competitive analysis and text mining: A case study in the pizza industry**, 33, n. 3, Junho 2013. 464–472. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0268401213000030>>. Acesso em: 25 Março 2015.

HOEL, L.; GARBER, N.; SADEK, A. **Transportation Infrastructure Engineering: A Multimodal Integration, SI Version**. Cengage Learning, 2011. 674 p. Disponível em: <[https://books.google.com.br/books?id=QkbGpZngUWwC&pg=PA15&lpg=PA15&dq=container+transportation+history+of+engineering&source=bl&ots=0O-1pJ\\_bRP&sig=o6NxsGOE4OGwbwko041PY7PSDNI&hl=pt-BR&sa=X&ved=0CEkQ6AEwBmoVChMIhliy\\_57jyAIVxBGQCh3CAQH#v=onepage&q=containe](https://books.google.com.br/books?id=QkbGpZngUWwC&pg=PA15&lpg=PA15&dq=container+transportation+history+of+engineering&source=bl&ots=0O-1pJ_bRP&sig=o6NxsGOE4OGwbwko041PY7PSDNI&hl=pt-BR&sa=X&ved=0CEkQ6AEwBmoVChMIhliy_57jyAIVxBGQCh3CAQH#v=onepage&q=containe)>. Acesso em: 12 Julho 2015.

HONG, T.-P. et al. **Using TF-IDF to hide sensitive itemsets**, 38, n. 4, Junho 2013. 502-510. Disponível em: <<http://link.springer.com/article/10.1007/s10489-012-0377-5>>. Acesso em: 25 Março 2015.

HUTCHINSON, B. G. **Princípios de Planejamento dos Sistemas de Transporte Urbano**. Rio de Janeiro: Guanabara Dois, 1979.

HWU, W.-M. **What is ahead for parallel computing**, 74, n. 7, Julho 2014. 2574–2581. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0743731514000331>>. Acesso em: 25 Março 2015.

IBGE. **Brasil em Números**, 2013. Disponível em: <[http://biblioteca.ibge.gov.br/visualizacao/periodicos/2/bn\\_2013\\_v21.pdf](http://biblioteca.ibge.gov.br/visualizacao/periodicos/2/bn_2013_v21.pdf)>. Acesso em: 17 mar. 2016.

JURISH, B.; WÜRZNER, K.-M. **Word and Sentence Tokenization with Hidden Markov Models**, 28, n. 2, 2013. 61-83. Disponível em: <[https://www.google.com.br/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0CB0QFjAAahUKEwjR16qNj-\\_IAhVJk5AKHbGDBDA&url=http%3A%2F%2Fwww.jlcl.org%2F2013\\_Heft2%2F3Jurish.pdf&usg=AFQjCNHS5Uco7Qh2fJBHq5o8JnsbwUisvw](https://www.google.com.br/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0CB0QFjAAahUKEwjR16qNj-_IAhVJk5AKHbGDBDA&url=http%3A%2F%2Fwww.jlcl.org%2F2013_Heft2%2F3Jurish.pdf&usg=AFQjCNHS5Uco7Qh2fJBHq5o8JnsbwUisvw)>. Acesso em: 25 Março 2015.

KÖCHE, J. C. **Fundamentos da Metodologia Científica**. Petrópolis: Vozes, 2011.

KUMARES, C.; BULLOCK D.; HENDRICKSON, C.; LEVINSON, H.; LYLES, R.; RADWAN, A.; LI, Z. **Development Transportation Engineering Research, Education, Practice and Changing Civil Engineering World**. Journal of Transportation Engineering, 2002. 301-3013. Disponível em: <<http://www.google.com.br/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&cad=rja&uact=8&ved=0CCYQFjABahUKEwit5ZC3hOPIAhXIEZAKHQJSJDE8&url=http%3A%2F%2Fwww.cmu.edu%2Fgdi%2Fdocs%2Fdevelopment-of-transportation.pdf&usg=AFQjCNGoAuzjz8miaaOOPpqjg0m55K1UzA&bvm=bv.10584>>. Acesso em: 15 Agosto 2014.

LESKOVEC, J.; RAJARAMAN, A.; ULLMAN, J. D. **Mining of Massive Datasets**. 5<sup>a</sup>. ed. Stanford: [s.n.], 2014. Disponível em: <<http://infolab.stanford.edu/~ullman/mmds/book.pdf>>. Acesso em: 14 Abril 2015.

LEVISON, D., Liu, H., GARRISON, W., HICKMAN, M., Danczyk, A., & CORBETT, M. **Fundamentals of Transportation: Transit**, 2015. Disponível em: <[http://en.wikibooks.org/wiki/Fundamentals\\_of\\_Transportation/Transit](http://en.wikibooks.org/wiki/Fundamentals_of_Transportation/Transit)>. Acesso em: 14 Abril 2015.



- LEYDESDORFF, L. **The Challenge of Scientometrics**. 2ª. ed. USA: Universal Publishers, 2001.
- LEYDESDORFF, L.; MILOJEVIC, S. Scientometrics. **School of Informatics and Computing, Indiana University**, 2013. Disponível em: <<http://arxiv.org/ftp/arxiv/papers/1208/1208.4566.pdf>>. Acesso em: 01 julho 2014.
- MACBRY, T. R. **A logística de transporte no cenário da globalização**. Ideias em destaque - I.H.C. da Aeronáutica, 2007.
- MACIAS-CHAPULA, C. A. **O papel da informetria e da cienciometria e sua perspectiva nacional e internacional**. Ciência da Informação, Brasília, 27, Maio/Agosto 1998. 134-140. Disponível em: <<http://dx.doi.org/10.1590/S0100-19651998000200005>>. Acesso em: 02 Abril 2014.
- MANHEIN, M. L. **Fundamentals of Transportation Systems Analysis - vol. 1**. Cambridge: MIT Press, 1979.
- MARCONI, M. A.; LAKATOS, E. M. **Fundamentos da Metodologia Científica**. 5ª. ed. São Paulo: Atlas, 2003.
- MARKSCHEFFEL, B. **An ontology based visualization approach for the joined interpretation of bibliometrics and webometrics data**. ACM Digital Library, 2011. Disponível em: <<http://dl.acm.org/citation.cfm?id=2077520>>. Acesso em: 12 Agosto 2015.
- MCGRATH, W. **What bibliometricians, scientometricians and informetricians study; a typology for definition and classification; topics for discussion**. International Conference on Bibliometrics, Scientometrics and Informetrics, Ontario, 1989. Acesso em: 20 Abril 2015.
- MENDEL, J. M.; McLAREN, R. W. **Reinforcement learning control and pattern recognition systems**. Academic Express, 1970.

- NASCIMENTO, H. A. D.; FERREIRA, C. B. R. **Visualização de Informações – Uma Abordagem**, 2005. Disponível em: <[https://www.researchgate.net/profile/Hugo\\_Nascimento/publication/267403645\\_Visualizacao\\_de\\_Informaes\\_\\_Uma\\_Abordagem\\_Prtica/links/5510a6940cf2ba84483f9704.pdf](https://www.researchgate.net/profile/Hugo_Nascimento/publication/267403645_Visualizacao_de_Informaes__Uma_Abordagem_Prtica/links/5510a6940cf2ba84483f9704.pdf)>. Acesso em: 23 mar. 2016.
- NEWMAN, M. E. J. **Modularity and community structure in networks**. Proc National Academy Science USA., 2006.
- NODE.JS. **Node.js**, 2015. Disponível em: <<https://nodejs.org>>. Acesso em: 16 Abril 2015.
- NOVAES, A. G. **Logística e gerenciamento da cadeia de distribuição**. 2ª. ed. Rio de Janeiro: Elsevier, 2007.
- ORTÚZAR, J. D.; WILLUMSEN, L. G. **Modelling Transport**. 2ª. ed. West Sussex: John Wiley and Sons, 1997.
- PAPACOSTAS, C. S.; PREVEDOUROS, P. D. **Transportation Engennering and Planning**. 2ª. ed. New Jersey: Prattice-Hall, 1993.
- PHRIDVIRAJA, M. S. B.; GURURAOB, C. V. **Data Mining – Past, Present and Future – A Typical Survey on Data Streams**. Petru Maior University of Tirgu Mures, Romania, 12, 2014. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S2212017313006683>>. Acesso em: 25 Março 2015.
- PRINCENTON UNIVERSITY. **Wordnet**, 2015. Disponível em: <<https://wordnet.princeton.edu/>>.
- PROBST, G.; RAUB, S.; ROMHART, K. **Gestão do conhecimento: os elementos construtivos do sucesso**. Porto Alegre: Bookman, 2002.

REED, C. A. **Gutenberg in Shanghai: Chinese Print Capitalism, 1876-1937**. UBC Press, 2011.

RESENDE, P. **Na área de infraestrutura logística, Brasil não investe o suficiente nem para recuperar perdas**, Fevereiro 2013. Disponível em: <<http://www.amcham.com.br/eventos/regionais/amcham-sao-paulo/noticias/2013/na-area-de-infraestrutura-logistica-brasil-nao-investe-o-suficiente-nem-para-recuperar-perdas-aponta-coordenador-da-fdc>>. Acesso em: 25 Agosto 2015.

REZENDE, S. -. **Sistemas Inteligentes. Fundamentos e aplicações**. São Paulo: Manole, 2005.

RODRIGUE, J.-P.; COMTOIS, C.; SLACK, B. **The Geography of Transport Systems**. 3<sup>a</sup>. ed. Routledge, 2013. Disponível em: <[https://books.google.com.br/books?id=WeocAAAQBAJ&hl=pt-BR&source=gbs\\_navlinks\\_s](https://books.google.com.br/books?id=WeocAAAQBAJ&hl=pt-BR&source=gbs_navlinks_s)>. Acesso em: 16 Abril 2015.

RUIZ, J. A. A.; JORGE, R. A. **Informetría, bibliometría y cienciometría: aspectos teórico-prácticos**, Havana, 10, 2002. Disponível em: <[http://scielo.sld.cu/scielo.php?pid=S1024-94352002000400004&script=sci\\_arttext](http://scielo.sld.cu/scielo.php?pid=S1024-94352002000400004&script=sci_arttext)>. Acesso em: 10 julho 2014.

RUSSEL, S.; NORVIG, P. **Inteligência Artificial**. São Paulo: [s.n.], 2004.

SAGE, A. P. **Concise Encyclopedia of Information Processing in Systems and Organizations**. Nova Iorque: Pergamon, 1990.

SANTOS, M. P. S.; BRAGA, M. G. C. **Research trends in urban transport science: some empirical evidence from academic research**. Transportation Research - A, 22A, 1986.

- SCI. **SCI**, 2014. Disponível em: <<http://thomsonreuters.com/en/products-services/scholarly-scientific-research/scholarly-search-and-discovery/web-of-science.html>>. Acesso em: 14 Abril 2015.
- SCIELO. **Hábitos de leitura de literatura científica entre pesquisadores**, SciELO em Perspectiva. 2014. Disponível em: <<http://blog.scielo.org/blog/2014/04/03/habitos-de-leitura-de-literatura-cientifica-entre-pesquisadores/>>. Acesso em: 25 ago. 2015.
- SILVA, J. A.; BIANCHI, M. D. L. P. **Cientometria: a métrica da ciência**, Pandéia - USP, Ribeirão Preto, 2001. Disponível em: <<http://dx.doi.org/10.1590/S0103-863X2001000200002>>. Acesso em: 25 Abril 2014.
- SMALL, H. G.; GRIFFITH, B. C. **The structure of scientific literatures, I: Identifying and graphing specialties**. Science Studies, 1974. 4, 17-40. Disponível em: <[http://www.jstor.org/stable/284536?seq=1#page\\_scan\\_tab\\_contents](http://www.jstor.org/stable/284536?seq=1#page_scan_tab_contents)>. Acesso em: 14 Abril 2015.
- SPINAK, E. **Diccionario enciclopédico de Bibliometria, cienciometría e informetría**. Caracas: Unesco -CII/II, 1996.
- TAGUE-SUTCLIFFE, J. **An introduction to informetrics**. Information Processing and Management: an International Journal - Special issue on Informetrics, 1992. Disponível em: <<http://dl.acm.org/citation.cfm?id=160642>>. Acesso em: 20 abril 2014.
- TEDESCO, G. M. I. **Metodologia para Elaboração do Diagnóstico de um Sistema de Transporte**. Brasília: UnB, 2008. (dissertação de mestrado).
- TENOPIR, C.; VOLENTINE, R.; CHRISTIAN, L. **Scholarly Reading by Academic Staff: Summary Results of a Study Conducted in 2012 at Two Universities in Australia**, 2013. Disponível em: <[http://libvalue.cci.utk.edu/sites/default/files/AU.Faculty.Combined.Final\\_.pdf](http://libvalue.cci.utk.edu/sites/default/files/AU.Faculty.Combined.Final_.pdf)>. Acesso em: 16 abril 2014.

- TENOPIR, C.; VOLENTINE, R.; CHRISTIAN, L. **Scholarly Reading by Faculty in the United States: Summary Results of a Study Conducted in 2012 in Five Universities**, 2013. Disponível em: <[http://libvalue.cci.utk.edu/sites/default/files/US.Faculty.Combined.Final\\_.pdf](http://libvalue.cci.utk.edu/sites/default/files/US.Faculty.Combined.Final_.pdf)>. Acesso em: 24 julho 2014.
- TRANSPORTATION RESEARCH. **Transportation Research**. Elsevier - ScienceDirect, 2015. Disponível em: <<http://www.sciencedirect.com/science/journal/00411647>>. Acesso em: 25 Agosto 2015.
- VAN NOORDEN, R. **Global scientific output doubles every nine years**, 07 maio 2014. Disponível em: <<http://blogs.nature.com/news/2014/05/global-scientific-output-doubles-every-nine-years.html>>. Acesso em: 11 dez. 2015.
- VASCONCELLOS, E. A. **Transporte urbano nos países em desenvolvimento: reflexões e propostas**. 3ª. ed. São Paulo: Annablume, 2000.
- VASCONCELLOS, E. A. **Transporte urbano, espaço e equidade: Análises de Políticas Públicas**. 2ª. ed. São Paulo: Annablume, 2001.
- WORLD Economic Forum. **World Economic Forum**, 2015. Disponível em: <[http://www3.weforum.org/docs/WEF\\_GlobalCompetitivenessReport\\_2014-15.pdf](http://www3.weforum.org/docs/WEF_GlobalCompetitivenessReport_2014-15.pdf)>. Acesso em: 25 Agosto 2015. pág. 134.
- YUE, X. et al. **Analysis of the Combination of Natural Language Processing and Search Engine Technology**. International Workshop on Information and Electronics Engineering (IWIEE), 2012. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1877705812001968>>. Acesso em: 25 Abril 2015.
- ZAVADSKAS, E. K.; SKIBNIEWSKI, M. J.; ANTUCHEVICIENE, J. **Performance analysis of Civil Engineering Journals based on the Web of Science (R) database**. Archives of Civil and Mechanical Engineering, 17 Maio 2014. Disponível em:

<[http://www.researchgate.net/publication/263317978\\_Performance\\_analysis\\_of\\_Civil\\_Engineering\\_Journals\\_based\\_on\\_the\\_Web\\_of\\_Science\\_%28R%29\\_database](http://www.researchgate.net/publication/263317978_Performance_analysis_of_Civil_Engineering_Journals_based_on_the_Web_of_Science_%28R%29_database)>. Acesso em: 15 Janeiro 2015.