

**UNIVERSIDADE DO VALE DO RIO DOS SINOS - UNISINOS
UNIDADE ACADÊMICA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS CONTÁBEIS
NÍVEL MESTRADO**

RICARDO DOS SANTOS CARDOSO

**A ADOÇÃO DA CIÊNCIA DE DADOS NA CONTABILIDADE:
UM ESTUDO DAS CARACTERÍSTICAS EVOLUTIVAS DA REDE DE CO-
PALAVRAS-CHAVES DAS DISSERTAÇÕES E TESES DA
PÓS-GRADUAÇÃO BRASILEIRA**

PORTO ALEGRE

2022

RICARDO DOS SANTOS CARDOSO

**A ADOÇÃO DA CIÊNCIA DE DADOS NA CONTABILIDADE:
Um Estudo das Características Evolutivas da Rede de Co-Palavras-Chaves das
Dissertações e Teses da Pós-Graduação Brasileira**

Dissertação apresentada como requisito parcial para obtenção do título de Mestre em Ciências Contábeis, pelo Programa de Pós-Graduação em Ciências Contábeis da Universidade do Vale do Rio dos Sinos (UNISINOS).

Orientador: Prof. Dr. Alexsandro Marian Carvalho

Porto Alegre

2022

C268a Cardoso, Ricardo dos Santos
A adoção da ciência de dados na contabilidade: um estudo das características evolutivas da rede de co-palavras-chaves das dissertações e teses da pós-graduação brasileira. / Ricardo dos Santos Cardoso -- 2022.
83 f. : il. ; color. ; 30cm.

Dissertação (Mestrado em Ciências Contábeis) -- Universidade do Vale do Rio dos Sinos, Programa de Pós-Graduação em Ciências Contábeis, 2022.

Orientador: Prof. Dr. Alexsandro Marian Carvalho.

1. Contabilidade. 2. Controle de gestão. 3. Ciência de dados. 4. Rede complexa. 5. Rede de palavras de co-palavras-chaves - Pós-Graduação. I. Título. II. Carvalho, Alexsandro Marian.

CDU 657

Ricardo dos Santos Cardoso

A adoção da ciência de dados na contabilidade:
Um Estudo das Características Evolutivas da Rede de Co-Palavras-Chaves das
Dissertações e Teses da Pós-Graduação Brasileira

Dissertação apresentada como requisito parcial para obtenção do título de Mestre em Ciências Contábeis, pelo Programa de Pós-Graduação em Ciências Contábeis da Universidade do Vale do Rio dos Sinos (UNISINOS).

Aprovado em 28 de abril de 2022.

BANCA EXAMINADORA

Prof. Dr. André Luis Korzenowski - UNISINOS

Prof. Dr. Tiago Wickstrom Alves - UNISINOS

Prof. Dr. Luis Felipe Dias Lopes - UFSM

AGRADECIMENTO À CAPES

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Mãe, pelo sinônimo de amor, força e resiliência frente às dificuldades da vida, uma pessoa inspiradora e de energia positiva que fez com que esse sonho pudesse se tornar realidade.

AGRADECIMENTOS

Agradeço a Deus, por me amparar em todos os momentos vivenciados, dando-me saúde, força, coragem, sabedoria, fé e inteligência para a conquista de meus objetivos e metas.

Agradeço a minha família, mãe, irmã, sobrinhos e sobrinhas que estiveram juntos vendo todo o esforço despendido, e sempre de uma forma ou outra dando o apoio com gestos, atitudes e palavras.

Agradeço aos ensinamentos vivenciados junto ao Programa de Pós-Graduação em Ciências Contábeis em nível de mestrado acadêmico da Unisinos, a todos professores, colegas, coordenadores de curso, secretaria, que me proporcionaram uma experiência ímpar com toda essa troca de conhecimento, destacando todo esse processo de ensino e aprendizagem que certamente levarei por toda minha vida.

Agradeço ao grupo de pesquisa Mercados e Competitividade de Instituições de Ensino Superior da Unisinos, pelo conhecimento, aprendizado e experiências vivenciadas que me aproximaram do mundo da pesquisa. E certamente pretendo continuar para aumentar meus conhecimentos.

Agradeço, ao meu orientador, Prof. Dr. Alexandro Marian Carvalho, que sempre me ajudou na construção dessa dissertação, de forma significativa, sempre me incentivando e apoiando na busca de novos conhecimentos, em especial sobre o universo das redes complexas, ciências de dados e contabilidade.

“O cientista não é o homem que fornece as verdadeiras respostas; é quem faz as verdadeiras perguntas” (LÉVI-STRAUSS, 2021).

RESUMO

A presente pesquisa investigou como a adoção da ciência de dados na Pós-Graduação Brasileira relaciona-se com a formação de mestres e doutores em contabilidade. A relevância da temática recai sobre a importância da disseminação de conhecimentos, que aproxima as universidades do mercado de trabalho contribuindo para o desenvolvimento da sociedade. Diante disso, a pesquisa teve como objetivo determinar os fatores de adoção de conhecimentos em ciência de dados e suas influências na formação de mestres e doutores em contabilidade, evidenciada na literatura baseada na difusão da inovação, redes complexas e ciências de dados na contabilidade. O percurso metodológico procurou explorar o Catálogo de Dissertações e Teses da Capes, referente ao período de 1987 até 2019, que posteriormente foram analisados com a utilização da linguagem computacional (Python e Gephi), mineração de dados, estatísticas descritivas e redes complexas, com objetivo de mapear e identificar as principais pesquisas e módulos criados e desenvolvidos pela rede de co-palavras-chaves, com intuito da expansão da pesquisa sobre a colaboração científica e a evolução da ciência de dados na contabilidade. Os resultados indicam a existência de uma evolução temporal nítida onde foi possível identificar que mais de 70% das pesquisas alinhadas com a área contábil estão concentradas em cinco módulos, formando um padrão a partir de 2013, sendo eles: aprendizagem de máquina, mineração de dados, inteligência artificial, grandes dados e inteligência de negócios. Além disso, a contabilidade financeira ganhou destaque por emergir como um módulo com termos essencialmente contábeis, apontando novos rumos e tendências. Por fim, foi efetuada a proposta estudos futuros, com a própria replicação dessa pesquisa em períodos futuros, e a inserção gradual das técnicas e ferramentas da ciência de dados nos currículos da pós-graduação da contabilidade brasileira, contribuindo de forma significativa na formação de mestres e doutores.

Palavras-chave: ciência de dados; redes complexas; rede de co-palavras-chaves; pós-graduação; contabilidade.

ABSTRACT

This research investigated how the adoption of data science in the Brazilian Graduate Program is related to the formation of masters and doctors in accounting. The relevance of the theme lies in the importance of the dissemination of knowledge, which brings universities closer to the labor market contributing to the development of society. Therefore, the research aimed to determine the factors of adoption of knowledge in data science and its influences on the formation of masters and doctors in accounting, evidenced in the literature based on the diffusion of innovation, complex networks and data sciences in accounting. The methodological path sought to explore the Catalog of Dissertations and Theses of Capes, referring to the period from 1987 to 2019, which were later analyzed with the use of computational language (Python and Gephi), data mining, descriptive statistics and complex networks, with the objective of mapping and identifying the main research emodules created and developed by the network of co-keywords, with the aim of expanding research on scientific collaboration and the evolution of data science in accounting. The results indicate the existence of a clear temporal evolution where it was possible to identify that more than 70% of the surveys aligned with the accounting area are concentrated in five modules, forming a pattern from 2013, being: machine learning, data mining, artificial intelligence, big data and business intelligence. In addition, financial accounting gained prominence for emerging as a module with essentially accounting terms, pointing out new directions and trends. Finally, future studies were proposed, with the replication of this research itself in future periods, and the gradual insertion of data science techniques and tools in the post-graduate curricula of Brazilian accounting, contributing significantly to the training of masters and doctors.

Key-words: data science; complex networks; co-keyword network; graduate studies; accounting.

LISTA DE FIGURAS

Figura 1 - Fluxo de pesquisa.....	20
Figura 2 - Forma S do crescimento de uma inovação.....	23
Figura 3 - Percurso metodológico	35
Figura 4 - Evolução da rede de co-palavras-chaves(1987-2019).....	55
Figura 5 - Redes de co-palavras-chave acumulada no tempo (1987-2019).....	60
Figura 6 - Principais módulos da rede de co-palavras-chaves (1987-2019).....	61
Figura 7 - Conexões da rede de co-palavras-chaves entre a ciências de dados e a ciências contábeis.....	65

LISTA DE GRÁFICOS

Gráfico 1 - Evolução das dissertações e teses de 1987 a 2019	43
Gráfico 2 - Tipo de titulação dos pesquisadores da ciência de dados (1987-2019)..	44
Gráfico 3 - Evolução das dissertações e teses por região para a ciência de dados (1987-2019).....	44
Gráfico 4 - Evolução das dissertações e teses da ciência de dados para a Região Norte (1987-2019).....	45
Gráfico 5 - Evolução das dissertações e teses da ciência de dados para a Região Centro-Oeste (1987-2019)	46
Gráfico 6 - Evolução das dissertações e teses da ciência de dados para a Região Nordeste (1987-2019)	46
Gráfico 7 - Evolução das dissertações e teses da ciência de dados para a Região Sudeste (1987-2019).....	47
Gráfico 8 - Evolução das dissertações e teses da ciência de dados para a Região Sul (1987-2019).....	48
Gráfico 9 - Evolução das dissertações e teses das ciências de dados nas ciências contábeis (1987-2019).....	49
Gráfico 10 - Tipo de titulação dos pesquisadores da ciência de dados nas ciências contábeis (1987-2019).....	49
Gráfico 11 - IES mestrado acadêmico - Produção de Dissertações na contabilidade que se apropriaram da ciências de dados	50
Gráfico 12 - IES mestrado profissional - Produção de Dissertações na contabilidade que se apropriaram da ciências de dados	51
Gráfico 13 - IES doutorado - Produção de Teses na contabilidade que se apropriaram da ciências de dados	52
Gráfico 14 - Número de nós e arestas da Rede de Co-Palavras-Chaves (1987-2019)	53
Gráfico 15 - Tipo de grau da rede de co-palavras-chaves (1987-2019).....	57
Gráfico 16 - Clusterização média da rede de co-palavras-chaves (1987-2019)	58
Gráfico 17 - Menor caminho médio da rede de co-palavras-chaves (1987-2019).....	58
Gráfico 18 - Cinco maiores módulos da rede de co-palavras-chaves (1987-2019)...	59
Gráfico 19 - Coeficiente de estabilidade da rede de co-palavras-chaves (1987-2019)	63

Gráfico 20 - Coeficiente de inovação da rede de co-palavras-chaves (1987-2019)..	64
Gráfico 21 - Coeficiente de estabilidade da rede de co-palavras-chaves na contabilidade.....	66
Gráfico 22 - Coeficiente de inovação da rede de co-palavras-chaves na contabilidade	66

LISTA DE QUADROS

Quadro 1 - Medidas de redes	30
Quadro 2 - Classificação das redes.....	31
Quadro 3 - Palavras-chaves.....	35

LISTA DE SIGLAS

BA	<i>Bussines Analytics</i>
BD	Big Data
BI	<i>Bussiness Intelligence</i>
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
FECAP	Fundação Escola de Comércio Álvares Penteado
FIPECAP	Fundação Instituto de Pesquisas Contábeis, Atuariais e Financeiras
FURB	Universidade Regional de Blumenau
FUCAPE	Fucape Bussiness School
FVC	Fundação Visconde de Cairu
IES	Instituições de Ensino Superior
OCDE	Organização para Cooperação e Desenvolvimento
PUC	Pontificia Universidade Católica de São Paulo
SNA	<i>Social Network Analysis</i>
TIC	Tecnologia da Informação e Comunicação
UFAM	Universidade Federal do Amazonas
UFC	Universidade Federal do Ceará
UNB	Universidade de Brasília
UFES	Universidade Federal do Espírito Santo
MACKENZIE	Universidade Presbiteriana Mackenzie
UFMG	Universidade Federal de Minas Gerais
UFPR	Universidade Federal do Paraná
UERJ	Universidade do Estado do Rio de Janeiro
UFSC	Universidade Federal de Santa Catarina
USP	Universidade de São Paulo
UNISINOS	Universidade do Vale do Sinos

SUMÁRIO

1 INTRODUÇÃO	15
1.1 Tema	16
1.2 Delimitação do tema	16
1.3 Problema	16
1.4 Objetivos	17
1.4.1 Objetivo geral	17
1.4.2 Objetivos específicos.....	17
1.5 Justificativa	17
1.6 Fluxo de pesquisa	20
2 FUNDAMENTAÇÃO TEÓRICA	22
2.1 Difusão da inovação	22
2.2 Ciências de dados na contabilidade	25
2.3 Redes complexas	30
2.3.1 Principais estruturas de redes	32
3 METODOLOGIA	34
3.1 Classificação da pesquisa	34
3.2 Dados: coleta e preparação	34
3.3 Rede de co-palavra-chave	38
3.4 Métricas de redes	38
4 RESULTADOS	42
4.1 Análise descritiva da ciência de dados.....	42
4.1.1 Análise descritiva na contabilidade.....	48
4.2 Visualização das redes	53
4.3 Características evolucionárias topológicas	56
4.4 Estabilidade e inovação	62
5 CONSIDERAÇÕES FINAIS	68
REFERÊNCIAS	72
ANEXO A - SCRIPT CO-PALAVRAS-CHAVES (CATÁLOGO CAPES)	79
ANEXO B - SCRIPT MATRIZ DE CO-PALAVRAS-CHAVES	83

1 INTRODUÇÃO

As organizações (governos, empresas ou entidades), de um modo geral, estão sendo impactadas na era moderna da computação. A utilização de novas tecnologias e o surgimento de grandes volumes de dados tornou-se um grande desafio da atualidade (WAZURKAR; BHADORIA; BAJPAI, 2017).

É notório o crescimento de ferramentas disponíveis que buscam dar esse suporte no processo de análise de dados, bem como a busca da otimização e melhor uso das informações geradas para o desenvolvimento dos negócios e da sociedade. É nesse contexto que surgem, por exemplo, *Business Intelligence* (BI) e o *Business Analytics* (BA) (MASHINGAIDZE; BACKHOUSE, 2017).

As informações disponibilizadas pela análise de grandes volumes de dados, é considerada fator determinante no processo de tomada de decisões para muitas organizações, que buscam a melhoria de seus processos internos e externos, e se preocupam com sua competitividade diante de cenários mercadológicos cada vez mais exigentes (DREMEL *et al.*, 2018). Esta posição das organizações frente ao tratamento dos dados tem alavancado substancialmente a demanda por profissionais competentes, que possuam a capacidade de extrair informações relevantes a partir da análise de dados econômicos, financeiros e sociais. Porém, a relação existente entre a demanda e a oferta desses profissionais ainda encontra desafios, em parte devido à escassez de mão de obra qualificada (BELLOUM *et al.*, 2019).

Atentas a isso, as instituições de ensino superior vêm adicionando cursos, em diferentes modalidades, que incluam conhecimentos em ciência de dados. Concentradas inicialmente em escolas de ciências exatas, estas atividades e programas rapidamente se difundiram em outras escolas (WIXOM *et al.*, 2014). Os programas existentes para a análise de dados são bastante heterogêneos, prevalecendo, muitas vezes, objetivos comuns para ensinar técnicas e habilidades para transformar dados em ideias para tomada de decisões (GORMAN; KLIMBERG, 2014).

A contabilidade, inserida nesse cenário de mudanças tecnológicas que afetam diretamente a profissão contábil, exige que esse profissional se atualize diante da economia em tempo real. Os cursos de graduação e pós-graduação precisam evoluir, pois os contadores necessitam de uma formação que ajude a superar os

novos desafios e, principalmente, que eles possam fazer uso dos métodos ou ferramentas para a extração de dados e informações da melhor forma possível (ALMEIDA, 2020).

No entanto, a rápida evolução das tecnologias e a pronta aceitação do conceito por parte dos setores público e privado deixou pouco tempo para que o discurso se desenvolvesse e amadurecesse no âmbito acadêmico. Isso traz a necessidade da investigação da forma como a colaboração científica, que visa à análise de dados, evolui nas instituições de ensino.

Uma das tendências é a formação de redes de colaboração científica, que objetivam aumentar a disseminação do conhecimento e informação, sendo que, neste estudo, serão investigadas, sob a ótica das redes, em especial, as redes de co-palavras-chaves.

Este estudo se destaca por identificar as redes de co-palavras-chaves em nível de pós-graduação *stricto sensu* que se apropriam de conhecimentos de ciências de dados e como essas informações são inseridas na Contabilidade.

1.1 Tema

Difusão de conhecimento.

1.2 Delimitação do tema

Difusão do conhecimento de ciências de dados na pós-graduação brasileira.

1.3 Problema

Como a adoção de conhecimentos em ciências de dados na pós-graduação brasileira relaciona-se com a formação de mestres e doutores em contabilidade?

1.4 Objetivos

1.4.1 Objetivo geral

Este estudo tem como objetivo principal determinar os fatores da adoção de conhecimentos em ciências de dados e suas influências na formação de mestres e doutores em contabilidade.

1.4.2 Objetivos específicos

Para obtenção da resposta perante a questão abordada, tem-se como objetivos específicos:

- a) determinar a estrutura da rede de palavras-chaves formada entre as pesquisas;
- b) investigar a evolução temporal das interações;
- c) identificar os mecanismos responsáveis pela formação de padrões na estrutura da rede.

1.5 Justificativa

A ascensão da ciência de dados nos últimos anos exigiu que o mercado de trabalho se aperfeiçoasse e investisse em mudanças, como a criação de um novo perfil profissional, dotado de competências no campo de análise de dados massivos para acompanhar o crescimento exponencial do Big Data (BD) (CARILLO, 2017). Em resposta a esse *boom* tecnológico, emergem os cientistas de dados, que buscam o desenvolvimento e aplicação das técnicas e ferramentas advindas das ciências de dados. Ainda, esse profissional contribui na disseminação da informação através de seus conhecimentos específicos para tratamento e análise de grandes volumes de dados (BELLOUM *et al.*, 2019).

Sabe-se que termos antes restritos ao campo da informática passaram a ser mais aplicados também em instituições educacionais e políticas públicas, como ciência de dados, big data, lei geral de proteção de dados pessoais, *data mining*, *learning analytics*, *deep learning*, *chatbots*, sistemas de tutoria inteligente e design

instrucional orientado a dados (FILATRO, 2020), e isso se reflete também em outras áreas, como o mercado de trabalho.

Em detalhes, o mercado de trabalho vem buscando cada vez mais profissionais interdisciplinares que, além de entenderem da linguagem dos negócios, sejam capazes também de entender de linguagem computacional (RIKHARDSSON; YIGITBASIOGLU, 2018). O intuito é que, por intermédio da análise de dados de maneira adequada, seja possível tomar decisões estratégicas e assertivas que ajudem no desenvolvimento dos mais diversos tipos de organizações, a saber: contabilidade, marketing digital, vendas, negócios, produtos e tecnologias (RIKHARDSSON; YIGITBASIOGLU, 2018).

Com a intenção de nivelar os esforços acadêmicos e as necessidades do mercado, nota-se um aumento no número de cursos de graduação e pós-graduação que incluem em seus currículos conhecimentos em ciência de dados (BELLOUM *et al.*, 2019). Nesse sentido, o discurso deve ser motivação de debate e amadurecimento junto às instituições de ensino superior, uma vez que estas são responsáveis pela transmissão do conhecimento científico às pessoas através da sua produção científica, especialmente por meio de seus programas de pós-graduação, de onde resultarão artigos, dissertações e/ou teses, patentes, entre outros.

Além disso, a contabilidade vem sendo impactada pela revolução tecnológica, afetando diretamente o mercado de trabalho e o cotidiano dos profissionais e docentes da área da contabilidade. Sendo assim, a ciência de dados deve avançar com o uso de novas tecnologias, programação em linguagens e *softwares* para aperfeiçoamento dos procedimentos metodológicos e pesquisas (ALMEIDA, 2020).

Na área de negócios, por exemplo, muitas atividades manuais foram substituídas por sistemas informatizados, o que possibilitou a integração com os diversos setores e atividades da empresa. Consequentemente, o papel do gestor também mudou. É necessário que ele saiba lidar com a tecnologia, utilizar *softwares* de gestão, que podem inclusive auxiliá-lo no processo de tomada de decisões (MIRANDA, 2018).

Se, no meio profissional, há diversos reflexos do avanço da tecnologia, é natural pensar que os profissionais da contabilidade, que estão atuando e que atuarão futuramente nesse meio, necessitem de uma formação diferenciada, de novos conhecimentos e habilidades (MIRANDA, 2018).

Sendo assim, o presente estudo torna-se relevante, pois busca o desenvolvimento da educação, em especial, da pós-graduação brasileira, visto que é nas universidades que se concentra a maior fonte de geração de conhecimentos, objetivando atender sua missão institucional e social no que se refere à pesquisa, ensino e extensão. Entretanto, não basta somente produzir informação e/ou conhecimento em ciência e tecnologia: é necessário que sua disseminação seja efetiva.

A contribuição desta pesquisa é a identificação das pesquisas e sua inserção e interação junto à comunidade científica, bem como abrir espaço para discussões sobre o processo de difusão da informação por intermédio de análise de redes de co-palavras-chaves e seus reflexos na contabilidade.

A observação e a compreensão da rede de co-palavras-chaves e suas relações são úteis para gerar conhecimentos a respeito da estrutura e desenvolvimento do sistema. No contexto da estrutura, uma topologia complexa está relacionada a uma heterogeneidade na capacidade e intensidade das conexões. Serrat (2017) corrobora ao explicar que os fenômenos de sincronização são relevantes para promover uma melhor compreensão dos mecanismos subjacentes à formação de comportamentos sociais coletivos, como o surgimento de modas e tendências.

A oportunidade do estudo recai sobre a questão do comportamento da oferta da mão de obra qualificada já existente (mestres e doutores) que podem ajudar nas discussões e desenvolvimento dessa temática em nosso país, com aumento e consolidação das pesquisas e abrir espaço para novos estudos, com objetivo da formação de recursos humanos qualificados, a fim de dar cobertura da demanda de mercado por cientistas de dados, em nível nacional e internacional.

Em relação à acessibilidade, a pesquisa é viável pela disposição dos dados junto ao Catálogo de Teses e Dissertações da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Os chamados Dados Abertos CAPES configuram-se como uma plataforma web que dispõe de informações (instituição, discente, orientador, banca, título, resumo etc.) sobre teses e dissertações apresentadas a programas de pós-graduação brasileiros.

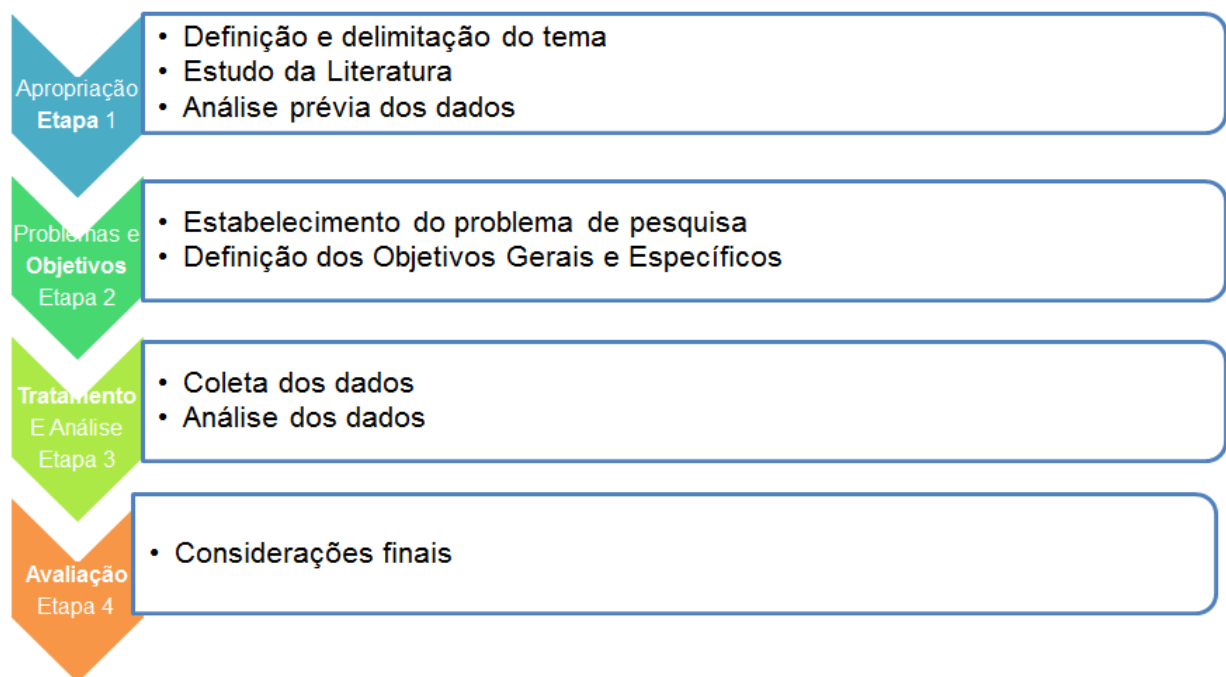
Em síntese, este estudo justifica-se por buscar identificar a dinâmica da oferta de profissionais com nível de pós-graduação (mestres e doutores) cuja formação possua uma projeção em ciências de dados. Para tanto, propõe-se um mapeamento

espacial e temporal dos discentes, docentes e instituições envolvidos no processo e, posteriormente, uma análise da rede de co-palavras-chaves e suas contribuições para a contabilidade brasileira.

1.6 Fluxo de pesquisa

Nesta seção, exibe-se as etapas do fluxo da pesquisa já realizados, conforme demonstra a figura abaixo.

Figura 1 - Fluxo de pesquisa



Fonte: Elaborada pelo autor (2021).

A etapa 1, denominada de apropriação, é onde aconteceu a definição e delimitação do tema da pesquisa que aborda a difusão dos conhecimentos em ciências de dados na pós-graduação brasileira. O estudo das literaturas é uma investigação permanente e constante de estudos que abordem a difusão da inovação, redes complexas por intermédio da rede de palavras-chaves, e a ciência de dados na contabilidade, além da análise prévia dos dados disponíveis no portal de dados abertos da CAPES no Catálogo de Dissertações e Teses (CAPES, 2021).

A etapa 2 é onde foi definido o problema de pesquisa que procura responder ao seguinte questionamento: Como a adoção de conhecimentos em ciências de

dados na pós-graduação brasileira relaciona-se com a formação de mestres e doutores em contabilidade?

A etapa 3, chamada de tratamento e análise, é onde ocorre a coleta e análise dos dados. Essa parte se desenvolverá no Catálogo de Dissertações e Teses da Capes com o levantamento das pesquisas relacionadas com o estudo das ciências de dados e, posteriormente, com a utilização de mineração de dados, linguagem computacional e análise da rede de co-palavras-chaves, suas principais características topológicas e métricas utilizadas.

A etapa 4 é a avaliação onde ocorre as considerações finais, discorrendo-se sobre os principais resultados, limitações e sugestões para estudos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, apresenta-se três seções que destacam os estudos nos quais se apoia a pesquisa. As primeiras três seções são destinadas à apropriação de conceitos associados à difusão da inovação, ciências de dados na contabilidade e redes complexas.

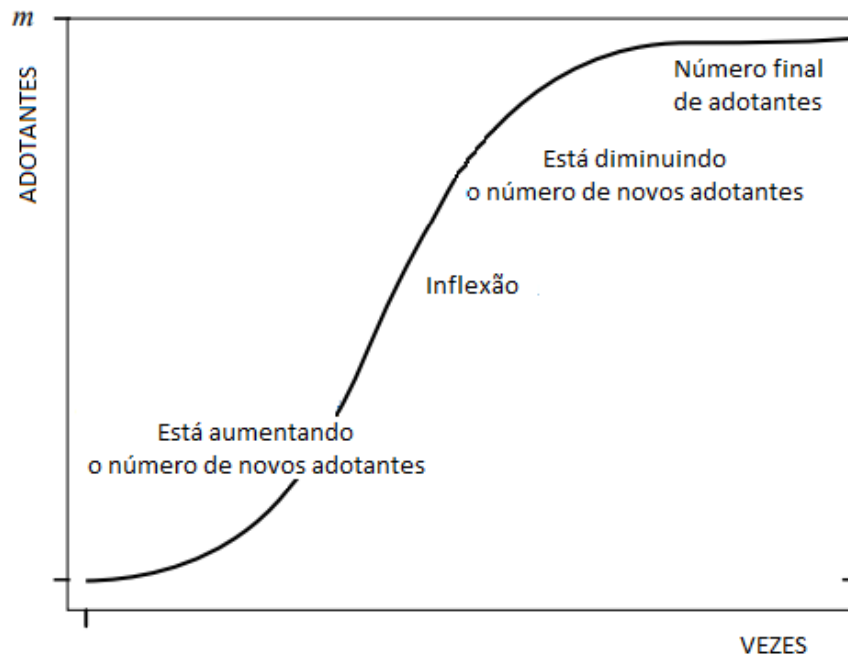
2.1 Difusão da inovação

Segundo Schumpeter (1912), a dinâmica da difusão da inovação divide-se em três fases, sendo elas: a invenção; a inovação e a difusão. A invenção, como uma ideia com potencial para ser explorada; a inovação, com fins voltados para a exploração comercial; e a difusão, alinhada com a propagação de novos produtos e processos pelo mercado (SCHUMPETER, 1912).

O principal conceito da adoção de inovações é dado pela Teoria de Difusão da Inovação (ROGERS, 2003). De acordo com Rogers (2003, p. 5), a difusão da inovação é dada como “[...] o processo pelo qual uma inovação é comunicada através de certos canais, ao longo do tempo, entre os membros de um sistema social”. Complementar a este conceito, Hall (2004, p. 17) traz, ainda, que a difusão de uma inovação é um fenômeno em que indivíduos (ou organizações) “[...] adotam uma nova tecnologia ou substituem uma velha tecnologia por uma nova”, sendo a escolha da adoção o resultado do reconhecimento individual quanto à necessidade da inovação.

De forma mais detalhada, os estudos de Rogers (2003) revelam que há um comportamento padrão na difusão das inovações, denominado curva *S*. Esta assinatura universal, evidenciada na Figura 2, deve-se aos diferentes agentes ao longo do processo. Na compreensão do autor, a população de adotantes está dividida em cinco classes: inovadores, adotantes precoces, maioria inicial, maioria final, retardados. As três primeiras classes configuram o que é denominado de adotantes iniciais e as duas últimas de adotantes atrasados. O sucesso da disseminação de uma inovação depende de sua condução no grupo dos adotantes iniciais. Isso porque é este conjunto de agentes que serão responsáveis por implantar e divulgar a inovação dentro do sistema (ROGERS, 2003).

Figura 2 - Forma S do crescimento de uma inovação



Fonte: Adaptada pelo autor de Rogers (2003).

Na primeira vez que esse modelo foi estudado, tratou-se sobre a adoção e difusão que ocorreu com produtores de milho híbrido no Estados Unidos, em estudo realizado por Griliches (1957). Esse estudo demonstrou que a expectativa dos produtores com a difusão do milho híbrido foi a lucratividade advinda dessa inovação, com intuito de ajudar no desenvolvimento dos negócios (VIEIRA FILHO; SILVEIRA, 2012).

Outro aspecto importante é a complementação de Mansfield (1961) ao estudo iniciado por Griliches (1957), que inclui dois pontos a serem aprofundados: o risco e as assimetrias da informação. O risco e a assimetria da informação são medidos pelo número de adotantes, sendo que, se o número for pequeno, indica riscos elevados e a difusão ocorre de forma lenta, e, com o crescimento desse número, as assimetrias e os riscos se reduzem e ocorre o aumento da difusão (FURTADO, 2006).

Furtado (2006) comenta que novos modelos com características mais dinâmicas foram propostos por David (1969), Davies (1979) e Stoneman (1983), modelos esses que levam em consideração as variáveis comportamentais heterogêneas da firma pelos adotantes, sejam elas o seu tamanho, produtos e processos de gestão, incertezas e a maximização da função utilidade.

Seguindo, outros modelos de maneira mista podem ser destacados, como os de Bass (1969) e Metcalfe (1981), que investigaram a adoção da difusão não mais sob a ótica da firma com sua decisão de ofertar, e, sim, pela ótica do consumidor final com sua decisão de consumir, apontando para dois momentos distintos da difusão, ou seja, a escolha das instituições em ofertar novas tecnologias e a escolha do consumidor final em sua adoção.

No modelo de Bass (1969), pode-se destacar que os inovadores buscam determinada autonomia na adoção ou não de um novo produto no meio onde estão inseridos. Já os imitadores são responsáveis pela massificação e pelas relações interpessoais no processo de difusão. A velocidade de propagação da difusão depende do padrão de comportamento estabelecido entre inovadores e imitadores.

Outros modelos considerados evolucionistas também podem ser destacados, como o de Rosenberg (1976), que aborda a difusão sob uma sequência de inovações incrementais, ou seja, com melhorias contínuas que modelam as inovações iniciais e sua adaptabilidade diante das diferentes formas de uso. Silveberg (1988), sob a ótica da oferta, defende que a difusão é estimulada pela aprendizagem pelo uso, afetando as expectativas de rentabilidade das organizações.

Diante disso, as teorias têm dividido o posicionamento dos estudos referentes entre a oferta e demanda, em que, para este estudo, destaca-se Rogers (1962) pelo formato dinâmico dado a difusão das inovações, em especial, ao comportamento dos consumidores finais. Conforme o autor, a difusão depende de quatro pontos de destaque, sendo eles: a) a inovação em si; b) a comunicação; c) o período de tempo; e d) o sistema social.

A inovação, sob ótica de Rogers (1962), possui cinco características, que são percebidas pelos agentes sociais, consideradas determinantes para o processo de adoção das difusões, sendo elas: a) vantagens comparativas da inovação diante de produtos existentes; b) compatibilidade com valores, normas e necessidades específicas; c) grau de complexidade da inovação; d) possibilidades de teste pelos adotantes iniciais; e e) facilidade de avaliação da inovação.

A relação dos estudos de Bass (1969) e Rogers (1962) são convergentes no que tange à questão das trocas interpessoais ou exposições midiáticas de forma positiva em relação à taxa de adoção.

Outro aspecto a ser destacado é que a variável tempo é determinante quando o indivíduo forma sua opinião sobre determinada inovação, podendo aceitá-la ou rejeitá-la, de forma que, ao implementar sua decisão, pode confirmar ou reformular sua escolha (ROGERS, 1962).

Ainda cabe ressaltar que o sistema social é outro ponto importante que traz à tona a questão dos grupos sociais aos quais os indivíduos estão inseridos conforme suas condutas, normas e padrões de comportamentos preestabelecidos, o que impõe regularidade ao comportamento individual. Nesse sentido, sabe-se que as decisões individuais e o comportamento influenciam o processo de difusão ao longo do tempo, sendo que determinados hábitos podem facilitar ou dificultar a difusão das inovações de maneira positiva ou negativa em relação à sua taxa de adoção (ROGERS, 1962).

2.2 Ciências de dados na contabilidade

Segundo Sá (1997, p. 16), “[...] a contabilidade nasceu com a civilização e jamais deixará de existir em decorrência dela; talvez, por isso, quase sempre seus progressos coincidiram com aqueles que caracterizam os da evolução do homem”.

Por outra ótica, Cosenza e Rocchi (2014) sustentam que a longevidade da contabilidade ultrapassa vários estágios de tempos e sempre esteve correlacionada com o nível de progresso econômico e tecnológico; e é a partir dessa evolução que são necessárias as mudanças, para que os processos se tornem mais seguros e ágeis.

Historicamente, sabe-se que a contabilidade passou por diversas transformações de processos realizados manualmente para o que hoje chama-se de padronização de informações. Essa padronização foi possível devido aos grandes avanços tecnológicos, em especial, no que tange à tecnologia da informação (TI). Com isso, contadores e empresários utilizam *softwares* adequados, capazes de suprir as necessidades impostas pela legislação contábil, societária, fiscal e trabalhista, para tornarem estas informações úteis para a tomada de decisão (RUSCHEL; FREZZA; UTZIG, 2011).

Sob a ótica de Arruda, Gomes e Santos (2013), a TI está cada vez mais presente na rotina do contador, pois o que antes era realizado completamente de

forma manuscrita, hoje em dia é lançado em sistemas, tornando-se necessário mais cuidado ao tratar de um dado ou informação.

Nesse sentido, a contabilidade evolui com o passar do tempo para acompanhar as transformações advindas da revolução tecnológica, que trouxe muitas facilidades para acompanhar o desenvolvimento das organizações, como, por exemplo: o big data, a mineração de dados e a aprendizagem de máquina (AGUIAR; GOUVEIA; RODRIGUES, 2021).

O BD terá implicações cada vez mais importantes para a contabilidade, mesmo que novos tipos de dados se tornem acessíveis. As informações de vídeo, áudio e texto disponibilizadas via BD podem fornecer melhores práticas de contabilidade gerencial, contabilidade financeira e relatórios financeiros. Na contabilidade gerencial, o BD contribuirá para o desenvolvimento e evolução de sistemas de controle gerencial e processos orçamentários eficazes. Na contabilidade financeira, o BD melhorará a qualidade e a relevância das informações contábeis, aumentando, assim, a transparência e a tomada de decisões das partes interessadas. Nos relatórios, o BD pode ajudar na criação e refinamento de padrões contábeis, ajudando a garantir que a profissão contábil continue a fornecer informações úteis à medida que a economia global dinâmica e em tempo real evolui (WARREN JR; MOFFITT; BYRNES, 2015).

A contabilidade é considerada a base de qualquer organização e com uma ampla gama de tarefas, incluindo relatórios internos e externos, custos, estimativas, avaliações, análises e auditorias. Sendo assim, a mineração de dados destaca-se por fornecer ferramentas avançadas para a análise de negócios e suporte para tomada de decisões (AMANI; FADLALLA, 2017).

Sabe-se que as aplicações da mineração de dados em contabilidade geralmente são focadas em uma determinada área específica, como: contabilidade financeira, contabilidade gerencial e/ou técnicas da mineração de dados, destacadas por alguns pesquisadores (CALDERON; CHEC, 2002; COAKLEY; BROWN, 2000; NGAI *et al.*, 2011; WANG, 2010; WANG; YANG, 2006).

Além disso, a mineração de dados ajuda as organizações a explorarem melhor suas informações e conhecimentos existentes em seus bancos de dados disponíveis. Mas é apenas uma ferramenta: não elimina a necessidade de se conhecer os negócios, entender os dados ou compreender os métodos analíticos envolvidos (JACKSON, 2002).

A aprendizagem de máquina é uma área importante da inteligência artificial. O objetivo da aprendizagem de máquina é descobrir conhecimento e tomar decisões inteligentes. Algoritmos de aprendizagem de máquina podem ser categorizados em supervisionado, não supervisionado e semi-supervisionado. Quando o big data está em causa, é necessário para escalar algoritmos de aprendizagem de máquina (CHEN; ZHANG, 2014; TARWANI; SAUDAGAR; MISALKAR, 2015).

Nesse sentido, a aprendizagem de máquina trata-se da construção de sistemas que possuem a capacidade de aprender, a partir de dados, a identificar padrões e ir adquirindo a capacidade de prever, a partir destes, os resultados futuros, possibilitando a tomada de decisão. Essa tecnologia usa de pouca intervenção humana, pois o seu processo de aprendizagem pode ou não acontecer sob supervisão humana (STODDER, 2018).

Sendo assim, a contabilidade, diante das transformações advindas pela revolução tecnológica e pelo grande volume de dados e informações disponíveis, enfrenta um grande desafio na atualidade, no que tange a seus processos, pessoas, tecnologias, tomada de decisões e, em especial, como isso pode afetar sua competitividade e inovação (MORRAR; ARMAN; MOUSA, 2017).

As inovações tecnológicas inseridas no mercado contábil fazem com que o profissional da contabilidade se prepare para atender as organizações em suas necessidades informacionais (ASSOCIATION TO ADVANCE COLLEGIATE SCHOOLS OF BUSINESS (AACSB), 2018; HOELSCHER; MORTIMER, 2018; PRICEWATERHOUSECOOPERS (PWC), 2015).

A apropriação de técnicas e ferramentas da ciência de dados demonstra a necessidade de tais conhecimentos serem inseridos gradualmente nos currículos dos cursos de graduação e pós-graduação em contabilidade, preparando o profissional contábil diante dessa nova realidade inserida pela revolução tecnológica (AACSB, 2018; PWC, 2015).

A literatura tem demonstrado que a inclusão pode ocorrer da seguinte maneira: (1) focada: com uma disciplina específica para lecionar tal conteúdo; (2) integrada: incluindo os conteúdos de análise de dados dentro das disciplinas contábeis; ou (3) híbrida: uma abordagem mista das duas anteriores (DZURANIN; JONES; OLVERA, 2018).

Isso ajudará no desenvolvimento do profissional contábil, além de possibilitar a criação de expertise ao lidar com os dados ou informações. Sabe-se que o

contador é responsável por registrar, filtrar, resumir e consolidar esses dados para fornecer informações para os tomadores de decisão internos e externos (JANVRIN; WATSON, 2017).

No meio empresarial, as organizações vêm gradativamente usando parte desses grandes dados à sua disposição, no entanto, muitas vezes, isso é pouco aproveitado para a tomada de decisão. Em primeiro lugar, isso ocorre porque muitas informações nem sempre são controladas nos sistemas de gestão e, em segundo lugar, é necessário treinamento para compreender como a ciência de dados pode contribuir com as profissões da área de negócios, em especial, para a contabilidade (ALMEIDA, 2020).

Diante disso, esse desafio é, então, estendido aos educadores, que têm a missão de ensinar e, ao mesmo tempo, preparar-se para essa nova realidade em fazer a integração da ciência de dados com a contabilidade. Em contrapartida, os profissionais da contabilidade precisam aprender a desenvolver habilidades para a extração de valor dos dados por meio de análises avançadas, disponíveis pelas técnicas ou ferramentas da ciência de dados (LAWLESS; PELLEGRINO, 2007).

A educação contábil precisa repensar sua forma de ensinar as ciências de dados, a fim de desenvolver profissionais que sejam capazes de aplicar uma gama de competências integradas (LAWSON *et al.*, 2014).

Em contrapartida, os graduandos e pós-graduandos em contabilidade precisam de novas perspectivas e devem preparar-se para o mercado de trabalho, que exige esse novo perfil profissional que saiba extrair *insights*, dados e informações econômicas e financeiras das organizações, a partir da análise de dados disponível pelas técnicas e ou ferramentas da ciência de dados (COYNE, J.; COYNE, E.; WALKER, 2016).

Estudos como os de Tysiac e Drew (2018) ressaltam que os líderes de companhias com visão de futuro na área contábil avaliam os avanços na análise de dados (*data analytics/big data*), inteligência artificial e da tecnologia *blockchain* em seus modelos de negócios.

No que tange à profissão contábil e à auditoria, por exemplo, Vasarhelyi Teeter e Krahel (2010) destacam a auditoria contínua em uma economia em tempo real. Os autores afirmam que o currículo dos cursos de graduação precisa evoluir, pois o auditor necessita de uma formação que o ajude a superar novos desafios, e a auditar em tempo real seus clientes.

Além disso, tratam do desenvolvimento de atividades essenciais, destacando três habilidades primárias: atitude (ética, habilidade para mudança e adaptação); competência técnica (conhecimentos básicos de tecnologia da informação, ter certificações etc.); e competência comportamental (relacionamento com cliente, trabalho em equipe, lidar com reguladores etc.) (VASARHELYI; TEETER; KRAHEL, 2010).

Outro estudo interessante é o de Ferreira *et al.* (2012), no qual os autores analisam a abordagem de diferentes conceitos de Inteligência Artificial, incluindo conceitos aplicáveis à Mineração de Dados (MD), apontando a importância em abordar tais conteúdos devido ao crescente uso de ferramentas de MD.

No que tange à utilização de ferramentas, no estudo de King e Stayanarayana (2013), são discutidas questões relevantes no ensino de MD em cursos de graduação, incluindo os principais tópicos que uma disciplina deve abordar.

Na pesquisa contábil, novas abordagens e interações com o uso de técnicas de extração, mineração e tratamento de dados podem agregar valor aos estudos, ao permitir interações entre bases de dados comumente utilizadas e novas variáveis, que seriam complexas para coleta manual ou tomariam muito tempo. As novas ferramentas e códigos desenvolvidos em sistemas de tratamento de dados e estatísticos ou linguagens de programação, por exemplo, Stata, Python e R, entre outros, com seus pacotes, podem auxiliar pesquisadores que utilizam tanto a abordagem quantitativa como a qualitativa, gerando gráficos, processando citações de artigos, rodando regressões com diversos estimadores, novas especificações de modelos e do design de pesquisa, análises textuais e de imagens, enfim, uma infinidade de aplicações que precisamos trazer para nossa área de pesquisa e que contribuem para a replicação dos estudos para verificação e aprendizagem de futuros pesquisadores (ALMEIDA, 2020).

Sendo assim, a profissão contábil precisa se adaptar às tecnologias e às novas demandas, para que os controles internos possam ser mais bem desenhados e mais eficazes, com mais riscos mapeados, e para que os relatórios elaborados pela contabilidade sejam mais transparentes e com maior utilidade aos usuários internos e externos. Esse cenário é plausível porque os profissionais contábeis precisam aplicar as bases de mensuração das IFRS, que exigem conhecimentos diversos, como matemática financeira, finanças e algumas aplicações de estatística (por exemplo, quando se discute provisões, contingências, garantias, riscos etc.) (ALMEIDA, 2020).

2.3 Redes complexas

A definição de rede é qualquer sistema que admite uma representação matemática abstrata, como um grafo cujos nós (vértices) identificam os elementos do sistema e em que o conjunto de elos de conexão (arestas) representa a presença de uma relação ou interação entre esses elementos (BARRAT; BARTHELEMY; VESPIGNANI, 2008). As redes estão por toda parte, e são exemplos: as redes de energia elétrica, a *world wide web*, a internet, as redes de telecomunicação, as rodovias, os sistemas de metrô, as redes neurais, as redes sociais e as redes de coautoria científica (BOCCALETTI *et al.*, 2006).

Historicamente, o estudo de redes foi difundido pela matemática desde seu nascimento em 1736, quando o matemático suíço Leonhard Euler introduziu a teoria dos grafos que ajudou na solução para o problema da ponte *Konigsberg* (BOCCALETTI *et al.*, 2006). Formalmente, uma rede complexa pode ser representada como um grafo.

Existe uma série de medidas para auxiliar no entendimento da rede. Mediante às métricas, os dados da rede podem ser organizados para quantificar a estrutura e os padrões das relações entre os nós (METZ *et al.*, 2007).

No Quadro 1, exibe-se algumas medidas e suas descrições.

Quadro 1 - Medidas de redes

(continua)

Medida	Descrição
Grau	Número de arestas conectadas a um nó.
Força	Soma dos pesos de todas as arestas conectadas a um nó.
Excentricidade	Número máximo de arestas entre um nó e qualquer outro.
Distância média	Número médio de ligações de um nó a todos os outros nós.
Triângulo	Número de vizinhos de um nó que também são vizinhos uns dos outros.
Coefficiente de agrupamento	Fração de triângulos presentes ao redor de um nó.
Centralidade de proximidade	Inverso do comprimento do caminho de um nó.
Centralidade de intermediação	Fração de todos os caminhos mais curtos no grafo que passam por um nó. Nós com altos valores de centralidade e intermediação participam de muitos caminhos curtos.
Eficiência global	Média do comprimento do caminho mais curto inverso de um nó para todos os outros nós.

(conclusão)

Medida	Descrição
Eficiência local	Eficiência global de um nó calculada no subgrafo criado pelos vizinhos do nó.
Modularidade	Até que ponto um grafo pode ser dividido em comunidades claramente separadas (ou seja, subgrafos ou módulos).
Pontuação z dentro do módulo	Extensão à qual um nó está conectado a outros nós na mesma comunidade. É uma versão de grau dentro do módulo.
Coefficiente de participação	Quantifica a relação entre o número de arestas conectando um nó fora de sua comunidade e seu número total de arestas.
Coefficiente de assortatividade	Está associado ao coeficiente de correlação entre os graus/forças de todos os nós em extremidades opostas de uma aresta.

Fonte: Adaptado pelo autor de Mijalkov *et. al.* (2017, tradução nossa).

O aprofundamento da teoria das redes complexas, que se fundamenta na Física Estatística, pesquisa operacional e matemática, auxilia na modelagem das redes presentes na natureza ou na sociedade, servindo como motivação para a exploração das medidas, tipos e principais estruturas e sua aplicabilidade (ALBERT; BARABASI, 2002; BOCCALETTI *et al.*, 2006; COSTA *et al.*, 2007). A classificação das redes é descrita sob diferentes aspectos, os quais são apresentados no Quadro 2.

Quadro 2 - Classificação das redes

Classificação das Redes	Descrição
Direcionada	Refere-se ao fato de as arestas expressarem sentido ao fluxo de informação.
Não direcionada	Refere-se ao fato de as arestas não expressarem sentido ao fluxo de informação.
Ponderada	Indica a existência de um número que representa cada conexão entre os nós da rede.
Não ponderada	Indica a não existência de um número que representa cada conexão entre os nós da rede, sendo que todas as conexões têm a mesma importância.
Esparsas	Possuem em cada vértice uma quantidade relativamente pequena de ligações, sendo que é comum que o número vizinho de cada vértice seja constante e independente do número total de vértices.
Densas	Cada vértice se liga a uma parcela significativa do total de vértices, sendo frequente o número de ligações de cada vértice ser proporcional ao número total de vértices da rede.
Conectadas	Não existem nós isolados ou mais de um <i>cluster</i> na rede, ou seja, toda a rede é conectada em um grande <i>cluster</i> .
Não conectadas	Existe mais de um cluster na rede, indicando existência de nós isolados na rede.
Dinâmica	É aquela que evolui ao longo do tempo, em que vértices/conexões são influenciados por diferentes períodos.
Estática	Não evolui ao longo do tempo, vértices/conexões não mudam.

Fonte: Adaptado pelo autor de Mello *et al.* (2010).

É oportuno complementar que as informações do quadro acima sobre a importância de se classificar as redes corretamente depende de outros conceitos interligados, como é o caso de *cluster* que é um conglomerado de nós em que todos os nós estão conectados (BOCCALETTI *et al.*, 2006).

Com intuito de colaborar com as discussões sobre redes complexas, destaca-se o estudo intitulado “Redes sociais científicas e inteligência artificial – uma revisão sistemática aplicada a reconhecimento de padrões”, que objetivou verificar estudos na área de reconhecimento de padrões com redes sociais científicas envolvendo estudos dos cientistas (SOMBRA *et al.*, 2020).

2.3.1 Principais estruturas de redes

O conhecimento de determinadas estruturas de redes é necessário para facilitar o entendimento sobre elas, destacando-se as redes aleatórias, redes de pequeno mundo e redes de livre escala.

As redes aleatórias são mais fáceis de compreender. Há um modelo proposto por Erdős e Rényi em que as arestas não direcionadas são adicionadas aleatoriamente entre um número fixo de N vértices (BARABASI; ALBERT, 1999). Por exemplo, os vértices de uma determinada rede podem ter conexões distribuídas de forma uniforme e com mesma probabilidade de receberem novas ligações.

As redes de pequeno mundo, baseadas nos estudos de Watts e Strogatz (1998), são grafos em que a maioria das arestas são compartilhadas por nós vizinhos e o número de conexões entre dois nós, eleitos ao acaso, percorre um pequeno número de nós. Trata-se de uma estrutura que combina características das redes regular e aleatória. Em nível local, aproxima-se de uma rede regular (variabilidade no número de conexões uniforme). Já na perspectiva global, comporta-se como uma rede aleatória (baixo número de vértices intermediando a conexão de dois nós arbitrários). A compreensão do efeito pequeno-mundo ajuda no entendimento da dinâmica dos processos das redes. Por exemplo, no experimento realizado por Stanley Milgram (METZ *et al.*, 2007) uma carta foi entregue a um indivíduo numa população de ordem de milhão de habitantes. Pediu-se a ele que repassasse a um outro e assim por diante, até que a carta chegasse ao destinatário. O resultado do exercício apontou que são necessários, em média, seis pessoas para atingir o indivíduo alvo. Em outras palavras, dois indivíduos arbitrários que não se

conhecem possuem seis graus de separação (o que reforça a ideia de que o mundo é pequeno).

As redes livres de escala possuem poucos vértices altamente conectados, conhecidos como *hubs*, além de muitos vértices com poucas conexões (METZ *et al.*, 2007). Uma das características principais é a tendência de um novo vértice se conectar a um vértice de rede que tem grau elevado de conexões. A utilização das redes livres de escala é observada em vários sistemas como: na internet, na web, em redes de metabolismo e nas redes de citação de artigos científicos (NEWMAN, 2003).

3 METODOLOGIA

Neste capítulo, apresenta-se o percurso metodológico empregado neste estudo. São apresentadas as quatro seções, que se subdividem em: classificação da pesquisa, em que foi possível classificá-la de maneira apropriada, posteriormente, a coleta e preparação dos dados, etapa em que foi efetuado o levantamento das informações necessárias para a investigação das palavras-chaves sobre ciências de dados e como a contabilidade se apropriou desses conhecimentos em suas áreas de atuação, bem como a formação e organização da base de dados. Seguindo, têm-se a rede de co-palavras-chaves e, finalmente, a análise da rede, com apresentação de algumas métricas, como: a) grau médio; b) grau médio ponderado; c) clusterização média; e d) menor caminho médio. Além disso, são apresentados dois coeficientes: a) estabilidade e b) inovação.

3.1 Classificação da pesquisa

A classificação da pesquisa é, quanto à natureza, uma pesquisa aplicada; quanto à abordagem, trata-se de uma abordagem quantitativa; quanto aos objetivos, exploratória; e, quanto ao procedimento técnico, tem-se um estudo de caso.

3.2 Dados: coleta e preparação

A coleta e a preparação dos dados estão apresentadas na Figura 3, evidenciando a base de dados utilizada, os critérios de seleção e a análise das palavras-chaves, a mineração de dados, a linguagem computacional, a análise descritiva e as redes de co-palavras-chaves.

Figura 3 - Percurso metodológico



Fonte: Elaborada pelo autor.

A base de dados do estudo é o Catálogo de Teses e Dissertações da CAPES (CAPES, 2021), uma base pública com informações das dissertações e das teses defendidas em programas de pós-graduação brasileiros. Trata-se de uma coleção estruturada que apresenta cinquenta e seis (número de colunas do documento) dados (autor, orientador, instituição etc.) nos anos de 1987 a 2019 – totalizando 1.232.795 (número de defesas) produções.

Os dados utilizados foram disponibilizados por intermédio da Plataforma Sucupira, em Dados e Estatísticas, Dados Abertos da CAPES (2021).

A escolha das palavras-chaves encontra-se baseada na literatura dos termos mais frequentes relacionados com ciência de dados. A lista reúne as principais expressões citadas nos trabalhos de Gibert *et al.* (2018), Larson e Chang (2016), Alloghani *et al.* (2020) e Nasution *et al.* (2020), as quais estão destacadas no Quadro 3.

Quadro 3 - Palavras-chaves

(continua)

Palavras-chaves	Palavras-chaves raiz
Inteligência artificial	inteligenc artific
Grandes dados	grand dad
Análise de negócios	anal negoci
Inteligência de negócios	inteligenc negoci
Ciências de dados	cienc dad

(conclusão)

Palavras-chaves	Palavras-chaves raiz
Análise de dados	analis dad
Engenharia de dados	engenh dad
Gestão de dados	gesta dad
Conjunto de dados	conjunt dad
Mineração de dados	minereca dad
Visualização de dados	visualizaca dad
Modelagem de dados	model dad
Organização de dados	organizaca dad
Árvore de decisão	arv decisa
Aprendizagem profunda	aprend profun
Aprendizado de máquina	aprend maquin
Aprendizagem por reforço	aprend reforc
Aprendizagem supervisionada	aprend supervision
Aprendizagem não supervisionada	aprend nao supervision
Dados não estruturados	dad nao estrutur

Fonte: Elaborado pelo autor.

Os arquivos utilizados podem ser obtidos em dois formatos, CSV ou Excel, e foram analisados e selecionados no formato de Excel, com utilização da mineração de dados, que objetivou a organização, padronização e estruturação dos dados.

Nesse sentido, inseriu-se a linguagem computacional com o uso do Python, que é uma linguagem de programação criada pelo holandês Guido Van Rossum, na década de 1980, sendo considerado um projeto de desenvolvimento de código-fonte aberto, ou seja, que permite o estudo de modificação livre. É uma linguagem que, de forma precisa, torna os programas mais legíveis, possui uma vasta biblioteca, que disponibiliza vários algoritmos que auxiliam no desenvolvimento de programas que permitem o desenvolvimento de aplicações robustas (PERVOVIK, 2016).

Sendo assim, com a utilização do Python, foi possível investigar 20 (vinte) palavras-chaves por meio de *script 1* (ver Anexo A) que encontrou os dados de forma mais rápida do que se fosse efetuada manualmente no Excel, por exemplo. Os termos aplicados na busca, em português, possibilitam a busca das palavras-chaves (formato raiz), além das palavras-chaves predeterminadas que estejam em inglês e que são traduzidas para português conforme indicadas no Quadro 3.

A linguagem computacional requer um padrão das palavras-chaves para que possam ser executadas nos *scripts* do Python de forma prática e ágil. Sendo assim,

a mineração de dados é importante, pois é nela que ocorrem as exclusões, como: palavras-chaves repetitivas, erros de digitação, sinais de pontuação, palavras-chaves em inglês, entre outras.

Em maiores detalhes, os dados gerados pelo Phyton auxiliaram na definição da amostra da pesquisa, que se refere ao período de 1987 a 2019, gerando arquivos no formato de Excel denominados de palavras-chaves, ano a ano, de onde foi possível extrair as informações referentes às palavras-chaves no que compete a ciência de dados.

Ainda assim, foi utilizada a Análise Descritiva, que por meio de métodos da estatística descritiva serviram para organizar, resumir e descrever os aspectos importantes de um conjunto de características observadas ou comparar tais características entre dois ou mais conjuntos. Além disso, as ferramentas descritivas são os muitos tipos de gráficos e tabelas e também medidas de síntese, como porcentagens, índices e médias (REIS, 2002).

Essas informações permitiram que fossem efetuadas algumas análises descritivas, como, por exemplo: evolução das dissertações e teses, tipos de titulação dos pesquisadores da ciência de dados, evolução das pesquisas por região para a ciência de dados e evolução das dissertações e teses das ciências de dados nas ciências contábeis.

Logo após, os arquivos formados pelas palavras-chaves, ano por ano, buscaram apoiar a construção e formação das redes de palavras-chaves e sua evolução ao longo do tempo, apresentando, dessa forma, as principais características topológicas da rede.

Para construção e visualização da rede, foi utilizado o *software* Gephi. Segundo Marquez *et al.* (2013, p. 74), “o Gephi é um programa gratuito e de código aberto, desenvolvido para analisar e criar visualizações específicas de quaisquer gráficos, mostrando importantes informações sobre eles”.

Sendo assim, foram criados via script 2 (ver Anexo B), no Phyton, arquivos denominados *togephi*, que posteriormente foram importados para o Gephi, ano a ano, que serviram para construção e visualização da rede de co-palavras-chaves. Isso possibilitou o acompanhamento da evolução das pesquisas e os principais módulos criados pelas palavras-chaves, que sustentam e apontam os principais resultados e tendências para o desenvolvimento da colaboração científica, formada

pelas palavras-chaves destacadas, além de apontar para as principais métricas e coeficientes utilizados.

3.3 Rede de co-palavra-chave

A rede de co-palavra-chave demonstrará as relações de afiliação entre as palavras-chaves e as dissertações e teses.

Pode-se destacar que, na rede de afiliação, predominam dois tipos típicos, como a rede de membros ou hiper rede.

A rede de afiliação de dois modos é composta por um conjunto de atores (palavras-chaves) e um conjunto de eventos (dissertações e teses).

Assumimos que A é a matriz primitiva $N \times M$ das relações de afiliação entre palavras-chaves e dissertações e teses, enquanto N é o conjunto de palavras-chaves, e M é o conjunto de dissertações e teses. Então, podemos usar a fórmula (1) para obter a derivada $M \times M$ matriz (X) da relação de palavras-chaves entre as dissertações e teses, e podemos usar a fórmula (2) para obter a derivada $N \times M$ matriz (Y) da relação das palavras chaves.

$$X = A'.A \quad (1)$$

$$Y = A.A' \quad (2)$$

3.4 Métricas de redes

Estuda-se a rede descrita utilizando o *software* de análise de rede Gephi (BASTIAN; HEYMANN; JACOMY, 2009), calculando medidas da estrutura das conexões. Em especial, mede-se a importância destas redes considerando: grau médio, grau médio ponderado e clusterização média. Outra medida importante que se emprega neste estudo é do menor caminho médio. Nos parágrafos que seguem apresenta-se, de forma sucinta, tais medidas. Mais detalhes podem ser obtidos nos trabalhos de Boccaletti *et al.* (2006) e Newman (2010).

O grau do nó indica o número dos nódulos em contato com ele. A regra $d(b_{ij} \geq 1) = 1$ é usada para alterar as redes ponderadas em não ponderadas, resultando

em Fórmula (3). Em seguida, usando Fórmula (4), o grau médio da rede é calculado medindo o grau de cada nó.

$$\begin{cases} d_{ij} = 1, \text{ sendo que o nó } i \text{ e o nó } j \text{ tem co - relacionam ento;} \\ d_{ij} = 0, \text{ sendo que o nó } i \text{ e o nó } j \text{ não tem co - relacionam ento.} \end{cases} \quad (3)$$

$$R = (R_i) = \left\langle \sum_{j=1}^n d_{ij} \right\rangle, i \neq j \quad (4)$$

Uma característica de destaque das redes ponderadas é a capacidade de medir a força de conexão de cada nó com outros nódulos.

Nas dissertações e teses, a identificação das palavras-chaves e as redes de palavras-chaves são importantes para determinar a proximidade das relações entre as dissertações e teses e a força de conexão entre as palavras-chave. Pode ser medida considerando o número de nódulos aos quais um determinado nó está conectado e considerando os pesos dos nódulos. Quanto maior o grau ponderado, maior a conectividade. O grau médio de peso do nó é calculado da seguinte forma:

$$S = \langle Si \rangle = \left\langle \sum_{j=1}^n b_{ij} \right\rangle, i \neq j \quad (5)$$

onde b_{ij} é formado pela Fórmula (1) e Fórmula (2).

Outra medida é o coeficiente de cluster, que serve para medir a conectividade de um determinado nó, onde usamos a clusterização média, que servirá para analisar a proximidade e a força das relações entre os nós, que podem ser calculadas pela Fórmula (6).

$$C = \langle C_i \rangle_i = \left\langle \frac{2E_i}{P_i(P_i - 1)} \right\rangle \quad (6)$$

onde E_i é o número de bordas existentes entre os primeiros vizinhos do nó i , e $1/2(P_i(P_i - 1))$ é o número de bordas potenciais ente os vizinhos do nó i .

Já o Menor Caminho Médio é o valor médio da menor quantidade de bordas entre qualquer um dos dois nós da rede, o que pode implicar a conectividade da rede, podendo ser calculado pela Fórmula (7).

$$D_{ij} = \frac{1}{N(N-1)} \sum_{i,j \in N, i \neq j} d_{ij} \quad (7)$$

Além dessas métricas, dois coeficientes destacaram-se: o de estabilidade e o de inovação, os quais podem ser medidos conforme a Fórmula (8).

Para medir a estabilidade da evolução da rede de co-palavras-chaves é necessário saber que $R_r(t=K)$ é o coeficiente de estabilidade da rede, N_K representa os nódulos na rede no momento em que N_K e N_{K-1} representam os nódulos na rede no tempo $K-1$. $N_K \cap N_{K-1}$ representam os nódulos na rede no tempo N_{K-1} e N_K , e $N_K \cup N_{K-1}$ é os nódulos em união de N_{K-1} e N_K .

A Fórmula (8) mostra o coeficiente de estabilidade, que só pode mostrar a dinâmica do grupo entre duas vezes diferentes, em ordem, para estimar a tendência das dissertações e teses ano após ano considerando os dados da série temporal.

Já o coeficiente de inovação da rede de co-palavras-chave é determinado pela razão entre o surgimento de novas palavras-chaves no presente tempo ($t=K$) e todas as palavras-chave do primeiro ano ($t=1$) até o presente ($t=K$).

$$\ln_c(K) = \frac{\left\{ N_K - \left[N_K \cap \left(\bigcup_{t=1}^{K-1} N_t \right) \right] \right\}}{\left(\bigcup_{t=1}^K N_t \right)} \quad (8)$$

onde N_K representa nós na rede no momento k . $\bigcup_{t=1}^K N_t$ representa os nós na união de

todas as redes de $t = 1$ a $t = k$, e a $\bigcup_{t=1}^{K-1} N_t$ representa os nós na união de todas as

redes de $t = 1$ a $t = k-1$. $\left\{ N_K - \left(\bigcup_{t=1}^{K-1} N_t \right) \right\}$ representa os novos nódulos que aparecem

em N_K e nunca apareceram de $t = 1$ a $t = k-1$.

Por intermédio das métricas e coeficientes de redes destacados, é possível identificar os principais achados e resultados da pesquisa, a fim de contribuir com a comunidade científica e, em especial, com a rede de colaboração, que, nesse caso, é a rede de co-palavras-chaves.

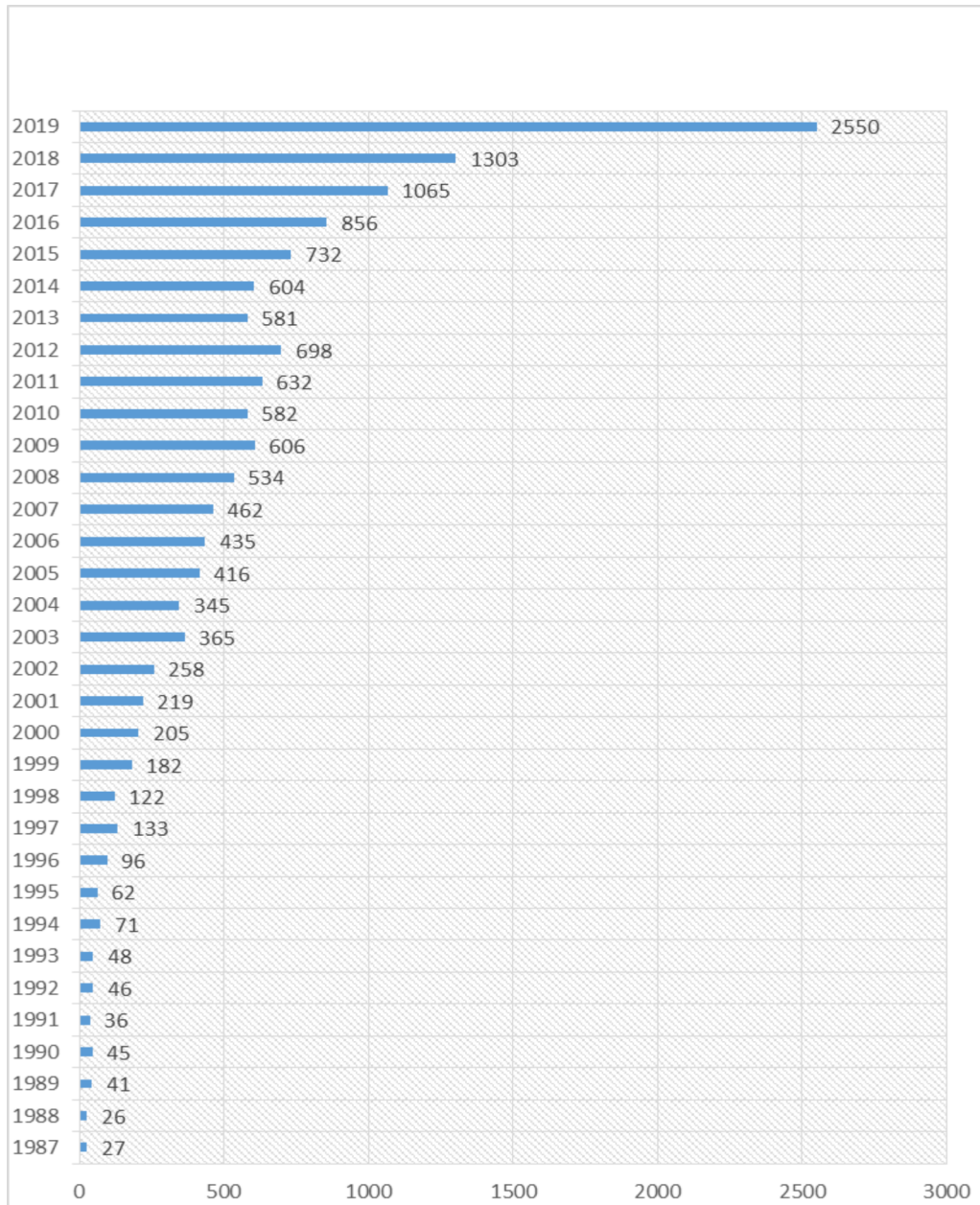
4 RESULTADOS

Neste capítulo, apresenta-se os resultados da pesquisa, o qual está dividido em quatro seções. A primeira trará uma análise descritiva, que será dividida em dois momentos, sendo que o primeiro abordará a ciência de dados de maneira geral, considerando a totalidade dos dados da amostra, e a segunda, mais específica, trazendo um recorte da ciência de dados na contabilidade. Posteriormente, tem-se a visualização das redes, seguida das características evolucionárias topológicas e suas métricas, e, por fim, serão evidenciados os coeficientes de estabilidade e inovação.

4.1 Análise descritiva da ciência de dados

A análise descritiva foi efetuada com o tratamento das palavras-chaves previamente definidas e investigadas no Catálogo de Dissertações e Teses da Capes, que se apropriaram da ciência de dados no Brasil. Para isso, ocorreu a utilização do Excel, o uso da mineração de dados e da linguagem computacional com utilização do Python. Sendo assim, houve uma mescla formada por planilhas em Excel, *scripts* do Python (ver Anexo A e B) e a mineração dos dados, que buscaram o aprofundamento e apresentação dos principais resultados da pesquisa.

Gráfico 1 - Evolução das dissertações e teses de 1987 a 2019

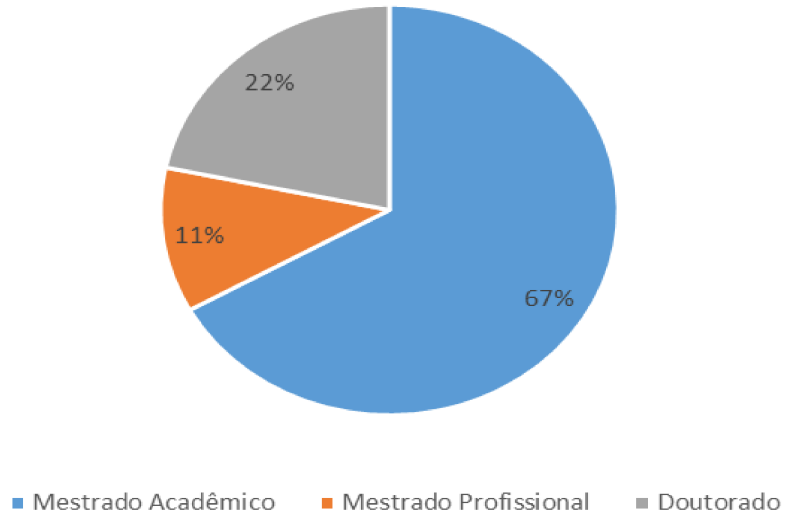


Fonte: Elaborado pelo autor.

No Gráfico 1, tem-se a evolução das dissertações e teses publicadas, que totalizam 14.383 pesquisas, podendo destacar-se que, na primeira década de pesquisas, o crescimento é considerado baixo, sendo que de 2005 a 2012 tem um crescimento médio e que a partir de 2013 o crescimento aumenta, destacando-se o ano de 2019, que tem 2550 pesquisas.

O Gráfico 2 apresenta o tipo de titulação predominante entre os pesquisadores, de maneira que há um destaque de 67% na modalidade de mestrado acadêmico, 11% no mestrado profissional e 22% no doutorado.

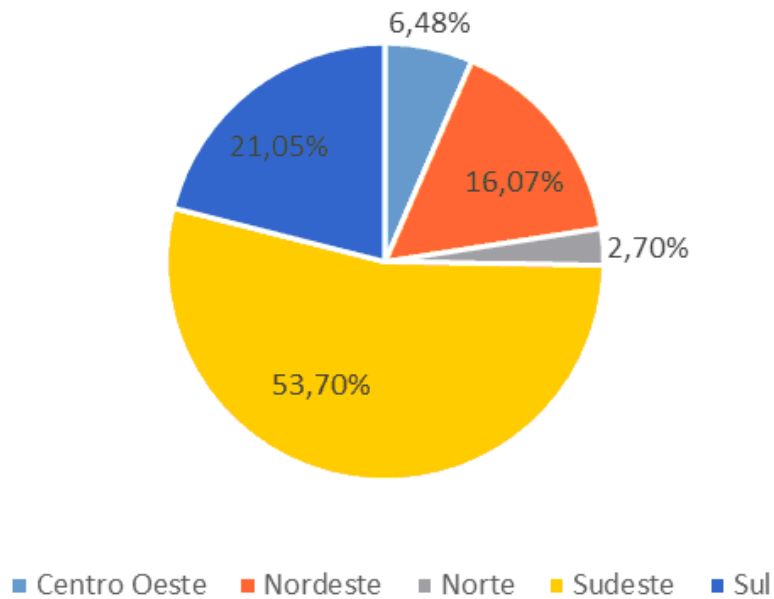
Gráfico 2 - Tipo de titulação dos pesquisadores da ciência de dados (1987-2019)



Fonte: Elaborado pelo autor.

No Gráfico 3, apresenta-se a evolução das dissertações e teses por regiões, sendo que a Região Norte apresenta o menor percentual, com apenas 2,70%, e em destaque tem-se a Região Sudeste, disparada com 53,70%. As regiões Nordeste e Sul são as que apresentam resultados semelhantes de crescimento.

Gráfico 3 - Evolução das dissertações e teses por região para a ciência de dados (1987-2019)

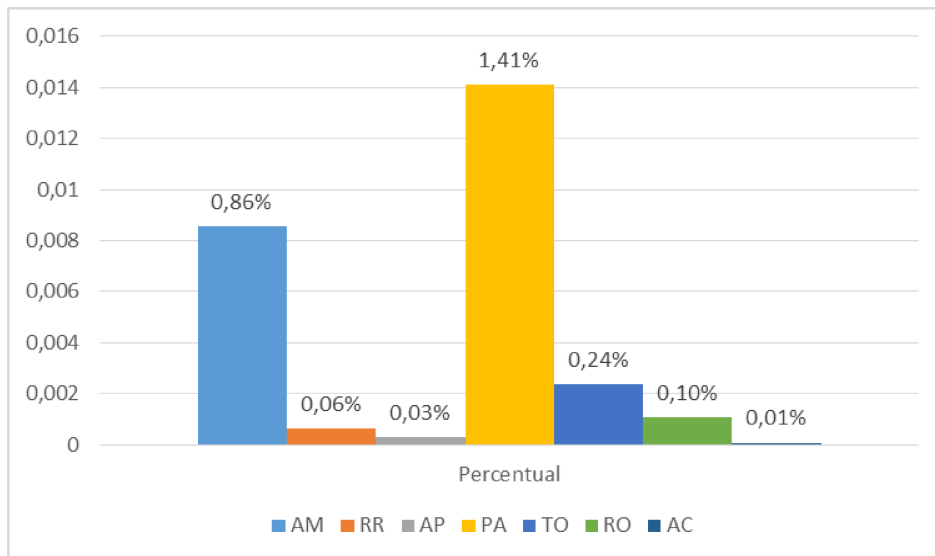


Fonte: Elaborado pelo autor.

Cabe destacar que essa análise considerou o aspecto de cada região do país, que possibilita a identificação e o aprofundamento, por exemplo, dos estados que se destacam com as dissertações e teses.

No Gráfico 4, que expõe os dados da Região Norte, é possível destacar que o estado do Pará, com 1,41%, e o estado do Amazonas, com 0,86%, são os estados que apresentaram maiores pesquisas sobre a temática. No entanto, destaca-se que as pesquisas nessa região devem ser incentivadas devido à disparidade com as demais regiões do país.

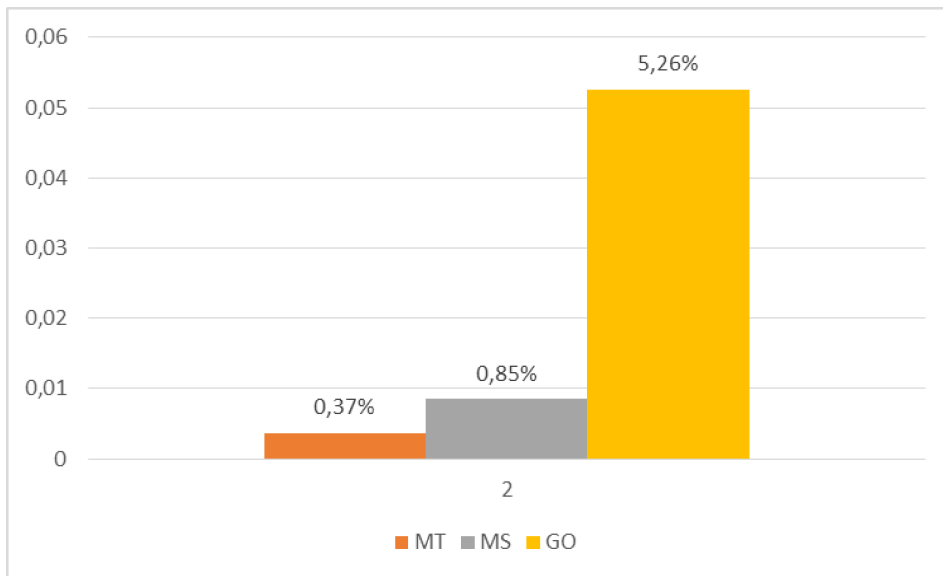
Gráfico 4 - Evolução das dissertações e teses da ciência de dados para a Região Norte (1987-2019)



Fonte: Elaborado pelo autor.

No Gráfico 5, é apresentada a situação das dissertações e teses na Região Centro-Oeste, onde se destaca o estado de Goiás, com 5,26%, que também incluiu informações referentes ao Distrito Federal, com a capital do país, Brasília.

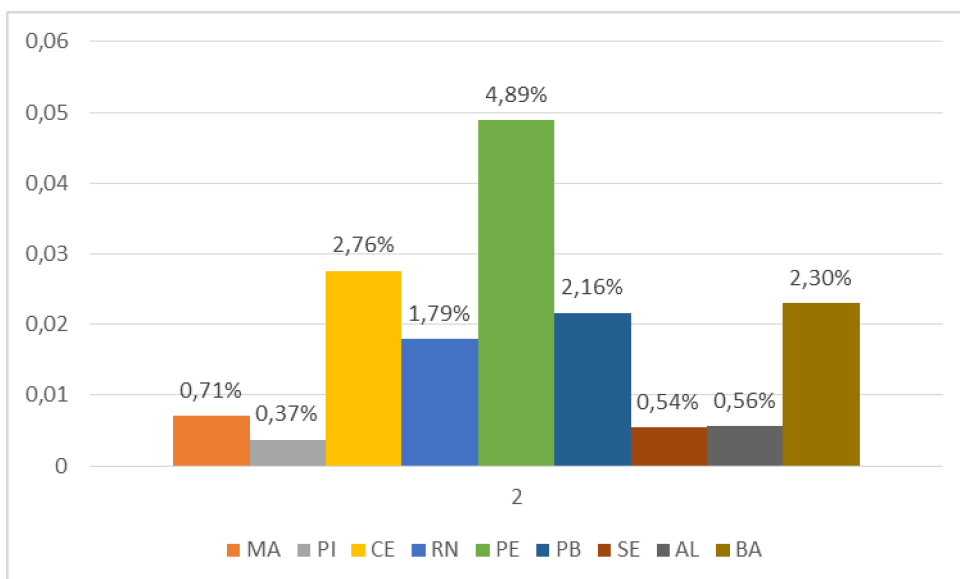
Gráfico 5 - Evolução das dissertações e teses da ciência de dados para a Região Centro-Oeste (1987-2019)



Fonte: Elaborado pelo autor.

No Gráfico 6, são demonstradas as dissertações e teses da Região Nordeste, onde destaca-se, com 4,89%, o estado de Pernambuco, seguido do estado do Ceará, com 2,76%. Por ser uma região composta com muitos estados, é possível identificar as disparidades entre eles, por exemplo, com índices baixos, o estado do Piauí, com 0,37%, Sergipe, com 0,54%, e Alagoas, com 0,56%.

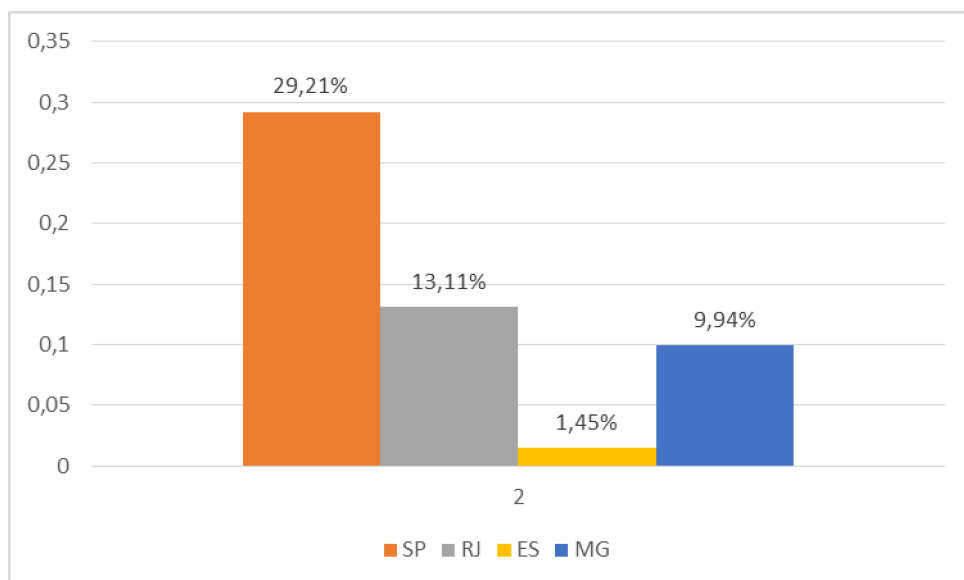
Gráfico 6 - Evolução das dissertações e teses da ciência de dados para a Região Nordeste (1987-2019)



Fonte: Elaborado pelo autor.

No Gráfico 7, é apresentada a Região Sudeste, que apresenta os maiores índices de dissertações e teses, onde se concentram algumas das principais capitais de nosso país, sendo que o estado de São Paulo, com 29,21%, seguido do estado do Rio de Janeiro, com 13,11%, se destacam com as pesquisas. Outra informação interessante é que, se comparado com outros estados, o Espírito Santo, com apenas 1,45%, fica inferior a outros índices de estados das demais regiões do país, como, por exemplo, 2,30% do estado da Bahia e 1,79% do Rio Grande do Norte, na Região Nordeste do Brasil.

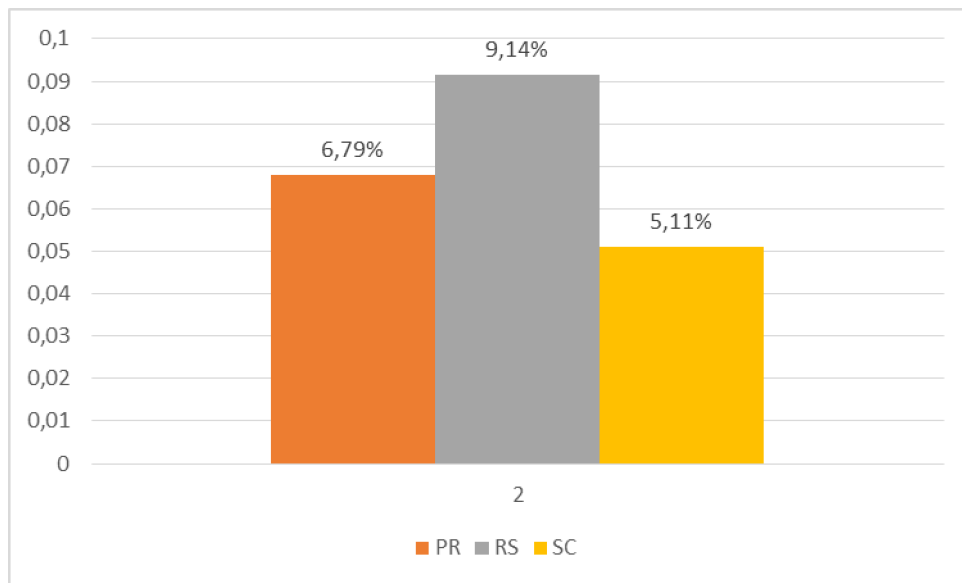
Gráfico 7 - Evolução das dissertações e teses da ciência de dados para a Região Sudeste (1987-2019)



Fonte: Elaborado pelo autor.

Conforme o Gráfico 8, a Região Sul se destaca com um número alto de dissertações e teses em comparação com as demais regiões do país, sendo que o estado do Rio Grande do Sul apresenta 9,14%, seguido pelo estado do Paraná, com 6,79%, e pelo estado de Santa Catarina, com 5,11%.

Gráfico 8 - Evolução das dissertações e teses da ciência de dados para a Região Sul (1987-2019)



Fonte: Elaborado pelo autor.

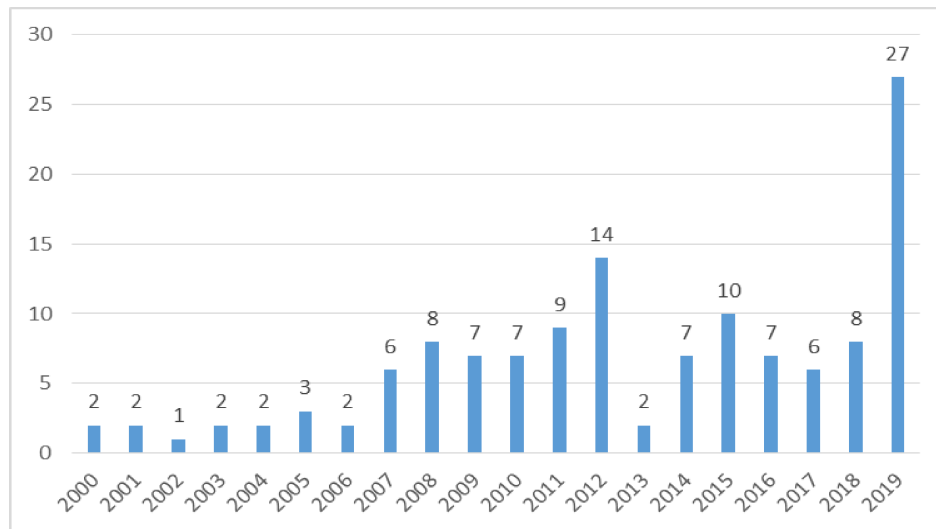
As disparidades entre as regiões do país são notórias, mas servem como parâmetros e também indicam os números das dissertações e teses, que podem evidenciar as que mais evoluíram e as que ainda necessitam de evolução em termos de publicações. Dessa forma, pode-se colaborar com o desenvolvimento dos estudos na comunidade científica no que se refere à presente temática.

4.1.1 Análise descritiva na contabilidade

Nesta pesquisa, cabe destacar, também, as dissertações e teses em contabilidade que se apropriaram da ciência de dados, evidenciando a sua evolução no decorrer do tempo e apresentando os principais resultados.

No Gráfico 9, é apresentada a evolução das dissertações e teses que se apropriaram da ciência de dados na contabilidade, totalizando 132 pesquisas da amostra de dados inicial, evidenciando a necessidade do crescimento das dissertações e teses sobre essa temática e, também, uma maior evolução, por se tratar de um assunto recente e que requer maior atenção dos pesquisadores.

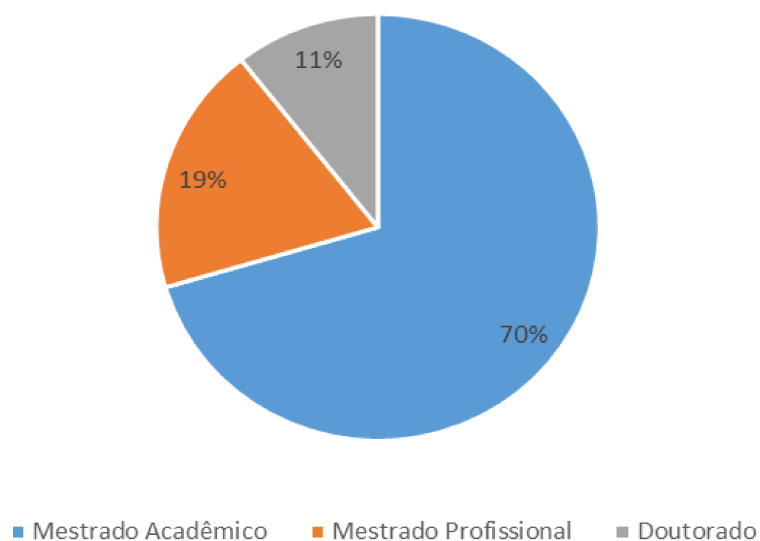
Gráfico 9 - Evolução das dissertações e teses das ciências de dados nas ciências contábeis (1987-2019)



Fonte: Elaborado pelo autor.

O Gráfico 10 demonstra o tipo de titulação predominante entre os pesquisadores na contabilidade, de maneira que há um destaque de 70% na modalidade de mestrado acadêmico, 11% no mestrado profissional e 19% no doutorado. Se comparado com o Gráfico 2, que evidenciou, de maneira global, todas as áreas, os dados são bastante semelhantes e convergentes, sinalizando ainda espaço para dissertações e teses sobre a temática proposta.

Gráfico 10 - Tipo de titulação dos pesquisadores da ciência de dados nas ciências contábeis (1987-2019)

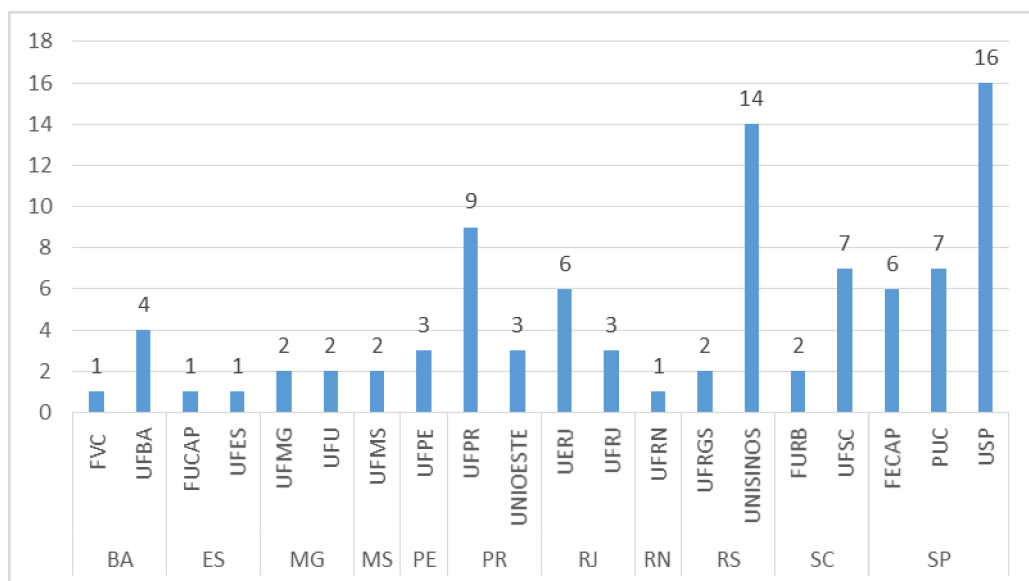


Fonte: Elaborado pelo autor.

No Gráfico 11, demonstra-se as Instituições de Ensino Superior em Nível de Mestrado Acadêmico que tiveram maior destaque, sendo que, com 16 pesquisas, a Universidade de São Paulo (USP), em São Paulo, seguida da Universidade do Vale dos Sinos (UNISINOS), no Rio Grande do Sul, com 14 pesquisas, e a Universidade Federal do Paraná (UFPR), com 9 pesquisas, são as que mais se destacaram. Há, ainda, outras instituições que permaneceram com 6 ou 7 pesquisas, a saber: Universidade do Estado do Rio de Janeiro (UERJ), no Rio de Janeiro, Fundação Escola de Comércio Álvares Penteado (FECAP), em São Paulo, Universidade Federal de Santa Catarina (UFSC), em Santa Catarina, e Pontifícia Universidade Católica de São Paulo (PUC), em São Paulo.

As demais instituições tiveram um baixo número, figurando entre 1 a 3 pesquisas, como, por exemplo: Fundação Visconde de Cairu (FVC), na Bahia, Universidade Federal do Espírito Santo (UFES), no Espírito Santo, Universidade Federal de Minas Gerais (UFMG), em Minas Gerais, além da Universidade Regional de Blumenau (FURB), em Santa Catarina. Com isso, foi possível corroborar que ainda existem muitas disparidades entre as Instituições de Ensino Superior no país no que tange às dissertações e teses relacionadas com essa temática, sugerindo uma grande lacuna existente, que pode ser preenchida com novos estudos.

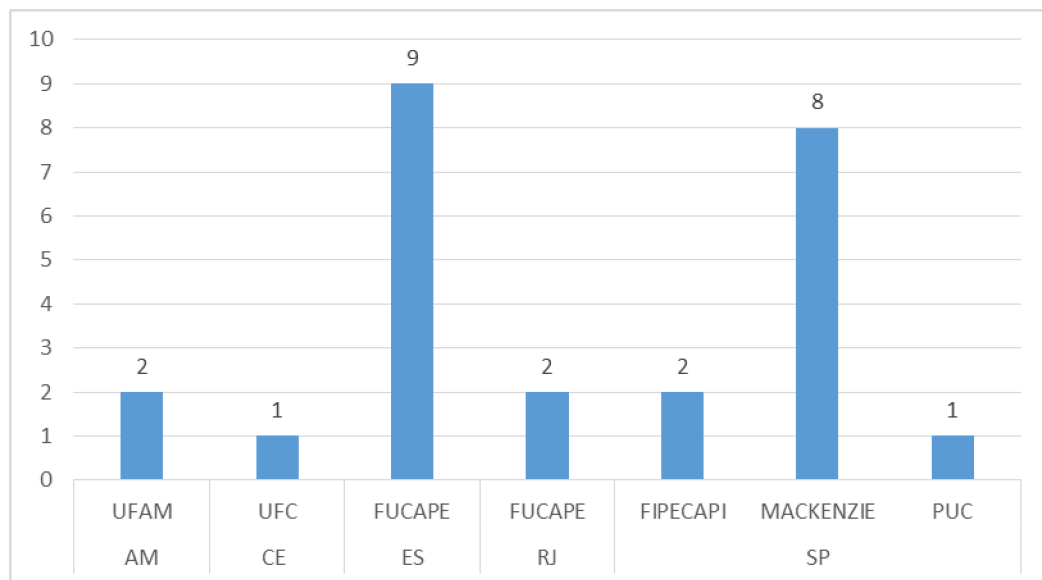
Gráfico 11 - IES mestrado acadêmico - Produção de Dissertações na contabilidade que se apropriaram da ciências de dados



Fonte: Elaborado pelo autor.

No Gráfico 12, são apresentadas as Instituições de Ensino Superior em nível de mestrado profissional que proporcionam uma interação entre as universidades e o meio empresarial. Em destaque, tem-se, com 9 pesquisas, a FUCAPE Business School (FUCAPE), no Espírito Santo, e a Universidade Presbiteriana Mackenzie (MACKENZIE), em São Paulo, com 8 pesquisas. As demais instituições apresentam dados entre 1 a 2 pesquisas, apontando que existe espaço para crescimento destes estudos: a Universidade Federal do Ceará (UFC), no Ceará, a Universidade Federal do Amazonas (UFAM), no Amazonas, e a Fundação Instituto de Pesquisas Contábeis, Atuariais e Financeiras (FIPECAPI), em São Paulo.

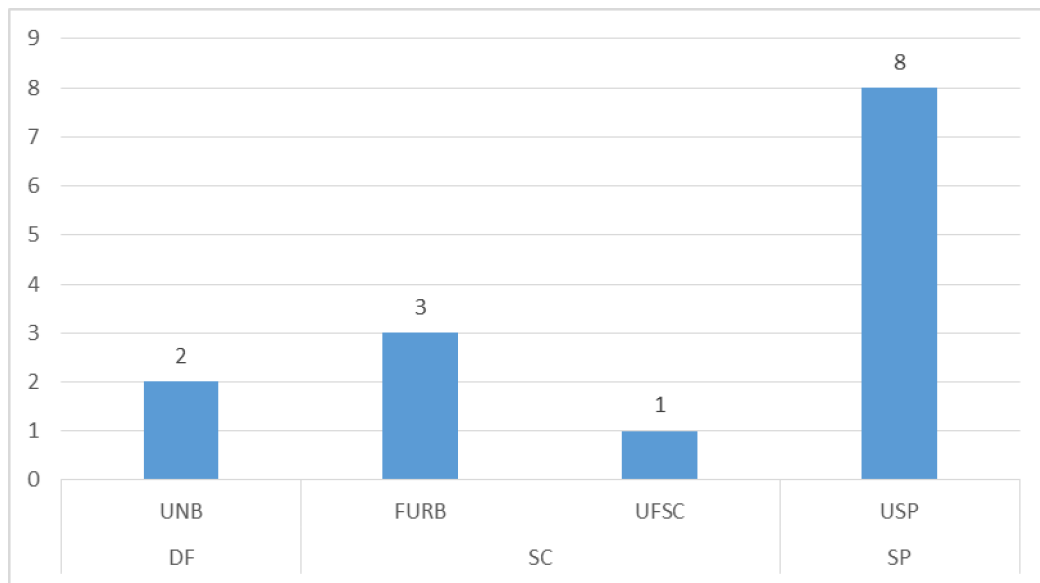
Gráfico 12 - IES mestrado profissional - Produção de Dissertações na contabilidade que se apropriaram da ciências de dados



Fonte: Elaborado pelo autor.

No Gráfico 13, são apresentadas as Instituições de Ensino Superior em nível de doutorado, sendo que, com 8 pesquisas, em destaque, tem-se a Universidade de São Paulo (USP), em São Paulo. Já as demais instituições ficaram entre 1 a 3 pesquisas, como a Universidade Federal de Santa Catarina (UFSC), em Santa Catarina, a Universidade de Brasília (UNB), no Distrito Federal, e a Universidade Regional de Blumenau (FURB).

Gráfico 13 - IES doutorado - Produção de Teses na contabilidade que se apropriaram da ciências de dados



Fonte: Elaborado pelo autor.

De maneira geral, pode-se confirmar a escassez de dissertações e teses e, por consequência, de pesquisadores que venham a trabalhar com essa temática, a fim de contribuir para o desenvolvimento das organizações da atualidade.

Desta forma, conclui-se que a Ciência de Dados (Data Science) é a combinação entre estatística, matemática, soluções computacionais para capturar dados, detectar padrões, juntamente com atividades de limpeza, preparação e organização de dados. É a definição mais abrangente que inclui estatística, matemática, mineração de dados, big data, visualização e aprendizado de máquina (HARDIN, 2015).

Essa definição inclui, também, a utilização de novas tecnologias e ferramentas advindas da ciência de dados e que já fazem parte do cenário da contabilidade nas suas mais diversas áreas de atuação, como: contabilidade financeira, contabilidade gerencial, contabilidade de custos, contabilidade pública, perícia contábil, auditoria interna, auditoria externa, controle interno e controladoria (VASARHELYI; TEETER; KRAHEL, 2015).

Esse crescimento da ciência de dados, em especial na contabilidade, pode auxiliar na extração de informações e *insights* dos dados de maneira significativa nos processos de trabalhos de uma organização (PROVOST, 2013).

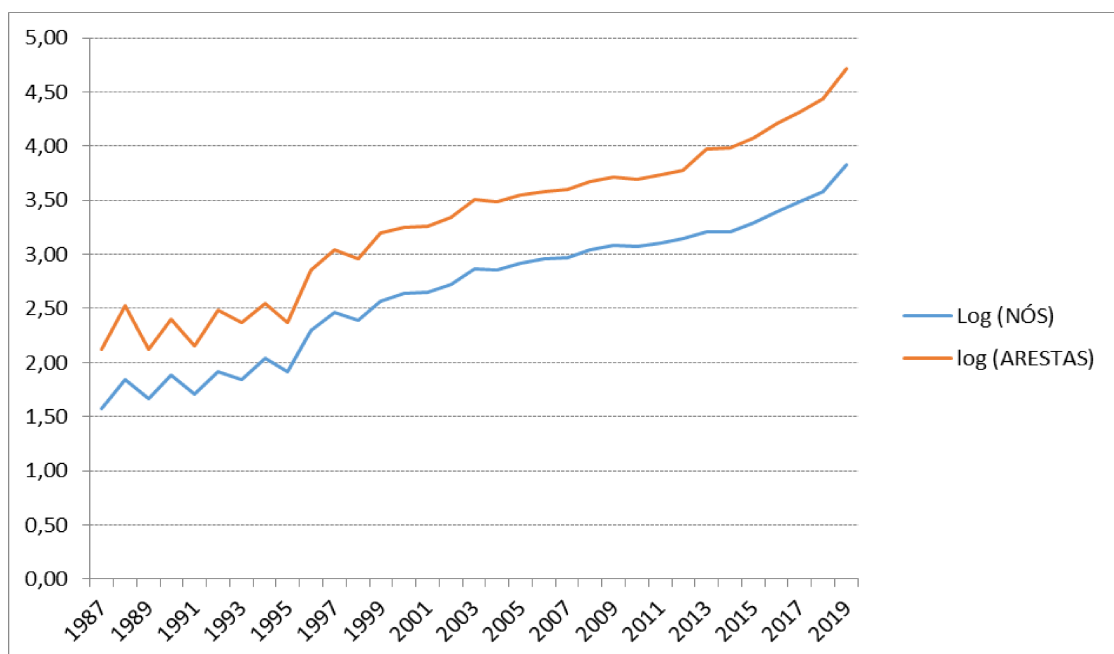
Até o presente momento, o enfoque dado a este trabalho foi direcionado para a análise descritiva, sendo que a parte “interessante” da pesquisa é a formação das redes de co-palavras-chaves, desde a sua visualização, suas características evolucionárias topológicas, suas métricas e os coeficientes de estabilidade e inovação. É o que veremos a seguir.

4.2 Visualização das redes

A fim de mapear a evolução temporal das palavras-chaves e sua identificação por intermédio dos principais módulos desenvolvidos durante o período que foi de 1987 até 2019, outros métodos e ferramentas apoiadas nas redes complexas, como a criação e visualização das redes com a utilização do software Gephi, foram utilizados na pesquisa.

Cabe destacar que a rede de co-palavras-chaves teve comportamento que evoluiu em três etapas distintas, sendo que a primeira vai de 1987 a 2005, a segunda, de 2006 até 2012, e a terceira, de 2013 até 2019. Nesse sentido, no Gráfico 14, apresenta-se o número de nós e arestas ao longo do tempo, apresentados por uma função logarítmica.

Gráfico 14 - Número de nós e arestas da Rede de Co-Palavras-Chaves (1987-2019)



Fonte: Elaborado pelo autor.

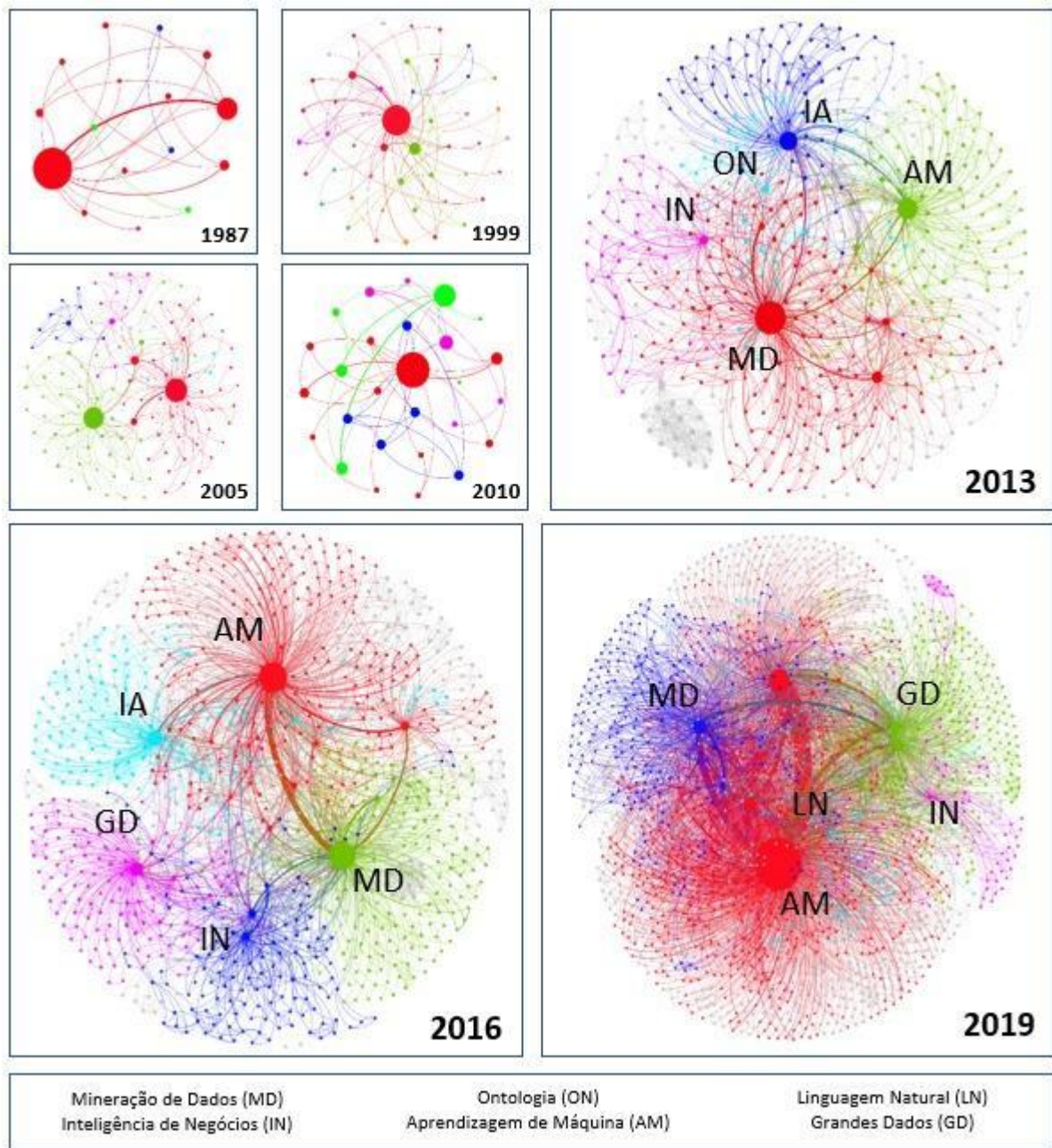
Na primeira etapa, ocorreram poucas conexões. O número de nós e arestas tem uma relação diretamente proporcional, destacando-se como palavras-chave a inteligência artificial e a análise de dados. Percebe-se a existência de uma correlação positiva entre os nós e arestas que se movem juntos, sendo que a relação é forte quando a correlação se aproxima de um. O ano de 2005 pode ser considerado o destaque, pois apresenta a maior correlação, além de apresentar o maior número de nós e arestas do período indicado.

Já na segunda etapa, ainda com poucas conexões, destacaram-se, também, palavras-chaves como inteligência artificial e análise de dados. Além disso, surgem outros módulos na rede, como nos anos de 2010 e 2012, que destacam as seguintes palavras-chaves: inteligência artificial, análise de dados, mineração de dados e aprendizagem de máquina.

Foi na terceira etapa que ocorreu o aumento das conexões, com aumento do número de nós e arestas. Sendo assim, o aumento dos módulos existentes é visível, sendo que podemos destacar que os anos de 2013, 2016 e 2019 trouxeram as seguintes palavras-chaves destacadas: ontologia, linguagem natural, inteligência artificial, aprendizagem de máquina, mineração de dados, grandes dados, inteligência de negócios. Aqui é nítido que o gráfico evidencia que aumentar/diminuir nós implica em aumentar/diminuir arestas e, como consequência destas relações, o surgimento das correlações, que, neste caso, são correlações positivas que podem ser fracas ou fortes.

A Figura 4, denominada de evolução da rede, demonstra as etapas percorridas ao longo do período (1987 a 2019), e os principais módulos criados pela rede de co-palavras-chaves.

Figura 4 - Evolução da rede de co-palavras-chaves(1987-2019)



Fonte: Elaborada pelo autor.

Cabe destacar que os módulos estão organizados por cores e identificados por uma legenda que auxilia em sua visualização e identificação, sendo que cada círculo representa uma palavra-chave em destaque de cada período, e que uma ligação (aresta) é uma associação formada pelas conexões formadas entre elas. A espessura de cada aresta está associada ao peso das ligações existentes, apoiando-se na identificação dos principais módulos formados pela evolução da rede, além de contribuir para as métricas utilizadas nesta pesquisa.

A evolução da rede se caracteriza por estruturas modulares denominadas *hubs* (nós com muitas conexões), sendo que isso determinará os módulos que serão denominados de acordo com seus hubs.

Na estrutura formada pelas palavras-chaves de campos de conhecimentos bastante diversos, é possível observar a formação de poucas estruturas regulares (módulos) dominantes ao longo do tempo.

Os dados apresentados no Gráfico 14 demonstram que, a partir do número de nós e arestas, combinados com a Figura 4, que apresenta a evolução da rede, é possível identificar informações relevantes sobre o comportamento da rede, como fora evidenciado nos anos de 2013, 2016 e 2019, destacados na terceira etapa da rede de co-palavras-chaves. Além disso, a estatística descritiva apontou o crescimento significativo das pesquisas, sendo que o destaque são para os anos de 2013, 2016, 2019, que apresentaram respectivamente: 581; 856 e 2550 dissertações e/ou teses.

Esses dados reforçam o crescimento das dissertações e teses a partir de 2013, que pode ser considerado o boom da rede formada pelas palavras-chaves que se apropriaram da ciência de dados, destacando os principais módulos apresentados pela evolução da rede.

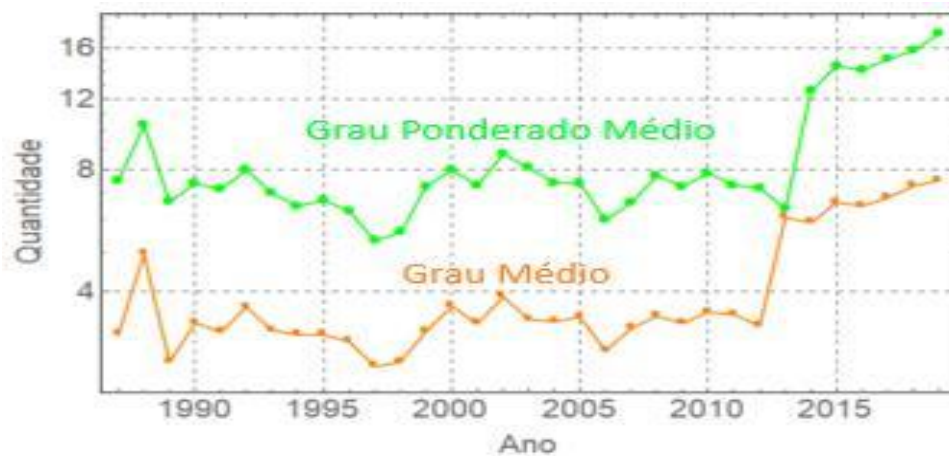
4.3 Características evolucionárias topológicas

Para apresentar as características evolucionárias topológicas, são apresentados o número de nós e arestas, além das principais medidas utilizadas por ano na pesquisa, como: grau médio, grau médio ponderado, clusterização média e menor caminho médio.

Nesse contexto, o grau médio do nó indica a probabilidade de dois nós estarem ligados entre si, representando o número de dissertações e teses que compartilham uma ou mais palavras-chaves. O grau médio ponderado, por sua vez, segundo Abassi e Altmann (2011), é a soma de todos os pesos das arestas ligadas a um nó, indicando tanto o número de dissertações e teses que compartilham uma ou mais palavras-chaves como o número de palavras-chaves que são compartilhadas.

O Gráfico 15, apresentado a seguir, apresenta essas duas medidas importantes da pesquisa: (1) grau médio e (2) grau médio ponderado.

Gráfico 15 - Tipo de grau da rede de co-palavras-chaves (1987-2019)



Fonte: Elaborado pelo autor.

Percebe-se que, em ambos os tipos de grau, é nítida uma regularidade por períodos que permanecem até o ano de 2012, com séries altamente correlacionadas entre si. A partir de 2013, percebe-se o boom em ambas as medidas apresentadas, corroborando o crescimento das dissertações e teses que evidenciam a ciências de dados na rede de co-palavras-chaves.

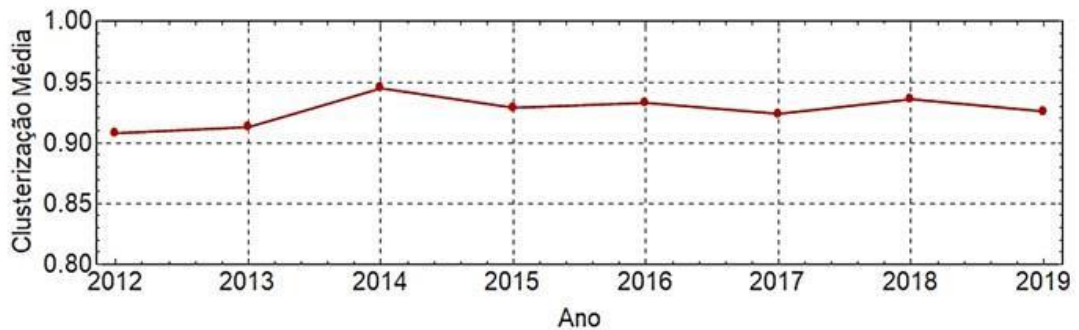
Se compararmos o período de 1987 até 2012, percebe-se um comportamento semelhante entre os dois tipos de medidas, caracterizando-se por conexões estáveis durante todo esse período, sem muito destaque para módulos criados.

O destaque do boom em 2013 é a formação de módulos bem distintos, que indicam as principais palavras-chaves desta pesquisa, elucidando o que ocorreu durante essa evolução temporal.

Entre outras medidas, tem-se a clusterização média e o menor caminho médio, aspectos importantes para análise das características topológicas, sendo que são popularmente utilizadas para medir a rede.

No Gráfico 16, denominado clusterização média, pode-se destacar que as análises começaram a ser efetuadas a partir de 2012, período em que as atividades na rede (nós e arestas) tornam-se mais intensas. O *clustering* nas redes complexas é usualmente utilizado para a detecção de módulos, além de servir para analisar a proximidade e a força das relações entre os nós, significando que, em média, em nível local, todas as palavras-chaves compartilham conexões, e reforçando a natureza modular da estrutura das conexões ao longo de todo o período de estudo.

Gráfico 16 - Clusterização média da rede de co-palavras-chaves (1987-2019)



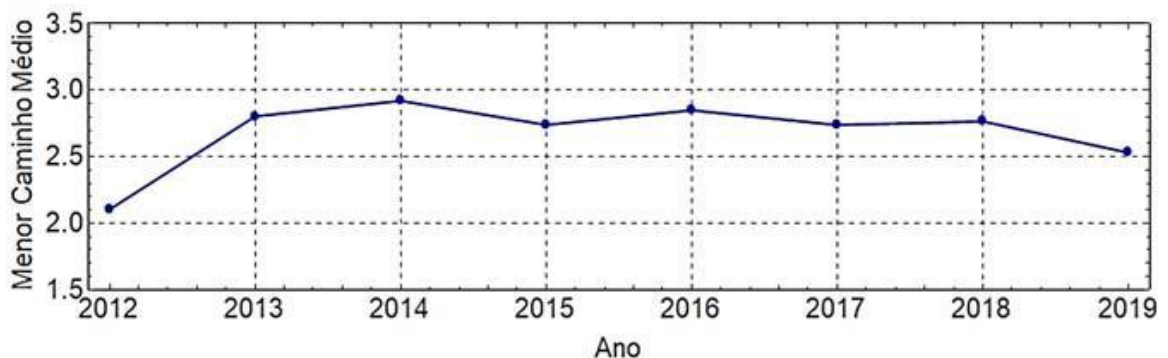
Fonte: Elaborada pelo autor.

No Gráfico 17, denominado menor caminho médio, pode-se destacar, de acordo com os resultados, os anos de 2013, com 2,803, 2014, com 2,92, e 2019, com 2,53. Sendo assim, pode-se constatar que qualquer dissertação e tese só precisa passar em torno de três dissertações e teses para chegar a outras dissertações e teses na rede de palavras-chaves.

No que tange à disseminação de conhecimento e sua associação (via sua distância) percebe-se um GAP conceitual/teórico/técnico baixo (menor caminho médio menor que três). No entanto, quanto mais distante, ou seja, mais palavras-chaves intermediárias, mais aumenta o GAP conceitual/teórico/técnico.

A tendência apontada pela série é de redução para período em estudo, confirmando o efeito pequeno-mundo que tem implicações óbvias na dinâmica de processo em redes, apontadas pelo menor caminho médio.

Gráfico 17 - Menor caminho médio da rede de co-palavras-chaves (1987-2019)



Fonte: Elaborado pelo autor.

Para acompanhar a evolução da rede, destaca-se que os cinco maiores módulos, destacados no Gráfico 18, ocupam, em média, 70% dos nós na rede (independente do período), criando um formato padrão, que destacou os principais

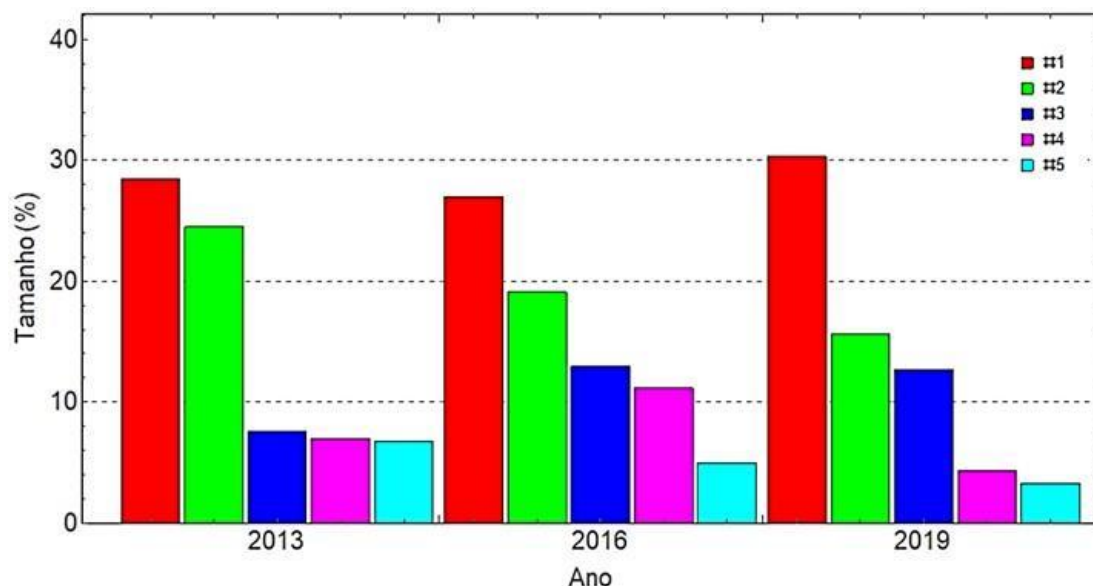
módulos formados pelas palavras-chaves referentes aos anos de 2013, 2016 e 2019.

Pode-se destacar que o módulo de maior destaque é o ano de 2013, quando ocorreu o aumento do número de dissertações e teses, conseqüentemente, o aumento do número de nós e arestas formados pela rede de co-palavras-chaves. Sendo assim, tem-se, no período supracitado, por exemplo: aprendizagem de máquina, mineração de dados e inteligência artificial.

A predominância de palavras-chaves, até então ligadas com o mundo computacional, abre espaço também para o mundo dos negócios, destacando que, no ano de 2016, módulos como aprendizagem de máquina, mineração de dados, inteligência de negócios e grandes dados sugerem essa sinergia entre a computação e a gestão e negócios.

Já no ano de 2019, percebe-se a consolidação de algumas palavras-chaves, as quais sugerem o aumento de pesquisas alinhadas com a gestão e negócios das organizações, com a maior inserção de palavras-chaves destacadas na rede, tais como: inteligência de negócios e grandes dados.

Gráfico 18 - Cinco maiores módulos da rede de co-palavras-chaves (1987-2019)



Fonte: Elaborado pelo autor.

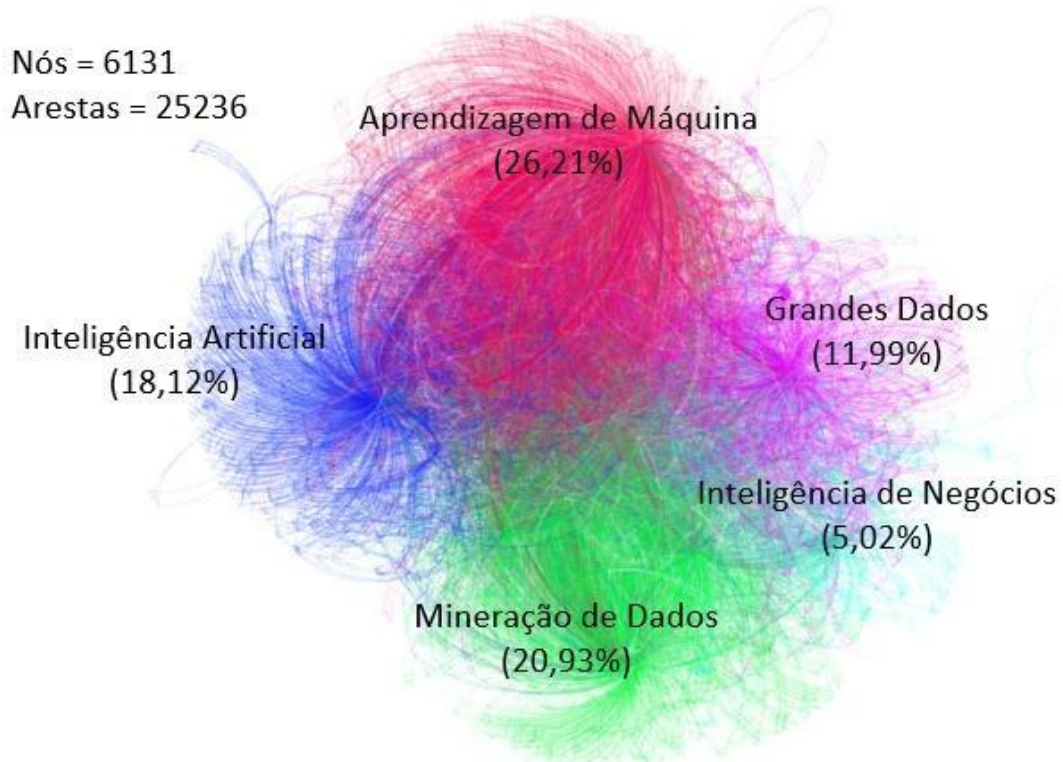
Nesse sentido, a Figura 5, denominada redes de co-palavras-chaves acumulada, indica o acumulado de todos os períodos desde (1987 a 2019), em que foi possível identificar o número total de nós (6131) e arestas (25236), sendo que os

principais módulos desenvolvidos foram os seguintes: com 26,21%, Aprendizagem de Máquina, 20,93%, Mineração de Dados, 18,12%, Inteligência Artificial, 11,99%, Grandes Dados e 5,02%, Inteligência de Negócios.

Se a rede de co-palavras-chaves for comparada com a evolução da rede demonstrada na Figura 4, percebe-se a existência de uma forte modularização que pode ser destacada em poucas unidades, como os anos de 2013, 2016 e 2019. As informações são convergentes e caminham no mesmo sentido, indicando o crescimento de dissertações e/ou teses que vem inserindo a ciência de dados, apropriando-se da linguagem computacional e também da linguagem de negócios, que são importantes para o mundo da gestão e negócios, em especial da contabilidade.

Os principais módulos destacados no acumulado, atualmente, são realidades em muitas organizações que se apropriam da ciência de dados, não como um achismo ou modismo, e, sim, como estudo de como essas ferramentas ou técnicas, como por exemplo: inteligência de negócios, os grandes dados e a mineração de dados, podem auxiliar no desenvolvimento das organizações, integrando a linguagem computacional com a linguagem de negócios.

Figura 5 - Redes de co-palavras-chave acumulada no tempo (1987-2019)



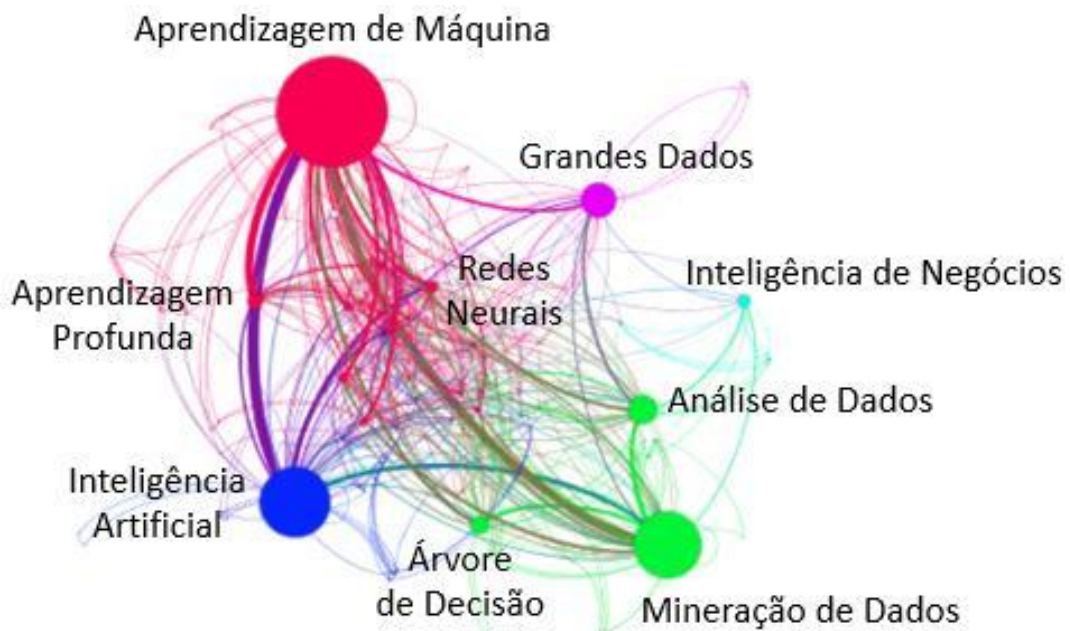
Fonte: Elaborada pelo autor.

Na Figura 6, denominada de principais módulos formados pelas palavras-chaves, foi possível reduzir as densidades das arestas indicadas na rede de co-palavras-chave (acumulada no tempo). Com isso, foi possível observar os *hubs* de cada módulo em maiores detalhes, como são feitas suas associações (em especial, os pesos das ligações), sendo possível visualizar o maior *hub* (aprendizagem de máquina) e menor *hub* (inteligência de negócios).

Os pesos das conexões, inicialmente concentrados na aprendizagem de máquina e inteligência artificial, reforçam a importância da linguagem computacional e sua predominância na análise da rede.

Em contrapartida, pode-se inferir que a inteligência de negócios, mais associada a grandes dados, análise de dados e mineração de dados, indica a mescla entre a linguagem computacional e a linguagem de negócios, extremamente necessária nas organizações da atualidade.

Figura 6 - Principais módulos da rede de co-palavras-chaves (1987-2019)



Fonte: Elaborada pelo autor.

Além das métricas acima, nas seções seguintes, destacam-se dois coeficientes: o de estabilidade e o de inovação. Ambos têm como objetivo ajudar na compreensão do processo da difusão da inovação na rede de co-palavras-chaves e, conseqüentemente, contribuir para os resultados em redes complexas que ainda estão experimentando períodos que podem aumentar seu desenvolvimento junto à comunidade científica.

4.4 Estabilidade e inovação

O coeficiente de estabilidade indica quais influências fazem com que a rede de co-palavras-chaves evolua ao longo do tempo, no período de 1987 até 2019. O Gráfico 19 trata-se de uma representação do coeficiente de estabilidade em tempo para seis medidas, ou seja, os cinco principais módulos mais a rede completa.

Nesse contexto, por um lado, um coeficiente de estabilidade maior indica que os módulos destacados na evolução da rede são mais semelhantes entre si. Ou seja, essa semelhança é identificada pelo interesse de pesquisa dos autores, que pode ser destacada, neste estudo, pelo aprofundamento das palavras-chaves das dissertações e teses.

Por outro lado, um coeficiente de estabilidade pequeno indica que as dissertações e teses tendem a explorar novas áreas de pesquisa.

Percebe-se que, de 1987 até 1997, o coeficiente de estabilidade é considerado lento, demonstrando que, no referido período, a inteligência artificial foi o principal módulo de destaque, sinalizando que, nesse primeiro momento, não ocorreu grandes avanços na evolução da rede.

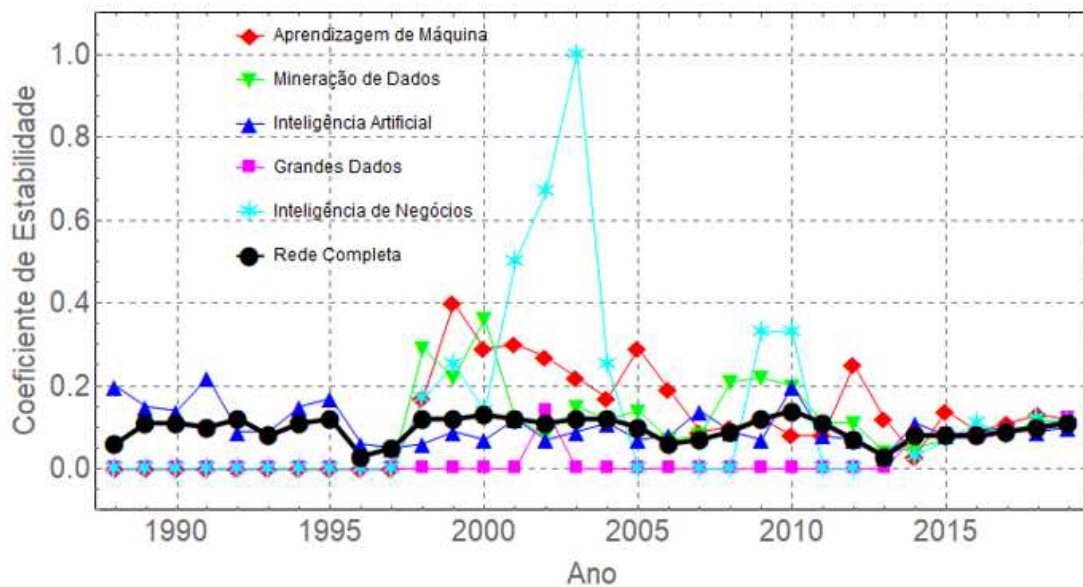
A partir de 1998 até 2012, o coeficiente de estabilidade apresenta as maiores variações, destacando alguns módulos, como: inteligência de negócios, aprendizagem de máquina e mineração de dados, sendo considerado o período em que ocorreram os maiores índices, indicando os principais avanços da evolução da rede.

Já de 2013 até 2019, o coeficiente de estabilidade pode ser considerado estável, sinalizando a continuidade de alguns módulos de destaque: aprendizagem de máquina, mineração de dados e inteligência de negócios.

Em comparação com a rede completa, percebe-se índices do coeficiente de estabilidade muito semelhantes e próximos entre si. Por exemplo, se verificarmos o período que vai de 2012 até 2019, percebemos a ocorrência de pequenas variações entre os módulos destacados.

Sendo assim, são pontos de destaque da pesquisa: os módulos que possu em valores ótimos (máximos e mínimos de estabilidade); os valores ótimos do coeficiente de estabilidade não são síncronos (não ocorrem ao mesmo tempo); e o comportamento do coeficiente de estabilidade dos módulos dominantes é diferente do coeficiente de estabilidade da rede.

Gráfico 19 - Coeficiente de estabilidade da rede de co-palavras-chaves (1987-2019)



Fonte: Elaborado pelo autor.

O Gráfico 20 trata-se de uma representação do coeficiente de inovação, que objetiva analisar os principais módulos de destaque e as principais tendências de palavras-chaves advindas das dissertações e teses sobre a ciência de dados.

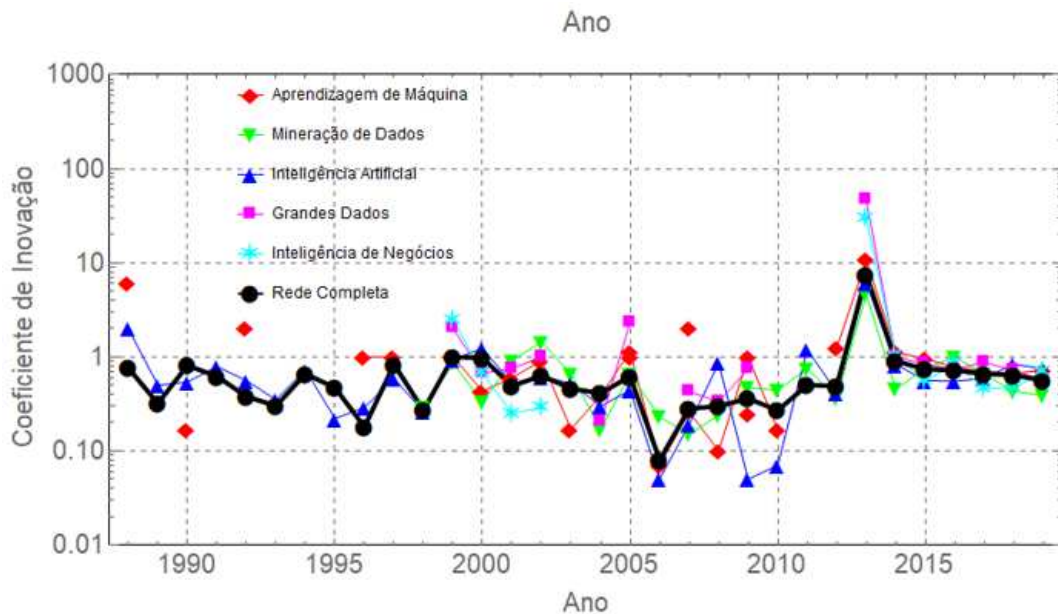
Percebe-se que, de 1987 até 1997, o coeficiente de inovação é considerado baixo, demonstrando que, nesse período, a inteligência artificial e a aprendizagem de máquina são os módulos de destaque, sinalizando que nesse primeiro momento não ocorreram grandes inovações na evolução da rede.

A partir de 1998 até 2012, o coeficiente de inovação é considerado médio, com destaque para alguns módulos, como: mineração de dados e grandes dados, sendo considerado o período em que os índices ainda permanecem baixos, porém com o surgimento de outros módulos que sinalizam a evolução da rede.

Já de 2013 até 2019, o coeficiente de inovação pode ser considerado maior, sinalizando altos índices, destacados por alguns módulos, como: grandes dados e inteligência de negócios.

Se comparar a rede completa, percebe-se que os índices do coeficiente de inovação, após o boom de 2013, vem caindo sensivelmente, sendo que, de 2014 para 2019, esse índice caiu mais de 60%. O nível de correlação entre o que ocorre entre os módulos e a rede completa é muito alto, e isto significa que a inovação acontece de forma síncrona e quase que uniforme durante toda a evolução da rede.

Gráfico 20 - Coeficiente de inovação da rede de co-palavras-chaves (1987-2019)



Fonte: Elaborado pelo autor.

No que tange à contabilidade, existe módulo com termos majoritariamente contábeis, com 43,47% de termos contábeis-financeiros, como, por exemplo: empresa de capital aberto, modelo de custeio, indicador econômico-financeiro e relação de equivalência, conselho administrativo e mix de produção.

Esse módulo formado pelas conexões da contabilidade é considerado pequeno, com apenas 0,39% na rede de co-palavras-chaves, indicando que ainda existe uma grande lacuna para dissertações e teses que versem sobre a ciência de dados em conjunto com a contabilidade.

Convém destacar que, no módulo de aprendizagem de máquina, palavras-chaves como análise discriminante, análise de crédito e indicadores financeiros apontam o caminho principal das pesquisas.

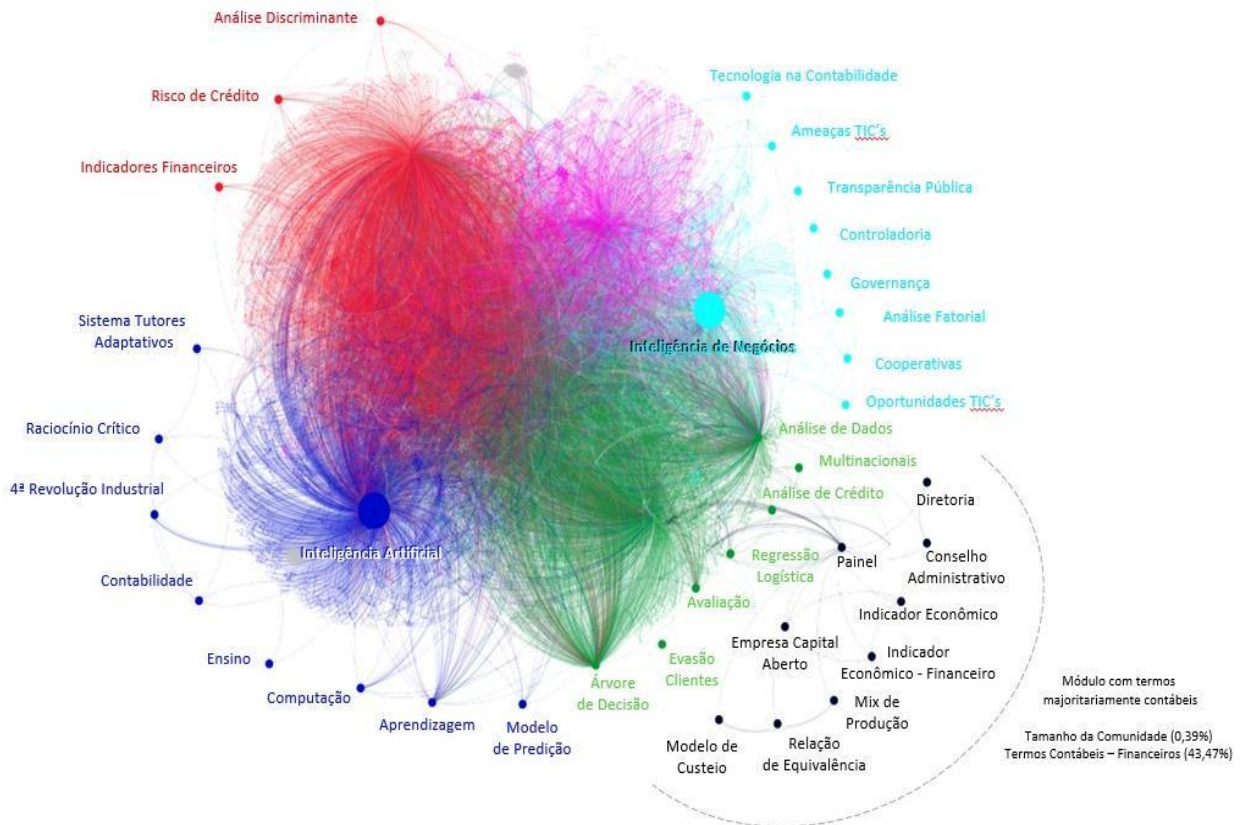
Já no módulo de inteligência artificial, algumas palavras-chaves são destacadas: 4ª Revolução Industrial, modelo de previsão, contabilidade, ensino e aprendizagem.

No módulo da mineração de dados, algumas palavras-chaves são destaque, como: análise de dados, evasão de clientes, multinacionais e análise de crédito, as quais apontam as conexões com a contabilidade.

E, ainda, no módulo de inteligência de negócios, algumas palavras-chaves de destaque apontam as tendências de pesquisa: tecnologia na contabilidade,

controladoria, governança, transparência pública, ameaças TIC's e oportunidades TIC's.

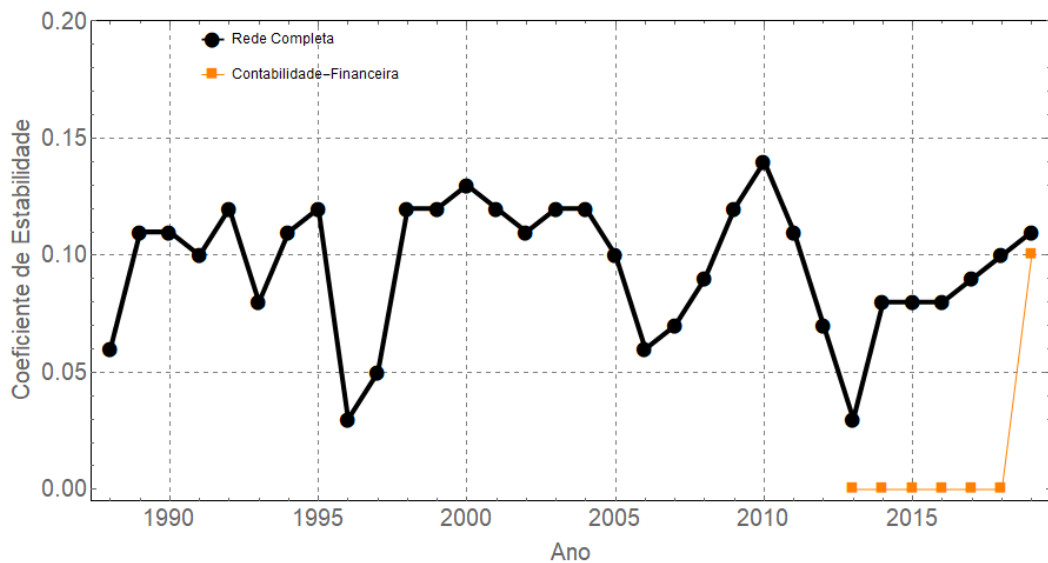
Figura 7 - Conexões da rede de co-palavras-chaves entre a ciências de dados e a ciências contábeis



Fonte: Elaborada pelo autor.

Sendo assim, apresenta-se o coeficiente de estabilidade no Gráfico 21, que demonstra que a partir de 2015 as pesquisas em contabilidade são principalmente na área da contabilidade-financeira, e que não houve nenhum aumento em outras áreas, o que corrobora a existência de espaço para pesquisas que objetivam explorar outras áreas da contabilidade.

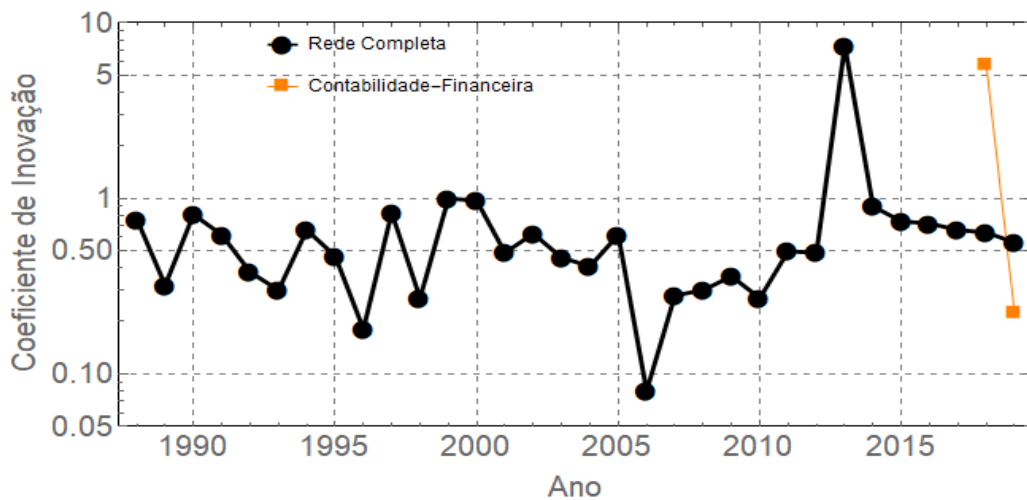
Gráfico 21 - Coeficiente de estabilidade da rede de co-palavras-chaves na contabilidade



Fonte: Elaborado pelo autor.

O coeficiente de inovação, destacado no Gráfico 22, aponta também que, a partir de 2015, a contabilidade-financeira foi a que mais se apropriou das ciências de dados em suas pesquisas, e aponta essa tendência para as futuras pesquisas.

Gráfico 22 - Coeficiente de inovação da rede de co-palavras-chaves na contabilidade



Fonte: Elaborado pelo autor.

Dessa forma, a integração das estatísticas descritivas e da análise da rede possibilitou a construção dos resultados que buscaram investigar na comunidade científica a adoção de pesquisas em ciência de dados por meio das dissertações

e/ou teses, em especial, suas relações com a contabilidade, e o destaque especial para as redes complexas, por intermédio das redes de co-palavras-chaves.

5 CONSIDERAÇÕES FINAIS

A difusão de conhecimentos em ciências de dados é um desafio da atualidade, sendo que a relevância da pesquisa foi o desenvolvimento da educação, em especial, com relação à pós-graduação. Para isso, ocorreu a investigação dos principais fatores de adoção de conhecimentos em ciência de dados e suas influências na formação de mestres e doutores em contabilidade.

A pesquisa apresentou uma proposta inovadora, pois os estudos existentes geralmente fazem a investigação da rede de co-palavras-chaves em artigos científicos, diferente do que ocorreu aqui, que buscou as palavras-chaves disponíveis no Catálogo de Dissertações e Teses da CAPES referentes ao período de 1987 até 2019.

O objetivo geral da pesquisa foi determinar os fatores de adoção de conhecimentos em ciências de dados e suas influências na formação de mestres e doutores. Esse objetivo foi atendido pois demonstrou que, ao longo da evolução temporal da série, foi possível identificar os principais módulos que influenciam a contabilidade brasileira e as pesquisas efetuadas. Sendo assim, foi constatado que 70% concentram-se em 5 módulos, onde houve um maior número de *hubs* e, por consequência, de conexões com a área contábil, sendo eles: aprendizagem de máquina, mineração de dados, inteligência artificial, grandes dados e inteligência de negócios.

Essa revolução tecnológica proposta pela ciência de dados impulsionou o mundo dos negócios, as inovações disruptivas, a contabilidade, o que afeta diretamente o mercado de trabalho, que busca profissionais com expertise em lidar com grandes dados e informações. Por outro lado, destaca-se que o mundo acadêmico precisa se reinventar no ensino e aprendizagem, com a inclusão gradual em seus currículos de técnicas e ferramentas da ciência de dados que busquem apoiar o desenvolvimento e formação dos mestres e doutores, para que possam replicar estes conhecimentos junto à comunidade científica e contribuir para a sociedade.

Em relação ao que foi proposto pelos objetivos específicos, convém lembrar que o primeiro foi determinar a estrutura da rede de palavras-chaves formada entre as pesquisas. Tal objetivo foi alcançado, pois foi possível identificar os principais módulos desenvolvidos ao longo da série, onde foi possível destacar três etapas

distintas: a) a primeira, de 1987 até 2005, com ocorrência de poucas conexões, mas com existência de correlações positivas fortes com o número de conexões; b) a segunda, de 2006 a 2012, com um número baixo de conexões e, por consequência, poucos módulos de destaque; e c) a terceira, de 2013 a 2019, onde ocorreu o aumento das conexões e o surgimento de módulos distintos, a partir do boom da rede em 2013, período em que ocorreu o aumento do número de dissertações e/ou teses. O segundo objetivo foi investigar a evolução temporal das interações, que pode ser confirmado por meio da rede de co-palavras-chaves formada ao longo do tempo, e que foram bem destacadas na Figura 4. E o terceiro objetivo, foi identificar os mecanismos responsáveis pela formação de padrões na estrutura da rede, que pode ser confirmado a partir do período de 2013, no qual ocorreu o boom da rede com o aumento do número de dissertações e/ou teses, conexões, *hubs* e módulos identificando um padrão a ser seguido pela rede.

Por intermédio da rede de co-palavras-chaves, foi possível evidenciar alguns achados importantes, como os principais módulos desenvolvidos pelas conexões diretas ou indiretas com a contabilidade ao longo da série.

Percebe-se a predominância de termos ligados inicialmente ao mundo computacional, que vem se expandindo de forma lenta e gradual para a área contábil, sendo que o módulo de aprendizagem de máquina (26,21%) é o maior módulo formado, mas que ainda não apresenta uma estrutura própria, trazendo termos ligados indiretamente com a contabilidade, a saber: indicadores financeiros, análise de crédito e análise discriminante.

Se olharmos por outra ótica, a inteligência de negócios (5,02%) apontou termos que se correlacionam diretamente com a contabilidade, como: transparência pública, controladoria e governança, apontando uma ligação com áreas de controles de gestão que são importantes para o desenvolvimento das organizações diante de um mercado cada vez mais competitivo e que carece de inovação.

Ainda assim, é perceptível a formação de um módulo exclusivo de termos contábeis na rede, com apenas 0,39% de termos ligados à Contabilidade Financeira, sendo eles: indicadores econômico-financeiros, modelo de custeio, diretoria, conselho administrativo, empresa de capital aberto. É considerado um módulo pequeno, mas que apontou um achado importante com a existência de um módulo formado estritamente de termos contábeis, diante dos demais módulos que se correlacionam com a contabilidade.

Os instrumentos de coleta de dados, já alinhados com a proposta inovadora que buscou agregar valor à pesquisa em termos de otimização do tempo, destacando a coleta automatizada ao invés da coleta manual, percorrendo um percurso metodológico, que foi desde a investigação no Catálogo de Dissertações e Teses da Capes, a seleção da 20 palavras-chaves baseadas na literatura, a mineração de dados, uso da linguagem computacional (Python, Gephi), a análise descritiva e as redes de co-palavras-chaves, podem ser avaliados positivamente, pois trouxeram maior confiabilidade, robustez e agilidade durante todo esse processo.

A pesquisa foi fundamentada pela teoria de difusão da inovação de Rogers, que pode ser confirmada nesta pesquisa, pois ficou evidenciada a evolução da rede ao longo do tempo, com a criação dos principais módulos, como foram criados, desenvolvidos e divulgados no sistema social, sendo que tais aspectos podem ser comparados a comunidades humanas que se conectam por sua linguagem, escrita, comportamento, compartilhamento de interesses e perspectivas comuns.

Outro achado importante foram os coeficientes de estabilidade e inovação, que trouxeram informações importantes no que se refere a tendências e rumos das pesquisas, em especial na contabilidade.

O coeficiente de estabilidade apontou que, a partir de 2015, a contabilidade financeira é quem obteve maior destaque, evidenciando uma grande lacuna ainda existente por meio dos termos encontrados, sendo possível que outras áreas tão importantes da área contábil busquem ganhar seu espaço, destacando-se, entre elas: controle de gestão, controladoria, auditoria, contabilidade de custos, contabilidade pública e perícia contábil.

O coeficiente de inovação sinaliza por meio da contabilidade financeira que pode ser considerada a pioneira na rede de co-palavras-chaves, inserindo a inovação na rede, com a formação de um módulo contábil independente de maneira ainda discreta, mas que já sinaliza a tendência e aponta o rumo para futuras pesquisas.

Quanto às limitações da presente pesquisa, menciona-se a relevância da definição das palavras-chaves de maneira apropriada, sendo que as mesmas serão indicadas nas dissertações e teses, sendo que devem ser muito bem pensadas pelos pesquisadores, para facilitar a compreensão, objetivando o aumento do índice de qualidade, desempenho e otimização entre elas, de forma a evitar algum viés na

escolha e evitando problemas semânticos, de subjetividade e até mesmo de entendimento dos pesquisadores, que, por consequência, podem afetar estudos da formação de redes de co-palavras-chaves.

Como proposta de estudos futuros, destaca-se a própria replicação da pesquisa junto ao Catálogo de Dissertações e Teses da Capes, a fim de verificar a evolução temporal da rede de co-palavras-chaves e o surgimento de novos módulos alinhados com a contabilidade. A inserção de métodos ou ferramentas advindos da ciência de dados na contabilidade também é incentivada, com intuito de contribuir para a formação dos profissionais e acadêmicos trazendo inovações para a literatura e o mercado.

Por fim, espera-se que este estudo tenha contribuído com informações relevantes sobre a adoção de conhecimentos em ciência de dados na pós-graduação brasileira, e sua contribuição na formação de mestres e doutores em contabilidade, aproximando as universidades, discentes e docentes de uma nova realidade demandada pelo mercado de trabalho.

REFERÊNCIAS

ABASSI, A., ALTMANN, J. On the correlation between research performance and social network analysis measures applied to research collaboration networks. *In: HAWAII INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES*, 41., 2011. **Proceedings [...]**. Waikoloa, HI: IEEE, 2011.

AGUIAR, Gisleise; GOUVEIA, Luis; RODRIGUES, Fabio. Accounting professionals and digital maturity: insight from the reflections of digital transformation. **Brazilian Journal of Business**, São José dos Pinhais, v. 3, n. 4, p. 3009-3029, 2021.
ALBERT, Réka; BARABASI, Albert-László. Statistical mechanics of complex networks. **Reviews of Modern Physics**, [S. l.], n. 74, p. 47-97, Jan. 2002.

ALLOGHANI, Mohamed *et al.* A systematic review on supervised and unsupervised machine learning algorithms for data science. *In: BERRY, Michael; MOHAMED, Azlinah; YAP, Bee Wah (ed.)*. **Unsupervised and semi-supervised learning**. Arkansas: Computer Science Department, 2020. p. 3-21. *E-book*. Disponível em: https://link.springer.com/chapter/10.1007/978-3-030-22475-2_1. Acesso em: 23 jun. 2021.

ALMEIDA, José Elias Feres. Revolução tecnológica no mundo dos negócios e algumas oportunidades e desafios na área contábil. **Revista de Contabilidade e Organizações**, Ribeirão Preto, v. 14, p. e165516-e165516, 2020.

AMANI, Farzaneh A.; FADLALLA, Adam M. Aplicações de mineração de dados em contabilidade: uma revisão da literatura e estrutura organizacional. **International Journal of Accounting Information Systems**, [S. l.], v. 24, p. 32-58, 2017.

ARRUDA, Danniely Cristiny Soares de; GOMES, Érika Zabala; SANTOS, Cleston Alexandre dos. Uma análise da percepção dos profissionais da área de contabilidade do município de Corumbá-MS sobre o SPED. **Revista Semana Acadêmica**. Fortaleza, 2013. Disponível em: https://semanaacademica.org.br/system/files/artigos/artigo_sped.pdf. Acesso em: 8 out. 2019.

ASSOCIATION TO ADVANCE COLLEGIATE SCHOOLS OF BUSINESS (AACSB). **Eligibility procedures and accreditation standards for accounting accreditation: engagement, innovation and impact**. Tampa, FL: AACSB, 2018. Disponível em: <https://www.aacsb.edu/-/media/aacsb/docs/accreditation/accounting/standards-and-tables/2018-accountingstandards.ashx?la=en&hash=8DCDA6CE3B0CEF6AB82D39CBF53995DA96111196>. Acesso em: 06 mar 2022.

BARABÁSI, Albert-László; ALBERT, Réka. Emergence of scaling in random networks. **Science**, [S. l.], v. 286, p. 509-512, Oct. 1999. Disponível em: https://science.sciencemag.org/content/286/5439/509.abstract?casa_token=Nv0BE04ab_EAAAAA:VMuwD8Sm7OI8iYBDqoagbYX6LQx01TLUZB9UeiBnhnQiXGAlv_OEyCwwOloWUGUHaeFDG3x7ldGeLQ. Acesso em: 08 jul. 2021.

BARRAT, Alain; BARTHELEMY, Marc; VESPIGNANI, Alessandro. **Dynamical processes on complex networks**. Cambridge: Cambridge University Press, 2008.

BASS, Frank M. A new product growth for model consumer durables. **Management Science**, Maryland, v. 15, n. 8, p. 215-227, Jan. 1969.

BASTIAN, Mathieu; HEYMANN, Sebastien; JACOMY, Mathieu. Gephi: an open source software for exploring and manipulating networks. *In*: INTERNATIONAL ICWSM CONFERENCE, 3., 2009, Califórnia. **Anais eletrônicos [...]**. Califórnia, CA, p. 361-362, 2009. Disponível em: <https://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154/1009>. Acesso em: 24 jun. 2021.

BELLOUM, Adam S. Z. *et al.* Bridging the demand and the offer in data science. **Concurrency and Computation: practice and experience**, [S. l.], v. 31, n. 17, p. e5200, 2019.

BOCCALETTI, Stefano *et al.* Complex networks: structure and dynamics. **Physics reports**, [S. l.], v. 424, n. 4-5, p. 175-308, 2006.

CALDERON, Thomas G.; CHEH, John J. Um roteiro para futuras pesquisas de redes neurais em auditoria e avaliação de risco. **International Journal of Accounting Information Systems**, [S. l.], v. 3, n. 4, p. 203-236, 2002.

CARILLO, Kevin Daniel André. Let's stop trying to be "sexy" – preparing managers for the (big) data-driven business era. **Business Process Management Journal**, [S. l.], v. 23, n. 3, p. 598-622, 2017.

CHEN, C. L. P.; ZHANG, C.-Y. Aplicações, desafios, técnicas e tecnologias com uso intensivo de dados: Uma pesquisa sobre Big Data. **Ciências da Informação**, Brasília, DF, v. 275, n. 10, p. 314-347, 2014.

COAKLEY, James R.; BROWN, Carol E. Redes neurais artificiais em contabilidade e finanças: questões de modelagem. **Sistemas Inteligentes em Contabilidade, Finanças e Gestão**, [S. l.], v. 9, n. 2, p. 119-144, 2000.

COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR (CAPES). **Dados aberto da CAPES**. Conjuntos de dados: catálogo de dissertações e teses: 1987-2019. Brasília, DF: CAPES, 2021. Disponível em: <https://dadosabertos.capes.gov.br/dataset?organization=diretoria-de-avaliacao>. Acesso em: 12 mar. 2021.

COSENZA, José Paulo; ROCCHI, Carlos Antonio de. Evolução da escrituração contábil: Desenvolvimento e utilização do sistema ficha tríplice no Brasil. **Revista de Contabilidade do Mestrado em Ciências Contábeis da UERJ**, Rio de Janeiro, v. 19, n. 1, p. 4-23, jan./abr. 2014. Disponível em: <http://www.spell.org.br/documentos/ver/30612/evolucao-da-escrituracao-contabil--desenvolvime>. Acesso em: 28 fev. 2022.

COSTA, Luciano D. *et al.* Characterization of complex networks: a survey of measurements. **Advances in Physics**, [S. l.], v. 56, p. 167-242, 2007.

COYNE, J. G.; COYNE, E. M.; WALKER, K. B. A model to update accounting curricula for emerging technologies. **Journal of Emerging Technologies in Accounting**, [S. l.], v. 13, n. 1, p. 161-169, 2016. DOI: <https://doi.org/10.2308/jeta-51396>.

- DAVID, Paul A. **A contribution to the theory of diffusion**. Stanford, CA: Research Center in Economic Growth, Stanford University, 1969.
- DAVIES, Stephen. **The diffusion of process innovations**. Cambridge: Cambridge University Press, 1979.
- DREMEL, Christian *et al.* Archetypes of data analytics providers in the Big Data Era. *In: AMERICAS CONFERENCE ON INFORMATION SYSTEMS*, 24., 2018, New Orleans. **Anais eletrônicos [...]**. New Orleans, LA, Aug. 2018. Disponível em: https://www.researchgate.net/publication/331963474_Archetypes_of_Data_Analytics_Providers_in_the_Big_Data_Era. Acesso em: 19 maio 2021.
- DZURANIN, Ann C.; JONES, Janet R.; OLVERA, Renee M. Infundindo a análise de dados no currículo contábil: uma estrutura e insights do corpo docente. **Journal of Accounting Education**, [S. l.], v. 43, p. 24-39, 2018.
- FERREIRA, Pedro Campos *et al.* How Computer Science Undergraduate Programs in Brazilian Public Universities Address Artificial Intelligence and its topics: a preliminary report. *Learning and Nonlinear Models*, [S. l.], v. 12, p. 73-85, 2012.
- FILATRO, Andrea. **Data science na educação: presencial, a distância e corporativa**. São Paulo: Saraiva, 2020.
- FURTADO, André. Difusão tecnológica: um debate superado? *In: PELAEZ, Victor; SZMRECSÁNYI, Tamás. Economia de inovação tecnológica*. São Paulo: Hucitec, 2006. p. 686-687.
- GIBERT, Karina *et al.* Environmental data science. **Environmental Modelling & Software**, [S. l.], v. 106, p. 4-12, Aug. 2018. Disponível em: https://www.science-direct.com/science/article/pii/S1364815218301269?casa_token=EwSo0ljy9rwAAAAA:X-AUvyFrXqnf1UmLXrICKP0zgUqK8ZDH7K-pWLFeyAYm9mzVaN3gzTyhcYttOoSLDpcYOCyjlpq Acesso em: 22 jun. 2021.
- GORMAN, Michael F.; KLIMBERG, Ronald K. Benchmarking academic programs in business analytics. **Interfaces**, [S. l.], v. 44, n. 3, p. 329-341, 2014.
- GRILICHES, Zvi. Hybrid corn: an exploration in the economics of technological change. **Econometrica**, New York, v. 25, n. 4, p. 501-522, Oct. 1957. Disponível em: <https://econ398.academic.wlu.edu/files/2017/01/Griliches-1957-Hybrid-Corn-An-Exploration-in-the-Economics-of-Te.pdf>. Acesso em: 17 jul. 2021.
- HALL, Bronwyn H. **Innovation and diffusion**. Cambridge, MA: The National Bureau of Economic Research, 2004. (Working Paper, n. 10212).
- HOELSCHER, Jamie; MORTIMER, Amanda. Using tableau to visualize data and drive decision-making. **Journal of Accounting Education**, [S. l.], v. 44, p. 49-59, 2018. DOI: <http://doi.org/10.1016/J.JACCEDU.2018.05.002>.
- JACKSON, Joyce. Data mining; a conceptual overview. **Communications of the Association for Information Systems**, [S. l.], v. 8, n. 1, p. 19, 2002.

- JANVRIN, Diane J.; WATSON, Marcia Weidenmier. "Big Data": a new twist to accounting. **Journal of Accounting Education**, [S. l.], v. 38, p. 3-8, 2017.
- KING, Brian R.; SATYANARAYANA, Ashwin. Teaching data mining in the Era of Big Data. *In*: PASEE ANNUAL CONFERENCE & EXPOSITION, 120., 2013. **Proceedings [...]**. [S. l.], 2013.
- LARSON, Deanne; CHANG, Victor. A review and future direction of agile, business intelligence, analytics and data science. **International Journal of Information Management**, [S. l.], v. 36, n. 5, p. 700-710, 2016. Disponível em: https://www.sciencedirect.com/science/article/pii/S026840121630233X?casa_token=w8tVEVxxQpQAAAAA:GTyo5xZ7fpSqOLX4XZ4EmUk13GQnSCak9pCs0OsyI6G_arlRtoa19av4f3lYegbMmaoCKLGrwg_. Acesso em: 23 jun. 2021.
- LAWLESS, Kimberly A.; PELLEGRINO, James W. Desenvolvimento profissional na integração da tecnologia no ensino e aprendizagem: conhecidos, desconhecidos e maneiras de buscar melhores perguntas e respostas. **Revisão da Pesquisa Educacional**, [S. l.], v. 77, n. 4, p. 575-614, 2007.
- LAWSON, Raef A. *et al.* Focusing accounting curricula on students' long-run careers: recommendations for an integrated competency-based framework for accounting education. **Issues in Accounting Education**, [S. l.], v. 29, n. 2, p. 295-317, 2014. DOI: <http://doi.org/10.2308/iace-50673>.
- MANSFIELD, Edwin. Technical change and the rate of imitation. **Econometrica**, Nova York, v. 29, n. 4, p. 741-766, 1961.
- MASHINGAIDZE, Kenneth; BACKHOUSE, Judy. The relationships between definitions of big data, business intelligence and business analytics: a literature review. **International Journal of Business Information Systems**, [S. l.], v. 26, n. 4, p. 488-505, 2017.
- MELLO, Bernardo Assunção *et al.* **Teoria de redes complexas e o poder de difusão dos municípios**. Brasília, DF: Instituto de Pesquisa Econômica Aplicada (IPEA), 2010. (Texto para Discussão, n. 1484).
- METCALFE, Stan. Impulse and diffusion in the study of technical change. **Futures**, [S. l.], v. 13, n. 5, p. 347-359, 1981.
- METZ, Jean *et al.* Redes complexas: conceitos e aplicações. **Relatórios Técnicos do ICMC**, São Carlos, n. 290, jan. 2007. Disponível em: https://web.icmc.usp.br/SCATUSU/RT/BIBLIOTECA_113_RT_290.pdf. Acesso em: 28 mar. 2021.
- MIJALKOV, Mite *et al.* Braph: a graph theory software for the analysis of brain connectivity. **Plos One**, [S. l.], v. 12, n. 8, p. e0178798, 2017.
- MORRAR, Rabeh; ARMAN, Husam; MOUSA, Said. A quarta revolução industrial (Indústria 4.0): uma perspectiva de inovação social. **Revisão da Gestão da Inovação Tecnológica**, [S. l.], v. 7, n. 11, p. 12-20, 2017.
- NASUTION, Mahyuddin K. M. *et al.* Data science: a review towards the Big Data Problems. **Journal of Physics: Conference Series**, Indonésia, v. 1898, p. 1-8, Dec.

2020. Disponível em: <https://iopscience.iop.org/article/10.1088/1742-6596/1898/1/012006/meta>. Acesso em: 23 jun. 2021.

NEWMAN, Mark. **Networks: an introduction**. New York: Oxford Scholarship Online, Sep. 2010.

NEWMAN, Mark. The structure and function of complex networks. **SIAM Review**, [S. l.], v. 45, p. 167-256, Aug. 2003.

NGAI, Eric WT *et al.* A aplicação de técnicas de mineração de dados na detecção de fraudes financeiras: uma estrutura de classificação e uma revisão acadêmica da literatura. **Sistemas de Apoio à Decisão**, [S. l.], v. 50, n. 3, p. 559-569, 2011.

PRICEWATERHOUSECOOPERS (PWC). **Data driven: what students need to succeed in a rapidly changing business world**. [S. l.], 2015. Disponível em: <https://www.pwc.com/us/en/facultyresource/assets/pwc-data-driven-paper-feb2015.pdf>. Acesso em: 06 mar. 2022.

RIKHARDSSON, Pall; YIGITBASIOGLU, Ogan. Business intelligence & analytics in management accounting research: status and future focus. **International Journal of Accounting Information Systems**, [S. l.], v. 29, p. 37-58, 2018.

ROGERS, Everett M. **Diffusion of innovation**. New York: The Free Press, 1983.

ROGERS, Everett M. **Diffusion of innovation**. New York: The Free Press, 2003.

ROGERS, Everett M. **Diffusion of innovations**. New York: The Free Press, 1962.

ROGERS-PETTIE, Christa; HERRMANN, Jeffrey. Information diffusion: a study of Twitter during large scale events. **IIE Annual Conference. Proceedings**, [S. l.], p. 1591, 2015.

ROSENBERG, Nathan. **Perspectives on technology**. Cambridge: Cambridge University Press, 1976.

RUSCHEL, Marcia Erna; FREZZA, Ricardo; UTZIG, Mara Jaqueline Santore. O impacto do SPED na contabilidade desafios e perspectivas do profissional contábil. **Revista Catarinense da Ciência Contábil**, Florianópolis, v. 10, n. 29, p. 09-26, 2011.

SÁ, A. L. de. **História geral e das doutrinas da contabilidade**. São Paulo: Atlas, 1997.

SCHUMPETER, Joseph Alois. **The theory of economic development**. Cambridge, MA: Harvard University Press, 1912.

SERRAT, Olivier. Social network analysis. *In*: SERRAT, Olivier. **Knowledge solutions**. Singapore: Springer, 2017. p. 39-43.

SILVERBERG, Gerald. Modeling economic dynamics and technical change: mathematical approaches to self-organization and evolution. *In*: DOSI, Giovanni;

FREEMAN, Christopher; NELSON, Richard; SILVERBERG, Gerald; SOETE, Luc. **Technical change and economic theory**. Londres: Pinter, 1988.

SOMBRA, Tobias Ribeiro *et al.* Redes sociais científicas e inteligência artificial – uma revisão sistemática aplicada a reconhecimento de padrões. **Brazilian Journal of Development**, São José dos Pinhais, v. 6, n. 3, p. 9957-9970, 2020. Disponível em: <https://www.brazilianjournals.com/index.php/BRJD/article/view/7321/6396>. Acesso em: 07 jul. 2021.

STODDER, David. BI e analytics na Era da IA e Big Data. **Relatório de Melhores Práticas TDWI**, [S. l.], p. 1-39, 2018.

STONEMAN, Paul. **The economic analysis of technological change**. New York: Oxford University Press, 1983.

TARWANI, K. M.; SAUDAGAR, S. S.; MISALKAR, H. D. Aprendizado de máquina em análise de big data: visão geral. **Revista Internacional de Pesquisa Avançada em Ciência da Computação e Engenharia de Software**, v. 5, n. 4, 270-274, 2015.

TYSIAC, Ken; DREW, Jeff. Accounting firms: the next generation. **Journal of Accountancy**, [S. l.], 1 June 2018. Disponível em: <https://www.journalofaccountancy.com/issues/2018/jun/next-generation-accounting-firms.html>. Acesso em: 06 mar. 2022.

VASARHELYI, Miklos A.; TEETER, Ryan A.; KRAHEL, J. P. Educação em auditoria e economia em tempo real. **Questões no Ensino Contábil**, [S. l.], v. 25, n. 3, p. 405-423, 2010.

VIEIRA FILHO, Jose Eustáquio Ribeiro; SILVEIRA, José Maria Ferreira. Mudança tecnológica na agricultura: uma revisão crítica da literatura e o papel das economias de aprendizado. **Revista de Economia e Sociologia Rural**, Brasília, v. 50, n. 4, p. 721-742, 2012. Disponível em: <https://www.scielo.br/j/resr/a/Pjz4mbbbKwDz8Vm4sbDY7mR/?lang=pt>. Acesso em: 07 jul. 2021.

WANG, João; YANG, James G. S. Técnicas de mineração de dados para auditoria atestam função e detecção de fraudes. **Journal of Forensic Investigative Accounting**, [S. l.], v. 1, n. 1, p. 4-10, 2006.

WANG, Shiguo. Uma pesquisa abrangente de pesquisa de detecção de fraude contábil baseada em mineração de dados. *In*: INTERNATIONAL CONFERENCE ON INTELLIGENT COMPUTATION TECHNOLOGY AND AUTOMATION, 2010. **Proceedings [...]**. [S. l.]: IEEE, 2010. p. 50-53.

WARREN JR, J. Donald; MOFFITT, Kevin C.; BYRNES, Paulo. Como o big data mudará a contabilidade. **Horizontes Contábeis**, [S. l.], v. 29, n. 2, p. 397-407, 2015.

WATTS, Duncan J.; STROGATZ, Steven H. Collective dynamics of small-world networks. **Nature**, [S. l.], n. 393, p. 440-442, 1998.

WAZURKAR, Parth; BHADORIA, Robin Singh; BAJPAI, Dhananjai. Predictive analytics in data science for business intelligence solutions. *In*: INTERNATIONAL

CONFERENCE ON COMMUNICATION SYSTEMS AND NETWORK TECHNOLOGIES (CSNT), 7., 2017, Índia. **Anais [...]**. Índia, 2017. p. 367-370.

WIXOM, Barbara *et al.* The current state of business intelligence in academia: The arrival of big data. **Communications of the Association for information Systems**, [S. l.], v. 34, n. 1, p. 1, 2014.


```

                                lambda                'utf-8' # remove acentuação

print "                :: OK!"
# -----

# --- PASSO 3: Busca Palavras-Chave --- #
print "PASSO 3 :: running ..."

                                # lista -> serie
str                               # sub. minuscula
lambda                            'utf-8' # remove acentuação
                                # Lista de Termos Normalizado

# Busca

for in
# Palavras - chave

# Titulo

                                list map lambda

# Resumo

                                list map lambda

                                list filter None      # index kw
                                list filter None      # index t
                                list filter None      # index r
                                list filter None      # termo t
                                list filter None      # termo r

print "                :: OK!"
# -----

# # --- PASSO 4: Palavras-Chaves (Gerador) --- #
print "PASSO 4 :: running ..."

# Conjuntos indices encontrados
set list
set list
set list
                                # não repetidos

# Filtra
                                list                # match linhas

# concatena termo (titulo, resumo) em Palavras-Chave

```

```

for i in range(len(PalavrasChave)):
    for j in range(len(PalavrasChave)):
        #customiza o tradutor
        # stematizador (português)
        # filtro (stopwords + pontuação)
        set
        set list
print
# formatação
for i in range(len(PalavrasChave)):
    # Remove Pontuação
    for i in range(len(PalavrasChave)):
        for j in range(len(PalavrasChave)):
            if not in
            # remove sequencia de virgulas
            # remove virgula inicio & fim
            if
            if
            # Tradução (ingles -> portugues)
            #conc_l=[maq.translate(x).strip() for x in conc_l]
            for i in range(len(PalavrasChave)):
                try:
                    # Normalização (remove stopwords + pontuação + acentos) + Stematização
                    for i in range(len(PalavrasChave)):
                        str 'utf-8' for i in range(len(PalavrasChave)):
                            if not in
                            # remove kw repetidas
                            list set
                            # remove virgula inicio & fim
                            if
                            if
                            # remove sequencia de virgulas
                            r'[, ]*(,|,)'

```

```

# substitui kw (antigas -> novas)
    'aprendiz' 'aprend'
    'big dat' 'grand dad'
    'megad' 'grand dad'
    'macrod' 'grand dad'
    'inteligenc empresario' 'inteligenc negoci'
'PalavrasChave'
print '          :: str round      100 s 'S'

print '          :: OK!'
# -----

# --- PASSO 5: Filtro + Lista Palavras - Chave (Saída) ---
print 'PASSO 5 :: running ...'

    'PalavrasChave'      'PalavrasChave_' str      '.xlsx'      False      False
    'Filtro_' str      '.xlsx'      False

print '          :: OK!'
print 'TEMPO :: str round
print '***** FIM *****'
# -----

```

ANEXO B - SCRIPT MATRIZ DE CO-PALAVRAS-CHAVES

```

#Matriz de Coocorrencia
import sys
import os

year = 2019
keywords = "PalavrasChave_" + str(year) + ".xlsx"
output = None

# Create a dictionary of keywords
keywords_list = list(set(sum(
    [os.listdir(os.path.join(year, keyword))
      for keyword in range(len(
        dict(zip(year, range(len(
          for year in enumerate(
            for keyword in
              if keyword is not None
                ]
            )
          )
        )
      )
    )
  )
))

# coocurrence matrix
#Xc = (X*X.T) # co-keyword matrix

# Save the matrix to a CSV file
together = str(year) + ".csv"

```