



**PROGRAMA DE PÓS-GRADUAÇÃO MULTI-INSTITUCIONAL
EM DIFUSÃO DO CONHECIMENTO**

SIMONE GONSALVES MENDES DE ARAÚJO

**MODELAGEM DE REDES DE PESQUISA E COLABORAÇÃO
CIENTÍFICA NA UEFS SOB A PERSPECTIVA MULTI E
INTERDISCIPLINAR**

Salvador

2025



SIMONE GONSALVES MENDES DE ARAÚJO

MODELAGEM DE REDES DE PESQUISA E COLABORAÇÃO CIENTÍFICA NA UEFS SOB A PERSPECTIVA MULTI E INTERDISCIPLINAR

Texto apresentado ao Programa de Pós-Graduação Multi-Institucional em Difusão do Conhecimento, como requisito ao Exame de Defesa de Tese.

Áreas de Concentração: Modelagem da Geração e Difusão do Conhecimento

Linha 02 - Difusão do Conhecimento: Informação, Comunicação e Gestão.

Orientadora: Prof^ª. Dr^ª. Débora Abdalla Santos

Coorientadora: Prof^ª. Dr^ª. Júlia Carvalho Andrade

Salvador

2025

Araújo, Simone Gonsalves Mendes de.

Modelagem de redes de pesquisa e colaboração científica na UEFS sob a perspectiva multi e interdisciplinar [recurso eletrônico] / Simone Gonsalves Mendes de Araújo. - Dados eletrônicos. - 2025.

Orientadora: Prof.^a Dr.^a Débora Abdalla Santos.

Coorientadora: Prof.^a Dr.^a Julia Carvalho Andrade.

Tese (Doutorado em Difusão do Conhecimento) - Programa de Pós-Graduação em Difusão do Conhecimento, Salvador, 2025.

Disponível em formato digital.

Modo de acesso: <https://repositorio.ufba.br/>

1. Ensino superior -Pesquisa. 2. Modelagem - Pesquisa científica. 3. Colaboração científica. 4. Disseminação do conhecimento. 5. Ciências - Pesquisa. 6. Análise de redes. I. Santos, Débora Abdalla. II. Andrade, Julia Carvalho. III. Programa de Pós-Graduação em Difusão do Conhecimento. IV. Título

CDD 378.0072 - 23. ed.




Universidade Federal da Bahia


**PROGRAMA DE PÓS-GRADUAÇÃO EM DIFUSÃO DO CONHECIMENTO
(DMMDC)**

ATA Nº 90

Ata da sessão pública do Colegiado do PROGRAMA DE PÓS-GRADUAÇÃO EM DIFUSÃO DO CONHECIMENTO (DMMDC), realizada em 12/02/2025 para procedimento de defesa da Tese de DOUTORADO EM DIFUSÃO DO CONHECIMENTO no. 90, área de concentração MODELAGEM DA GERAÇÃO E DIFUSÃO DO CONHECIMENTO - CONSTRUÇÃO DO CONHECIMENTO: COGNIÇÃO, LINGUAGENS E INFORMAÇÃO, do(a) candidato(a) SIMONE GONSALVES MENDES DE ARAÚJO, de matrícula 217123226, intitulada MODELAGEM DE REDES DE PESQUISA E COLABORAÇÃO CIENTÍFICA NA UEFS SOB A PERSPECTIVA MULTI E INTERDISCIPLINAR. Às 09:00 do citado dia, <https://conferenciaweb.rnp.br/sala/debora-abdalla-santos>, foi aberta a sessão pelo(a) presidente da banca examinadora Profª. Dra. DEBORA ABDALLA SANTOS que apresentou os outros membros da banca: Prof. Dr. ANTONIO CARLOS DOS SANTOS SOUZA, Profª. Dra. REINALDA SOUZA OLIVEIRA, Prof. Dr. JOSÉ DAMIÃO DE MELO, Profª. Dra. Júlia Carvalho Andrade e Prof. Dr. PABLO VIEIRA FLORENTINO. Em seguida foram esclarecidos os procedimentos pelo(a) presidente que passou a palavra ao(à) examinado(a) para apresentação do trabalho de Doutorado. Ao final da apresentação, passou-se à arguição por parte da banca, a qual, em seguida, reuniu-se para a elaboração do parecer. No seu retorno, foi lido o parecer final a respeito do trabalho apresentado pela candidata, tendo a banca examinadora aprovado o trabalho apresentado, sendo esta aprovação um requisito parcial para a obtenção do grau de Doutor. Em seguida, nada mais havendo a tratar, foi encerrada a sessão pelo(a) presidente da banca, tendo sido, logo a seguir, lavrada a presente ata, abaixo assinada por todos os membros da banca.

Documento assinado digitalmente
 JOSE DAMIAO DE MELO
Data: 07/03/2025 20:25:39-0300
Verifique em <https://validar.itl.gov.br>

Dra. REINALDA SOUZA OLIVEIRA, UEFS

Documento assinado digitalmente
 REINALDA SOUZA OLIVEIRA
Data: 07/03/2025 09:53:45-0300
Verifique em <https://validar.itl.gov.br>

Dr. JOSÉ DAMIÃO DE MELO, IFSE

Examinador Interno
Documento assinado digitalmente
 JULIA CARVALHO ANDRADE
Data: 20/02/2025 14:53:45-0300
Verifique em <https://validar.itl.gov.br>

Dra. Júlia Carvalho Andrade, UFBA

Examinadora Externa à Instituição
Documento assinado digitalmente
 PABLO VIEIRA FLORENTINO
Data: 07/03/2025 16:10:29-0300
Verifique em <https://validar.itl.gov.br>

Dr. PABLO VIEIRA FLORENTINO

Examinador Externo à Instituição



Universidade Federal da Bahia

**PROGRAMA DE PÓS-GRADUAÇÃO EM DIFUSÃO DO CONHECIMENTO
(DMMDC)**

Documento assinado digitalmente
gov.br **ANTONIO CARLOS DOS SANTOS SOUZA**
Data: 07/03/2025 20:32:19-0300
Verifique em <https://validar.itl.gov.br>

ANTONIO CARLOS DOS SANTOS SOUZA, IFBA

Examinador Interno

Documento assinado digitalmente
gov.br **DEBORA ABDALLA SANTOS**
Data: 20/02/2025 15:16:28-0300
Verifique em <https://validar.itl.gov.br>

Dra. DEBORA ABDALLA SANTOS, UFBA

Presidente

Documento assinado digitalmente
gov.br **SIMONE GONSALVES MENDES DE ARAÚJO**
Data: 10/03/2025 11:17:22-0300
Verifique em <https://validar.itl.gov.br>

SIMONE GONSALVES MENDES DE ARAÚJO

Doutorando(a)

Para poder sobreviver, um organismo necessita ou blindar-se (como uma árvore ou um marisco) e “esperar pelo melhor”, ou então desenvolver métodos para escapar de possíveis danos, indo para vizinhas melhores. Se você escolher esta última opção, você se confrontará com o problema primordial que todo agente precisa continuamente resolver: e agora, o que devo fazer?

(A Consciência Explicada. Dennett, Daniel C.. A Consciência Explicada. Porto Alegre: Bookmann, 2004, p. 95)

AGRADECIMENTOS

Nesse texto falamos o tempo todo sobre colaborações, redes, conexões.... Nesse momento, e antes de falar teoricamente sobre tudo isso, gostaria de falar emocionalmente sobre este assunto, sobre as minhas redes, as minhas relações, conexões, colaborações... Poderia dizer tantos sinônimos e ainda assim não conseguiria descrever a felicidade que sinto em poder contar com muitas pessoas a quem eu tenho grande apreço e gratidão.

Começando o agradecimento pelo grande Ser responsável pela perfeita sincronia do universo, que me faz feliz a cada dia, e não deixa que eu me sinta só. Aos meus pais, que ajudaram a moldar o meu comportamento, em especial o meu paizinho que se foi nesse período tão turbulento para mim, e que hoje sinto ele ainda mais próximo, do meu lado, todos os dias. Gratidão!

Ao marido e aos filhos, agradeço a compreensão em entender (nem sempre!) que, ao menos por alguns longos anos, eu estava, por vezes, distante, mergulhada em dúvidas, angústias e medos. À minha família, irmãos, tios, primos, até os mais distantes, que ainda assim, tinham sempre uma palavra de incentivo, um abraço em cada encontro.

Agradeço o afeto das amigas Claudia, Cleide e Gabi, sempre compreensivas e disponíveis ao auxílio, proporcionando momentos de alegrias diante de tantos percalços. A todos os amigos e amigas que me fizeram companhia, compreenderam a minha ausência durante todo esse tempo e ainda assim se mantiveram fiéis, com a alegria de sempre.

Agradeço ao meu ex orientador, professor Trazíbulo, por trazer para mim, pela primeira vez, o tema “Redes Complexas”. À minha orientadora, professora Débora, por me aceitar como orientanda, mesmo já estando num período de finalização do trabalho, e com sua tranquilidade característica, ter confiado e encaminhado o meu estudo, sempre dizendo: “vai dar tudo certo!”. A minha querida coorientadora Júlia, que tem sido um “nó” de destaque na minha rede, nas três centralidades; teórico-prático-psicológico, e como amiga tem aguentado as minhas lamentações e desabafos durante esse período de “gestação”.

Aos amigos fofíssimos Soltan e Lívia, sempre atentos e dispostos a ajudar com a maior torcida e vibração que ninguém pode imaginar! Ao meu novo amigo Toni, a quem conheci e com quem quero continuar trabalhando nos próximos anos. A Denise, uma amiga que, literalmente, acompanhou toda a minha pesquisa (risos).

Agradeço à UEFS, que me apoiou no decorrer da pesquisa. Ao DEXA e à Área de Matemática Aplicada, que por vezes entenderam a minha ausência, em especial à professora e ex-coordenadora da área, Olga Claro, que “brigou” pelo meu direito de afastamento, mesmo diante das dificuldades com a falta de professores. Aos demais colegas da UEFS, sempre

disponíveis para debates teóricos. À UFBA, ao DMMDC e a todos os professores e funcionários sempre tão solícitos!

Um agradecimento especial a toda a Banca. Vocês tornaram a minha experiência menos dolorosa e bem mais rica, agindo como verdadeiros colaboradores/disseminadores de conhecimento, me fizeram sentir ainda mais vontade de fazer mais e melhor como professora e como gente!

Por fim, agradeço a todas as experiências e a todas as pessoas que contribuíram de alguma forma para que conseguisse chegar até aqui.

Obrigada! Obrigada! Obrigada!

RESUMO

No Brasil, as universidades têm se destacado na busca pela produção acadêmico-científica, principalmente nas pós-graduações. Tais instituições incentivam as publicações e a participação em projetos de pesquisa pelos principais órgãos de fomento do Brasil e do mundo. Apesar da representatividade das Instituições de Ensino Superior em pesquisa e publicações científicas, o modelo disciplinar ainda é dominante e se propaga desde a construção dos trabalhos de pesquisa até a sua disseminação. Com o objetivo de atender às demandas e complexidade de questões reais que fazem parte da vida de toda sociedade, nossa proposta é modelar a comunidade de pesquisa em redes, descrevendo o funcionamento dessa comunidade, suas relações de cooperação e colaboração científica, identificando elementos multi e interdisciplinares, trazendo à tona atores da rede que podem ser identificados para otimização e enriquecimento da pesquisa. A validação do modelo será realizada por meio de um estudo de caso com a rede de pesquisa formada por professores e pesquisadores da Universidade Estadual de Feira de Santana (UEFS), utilizando dados públicos, coletados dos currículos lattes e dos grupos de pesquisas da UEFS que foram cadastrados no Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq do Ministério da Educação - MEC. Tais dados foram tratados e organizados em planilhas, as quais foram utilizadas para construção das redes com auxílio de softwares para sua visualização e cálculos das métricas, auxiliando na identificação de grandes áreas e atores estratégicos da rede. O modelo desenvolvido contribui para ações efetivas de estímulo à integração multi/interdisciplinar, fortalecimento e enriquecimento da produção acadêmica, sua disseminação, projeção políticas de incentivos, além de facilitar a democratização de informações e compartilhamento de conhecimento.

Palavras-chave: Pesquisa. Colaboração Científica. Disseminação do Conhecimento. Ciência das Redes. Multi/Interdisciplinaridade.

ABSTRACT

In Brazil, universities have stood out in the pursuit of academic and scientific production, especially in postgraduate courses. These institutions encourage publications and participation in research projects by the main funding agencies in Brazil and around the world. Despite the representativeness of Higher Education Institutions in research and scientific publications, the disciplinary model is still dominant and spreads from the construction of research works to their dissemination. In order to meet the demands and complexity of real issues that are part of the life of every society, our proposal is to model the research community in networks, describing how this community works, its relations of cooperation and scientific collaboration, identifying multi and interdisciplinary elements, and bringing to light actors in the network that can be identified to optimize and enrich research. The model will be validated through a case study with the research network formed by professors and researchers from the State University of Feira de Santana (UEFS), using public data collected from the Lattes CVs and research groups of UEFS that were registered with the National Council for Scientific and Technological Development (CNPq) of the Ministry of Education (MEC). These data were processed and organized in spreadsheets, which were used to build the networks with the help of software for visualization and calculation of metrics, helping to identify large areas and strategic actors of the network. The model developed contributes to effective actions to stimulate multi/interdisciplinary integration, strengthening and enriching academic production, its dissemination, projection of incentive policies, in addition to facilitating the democratization of information and sharing of knowledge.

Keywords: Research. Scientific Collaboration. Knowledge Dissemination. Network Science. Multidisciplinarity. Interdisciplinarity.

LISTA DE SIGLAS E ABREVIATURAS

ACCs - Atividades Complementares

ANT - Teoria Ator-Redes

ARS - Análise de Redes Sociais

Capes - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

CEB - Câmara de Educação Básica

CEFET-BA - Centro Federal de Educação Tecnológica da Bahia

CEFET-MG - Centro Federal de Educação Tecnológica de Minas Gerais

CEP/CONEP - Comitê de Ética em Pesquisa/Comissão Nacional de Pesquisa

CEP/ICS - Comitê de Ética em Pesquisa/Instituto de Ciência da Saúde

CG - Componente Gigante

CGr - Centralidade de Grau

CGr(N) - Centralidade de Grau Normalizado

CGRF - Componente Gigante da Rede Filha

CGRM - Componente Gigante da Rede Mãe

CI - Centralidade de Intermediação

CP - Centralidade de Proximidade

CNE - Comissão Nacional de Educação

CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico

CNS - Conselho Nacional de Saúde

CONSEPE - Conselho Superior de Ensino, Pesquisa e Extensão

CPC - Conceito Preliminar do Curso

CSV - Comma - Separed - Values

CT&I - Ciência, Tecnologia e Inovação

DCBIO - Departamento de Ciências Biológicas

DCHF - Departamento de Ciências Humanas e Filosóficas

DCIS - Departamento de Ciências Sociais Aplicadas

DEDU - Departamento de Educação

DEXA - Departamento de Ciências Exatas

DFIS - Departamento de Física

DLA - Departamento de Letras e Artes

DSAU - Departamento de Saúde

DTEC - Departamento de Tecnologia

ENADE - Exame Nacional de Desempenho dos Estudantes

FAPDF - Fundação de Apoio à Pesquisa do Distrito Federal

FINEP - Financiadora de Estudos e Projetos do Ministério da Ciência, Tecnologia e Inovação

Fiocruz - Fundação Oswaldo Cruz

FUFS - Fundação Universidade de Feira de Santana

GT - Grounded Theory - GT

Ibict - Instituto Brasileiro de Informação em Ciência e Tecnologia

IDD - Indicador de Diferenças entre os Desempenhos Observados e Esperados

IES - Instituição de Ensino Superior

IFBA - Instituto Federal de Educação Tecnológica da Bahia

IFbaiano - Instituto Federal Baiano

Inep - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira

KDD - Knowledge Discovery in Databases

LDB - Lei de Diretrizes e Bases da Educação Nacional

MEC - Ministério da Educação

MCTI - Ministério da Ciência, Tecnologia e Inovação

OCTI - Observatório da Ciência, Tecnologia e Inovação

ORCID - Open Researcher and Contributor ID

OS - Open Science

PIBIC - Programa Institucional de Bolsas de Iniciação Científica

PPC - Projeto Pedagógico do Curso

PUC - RJ - Pontifícia Católica do Rio de Janeiro

QS - Quacquarelli Symonds

RA - Rede Aleatória

RF - Rede “Filha”

RM - Rede “Mãe”

SINAES - Sistema Nacional de Avaliação da Educação Superior

SISU - Sistema de Seleção Unificado

TFD - Teoria Fundamentada em Dados

UEBA - Universidades Estaduais da Bahia

UEFS - Universidade Estadual de Feira de Santana

UESC - Universidade Estadual de Santa Cruz

UFABC - Universidade Federal do ABC

UFBA - Universidade Federal da Bahia

UFC - Universidade Federal do Ceará

UFRB - Universidade Federal do Recôncavo Baiano

UFRGS - Universidade Federal do Rio Grande do Sul

UFRJ - Universidade Federal do Rio de Janeiro

UFSB - Universidade Federal do Sul da Bahia

UFSC - universidade Federal de Santa Catarina

UFSCar - universidade Federal de São Carlos

UnB - Universidade de Brasília

UNEB - Universidade do Estado da Bahia

UNIVASF - Universidade Federal do Vale do São Francisco

VOSviewer - Visualizing Scientific Landscapes

WoS - Web of Science

XML - Extensible Markup Language

LISTA DE FIGURAS

Figura 1. As sete pontes de Königsberg (A) e Modelo de Multigrafo da cidade de Königsberg (B)	46
Figura 2. Transformação dos dados em informação /conhecimento	55
Figura 3. Percentual das Grandes Áreas na UEFS	59
Figura 4. Processo de captura e extração de dados dos Currículos Lattes	69
Figura 5. Tela inicial de consulta do Diretórios dos Grupos de Pesquisa no Brasil (DGP)	74
Figura 6. Processo de captura e extração de dados dos Grupos de Pesquisa da UEFS/CNPq	76
Figura 7. Grafo orientado X grafo não orientado	83
Figura 8. Uma rede de computadores representada por um grafo simples não orientado	84
Figura 9. Uma rede de computadores representada por um grafo não orientado com arestas múltiplas e sem laços (Multigrafo)	85
Figura 10. Uma rede de computadores, com laços e arestas múltiplas, representada por grafo não orientado	85
Figura 11. Uma rede de computadores representada por um grafo orientado com arestas múltiplas e sem laços	85
Figura 12. Grafo Ponderado G (A) e Sub grafo Ponderado do Grafo G (B)	86
Figura 13. Grafo não orientado (A) e sua Matriz Adjacência (B)	87
Figura 14. Grafo não orientado e a sua matriz de incidência	88
Figura 15. Quantidade de trabalhos publicados na UEFS por ano	100
Figura 16. Recorte da planilha de coautoria	102
Figura 17. As conexões da Rede “Mãe” (RM) nos intervalos de tempo A, B e C	103
Figura 18. (A) Rede “Mãe” (RM) e (B) Sua Componente Gigante (CGRM)	104
Figura 19. (A) Rede de Coautoria “Filha” (RF); (B) Sua Componente Gigante (CGRF); (C) RF com os rótulos	107
Figura 20. Nós isolados por Departamento da UEFS	109
Figura 21. Gráfico Log-Log da Rede de Coautoria Filha (RF)	111
Figura 22. Representatividade dos professores na UEFS x “10%+” nas três Centralidades	114
Figura 23. Rede com filtro de 10% dos nós com maior grau da rede de coautoria RF	118
Figura 24. Divisão em comunidades com nós (professores) de departamentos distintos	119
Figura 25. Instituições parceiras da UEFS em grupos de pesquisa CNPq	124
Figura 26. Grupos de Pesquisa da UEFS dividido por “Grandes Áreas” do Conhecimento	125
Figura 27. Rede de professores participantes de grupos de pesquisa cadastrados no CNPq	127
Figura 28. Fragmento da Tabela de Palavras referente a Rede Títulos no Software VOSviewer	133
Figura 29. Fragmento do Tesouro (.txt) anexado ao projeto ‘Títulos’ no Software VOSviewer	134
Figura 30. Rede de Títulos do software VOSviewer com o mínimo de 3 ocorrências da palavra	135

Figura 31. Componente gigante da rede de títulos com relevância de 60%	136
Figura 32. Gráfico de Dispersão, Reta de Regressão, Equação Linear e o R^2	138
Figura 33. Nós “tratamento”, “apocynaceae” e “baixo peso ao nascer” da Rede de Títulos	141
Figura 34. Componente Gigante da Rede de Palavras-Chave dos artigos/livros/capítulos de livros dos professores/pesquisadores da UEFS com relevância de 60%	144
Figura 35. Gráfico de Dispersão. Reta de regressão, Equação Linear e R^2	148
Figura 36. Conexões nós “saude” e “ensino”	148
Figura 37. Rede dos Nomes dos Grupos de Pesquisa da UEFS cadastrados no CNPq	152
Figura 38. Rede das Linhas de Pesquisa dos Grupos de Pesquisa da UEFS cadastrados no CNPq	153
Figura 39. Modelo de análise de Redes de Pesquisa e Colaboração da UEFS	157
Figura C1. Tela do Gephi versão 0.10.1	206
Figura D1. Janela principal do VOSviewer	207
Figura D2. “Density Visualizacion”	208

LISTA DE QUADROS

Quadro 1. Questões e objetivos específicos	30
Quadro 2. Seleção de artigos e capítulos da LDB nº 9.394/96 relacionados à Educação Superior e características buscadas pelo modelo proposto	38
Quadro 3. Conceito da Análise de Conteúdo	51
Quadro 4. Áreas e Grandes Áreas do Conhecimento	58
Quadro 5. Seleção, coleta e extração dos dados	63
Quadro 6. Planilhas geradas via script	71
Quadro 7. Informações sobre os códigos das planilhas dos Currículos Lattes	72
Quadro 8. Planilha dos Grupos de Pesquisa UEFS/CNPq	77
Quadro 9. Medidas Estatísticas para grafos não orientados (Visão Geral da Rede)	89
Quadro 10. Medidas Estatísticas para grafos não orientados (Detecção de Comunidades)	92
Quadro 11. Medidas Estatísticas para grafos não orientados (Visão Geral dos nós)	93
Quadro 12. Medidas Estatísticas para grafos não orientados (Visão Geral das Arestas)	95
Quadro 13. Métricas utilizadas para análise das redes	97
Quadro 14. Representação de cada departamento da UEFS por cores	120
Quadro 15. Propostas/pontos a serem avaliados na rede de professores dos grupos de pesquisa da UEFS/CNPq de autoria filha (RF)	122
Quadro 16. Propostas/pontos a serem avaliados da rede de professores dos grupos de pesquisa da UEFS/CNPq	130
Quadro 17. Propostas/pontos a serem avaliados da rede de Títulos e Palavras-chave	150

LISTA DE TABELAS

Tabela 1. Estatísticas da Rede Mãe (RM) e sua Componente Gigante (CGRM)	104
Tabela 2. Estatísticas da Rede Filha (RF) e sua Componente Gigante (CGRF)	108
Tabela 3. Métricas da Rede Aleatória (RA)	110
Tabela 4. Os nós 10% mais nas Centralidade de Grau (CGr), Centralidade de Proximidade (CP) e Centralidade de Intermediação (CI)	112
Tabela 5. Estatísticas do Gephi para a Rede de Títulos	136
Tabela 6. Estatísticas Rede Aleatória (RA)	137
Tabela 7. As “20+” dos Títulos nas três Centralidades	139
Tabela 8. As “20+” Palavras-Chave nas três Centralidades	145
Tabela 9. Estatísticas do Componente Gigante da Rede de Palavras-chave dos artigos/livros/capítulos de livro	146
Tabela 10. Estatísticas da Rede Aleatória (RA)	147

SUMÁRIO

1. INTRODUÇÃO	21
1.1 TEMA, PROBLEMA E JUSTIFICATIVA	27
1.2 OBJETIVOS	29
1.2.1 Objetivo Geral	29
1.2.2 Objetivos Específicos	29
2. FUNDAMENTAÇÃO TEÓRICA	31
2.1 PROCESSO TEÓRICO-EPISTEMOLÓGICO	31
2.2 AS INSTITUIÇÕES DE ENSINO SUPERIOR E A DISSEMINAÇÃO DO CONHECIMENTO	32
2.3 MULTIDISCIPLINARIDADE E INTERDISCIPLINARIDADE	33
2.3.1 Colaboração e Cooperação	43
2.4 CIÊNCIA DAS REDES	44
3. METODOLOGIA	49
3.1 TEORIA FUNDAMENTADA DE DADOS (TFD)	49
3.2 ANÁLISE DE CONTEÚDO	50
3.3 MÉTODOS QUANTITATIVOS	53
3.3.1 Análise de Redes Sociais e Complexas (ARS)	53
3.4 PERCURSO INVESTIGATIVO	54
3.4.1 KDD (Knowledge Discovery in Databases) e ETL (Extract Transform Load)	54
3.4.2 Dados: Escolha da amostra para o desenvolvimento do modelo: Universidade Estadual de Feira de Santana (UEFS)	56
3.4.3 Seleção, Coleta e Extração dos dados	61
3.4.3.1 Aspectos éticos da seleção dos dados	64
3.4.3.2 Currículo Lattes	67
3.4.3.3 Grupos de Pesquisa UEFS/CNPq	73
3.4.4 Reprocessamento e Transformação dos dados	78
3.4.4.1 Currículo Lattes	78
3.4.4.2 Grupos de Pesquisa UEFS/CNPq	80
3.4.5 Mineração e Interpretação dos Dados	81
3.4.5.1 Teoria dos Grafos como Metodologia de Análise	82
3.4.5.2 Métricas Utilizadas para Análise das Redes Propostas (Grafos não Orientados e não Ponderados)	88
3.4.5.3 Análise Topológica da Rede	96
3.4.5.4 Análise das Centralidades da Rede	97
3.4.6 Validação e Conhecimento	98
4. ESTUDO DAS REDES DA UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA	99

4.1 REDES SOCIAIS QUE FAZEM PARTE DO MODELO	101
4.1.1 Rede de Coautoria Mãe (RM)	101
4.1.2 Rede de Coautoria Filha (RF)	105
4.1.3 A Rede Social Formada por Professores dos Grupos de Pesquisa da UEFS Cadastrados no CNPq	124
4. 2 REDES SEMÂNTICAS	130
4.2.1 Redes de Títulos dos Artigo/Livros/Capítulos de Livros	131
4.2.2 Redes de Palavras-Chave dos Artigos/Livros/Capítulos de Livros	142
4.2.3 Redes Semânticas dos Grupos de Pesquisa	150
4.2.3.1 Redes dos Nomes dos Grupos de Pesquisa da UEFS cadastrados no CNPq	151
4.2.4 Rede formada pelas Linhas de Pesquisa dos Grupos de Pesquisa cadastrados pelo CNPq	153
5. MODELO DE ANÁLISE DE REDES DE PESQUISA E COLABORAÇÃO CIENTÍFICA	155
6. CONTRIBUIÇÕES E PESQUISAS FUTURAS	160
7. LIMITAÇÕES DO ESTUDO	162
8. CONSIDERAÇÕES FINAIS	163
REFERÊNCIAS	165
APÊNDICE A – CURSOS OFERECIDOS PELA UEFS/GRANDE ÁREA A QUE PERTENCE	175
APÊNDICE B – SCRIPT 01	178
APÊNDICE C – SOFTWARE GEPHI	206
APÊNDICE D – SOFTWARE VOSVIEWER	207
APÊNDICE E – SCRIPT 02	209

1. INTRODUÇÃO

A pesquisa científica é um dos pilares das instituições de ensino superior (IES) e uma ferramenta importante para o desenvolvimento da sociedade e disseminação do conhecimento. Propomos, neste trabalho, um modelo de análise da colaboração da pesquisa científica a ser utilizado num ambiente de ensino superior, que será aplicado na UEFS (Universidade Estadual de Feira de Santana) para mediação de uma proposta multi e interdisciplinar que traga para as diferentes áreas da instituição, maior conexão, novas discussões, ideias e, principalmente, um conhecimento mais amplo dos objetos de investigação e pesquisa, consequentemente, enriquecimento dessa produção. Buscamos entender a interação social de pesquisadores tomando por base os princípios de observação, experimentação, medição, análise crítica e revisão para análise de pesquisa.

Segundo Silva (1987, p. 194) “Nenhum conhecimento precede a experiência e todo conhecimento começa com ela”. O conhecimento científico é fundamental para avançar nossa compreensão do mundo e para resolver problemas complexos em diversas áreas, dentro de uma abordagem rigorosa e confiável para adquirir conhecimento. Neste caso, sobre as relações sociais que acontecem no ambiente acadêmico, experimentando suas nuances e o impacto direto na comunidade ao redor da instituição.

Entendemos que as universidades podem e devem cumprir sua função social, sendo fonte de conhecimento para outras instituições de ensino, promovendo a inclusão, firmando laços com a comunidade, sendo parte desta comunidade e junto com ela ir se construindo, se modificando, se adequando às necessidades da sociedade (Gimenez; Bonacelli, 2013). Por tudo isso, acreditamos que as universidades podem oferecer o embasamento necessário para construção de propostas mais interdisciplinares, cooperativas e colaborativas, propondo mudanças curriculares no ensino no Brasil, e mais identificadas com as necessidades da sociedade.

É sabido que a Universidade tem sido chamada a promover a inclusão social, a dialogar com o setor produtivo, a reformular as suas grades curriculares, a criar novos cursos, a estreitar mais os laços com a sociedade em geral e a participar mais ativamente do desenvolvimento econômico. Percebe-se, portanto, que a universidade contemporânea tem se deparado com diferentes demandas e isso a tem forçando a repensar e a redefinir o seu papel, em

busca de um “modelo” que dê conta das suas missões, bem como de expectativas da sociedade (Gimenez; Bonacelli, 2013, p. 10).

As universidades desempenham um papel fundamental na construção e disseminação do conhecimento em nossa sociedade, pois são responsáveis por várias funções relacionadas à geração, disseminação e difusão do conhecimento. Através das publicações oriundas destas instituições, elas disseminam o conhecimento gerado por meio da interação entre todo público que tenha interesse no assunto, transcendendo os muros institucionais, transformando e construindo novos conhecimentos. Muitas ações voltadas para construção de conhecimento são promovidas nas universidades, algumas delas são: pesquisa, inovação, transferência de tecnologia, incentivo ao empreendedorismo, preservação e conservação do conhecimento (através de suas bibliotecas e livrarias), formação de recursos humanos qualificados e críticos, promoção da discussão e debate de ideias, incentivo a criatividade e a inovação, entre outras. Todos estes pontos tornam as universidades instituições cruciais para a construção, preservação e disseminação do conhecimento em nossa sociedade, cumprindo a sua missão não só de oferecer ensino, pesquisa e extensão, mas também de formar cidadãos para atuar nas demandas sociais. As Instituições de Ensino Superior são atores principais no avanço da ciência, cultura e desenvolvimento econômico, além de preparar indivíduos para enfrentar os desafios do mundo contemporâneo, elas têm a autonomia para promover mudanças na forma de gerar conhecimento, não apenas no ensino superior, mas em todo sistema educacional.

A relevância das universidades no meio social é notável quando seus egressos levam à sociedade seus aprendizados, pois perpassam um estilo ideológico, embutido em suas ações. Aprenderam quando transpuseram os níveis acadêmicos, superaram os desafios durante a aquisição de conhecimentos que, gradualmente, foram propostos, durante o período de formação, até se tornarem profissionais (Gomes, 2014, p. 4).

Em relação ao campo de pesquisa nas universidades, Ortega (2020) reforça que, nos dias de hoje, o espaço social legitimado no campo educacional para a produção de pesquisa em Educação é o da pós-graduação. O fato se deve a formação desses professores, geralmente mestres e doutores, e ao currículo mais flexível para diferentes formas de avaliação.

No cenário da produção científica, partimos para a investigação sobre como o Brasil aparece no cenário mundial. O 3º (terceiro) Boletim Anual do Observatório de Ciência, Tecnologia e Inovação (OCTI) de 2022 divulga informações sobre Ciência, Tecnologia e

Inovação (CT&I) no mundo e descreve a situação do Brasil em relação aos demais países, levando em conta a produção científica nacional referente aos anos de 2019 a 2022, identificando o Brasil dentre os 15 países com maior volume de artigos. A análise foi feita em relação à base de indexação Web of Science (WoS) (Boletim OCTI 2022) e demonstra o potencial que tem a pesquisa no Brasil, apesar das dificuldades em investimentos que aumentaram ainda mais nos últimos anos.

Com o intuito de alavancar a pesquisa no Brasil, o CNPq tem trabalhado para acompanhar o mundo no sentido de divulgar e democratizar as informações, para que sejam cada vez mais acessíveis a um número maior de pessoas. A proposta “Ciência Aberta” ou “Open Science” (OS) tem o objetivo de tornar o conhecimento científico mais transparente e aberto. Esse movimento é mundial e começou a ser discutido em 1991. Para Albagli (2014, p.2) “a ciência aberta promove o aumento dos estoques de conhecimento público, propiciando não apenas a ampliação dos índices gerais de produtividade científica e de inovação, como também das taxas de retornos sociais”. Albagli ainda ressalta a importância do compartilhamento, e afirma que é na coletividade que se promove a criatividade e inovação, o que não tem como acontecer de forma individual.

Nessa onda de “Ciência Aberta”, o CNPq, juntamente com o Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) anunciaram em junho de 2022, o repositório de dados denominado ‘Lattes Data’, que disponibiliza para o público, os grupos de pesquisa que recebem ou receberam fomento do CNPq. Outra novidade criada pelo Ibict e implementada em 2020, é o BrCris¹. Dias *et al.* (2022) ratificam a importância de um sistema como o BrCris, destacando que as informações são organizadas de forma semântica, permitindo que sejam geradas visualizações e análises cientométricas (que avaliam a produção científica e o impacto da pesquisa) de conexões relacionadas à produção científica no Brasil.

O projeto BrCris reúne dados de várias plataformas de pesquisa e oferece o acesso, análise e até visualização de dados de pesquisa, pesquisadores e instituições. O sistema visa proporcionar maior transparência e visibilidade para a produção científica e tecnológica do país, facilitando a busca e o acesso a informações sobre pesquisadores e suas contribuições, bem como o acompanhamento de atividades de pesquisa em instituições de ensino superior e

¹ O BrCris (acrônimo para Current Research Information System), foi desenvolvido pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT). Trata-se de uma plataforma computacional para integração, visualização e prospecção de dados científicos com a finalidade de estabelecer um modelo único de organização da informação científica de todo o ecossistema da pesquisa brasileira. Disponível em: <https://brcris.ibict.br/pt-BR/about>.

de pesquisa. No decorrer deste trabalho, nos sentimos parte deste processo e, à medida que nos deparamos com estas novidades, fomos acrescentando informações destes novos recursos à pesquisa, no intuito de contribuir para a colaboração e disseminação do conhecimento, auxiliando a disponibilizar ciência de forma mais fácil e democrática.

Relatos sobre os problemas que passam as Instituições Públicas de Ensino no Brasil (municipais, estaduais e/ou federais) são frequentes. Em sua grande maioria, faltam recursos humanos, infraestrutura, verbas para pesquisa e extensão, entre outras dificuldades. Como docente de uma instituição estadual de ensino superior da Bahia, ressaltamos o esforço diário dos profissionais envolvidos, dos funcionários, professores e gestores, em tornar a universidade cada vez maior e melhor para a comunidade acadêmica e a população ao seu redor.

No Brasil, a educação deve ser compromisso de todos nós como sociedade, mas são os órgãos governamentais os principais responsáveis; os governos, federal, estadual e municipal, além dos professores/pesquisadores, da sociedade civil e dos conselhos de educação. O trabalho em conjunto dos diferentes níveis de governo e a sociedade é essencial para enfrentar os desafios e melhorar o sistema educacional de um país com tanta desigualdade social e má distribuição de recursos. No campo da pesquisa, tema central deste trabalho, o compromisso é compartilhado entre diversas instituições e atores, tendo como principais responsáveis, além da sociedade civil, o setor privado, professores e pesquisadores. Destacamos o papel do Ministério da Ciência, Tecnologia e Inovação - MCTI formulando e executando políticas públicas relacionadas à pesquisa, ciência, tecnologia e inovação. As agências de fomento à pesquisa, como o Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq e a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES, são fundamentais no apoio e fomento à pesquisa, ciência, tecnologia e inovação no Brasil. Além das universidades e dos institutos de pesquisa, que mantêm laboratórios, grupos de pesquisa e equipes de pesquisadores que contribuem significativamente para a produção de conhecimento.

A pesquisa é fundamental para o avanço científico e tecnológico, e para o desenvolvimento econômico, porém, ressaltamos a importância da cooperação e coordenação entre esses diferentes atores para promover uma pesquisa de alta qualidade e com impacto social significativo. A multi e interdisciplinaridade são conceitos muito buscados em todas as esferas da educação, apesar das dificuldades que ainda são encontradas para implementação

em sala de aula. Para Pombo (2006), o conceito ainda não é totalmente compreendido. Ela enfatiza três conceitos que têm por trás o mesmo intuito; pluridisciplinaridade (ou multidisciplinaridade), interdisciplinaridade e transdisciplinaridade. Entendemos a multidisciplinaridade como várias disciplinas que não precisam de conexão, e é isso que fazemos, muitas vezes, na prática de ensino no Brasil. Para chegar à transdisciplinaridade, o segundo passo é a interdisciplinaridade, que propõe a conexão de, ao menos, duas disciplinas em torno de um mesmo objeto de estudo. Nas nossas instituições, ainda não conseguimos a melhor forma de aplicar este segundo passo, e as discussões sobre como devemos fazer continuam acontecendo. Se nas universidades ainda podemos observar dificuldades em utilizar a interdisciplinaridade, imagine o terceiro passo proposto por Pombo, a transdisciplinaridade, ainda distante na prática das universidades, pois é uma abordagem mais complexa, que busca integrar diferentes disciplinas de forma a transcender os limites tradicionais de cada uma delas. Ela não apenas combina conhecimentos de várias áreas, como ocorre na interdisciplinaridade, mas também busca criar um entendimento unificado e holístico de um problema ou fenômeno. Apesar disso, vemos iniciativas para a discussão do tema, e alguns grupos de pesquisa da universidade escolhida para o nosso estudo de caso, propõem temas multi/inter e transdisciplinares.

Para entender o processo de disseminação produzido na universidade, buscamos fazer uma análise das redes de colaboração e pesquisa para que possamos ter uma visão da cooperação, colaboração e troca de conhecimento existente na instituição. Observamos comportamento dessas redes, suas relações, suas características, identificando as possibilidades de otimizá-las, propondo novas conexões (relações entre os professores), para obtermos redes mais conectadas, colaborativas e mais interdisciplinares, contribuindo não só para a otimização da colaboração, mas também para melhora na qualidade da pesquisa e sua disseminação, através de um trabalho mais multi e interdisciplinar.

Como fonte de dados, utilizamos dados de pesquisa do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), incluindo a Plataforma Lattes, que é a maior base de dados sobre ciência no Brasil, e reúne informações sobre pesquisadores, professores e cientistas no Brasil. A Plataforma Lattes foi desenvolvida pelo CNPq e lançada em 1999 (Brasil, 2023b). É uma ferramenta importante para a gestão da pesquisa e da carreira acadêmica no país. Permite que os pesquisadores registrem e atualizem informações sobre suas qualificações, produção acadêmica, projetos de pesquisa, orientação de estudantes e outras atividades relacionadas à pesquisa e ensino. Essas informações podem ser utilizadas

para avaliar a produtividade e a qualidade da pesquisa acadêmica no país, para a distribuição de bolsas e financiamentos de pesquisa, assim como para a colaboração acadêmica nacional e internacional, promovendo transparência e visibilidade da pesquisa no Brasil e tem sido fundamental na avaliação e no acompanhamento da produção acadêmica, por isso, desempenha um papel central na gestão da pesquisa e na construção de perfis acadêmicos de pesquisadores brasileiros. Os dados de grupos de pesquisa foram obtidos pelo Diretório dos Grupos de Pesquisa no Brasil (Lattes/CNPq), esses dados também estão disponibilizados de forma livre para qualquer usuário.

Para ratificar o nosso modelo e validá-lo, utilizamos dados da Universidade Estadual de Feira de Santana - UEFS. Ela é uma das quatro Universidades Estaduais da Bahia - UEBA mantida pelo governo do Estado da Bahia. Além da UEFS, temos a Universidade do Estado da Bahia - UNEB, a Universidade Estadual de Santa Cruz - UESC e a Universidade Estadual do Sudoeste da Bahia - UESB. Além das universidades estaduais, a Bahia possui as seguintes universidades federais: Universidade Federal da Bahia - UFBA, Universidade Federal do Recôncavo Baiano - UFRB, Universidade Federal do Vale do São Francisco - UNIVASF e a Universidade Federal do Oeste da Bahia - UFOB. Temos ainda dois Institutos Federais: o Instituto Federal de Educação, Ciência e Tecnologia da Bahia - IFBA, antigo Centro Federal de Educação Tecnológica da Bahia - CEFET/BA, e o IF Baiano (que são antigas Escolas Agrotécnicas Federais e Escolas Médias de Agropecuária Regionais).

A UEFS foi eleita a melhor instituição pública estadual de ensino superior da Bahia, no ranking 2024 das universidades da América Latina e Caribe. O ranking é elaborado pela Quacquarelli Symonds (QS), uma empresa britânica especializada na análise de instituições de ensino superior em todo o mundo, que utiliza cinco critérios na avaliação: impacto e produtividade da pesquisa; compromisso docente; empregabilidade; impacto online; e internacionalização (UEFS, 2023).

Com a escolha da universidade, seguimos para a captura dos dados. Através da Plataforma Lattes, utilizamos os currículos lattes dos professores da UEFS e a Teoria das Redes para traçar as seguintes redes para estudo: a Rede de Coautoria, dos autores que têm publicações em comum de artigos em periódicos, de livros ou capítulos de livros. A Rede de Títulos, com os títulos destas publicações e a Rede de Palavras-Chaves, com as palavras-chave das publicações dos artigos/capítulos/livros, buscando identificar as principais propriedades que emergem destas redes. Todo o processo de captura dos dados dos currículos foi feito por meio de mineração - extração (Script 01, do APÊNDICE B) é devidamente

descrita na metodologia, bem como toda organização das redes propostas. A segunda fonte de dados utilizada será a dos pesquisadores dos grupos de pesquisas, também via Plataforma Lattes, sendo que os dados foram obtidos diretamente do site do CNPq, pelos registros dos grupos que são cadastrados e recebem algum financiamento ou apoio do referido órgão. Foram selecionados os grupos de pesquisa e pesquisadores com vínculo com a UEFS e cadastrados no CNPq. Encontramos as redes dos nomes destes grupos de pesquisa, as linhas de pesquisa e os professores envolvidos no processo de pesquisa.

1.1 TEMA, PROBLEMA E JUSTIFICATIVA

A escolha do tema se dá devido à sua importância no contexto das universidades e sua viabilidade para o ambiente em que se discute pesquisa científica (Severino, 2014). A temática que será desenvolvida na nossa pesquisa é o estudo da colaboração científica através de redes de colaboração, cooperação e de pesquisas na UEFS. Com o auxílio da Teoria das Redes, buscamos identificar a colaboração/cooperação científica e sua disseminação na UEFS, para avaliação de propostas mais multi e interdisciplinares que possam colaborar com a pesquisa científica. Para isso, a nossa pesquisa irá explorar, analisar, questionar e descrever características e padrões da pesquisa científica na UEFS. Dentro da temática proposta, identificamos algumas questões norteadoras para a pesquisa: Como se articulam os elementos estruturais da rede de professores/pesquisadores da UEFS em termos de colaboração e produção científica? Como acontecem as colaborações nas pesquisas? Emergem características de disseminação/difusão nas redes de pesquisa? Emergem características de multi e interdisciplinaridade dessas relações? Diante destes questionamentos, buscamos identificar o que podemos melhorar nessas conexões, quais pontos poderiam ser modificados na proposta de redes, para torná-las mais interdisciplinares e colaborativas e propícias à disseminação do conhecimento.

O problema da nossa pesquisa surge como uma dificuldade apresentada pelas universidades (Gil, 2008). Observamos, com um olhar muito próximo da UEFS, que a complexidade das questões reais encontradas na vida cotidiana é impossível de ser compreendida pela abordagem disciplinar, ainda tão utilizada nas universidades. Essa postura fragmentada diminui o potencial de colaboração multi e interdisciplinar nestas instituições e

implica na perda de oportunidades de pesquisa, captação de recursos, de um maior e melhor desenvolvimento da produção científica e da disseminação do conhecimento construído na Instituição.

É preciso saber ligar os elementos das diferentes disciplinas. Eu dou mais um exemplo, falo da identidade humana. Nós, como seres humanos, somos indivíduos, membros de uma sociedade, e fazemos parte de uma espécie biológica, a espécie humana. Esses três aspectos são absolutamente inseparáveis. Mas como é que eles são ligados? (Morin, 2003, p. 23).

Teixeira (2007) considera que a introdução da interdisciplinaridade e da transdisciplinaridade nas universidades é uma questão urgente. Para ele, a interdisciplinaridade é uma necessidade epistemológica e política para que se possa organizar o conhecimento e institucionalizar a ciência. Teixeira ainda enfatiza que a Ciência e a tecnologia da atualidade trazem um modelo circular e de auto-organização, com possibilidades que as estruturas pedagógicas tradicionais lineares não têm mais como acompanhar.

A universidade atual, fortemente voltada para as necessidades da sociedade, sucumbe às exigências do mercado e aos seus interesses passageiros. Nessa situação, a fragmentação excessiva da ciência é uma das causas da desagregação da universidade. Esse mesmo ciclo inflacionário faz com que as disciplinas clássicas, aquelas que estão na base dos conhecimentos teóricos, sejam muitas vezes postas em segundo plano, cedendo lugar aos conhecimentos fugazes. Embora sem pretender ser exaustivo e definitivo (Teixeira, 2007, p. 60).

Como justificativa do nosso trabalho, identificamos a importância da pesquisa, as razões (fundamentação teórica) e os motivos (práticos) para realizá-la (Gil, 2008; Severino, 2014; Marconi & Lakatos, 2017). A Teoria das Redes nos oferece o arcabouço para o entendimento das relações entre os pesquisadores ou grupos em diferentes áreas do conhecimento. A Análise das Redes Sociais (ARS) (Recuero, 2017) será utilizada para mapear a construção do conhecimento nas redes, ajudando a entender como esse conhecimento é compartilhado, distribuído e produzido, por meio do estudo das estruturas das redes por métricas que descrevem a importância e a posição dos nós nas redes (Guedes, 2010). A multi/interdisciplinaridade (Pombo, 2006) fornecerá elementos para análise das conexões entre os pesquisadores, a colaboração/cooperação entre áreas distintas, levando em

conta que um trabalho mais colaborativo deve resultar em maior capacidade criativa, fruto das trocas entre os pares.

Entre os motivos práticos destacamos a análise e incentivo à colaboração/cooperação e à interdisciplinaridade num ambiente de ensino que pode proporcionar parcerias e inovação nos três pilares da universidade: ensino, pesquisa e extensão. Uma maior colaboração multi/interdisciplinar incentiva a produção científica com resultados mais robustos e relevantes, além de promover o capital social institucional. A identificação de possíveis obstáculos à disseminação do conhecimento e relação entre as diferentes áreas podem levar a discussões sobre políticas de compartilhamento de informações e de incentivo às produções multi/interdisciplinares. A pesquisa poderá ainda trazer para a instituição, informações para tomadas de decisão direcionadas para políticas de pesquisa e inovação, apoio a projetos interdisciplinares, além de possibilitar financiamentos e parcerias com outras instituições.

1.2 OBJETIVOS

1.2.1 Objetivo Geral

O objetivo principal desta pesquisa é desenvolver um modelo de análise de cooperação e colaboração multi e interdisciplinar entre professores utilizando redes de pesquisa (redes de colaboração científica e redes semânticas) para compreender as dinâmicas de colaboração e identificar padrões, estruturas e relações que possam influenciar a produção científica e o fortalecimento da rede multi/interdisciplinar e de disseminação de conhecimento na instituição.

1.2.2 Objetivos Específicos

Em busca do objetivo geral, modelamos os objetivos específicos em resposta às questões da nossa temática de pesquisa (Gil, 2008) (Quadro 1).

Quadro 1. Questões e objetivos específicos

QUESTÕES	OBJETIVO ESPECÍFICO (O QUE PRETENDEMOS FAZER PARA RESPONDER?)
<p>1. Como se articulam os elementos estruturais da rede de professores da UEFS em termos de colaboração e produção científica?</p> <p>2. Como acontecem as colaborações nas publicações e grupos de pesquisa?</p>	<p>I. Mapear a rede de professores da UEFS para compreender o processo de colaboração entre estes professores.</p> <p>II. Propor possibilidades de aumentar a colaboração, diante da posição dos professores nas redes e/ou grupos de pesquisas.</p>
<p>3. Emergem características de disseminação/difusão nas redes de pesquisa?</p>	<p>III. Compreender como os professores e suas temáticas se posicionam e se conectam nas redes de pesquisas da UEFS.</p>
<p>4. Emergem características de multi e interdisciplinaridade dessas relações?</p>	<p>IV. Identificar os atores das comunidades que colaboram na rede, e seus departamentos/grandes áreas/áreas da UEFS.</p> <p>V. Identificar temáticas de destaque na rede e/ou que são comuns a pesquisadores de grupos distintos.</p>

Fonte: Elaboração própria (2021).

2. FUNDAMENTAÇÃO TEÓRICA

2.1 PROCESSO TEÓRICO-EPISTEMOLÓGICO

Para modelar todo processo filosófico que permita a construção do nosso trabalho de forma mais fiel possível aos nossos questionamentos e objetivos, buscamos por teóricos que abordem as relações dinâmicas entre conhecimento, sociedade e redes. Este estudo, portanto, alia a sociedade em rede, a ciência das redes e a interdisciplinaridade permitindo a compreensão das interações, semânticas e a disseminação e organização do conhecimento em ambientes complexos.

Adotamos o referencial de Castells (2004; 2006) para ampliar a visão de Sociedade em Rede, em que as interações se estruturam dinamicamente por meio de redes digitais e sociais. As universidades, como agentes centrais nessa configuração, desempenham papel crucial na produção e disseminação do conhecimento, atuando como pontos de conexão para diferentes áreas de pesquisa. Nesse contexto, as interações acadêmicas são analisadas como parte de um sistema interconectado e complexo.

No campo da Ciência das Redes, os autores Barabási & Albert (1999), Barabási (2002), Newman (2001) e Watts & Strogatz (1998) oferecem os instrumentos necessários para melhor compreensão das redes complexas e o entendimento das relações entre os elementos destes sistemas, no caso da nossa pesquisa, as redes sociais e semânticas. Para análise das propriedades das redes, formalização de métricas de rede, centralidades e estruturas topológicas, utilizaremos Freeman (1978), Newman & Girvan, (2004) e Watts & Strogatz (1998). Em relação aos métodos quantitativos e aplicações da análise de redes sociais, nossas referências são Freeman (1979); Newman (2001), Wasserman & Faust (1994) que oferecem uma visão detalhada do desenvolvimento e aplicação dessas metodologias no estudo das interações das redes complexas e seus atores.

Os conceitos de multidisciplinaridade e interdisciplinaridade são essenciais para abordar os desafios complexos enfrentados pelas universidades, Japiassu (1976), Fazenda (1994) e Pombo (2004; 2006) fornecem a base para compreender as relações multi e interdisciplinares de colaboração entre os atores da rede ao investigar a percepção e organização do conhecimento que emergem das dinâmicas do aprendizado coletivo.

Dessa forma, este modelo teórico oferece uma abordagem integrada para compreender

como as interações humanas e temáticas se organizam e disseminam o saber, conectando diferentes áreas de pesquisa em um esforço colaborativo para produzir conhecimento.

2.2 AS INSTITUIÇÕES DE ENSINO SUPERIOR E A DISSEMINAÇÃO DO CONHECIMENTO

(...) a educação é um processo de humanização, que ocorre na sociedade humana com a finalidade explícita de tornar os indivíduos em participantes do processo civilizatório e responsáveis por levá-lo adiante. Enquanto prática social, é realizado por todas as instituições da sociedade. Enquanto processo sistemático e intencional, ocorre em algumas, dentre as quais se destaca a escola. A educação escolar, por sua vez, está assentada fundamentalmente no trabalho dos professores e dos alunos. A finalidade desse trabalho - de caráter coletivo e interdisciplinar e que tem como objetivo o conhecimento - é contribuir como processo de humanização de ambos, numa perspectiva de inserção social crítica e transformadora (Pimenta & Anastasiou, 2005, p. 80-81).

As universidades têm o compromisso de atender estudantes no ensino, pesquisa e extensão no Brasil, e ocupam um papel importante na construção do conhecimento e na disseminação desse conhecimento. Mas a missão dessas instituições vai muito além, uma vez que tem compromisso direto com a formação de cidadãos, garantindo que os estudantes saiam preparados para atender não apenas ao mercado, mas às questões do mundo, que são, a cada dia, mais complexas. (Pimenta & Anastasiou, 2005).

A posição estratégica da universidade em relação ao acesso a todos os outros graus de ensino no Brasil e a sua proximidade, não só da comunidade acadêmica, mas de toda comunidade ao seu redor, aumenta a responsabilidade dessas instituições. O ensino e a pesquisa que acontece neste ambiente é um instrumento para, junto com os colegas pesquisadores e alunos, amadurecer e ampliar o conhecimento, resultando não só na divulgação de informações, mas também no cumprimento dos objetivos das ações planejadas, tornando essas informações úteis (Bortoliero, 2009). Para que isso ocorra, é imprescindível que possamos fazer uso, da melhor forma, da disseminação e difusão dessas informações.

A disseminação e a difusão do conhecimento são dois processos relacionados, porém, são distintos. Envolvem a propagação de informações, ideias, inovações ou práticas; a disseminação está relacionada à divulgação e disponibilidade de informações e conhecimento a um grupo específico, ou de forma mais ampla por diferentes meios de propagação. Porém

disseminar uma informação não implica necessariamente que a informação seja feita de forma significativa. Na difusão do conhecimento o foco é a propagação desta informação de um grupo para outro (grupos distintos), uma forma de disseminar estas informações de maneira que cheguem a ser utilizadas por outros grupos, ou seja, vai além da necessidade da divulgação, há neste processo, a incorporação de ideias. Ambos os processos são importantes para mudanças efetivas em todo e qualquer processo de construção do conhecimento. Nas universidades, a distribuição e compartilhamento acontecem através dos grupos de pesquisa e extensão e das publicações realizadas pela universidade e disponibilizadas para toda a comunidade. Em contextos interdisciplinares e multidisciplinares, a disseminação de conhecimento é um processo essencial para facilitar a comunicação e a colaboração entre campos de estudo distintos (Tuckman, 1987).

A sociedade necessita de conhecimento! Laufer (2008) ressalta que a educação, a formação e a capacidade de cada pessoa serão determinantes para compreensão e manipulação da informação. As universidades podem oferecer uma gama de oportunidades para construção, disseminação e difusão do conhecimento e a sociedade tem cobrado das instituições o atendimento de suas demandas (Gomes, 2014). Precisamos, como instituição de ensino superior, potencializar a produção científica e compartilhar os saberes.

Nesse sentido, para aprofundar a percepção de como funciona a pesquisa científica nas universidades, analisamos as redes de pesquisa da Universidade Estadual de Feira de Santana (UEFS), buscando fortalecer a colaboração científica e trazer resultados mais positivos para a disseminação do conhecimento.

2.3 MULTIDISCIPLINARIDADE E INTERDISCIPLINARIDADE

A construção do conhecimento é um processo complexo e multifacetado que envolve a aquisição, organização, interpretação e aplicação de informações e experiências. Existem várias teorias e abordagens para analisar como o conhecimento é construído. A abordagem multi e interdisciplinar vai entrar no processo da construção do conhecimento para o entendimento dos diferentes campos que contribuem para essa construção. Para Werneck (2006), no século XX, o papel da universidade não é somente o ensino, mas também a pesquisa como produção de conhecimento e aplicação desse conhecimento a serviço da sociedade.

O prefixo “inter” significa posição ou ação intermediária, o sufixo “dade” atribui o sentido de ação ou resultado de ação ao termo, já “disciplina”, núcleo do termo estudado, significa a epistemé, podendo caracterizar como ordem que convém ao funcionamento de uma organização, ou ainda regime de ordem imposta ou livremente consentida (Favaro & Araújo, 2004, p. 105).

Os conceitos ‘construção do conhecimento’ e ‘multi e interdisciplinaridade’ estão ligados e frequentemente se complementam. A interdisciplinaridade envolve o conceito e a integração de conhecimentos de diferentes disciplinas para abordar problemas complexos e questões multifacetadas, isso significa que os estudantes e pesquisadores precisam construir um entendimento multi e interdisciplinar, unindo conceitos e métodos de diversas áreas do conhecimento para compreender e resolver desafios complexos.

A multidisciplinaridade é trabalhada em cada disciplina ou área específica, e cada uma delas busca uma análise específica do objeto. Segundo Ferreira de Silva (2018, p. 17), o conceito final formado sobre o objeto “independe das relações existentes entre os campos que realizam as observações, pois, para este conceito o que está em prioridade são as diferentes análises realizadas neste processo sobre o mesmo objeto”. Assim, a multidisciplinaridade pode ser entendida como a justaposição de diferentes disciplinas que atuam em torno de um mesmo problema, sem que haja necessariamente uma integração entre seus métodos e conceitos. De acordo com Japiassu (1976), trata-se de uma abordagem em que as disciplinas mantêm sua autonomia, cooperando de forma paralela, cada uma com sua perspectiva própria.

Para Rodrigues Silva (2011, p. 589), “a interdisciplinaridade, em sentido restrito, caracteriza-se pela utilização de elementos ou recursos de duas ou mais disciplinas para a operacionalização de um procedimento investigativo”. Parece simples, mas desafia os limites das disciplinas tradicionais, incentivando a flexibilidade de pensamento e a capacidade de integrar diferentes perspectivas. Isso, por sua vez, contribui para uma construção de conhecimento mais flexível e adaptável. A interdisciplinaridade representa uma interação mais profunda entre os campos do saber, com trocas metodológicas e conceituais. Fazenda (1994) define a interdisciplinaridade como um processo de diálogo e construção conjunta de conhecimento, no qual há interferência recíproca entre as disciplinas, resultando em novas compreensões e abordagens integradas. Dessa forma, o conceito de interdisciplinaridade pode ser uma abordagem eficaz para promover uma construção de conhecimento mais completa e relevante, especialmente quando se lida com problemas complexos e multifacetados. A

multidisciplinaridade e a interdisciplinaridade têm desempenhado um papel crucial na abordagem de problemas complexos que exigem a colaboração de várias áreas do conhecimento.

A forma disciplinar como se trata as disciplinas no ensino superior pode ser considerada um desafio por vários motivos. Por muito tempo a abordagem disciplinar serviu não só para construção do conhecimento, como também contribuiu grandemente para todo desenvolvimento científico. Embora a disciplinaridade seja fundamental para aprofundar o conhecimento em campos específicos, também pode apresentar limitações quando não é combinada de maneira adequada com abordagens interdisciplinares. A forma disciplinar leva à fragmentação do conhecimento, onde tópicos específicos são vistos superficialmente, sem conexão entre si. Essa falta de conexões entre as diferentes áreas do conhecimento atribui soluções limitadas para problemas complexos, sem visão holística, sem múltiplas e diferentes perspectivas para compreensão do todo, levando à incapacidade de integrar diversas habilidades e de estimular a competição entre áreas ou departamentos, ao invés de colaboração e compartilhamento de conhecimentos.

Para Nascimento *et al.* (2021), a universidade adquiriu e desenvolveu vários sentidos na consolidação do saber, interferiu e ainda interfere nos contextos, além de ser afetada por eles. Esta relação entre universidade e sociedade, sujeito e conhecimento, provoca expectativas e exigências para a ampliação de produção de conhecimentos que atendam ao contexto atual em todas as suas dimensões. Assumindo a responsabilidade do cidadão deste século, devemos dar conta de provocar uma mudança profunda de fazer ciência, e as universidades têm as ferramentas necessárias para iniciar este processo e provocar não apenas os questionamentos e provocações que o tema “interdisciplinaridade” exige, como também impulsionar o movimento, que já permeia e incomoda há algum tempo. Pombo (2006, p.14) acrescenta que devemos fazer um esforço para ultrapassar as fronteiras disciplinares, de forma a aproximar-se de um modelo novo de compreensão, que seja não analítico. Ressaltamos a importância da forma analítica de pensar. Através dela o homem descobriu a ciência e tem compreendido o mundo, porém, sua forma segmentada deixou a ciência presa em especialidades, tornando as instituições fragmentadas, uma forma totalmente contrária ao espírito científico, de característica cooperativa e dialógica (Pombo 2004; 2006). Assim, às universidades cabe a defesa da multi e interdisciplinaridade seguindo para a transdisciplinaridade, onde diferentes linguagens devem resultar no progresso científico.

O caráter eminentemente multidisciplinar da divulgação científica, onde está situado o campo do jornalismo científico, vem reunindo um conjunto de profissionais e acadêmicos de distintas áreas do conhecimento. São pessoas que comungam da ideia de que a divulgação pode contribuir com a democratização do conhecimento científico, facilitada pelo uso de uma linguagem acessível à maioria, levando-se em consideração não o nível de escolaridade, mas o entendimento de que o acesso às informações científicas e tecnológicas pode contribuir com a melhoria da qualidade de vida e com a tomada de decisões (Bortoliero, 2009, p. 49).

A preocupação com o modelo linear de ensino tem levado muitas instituições de ensino superior a buscar abordagens mais interdisciplinares, combinando conhecimentos e perspectivas de diferentes disciplinas. Isso pode enriquecer a formação dos estudantes, permitir uma compreensão mais abrangente dos problemas e prepará-los para enfrentar desafios do mundo real. No entanto, precisamos ressaltar que a disciplinaridade também tem sua importância, especialmente quando se trata de aprofundar o conhecimento em áreas específicas. Esse equilíbrio entre abordagens disciplinares e interdisciplinares é o desafio a ser ultrapassado. O que temos utilizado muito nas salas de aula é o conceito mais multidisciplinar, onde várias disciplinas diferentes falam sobre uma parte do objeto do estudo, mas sem a integração necessária para chegar no conceito interdisciplinar (Petraglia, 1993).

Efetivamente, a inteligência que só sabe separar fragmenta o complexo do mundo em pedaços separados, fraciona os problemas, unidimensionaliza o multidimensional. Atrofia as possibilidades de compreensão e de reflexão, eliminando assim as oportunidades de um julgamento corretivo ou de uma visão a longo prazo. (...) Assim, os desenvolvimentos disciplinares das ciências não só trouxeram as vantagens da divisão do trabalho, mas também os inconvenientes da superespecialização, do confinamento e do despedaçamento do saber. Não só produziram o conhecimento e a elucidação, mas também a ignorância e a cegueira (Morin, 2003, p. 14-15).

Até aqui falamos das muitas vantagens do trabalho interdisciplinar. Mas quais seriam as dificuldades em tornar os currículos mais interdisciplinares? Os professores nem sempre se mostram dispostos a ultrapassar as barreiras da sua disciplina e introduzir outros conceitos do mesmo objeto de estudo, porém bastante diferentes dos seus. Uma abordagem interdisciplinar exige que o professor se mostre aberto às novas perspectivas, que tenha coragem para navegar em mares desconhecidos, se dedique ao estudo e seja persistente. Para Rodrigues Silva (2011, p. 589), “a resistência à interdisciplinaridade, tanto no âmbito do ensino quanto no da

pesquisa, demonstra temor da perda do exclusivismo ou da autonomia”.

A preocupação das universidades quanto às mudanças na forma de ensino tem motivo, pois as universidades têm autonomia e colocam no mercado de trabalho milhares de profissionais das mais diversas áreas. No Brasil, a Lei de Diretrizes e Bases da Educação Nacional - LDB é uma legislação que estabelece as normas gerais para a educação no país, incluindo o ensino superior. A LDB define princípios e diretrizes para a organização e funcionamento das instituições de ensino, critérios para criação e funcionamento de cursos, estrutura dos sistemas de ensino, entre outros aspectos. Uma seção específica da LDB trata das diretrizes e bases da educação superior. Essa seção abrange a organização, a estrutura curricular, a avaliação, a autonomia universitária e outros temas relevantes para essas instituições e seus cursos (Brasil, 1996).

A educação superior é regulamentada principalmente nos artigos 43 a 57 da Lei nº 9.394/96, sob o título “Da Educação Superior”, no seu capítulo IV. Apesar da lei já ter 27 anos, destacamos aqui, alguns pontos importantes para o nosso estudo e os respectivos tópicos que estão relacionados à modelagem proposta (Brasil, 1996) (Quadro 2).

Quadro 2. Seleção de artigos e capítulos da LDB nº 9.394/96 relacionados à Educação Superior e características buscadas pelo modelo proposto

ARTIGO	CAPÍTULOS SELECIONADOS	CARACTERÍSTICAS BUSCADAS NO MODELO
Art. 43: A educação superior tem por finalidade:	<p>II - Formar diplomados nas diferentes áreas de conhecimento, aptos para a inserção em setores profissionais e para a participação no desenvolvimento da sociedade brasileira, e colaborar na sua formação contínua;</p> <p>III - Incentivar o trabalho de pesquisa e investigação científica, visando o desenvolvimento da ciência e da tecnologia e da criação e difusão da cultura, e, desse modo, desenvolver o entendimento do homem e do meio em que vive;</p> <p>VI - Estimular o conhecimento dos problemas do mundo presente, em particular os nacionais e regionais, prestar serviços especializados à comunidade e estabelecer com esta uma relação de reciprocidade.</p>	<p>II - Interdisciplinaridade; Colaboração. Cooperação.</p> <p>III - Estímulo à pesquisa científica; Disseminação/Difusão do Conhecimento; Multi/interdisciplinaridade.</p> <p>VI - Multi/interdisciplinaridade. Disseminação/Difusão do Conhecimento; Colaboração. Cooperação.</p>
Art. 52: As universidades são instituições pluridisciplinares de formação dos Quadros profissionais de nível superior, de pesquisa, de extensão e de domínio e cultivo do saber humano, que se caracterizam por:	I - Produção intelectual institucionalizada mediante o estudo sistemático dos temas e problemas mais relevantes, tanto do ponto de vista científico e cultural, quanto regional e nacional.	I - Trabalho em conjunto com temáticas afins; Disseminação.
Art. 53: No exercício de sua autonomia, são asseguradas às universidades, sem prejuízo de outras, as seguintes atribuições:	<p>II - Fixar os currículos dos seus cursos e programas, observadas as diretrizes gerais pertinentes;</p> <p>III - Estabelecer planos programas, observadas as diretrizes gerais pertinentes;</p> <p>V - Elaborar e reformar os seus estatutos e regimentos em consonância com as normas gerais atinentes.</p>	II, III, V - Autonomia da universidade para criar o seu programa em moldes mais multi/interdisciplinares.

Fonte: Elaboração própria, (2023).

Esses pontos relacionados ao ensino superior abordam vários aspectos importantes, incluindo: a necessidade de formação de profissionais nas mais diversas áreas, o incentivo à pesquisa, estímulo ao conhecimento e investigação de problemas reais, formação pluridisciplinar, promoção da qualidade na formação acadêmica, pesquisa e extensão e a autonomia das universidades reconhecida e garantida, permitindo que elas definam seus projetos acadêmicos, planos de desenvolvimento institucional e outros aspectos relacionados à gestão (Brasil, 1996). O último ponto, a autonomia, torna a universidade protagonista no processo de difusão/disseminação e da multi e interdisciplinaridade para os demais níveis de ensino. Embora a LDB não mencione diretamente a interdisciplinaridade nas universidades, a autonomia universitária permite que as instituições desenvolvam suas próprias políticas e práticas acadêmicas, incluindo a promoção da interdisciplinaridade. Muitas universidades brasileiras, apesar de ainda muito superficialmente, têm se preocupado com abordagens mais interdisciplinares em seus currículos e programas de pesquisa como uma forma de estimular a integração de diferentes áreas de conhecimento. Trevisol (2021) evidencia que há uma relação entre a autonomia universitária e a democracia na universidade. Em sua pesquisa, ele analisa historicamente a trajetória feita pelas universidades brasileiras e destaca:

A pesquisa evidenciou uma estreita correlação entre democracia e autonomia. A trajetória de uma se confunde com a outra. Como a democracia no Brasil tem sido mais a exceção que a regra, as universidades brasileiras passaram a maior parte de sua história sob a tutela e o controle direto ou indireto dos governos autoritários. Os regimes autoritários, invariavelmente, moveram ações para suprimir as liberdades acadêmicas e de cátedra, assim como restringiram o direito das instituições de definirem as suas regras relativas à gestão administrativa e financeira (Trevisol, 2021, p. 322).

Essa relação entre autonomia e democracia parece lógica quando Trevisol a expõe, porém, não é algo que observamos logo de início. Devemos ressaltar a importância que a democracia tem na vida não só de uma universidade, mas de cada um de nós, de todos nós. Sem ela, não adianta professor preparado, nem aluno interessado. A autonomia é um bem que a universidade deve buscar e exigir sempre, democraticamente, para que tenhamos direito de mudar e propor mudança à toda sociedade, de acordo com as suas demandas. E a interdisciplinaridade e a multidisciplinaridade podem ser instrumentos importantes para esta forma democrática e autônoma de pensar e discutir ideias, pois são características inerentes

delas.

A multidisciplinaridade é frequentemente considerada, em conjunto com a interdisciplinaridade, como parte das diretrizes para o desenvolvimento de currículos de qualidade. Para buscar requisitos que alicerçam a multi e interdisciplinaridade para Instituições de Ensino Superior (graduação e pós-graduação) buscamos informações do Ministério da Educação - MEC, do Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq, ligado ao Ministério da Ciência, Tecnologia e Inovação - MCTI para incentivo à pesquisa no Brasil, e da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES, que avalia a pós-graduação no Brasil. A avaliação da multidisciplinaridade e interdisciplinaridade em cursos de graduação está mais relacionada a diretrizes curriculares do MEC do que diretamente à CAPES, apesar desta estar envolvida em programas de apoio à formação de professores e à educação básica, fatores diretamente relacionados ao trabalho multi e interdisciplinar (Brasil, Ministério da Educação).

O intuito de utilizar a interdisciplinaridade não é acabar com as disciplinas, e sim utilizar os conteúdos de todas essas disciplinas, para melhor entender o problema e buscar a resolução nas diferentes visões de cada uma das disciplinas. Assim, a multidisciplinaridade e a interdisciplinaridade em cursos superiores são ferramentas importantes para promover uma formação mais abrangente e preparar os estudantes para lidar com problemas complexos do mundo real, mesmo que, na prática, ainda exista muita dificuldade na implementação de currículos interdisciplinares. Embora esses documentos ofereçam diretrizes gerais, a implementação da interdisciplinaridade muitas vezes ocorre de forma mais específica nas políticas e práticas das instituições de ensino. Cada instituição pode desenvolver abordagens interdisciplinares de acordo com seus objetivos e contextos educacionais.

Esta autonomia que é dada aos cursos de graduação permite que as universidades possam elaborar seu planejamento de acordo com o curso de cada uma. Porém, para pós-graduação, a CAPES, vinculada ao MEC, é responsável pelo fomento e pela coordenação de atividades de ensino superior, pesquisa e pós-graduação. A CAPES financia e oferece bolsas para incentivo aos projetos de pesquisa, avalia periodicamente cursos de graduação e pós-graduação, incentiva parcerias nacionais e internacionais e oferece programas específicos para a educação básica e formação de professores, desempenhando um papel estratégico no desenvolvimento da educação e da pesquisa no Brasil, contribuindo para a formação de recursos humanos qualificados, na produção de conhecimento científico e a promoção da qualidade na educação superior e na pesquisa acadêmica (Brasil, 2022).

Além de avaliar, a CAPES é responsável também por classificar os programas de pós-graduação em diferentes áreas do conhecimento no Brasil. Essa classificação é feita com o intuito de que as instituições possam organizar as informações que se referem aos projetos de pesquisa. A área “interdisciplinar” fica dentro da Grande Área “multidisciplinar”. (CAPES, Sobre as áreas de avaliação; CAPES, Tabela de Áreas de Conhecimento/Avaliação). Programas classificados como “multidisciplinar” abordam temas que não se encaixam estritamente em uma única disciplina tradicional. Eles podem envolver múltiplas disciplinas, abordagens e métodos para questões complexas e interconectadas. Já os programas classificados como “Interdisciplinar” também transcendem as disciplinas tradicionais, mas têm um foco mais intenso na integração e colaboração entre diferentes campos de conhecimento. Eles buscam abordar problemas complexos por meio da combinação de perspectivas, teorias e metodologias de diversas áreas. A denominada “Área Interdisciplinar” tem o maior número de cursos hoje e tem crescido muito mais do que as demais áreas, o que demonstra um interesse crescente na interdisciplinaridade. Essas áreas são reconhecidas pela CAPES por acomodar programas de pós-graduação que não se enquadram em uma única disciplina ou que têm um foco específico na integração de múltiplas disciplinas. A classificação em uma dessas áreas reflete a abordagem e os objetivos do programa em questão (CAPES, Sobre as áreas de avaliação).

As universidades devem ter diretrizes curriculares, projetos de pesquisa e programas acadêmicos que busquem fomentar a colaboração entre diferentes áreas de conhecimento, promover a interdisciplinaridade como um meio de abordar questões complexas e desafios contemporâneos. Os requisitos do Ministério da Educação - MEC do Brasil para avaliar a interdisciplinaridade em cursos superiores estão sujeitos a políticas e diretrizes em constante evolução, mas existem algumas informações gerais sobre como a interdisciplinaridade pode ser avaliada no contexto da educação superior no Brasil. As Diretrizes Curriculares Nacionais - DCNs estabelecem os requisitos mínimos para a organização dos cursos de graduação no Brasil, porém tem caráter específico, dependendo do curso. Elas podem incluir diretrizes sobre a integração de conteúdos e a promoção da interdisciplinaridade nos currículos dos cursos (Conselho Nacional de Educação, 2007).

O MEC realiza avaliações periódicas dos cursos superiores no Brasil por meio do Sistema Nacional de Avaliação da Educação Superior - SINAES. Durante essas avaliações, a interdisciplinaridade pode ser um dos critérios considerados. Os avaliadores podem verificar, por exemplo, como os cursos integram conhecimentos de diferentes disciplinas para promover

uma formação mais abrangente e contextualizada. Um dos itens avaliados é o Projeto Pedagógico do Curso - PPC, que descreve a estrutura curricular de cada curso, seus objetivos e os métodos de ensino. O PPC deve refletir o compromisso com a interdisciplinaridade, descrevendo como ela será incorporada ao longo do curso. Entrevistas feitas pelos avaliadores, com alunos e professores, também podem identificar multi e interdisciplinaridade destes cursos (Brasil, SINAES).

Diante das avaliações realizadas, a instituição obtém o Conceito Preliminar de Curso (CPC) dos cursos e instituições, elaborado pelo MEC e o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). O ciclo de avaliação mais recente é chamado de Ciclo de Avaliação do ENADE - Exame Nacional de Desempenho dos Estudantes, que é realizado a cada três anos. Durante esse ciclo, diferentes áreas do conhecimento passam pela avaliação, e os cursos pertencentes a essas áreas participam do ENADE e têm o CPC calculado. O desempenho de um curso no CPC pode ser influenciado pela sua capacidade de promover a interdisciplinaridade e proporcionar uma formação de qualidade aos alunos (Brasil, ENADE).

No ensino médio, as diretrizes são mais específicas quanto à interdisciplinaridade, e as instituições de ensino superior devem estar em consonância com todos os graus de ensino, pois são responsáveis por grande parte dos profissionais que levarão a educação aos mais diversos setores da sociedade. Ressaltamos o artigo 8º, na resolução CEB (Câmara de Educação Básica do Conselho Nacional de Educação - CNE) nº 3, de 26 de junho de 1998 (Brasil, 2018), que institui as Diretrizes Curriculares Nacionais para o Ensino Médio (DCNs), onde são estabelecidos pontos importantes, como: a necessidade de diálogo permanente entre conhecimentos; o estudo de problemas concretos que promovam a capacidade de análise do estudante e seu desenvolvimento intelectual, social e afetivo integrado; a busca da interação entre diferentes áreas do conhecimento; e a responsabilidade da escola em integrar conhecimentos e competências para pleno exercício da cidadania. Por serem pontos a serem evidenciados no ensino médio, devem ser trabalhados nas universidades, já que elas são responsáveis pela formação dos professores que estarão nas salas de aula do Ensino Fundamental e Médio. Tais instituições, além de formar esses professores, continuam em contato através dos programas de educação continuada.

2.3.1 Colaboração e Cooperação

Falar em interdisciplinaridade e não falar de colaboração parece contraditório, e é, pois todos sabemos da importância do trabalho em equipe, e nos dias de hoje, são recorrentes as falas e análises feitas em relação à socialização. Ressaltamos aqui que, além dos pontos citados, foram analisados os resultados da interação dos grupos, que por vezes chamamos de grupos “multi-áreas” ou “multi departamentais”, nos quais pessoas de diferentes cursos e departamentos ou áreas podem realizar trocas e construir juntos novos conhecimentos que podem ser melhor disseminados e difundidos (dentro e fora da comunidade acadêmica). Os termos imprescindíveis para essa interação são colaboração e cooperação, pois descrevem diferentes tipos de relações entre indivíduos, grupos, organizações ou entidades.

É importante ressaltar a questão da competição que existe em grupos de produção científica. Pombo (2008) afirma que a ciência hoje parece ser uma enorme instituição, com diferentes comunidades competitivas entre si, mas que não conseguem colaborar entre elas. Os grupos parecem brigar por posições, competindo e estabelecendo entre si uma concorrência, contra o verdadeiro ideal científico da comunicação. Para ilustrar a sua posição, ela fala da “patentificação” na ciência, onde, mesmo na investigação inicial, os pesquisadores patenteiam o “produto”, impossibilitando que outras instituições possam seguir o caminho já percorrido, o que certamente empobrece a pesquisa, pois impede a troca de informações e ideias, reduzindo as possibilidades. Esse fato é comum, mas vai contra o que sugerimos neste trabalho. A nossa proposta é que haja um trabalho multi/interdisciplinar cooperativo e colaborativo, fatores que julgamos essenciais para obtermos resultados muito mais colaborativos e participativos.

Essa colaboração deve envolver indivíduos ou grupos que trabalham juntos para alcançar um objetivo comum. Nesse contexto, eles compartilham recursos, conhecimentos, habilidades e esforços para benefício mútuo. Na colaboração, todas as partes envolvidas buscam resultados positivos para todos, é uma forma de interação em que indivíduos ou grupos trabalham juntos de maneira coordenada, mesmo que não estejam buscando o mesmo objetivo. Os participantes compartilham conhecimentos, habilidades e recursos, contribuindo coletivamente para o sucesso do projeto ou meta. As partes dependem umas das outras para chegarem a uma decisão conjunta. Gray (1989) define colaboração como a gestão construtiva

de diferenças, um processo de embates e conflitos. A colaboração seria um método para resolução de disputas entre colaboradores. Apesar da impressão de resultar numa situação conflituosa, Gray afirma que é importante manter um clima de confiança entre as partes, para que os envolvidos possam se sentir seguros suficientemente para explorar suas diferenças e fraquezas.

Em um estudo feito na Inglaterra, Creese, Norwich & Daniels (1997) apontam evidências de que escolas com predominância de trabalhos colaborativos são mais inclusivas, têm menos evasão e percebe-se uma maior aproximação efetiva na resolução de problemas de estudantes secundaristas. Os resultados refletem características de socialização e apontam habilidades que são adquiridas e trocadas pelos envolvidos no processo.

Ainda falando de colaboração, Martins & Ferreira (2013) apontam a necessidade de entendimento das causas e as principais razões que influenciam os pesquisadores na forma de articulação e construção de suas redes de colaboração científica, o que pode influenciar o desenvolvimento de novos indicadores e modos de avaliação da produção científica. Para eles, “o conceito de redes sociais permite operar novos planos de análise, contribuindo com seus aspectos estruturais e dinâmicos ao estudo dos mecanismos e gatilhos causais que levam à constituição dessas redes de colaboração científica” (Martin & Ferreira, 2013, p. 1).

Outro conceito utilizado como base em nossa pesquisa é a cooperação. A cooperação tem uma conotação positiva, já que envolve a disposição de ajudar e contribuir. Os grupos podem cooperar ao compartilhar informações e/ou recursos para atingir seus próprios objetivos separados. Cooperação é o ato de trabalhar em conjunto com outros, mas com menos integração e interdependência do que na colaboração. Na cooperação, as partes ajudam-se mutuamente, mas podem ter objetivos individuais ou independentes.

O processo de construção do conhecimento científico ocorre a partir de uma ação coletiva, orquestrada pela conjugação de esforços individuais e institucionais, ora cooperando, ora concorrendo, em um jogo que se dá em torno da disputa por recursos materiais finitos e, muitas vezes, escassos, em situação de cooperação/concorrência que, via de regra, traz avanços para o conjunto da sociedade (Melo, 2022, p. 21).

2.4 CIÊNCIA DAS REDES

A estrutura mundial nos dias de hoje tem o formato de rede. Tudo que faz parte do

mundo real possui relações que podem ser representadas em rede. É o que Castells (2006) chama de “Sociedade em Rede” e cuja “nova” economia do século XX tem característica informacional, globalizada e em rede. Tais características se encaixam também no funcionamento das universidades, que são informacionais, à medida que funcionam como centros de produção, disseminação e gestão da informação e do conhecimento. São globais, à medida que a colaboração científica acontece através de parcerias internacionais. E as conexões acadêmicas entre professores, pesquisadores, alunos e até mesmo instituições, formam redes importantes para entender as relações e tendências (Castells, 2006).

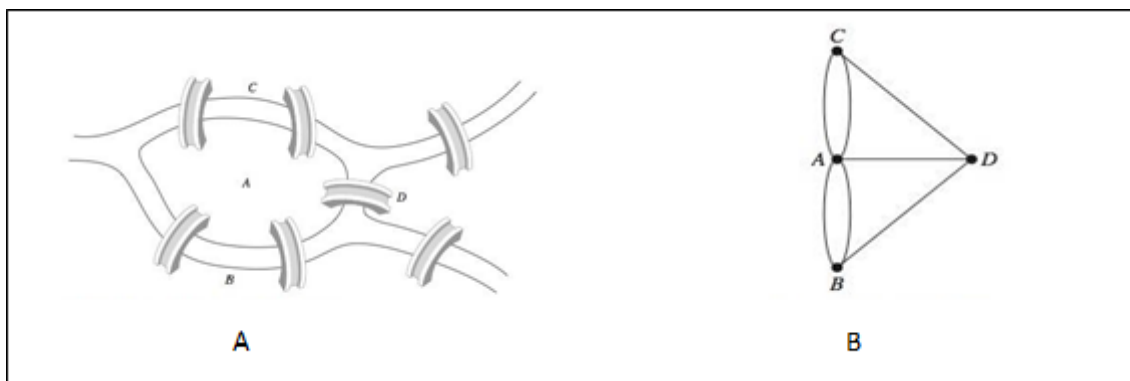
As universidades têm o conhecimento como seu bem mais valioso, mas para entender como acontece o processo de produção de conhecimento nessas instituições, precisamos identificar a forma que se produz esse conhecimento, como esses grupos se organizam, quem faz parte desses grupos, a quais áreas pertencem, os temas que discorrem, a forma como é feita colaboração. Estas características devem emergir das redes e descrever o funcionamento da produção científica da instituição, e são pontos a serem avaliados para que possamos propor mudanças para fortalecer a construção e disseminação do conhecimento. Todos os modelos de organização têm características de redes, o mesmo acontece nas universidades. Para Castells (2006, p. 565), “redes constituem a nova morfologia social de nossas sociedades e a difusão da lógica de redes modifica de forma substancial a operação e os resultados dos processos produtivos e de experiência, poder e cultura”.

Uma estrutura social é definida por ele como um sistema aberto e dinâmico suscetível de inovação, sem ameaças ao seu equilíbrio. São estruturas capazes de expandir de forma ilimitada, integrando novos nós desde que consigam comunicar-se dentro da rede. Para melhor definição, rede é um conjunto de nós interconectados, os quais são os sujeitos relacionados pela ação em estudo. Castells (2006) entende que a topologia definida por redes determina que a distância (ou intensidade e frequência da interação) entre dois pontos (ou posições) é menor (ou mais frequente, ou mais intensa) se ambos os pontos forem nós de uma rede do que se não pertencesse à mesma rede e essas conexões representam as relações diretas.

Para representar da melhor maneira as redes sugeridas, recorreremos a “A ciência das Redes”, também conhecida como teoria das redes, cujo estudo teve início em 1736, quando Leonhard Euler publicou a solução das pontes de Königsberg fundamentando a teoria dos grafos. A cidade de Königsberg ficava na Prússia e hoje é chamada Kaliningrad, localizada na República Russa. A cidade era dividida em quatro partes (A, B, C e D) (Figura 1); O Rio

Pregel (B e C), a Ilha de Kneiphof (A) e a região entre os dois braços (D) e sete pontes ligavam as regiões. Todos questionavam se seria possível sair de um dos pontos da cidade (A, B, C ou D), passar por todas as pontes (sem repeti-las) e finalizar a caminhada no ponto de partida. Euler representou a situação utilizando, pela primeira vez que se tem notícia, o conceito de grafos, e desenhou um multigrafo para fazer esta representação, onde, cada ponto de partida representa os nós (vértices) e as pontes representam as arestas (Figura 1). Dessa forma, Euler provou a impossibilidade de sair de qualquer um dos vértices, passar por todas as arestas apenas uma vez, e retornar ao vértice de partida (Rosen, 2010).

Figura 1. As sete pontes de Königsberg (A) e Modelo de Multigrafo da cidade de Königsberg (B)



Fonte: Elaboração própria, baseado em (Rosen, 2010, p. 634)

A partir deste problema, surgiram os ciclos eulerianos, os caminhos eulerianos e muitos teoremas e definições ligados aos grafos, que fazem parte hoje da matemática discreta, representando situações reais de logística, aplicativos de rotas etc.

Historicamente, alguns fatos importantes marcaram o desenvolvimento do estudo dos grafos. Na década de 1930, pesquisadores das ciências sociais, como Jacob Moreno, e antropólogos britânicos, como Radcliffe-Brown e Max Gluckman, aplicaram métodos de análise de redes sociais para investigar as relações entre membros de grupos sociais. Embora não sejam considerados figuras centrais na ciência das redes modernas, suas contribuições para a sociologia e a antropologia influenciaram significativamente a forma como os estudiosos analisam as relações e estruturas sociais (Russo, 2010; Ingold, 2016; Gluckman, 1955).

Era 1950 quando os matemáticos húngaros Paul Erdős e Alfréd Rény criaram o modelo de Erdős-Rény, introduzindo o conceito de grafos aleatórios. O modelo descreve a

formação de grafos aleatórios através de uma probabilidade de inclusão de arestas entre os nós de um grafo. Esses modelos de grafos aleatórios ajudaram a estabelecer as bases para o estudo de propriedades probabilísticas em grafos e foram fundamentais para o desenvolvimento da teoria dos grafos probabilísticos. O trabalho de Erdős e Rényi influenciou significativamente o campo da teoria dos grafos e teve aplicações em várias áreas, incluindo ciência da computação, sociologia, biologia e física estatística (Abbas, 2017).

Além do estudo da teoria dos grafos aleatórios, Paul Erdős e Alfréd Rényi desenvolveram métricas como o grau de um vértice, precursoras das métricas de centralidade. No entanto, foi somente mais tarde que os conceitos de centralidade em redes sociais foram formalizados e popularizados por pesquisadores, destacando-se Linton Freeman, que publicou trabalhos significativos sobre centralidade em redes sociais na década de 1970. A partir desse ponto, a análise de centralidade em redes sociais se tornou uma área importante de pesquisa em sociologia, ciência da computação e outras disciplinas relacionadas. Em Freeman (1978), podemos encontrar uma das primeiras explicações detalhadas do conceito de centralidade em redes sociais.

A partir do ano de 1960, os pesquisadores Stanley Milgram, Duncan Watts, Steven Strogatz e outros começaram a explorar as propriedades estruturais e dinâmicas das redes complexas. Neste mesmo ano, Milgram ficou famoso por conduzir o “Experimento de Seis Graus de Separação”, que tinha o objetivo de investigar a “pequenez do mundo”, ou seja, ele queria comprovar que qualquer pessoa no planeta pode ser conectada a outra por uma cadeia de conhecidos relativamente curta. Os resultados do experimento indicaram que as cadeias sociais de conhecidos eram, em média, muito curtas, apoiando a noção de que o mundo social é interconectado em uma escala surpreendentemente pequena (Milgram, 1967).

Watts é conhecido por seus trabalhos na teoria dos grafos e na ciência das redes. Ele colaborou com Strogatz introduzindo o conceito de “mundo pequeno” em redes. Nesse trabalho, intitulado “Collective dynamics of 'small-world' networks” (1998) eles apresentaram um modelo de rede que exhibe características de “mundo pequeno”, onde a maioria dos nós está a uma distância relativamente curta um do outro. Juntos, Watts e Strogatz desenvolveram o modelo Watts-Strogatz, projetado para capturar características de “mundo pequeno” (Watts & Strogatz, 1998).

Estudos mais recentes foram feitos pelo físico Albert-László Barabási e seus colaboradores, que descobriram que muitas redes naturais seguem o modelo livre de escala ou Scale-Free Network, onde alguns nós (hubs) têm um número muito maior de conexões do que

a maioria dos outros nós. Barabási mostrou como a distribuição de graus seguindo uma lei de potência é uma característica comum em muitos sistemas complexos, desafiando as ideias tradicionais de distribuição de graus (Barabási, 2000; 2002; 2016; Barabási & Albert, 1999).

Outro nome de destaque para ciência das redes atualmente é Mark Newman, um renomado físico teórico e estatístico, com muitas contribuições nesta área, das quais destacamos aqui, duas delas, que foram importantes para a nossa pesquisa. Newman desenvolveu métodos avançados para detecção de comunidades em redes complexas, utilizando métodos estatísticos (Newman, 2006). Ele também desenvolveu métricas e medidas para avaliar a centralidade, que podem identificar nós importantes nas redes (Newman, 2005).

A ciência das redes é uma área interdisciplinar em rápido crescimento, que explora as propriedades e os padrões emergentes em uma variedade de sistemas complexos interconectados. Atualmente é comum o uso de softwares que auxiliam a visualização e análise das redes, através de métricas e recursos que fazem os cálculos estatísticos e organizam os dados conforme a necessidade do usuário.

3. METODOLOGIA

3.1 TEORIA FUNDAMENTADA DE DADOS (TFD)

A Teoria Fundamentada de Dados - TFD (ou *Grounded Theory* - GT) foi desenvolvida na década de 60, nos Estados Unidos, por Barney Glaser e Anselm Strauss como um método de pesquisa (Tarozzi, 2011). A TFD tem foco na análise qualitativa dos dados, com uma abordagem interpretativa. Seguindo um processo iterativo e indutivo de análise, onde há uma análise dos dados, padrões e categorias são identificados, e teorias são construídas a partir desses padrões. No nosso trabalho, optamos pela linha construtivista.

Para Charmaz (2009), a Teoria Fundamentada de Dados Construtivista (Constructivist Grounded Theory) é uma variação da abordagem da Teoria Fundamentada nos Dados (Grounded Theory) e foi desenvolvida por ele nos anos 1980, como uma resposta à abordagem original de Glaser e Strauss. Enquanto a abordagem original era mais objetiva e centrada na descoberta das categorias e teorias diretamente dos dados, a abordagem construtivista ressalta a interpretação ativa do pesquisador e o papel do contexto social e cultural na construção de significados. Para Nascimento *et al.* (2021), a Teoria Fundamentada é uma metodologia de pesquisa sistemática, flexível, estruturada em várias fases interdependentes e uma dinâmica circular de comparações constantes dos dados. As fases principais envolvem: 1. Coleta dos Dados; 2. Codificação dos Dados; 3. Construção de Categorias e relações entre elas e 4. Interpretações. As fases não são lineares, mas interdependentes e iterativas, onde o pesquisador retorna aos dados constantemente para refinar as categorias e relações. Tal flexibilidade permite que novas interpretações surjam à medida que o processo evolui.

A escolha da Teoria Fundamentada de Dados Construtivista deve-se ao fato dela ser especialmente útil em pesquisas que lidam com questões complexas e multifacetadas, onde a interpretação pessoal e a compreensão do contexto social são essenciais para a geração de teorias significativas. O construtivismo considera que a realidade é construída socialmente e que as interpretações e significados são influenciados pelas interações entre os pesquisadores e os participantes do estudo. A TFD construtivista valoriza a colaboração e a interação entre pesquisadores e participantes, buscando uma compreensão mais profunda das experiências e significados subjacentes aos dados coletados, permitindo que os pesquisadores possam

abordar a construção de teorias de maneira mais colaborativa, envolvendo os participantes no processo de análise e validação das conclusões, enfatizando a co-construção de significados entre os pesquisadores e os participantes, levando em consideração as perspectivas individuais e os contextos sociais em que os fenômenos ocorrem.

Em nosso trabalho, combinamos a Teoria Fundamentada nos Dados- TFD Construtivista com Análise de Redes Sociais - ARS, essa combinação da TFD foi utilizada para examinar a dinâmica das interações sociais e como as relações influenciam o desenvolvimento de conceitos emergentes.

3.2 ANÁLISE DE CONTEÚDO

Partindo do interesse nas informações das redes semânticas, percebemos a necessidade de utilizar a análise de conteúdos na pesquisa dessas redes para interpretar os dados extraindo das palavras os significados e tendências. Dessa forma, usamos uma categorização temática baseada na análise de conteúdos de Laurence Bardin (2011). Trata-se de um conjunto de técnicas de análise de comunicações com o objetivo de interpretar e categorizar o conteúdo de textos (ou outras formas de comunicação) de maneira sistemática e objetiva. Os conceitos fundamentais de Bardin foram ressignificados e representados no Quadro 3.

Quadro 3. Conceito da Análise de Conteúdo

ANÁLISE DE CONTEÚDO	
DEFINIÇÃO	OBJETIVO
É definida como um conjunto de técnicas de análise das comunicações, que visa descrever o conteúdo das mensagens, de forma sistemática, objetiva e quantitativa ou qualitativa.	Transformar o conteúdo qualitativo (como textos, discursos, imagens) em dados analisáveis, revelando padrões, tendências e inferências que não são imediatamente visíveis.
ETAPAS	
Pré-análise	<ul style="list-style-type: none"> • Organização e sistematização • Seleção do material • Escolha do que vai ser utilizado (palavras, frases etc.)
Exploração	<ul style="list-style-type: none"> • Codificação do material • Fragmentação em unidades menores • Aplicação de regras (classificação, frequência etc.)
Tratamento, Inferência e Interpretação	<ul style="list-style-type: none"> • Interpretação dos dados • Comparação dos elementos encontrados • Extração de significados, tendências ou implicações
UNIDADES DE ANÁLISE	
	<ul style="list-style-type: none"> • Unidades de registro (a menor unidade analisada, como uma palavra ou expressão);

Unidades de Análise	<ul style="list-style-type: none">• Unidades de Contexto (são mais amplas, utilizam o contexto, como um parágrafo, para explicar as unidades de registro).
---------------------	--

Fonte: Elaboração própria, baseado em (Bardin, 2011), (2024).

Para Bardin as “unidades de análise” são partes específicas do trabalho e serão examinadas para responder às questões da pesquisa. Elas servem para delimitar o objeto de estudo e organizar o processo de análise de forma sistemática. Dessa forma, o material coletado deve ser agrupado em categorias, que são classes temáticas organizadas com base nos objetivos da pesquisa, e então, é feita a análise de conteúdo, que pode ser quantitativa, como a frequência de termos ou qualitativa, obtendo seus significados e interpretações. A análise de conteúdo permite identificar categorias e padrões emergentes em dados textuais, sendo apropriada para explorar redes semânticas da pesquisa (Bardin, 2011; Krippendorff, 1980).

3.3 MÉTODOS QUANTITATIVOS

As estatísticas descrevem matematicamente o comportamento das redes encontradas e oferecem subsídios para possíveis alterações na rede, no intuito de torná-la mais conectada e multi/interdisciplinar. As análises estatísticas podem trazer características das redes que não aparecem nas suas visualizações. Esses métodos são amplamente utilizados para analisar e interpretar características das redes sociais, semânticas ou de outros tipos, permitindo uma avaliação objetiva e mensurável.

3.3.1 Análise de Redes Sociais e Complexas (ARS)

A análise de Redes Sociais - ARS (Hanneman *et al*, 2005; Recuero, 2017) é uma metodologia quantitativa (análise de métricas) e qualitativa (significado das relações) e por isso pode oferecer uma visão ampla das conexões da rede. A ARS examina as interações e relações entre indivíduos, grupos ou entidades em um sistema.

No caso da rede de coautoria, os autores considerados mais influentes são os que contribuem mais, aparecem mais, colaboram mais na rede por terem mais conexões ou por estarem mais próximos de outros nós, ou ainda, por facilitarem a conexão entre grupos distintos na rede. Os padrões de coautoria devem ser identificados para entender como está funcionando a colaboração da rede. Isso pode revelar comunidades, indicar grupos de

pesquisa, redes de colaboração, disseminação e conexões de multi e interdisciplinaridade. A análise topológica da rede pode trazer mais características das relações na rede.

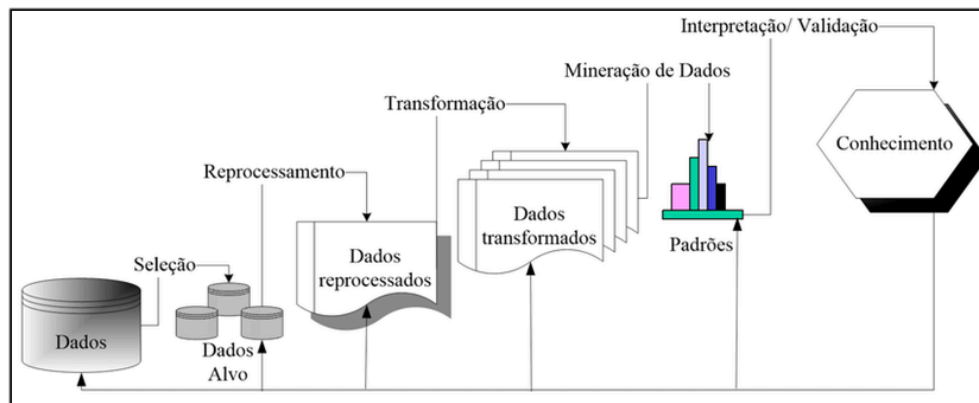
No caso das redes semânticas, as análises são feitas pelas palavras e sua posição na rede, indicando principais temáticas publicadas pelos professores através destas palavras ou junções de palavras. Essa abordagem quantitativa é frequentemente usada para avaliar a relevância e o impacto da produção acadêmica em uma determinada área de pesquisa (Bornmann & Leydesdorff, 2014).

3.4 PERCURSO INVESTIGATIVO

3.4.1 KDD (Knowledge Discovery in Databases) e ETL (Extract Transform Load)

Como caminho de pesquisa utilizamos como referência o processo KDD (Knowledge Discovery in Databases). O conceito de KDD surgiu na década de 1980 como uma resposta à crescente disponibilidade de grandes volumes de dados armazenados em bancos de dados. Embora o termo "Knowledge Discovery in Databases" (KDD) tenha sido popularizado na década de 1990, a ideia de explorar dados para descobrir padrões e conhecimentos remonta a décadas anteriores. A formalização do processo KDD como um campo de estudo veio com o crescimento exponencial da quantidade de dados digitais disponíveis e a necessidade de métodos sistemáticos para lidar com esses dados. Isso inclui dados estruturados armazenados em bancos de dados. Dentro desse contexto, pesquisadores de áreas como banco de dados, inteligência artificial, aprendizado de máquina e estatística começaram a colaborar e desenvolver abordagens mais sistemáticas para explorar esses dados em busca de conhecimento. Isso resultou na formalização do processo KDD como uma abordagem iterativa e interdisciplinar para descobrir conhecimento útil e compreensível em grandes volumes de dados (FAYYAD, 1996). As etapas do processo estão descritas na Figura 2.

Figura 2. Transformação dos dados em informação /conhecimento



Fonte: Camilo & Silva (2009), adaptado de Fayyad *et al* (1996).

As fases a serem descritas na pesquisa são:

1. Seleção dos Dados: envolve escolha e coleta do conjunto de dados relevantes para a análise.
2. Reprocessamento dos Dados (ou pré-processamento): os dados brutos são limpos e processados para eliminação de ruído, tratamento de valores ausentes, normalização e transformação para formatos adequados para análise.
3. Transformação dos Dados: os dados são transformados em formatos mais adequados para análise e modelagem. A transformação dos dados foram
4. Mineração de Dados: é a fase central do processo KDD, onde algoritmos de mineração de dados são aplicados aos dados transformados para identificar padrões, tendências ou relações interessantes.
5. Interpretação e Utilização dos Resultados: os padrões úteis e compreensíveis são interpretados e utilizados para tomar decisões, fazer previsões ou gerar novos conhecimentos. Isso pode envolver a criação de relatórios, visualizações ou modelos preditivos.

O KDD busca identificar padrões e conhecimentos úteis em grandes volumes de dados. As etapas iterativas envolvem ajustes nos modelos de mineração de dados e na interpretação dos resultados. A utilização de técnicas do KDD tem como intuito obter uma compreensão mais rica e abrangente de um problema de pesquisa, onde os dados podem ser usados para validar ou enriquecer insights obtidos através da mineração de dados.

No processo de descoberta de conhecimento em bases de dados (KDD), a etapa de pré-processamento é fundamental para garantir a qualidade dos dados utilizados na mineração. Para fase de preparação desta pesquisa adotamos o processo ETL (Extração, Transformação e Carga). Esta técnica é utilizada para preparar dados brutos para análise, garantindo que estejam organizados corretamente para as etapas posteriores, como a mineração de dados (KIMBALL & CASERTA, 2004). Dessa forma, o processo KDD envolve a preparação dos dados antes da mineração, utilizando os dados processados pelo ETL, encontrar padrões e gerar conhecimento. Assim, o ETL é uma abordagem essencial na nossa pesquisa e segue os seguintes passos:

1. Extração: coleta de dados
2. Transformação: processamento e organização dos dados
3. Carga: inserção dos dados transformados em um formato adequado para análise

A fase de pré-processamento no KDD se relaciona fortemente com o ETL, pois envolve limpeza, normalização e preparação dos dados para a análise.

3.4.2 Dados: Escolha da amostra para o desenvolvimento do modelo: Universidade Estadual de Feira de Santana (UEFS)

As universidades públicas ainda lideram a preferência do mercado de trabalho no Brasil. Isso se deve à boa qualidade do ensino, espaço aberto para discussões e formação humanizada. “O fato de as universidades públicas serem mais concorridas reside no fato delas serem reconhecidas no mercado pela qualidade do ensino e geralmente aparecem nas listas das melhores universidades do Brasil e até mesmo do mundo” (Educa mais Brasil, 2019a). São as Instituições públicas que mais publicam no Brasil, onde mais de 90% da produção científica nacional são de universidades públicas, realizada por docentes e discentes de programas de pós-graduação (Carta Campinas, 2019).

Num ranking das 50 instituições brasileiras que mais produziram trabalhos científicos nos últimos cinco anos, 43 são universidades públicas, e apenas uma é universidade privada. As top 10 da lista são todas públicas, e apenas uma não é universidade: a Embrapa (Jornal da USP, 2019).

Reconhecendo a importância das universidades públicas no Brasil em relação à produção científica, para o estudo de caso da nossa pesquisa, decidimos escolher uma universidade pública estadual, a Universidade Estadual de Feira de Santana (UEFS). Essa instituição está localizada na Av. Transnordestina, s/n, no bairro Novo Horizonte em Feira de Santana, na Bahia. A intenção da criação de uma “faculdade”, na época, era interiorizar o ensino superior na Bahia, já que as universidades estavam restritas à Salvador, capital do Estado. Feira de Santana é um importante entroncamento rodoviário e vetor comercial. Com o objetivo de aumentar o número de professores formados, expandir a educação e a industrialização da cidade, foi criada, em 1968, uma Faculdade de Educação, que mais tarde passou a ser a Fundação Universidade de Feira de Santana - FUFs, através da Lei Estadual nº 2.784, de 24 de janeiro de 1970. A Universidade foi instalada e autorizada em 31 de maio de 1976 e tinha os seguintes cursos disponíveis: Licenciatura de 1º e 2º graus em Letras – Inglês/Francês; Licenciatura Plena em Ciências, com habilitação em Matemática e Biologia e em Ciências 1º grau; Licenciatura Plena em Estudos Sociais, com habilitação em Educação Moral e Cívica e em Estudos Sociais 1º grau; e mais os cursos de Enfermagem, Engenharia de Operações – Modalidade Construção Civil, Administração, Economia e Ciências Contábeis (UEFS, Nossa História). A UEFS cresceu muito nesses últimos anos e hoje tem quase dois mil servidores e mais de 10 mil estudantes matriculados de graduação e pós-graduação, totalizando mais de 25 mil alunos já formados pela Instituição. Com estrutura baseada nos três pilares; ensino, pesquisa e extensão, cumpre a sua missão de oferecer os instrumentos necessários para formação do cidadão, não só da cidade, mas também de todo o seu entorno, sendo protagonista no crescimento da cidade desde a sua implantação.

A UEFS oferece 31 (trinta e um cursos), 29 semestrais e 2 anuais (Medicina e Música). Sendo 14 (quatorze) licenciaturas e 17 (dezessete) bacharelados. São oferecidos 9 (nove) cursos de especialização (Lato Sensu) e 27 (vinte e sete) cursos de pós-graduação (Stricto Sensu), 14 (quatorze) mestrados acadêmicos, 6 (seis) doutorados e mestrados profissionais e 4 (quatro) mestrados em rede nacional. A universidade tem ainda 2 (dois) Programas Interinstitucionais, sendo 1 (um) doutorado e 1(um) mestrado e doutorado (UEFS, 2020; UEFS, Cursos Stricto Sensu; UEFS, Cursos Lato Sensu). A referida universidade aderiu ao Sistema de Seleção Unificada - SISU como processo seletivo para os cursos de graduação, oferecendo 1.090 (mil e noventa) vagas no primeiro semestre de cada ano e 1.097 (mil e

noventa e sete) no segundo semestre, totalizando 2.187 (duas mil cento e oitenta e sete) vagas por ano (UEFS, 2020).

A UEFS é hoje uma universidade de destaque na Bahia e no Brasil, oferecendo um ensino gratuito e de qualidade. A universidade tem o total de 9 (nove) departamentos: Departamento de Ciências Exatas - DEXA; Departamento de Ciências Biológicas - DCBIO; Departamento de Letras e Artes - DLA; Departamento de Ciências Humanas e Filosofia - DCHF; Departamento de Tecnologia - DTEC; Departamento de Ciências Sociais Aplicadas - DCIS; Departamento de Física - DFIS; Departamento de Saúde - DSAU e Departamento de Educação - DEDU. Nas nossas redes, cada professor da instituição foi identificado quanto ao Departamento ao qual pertence, com o intuito de quantificar e qualificar a relação entre as áreas (UEFS, Departamentos). No Apêndice A, apresentamos todos os cursos oferecidos pela UEFS e sua respectiva Grande Área. Este relaciona todos os cursos de graduação oferecidos na UEFS e sua respectiva “Grande Área”, de acordo com a CAPES (CAPES, Tabelas de Áreas de Conhecimento/Avaliação).

Os cursos representados na UEFS contemplam todas as nove grandes áreas determinadas pela CAPES conforme visto no Quadro 4 (Áreas e Grandes Áreas do Conhecimento).

Quadro 4. Áreas e Grandes Áreas do Conhecimento

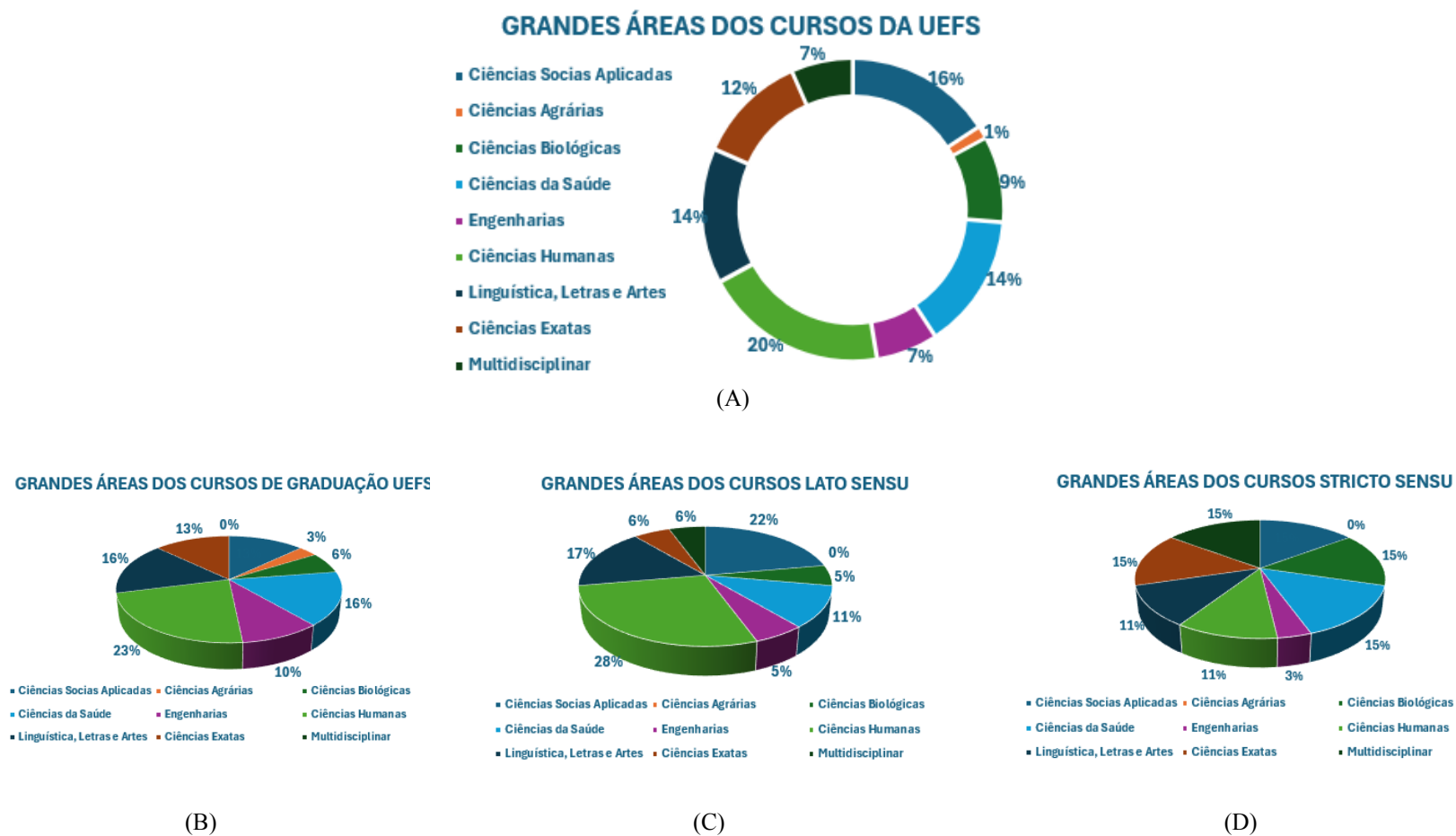
ÁREAS	GRANDE ÁREA
Agronomia. Recursos Florestais e Engenharia Florestal. Engenharia Agrícola. Zootecnia. Recursos Pesqueiros e Engenharia de Pesca. Medicina Veterinária. Ciência e Tecnologia de Alimentos.	Ciências Agrárias
Biologia Geral. Genética. Morfologia. Fisiologia. Bioquímica. Biofísica. Farmacologia. Imunologia. Microbiologia. Parasitologia. Ecologia. Oceanografia. Botânica. Zoologia.	Ciências Biológicas
Medicina. Nutrição. Odontologia. Farmácia. Enfermagem. Saúde Coletiva. Educação Física. Fonoaudiologia. Fisioterapia. Terapia Ocupacional.	Ciências da Saúde
Filosofia. Teologia. Sociologia. Antropologia. Arqueologia. História. Geografia. Psicologia. Educação. Ciência Política.	Ciências Humanas
Direito. Administração. turismo. Economia. Arquitetura e Urbanismo. Desenho Industrial.	Ciências Sociais Aplicadas

Planejamento Urbano e Regional. demografia. Ciência da Informação. Museologia. Comunicação. Serviço Social.	
Linguística, Letras e Artes.	Linguística, Letras e Artes
Matemática. Probabilidade e Estatística. Ciência da Computação. Astronomia. Física. Química. Geociências.	Ciências Exatas e da Terra
Engenharia Civil. Engenharia Sanitária. Engenharia de Transportes. Engenharia de Minas. Engenharia de Materiais e Metalúrgica. Engenharia Química. Engenharia Nuclear. Engenharia Mecânica. Engenharia de Produção. Engenharia Naval e Oceânica. Engenharia Aeroespacial. Engenharia Elétrica. Engenharia Biomédica.	Engenharias
Interdisciplinar. Ensino (de Ciências e Matemática). Materiais (Biotecnologia). Biotecnologia (Ciências Ambientais). Ciências Ambientais.	Multidisciplinar

Fonte: Elaboração própria, baseado em (CAPES, Tabela de Áreas de Conhecimento/Avaliação) (2024).

Sua representação percentual é mostrada nos gráficos da Figura 3 (A). A maior representação é da grande área de Ciências Humanas. Na segunda posição, Ciências Sociais Aplicadas. Em terceiro, empatadas, Ciências da Saúde e Linguística, Letras e Artes. A próxima é Ciências Exatas e da Terra, seguida por Ciências Biológicas. Empatadas, as próximas são Engenharias e Multidisciplinar, e por último, aparece apenas um curso das Ciências Agrárias. Os demais gráficos B, C e D mostram os percentuais dos cursos, respectivamente, de Graduação, Lato Sensu e Stricto Sensu. Observe que não temos curso na grande área “Multidisciplinar” nos cursos de graduação, tal área é encontrada nos cursos Stricto Sensu e Lato Sensu. A grande área de Ciências Humanas lidera no total, nas graduações e nos cursos Lato Sensu, só fica na segunda posição nos cursos Stricto Sensu. Depois de Ciências Humanas, cursos da área das Ciências Sociais Aplicadas tomam a frente na maioria dos cursos com um percentual muito próximo das Ciências da Saúde. Ciências Agrárias fica em último lugar. Vamos entender se esses resultados refletem nas pesquisas da UEFS [Figura 3 (A), (B), (C) e (D)].

Figura 3. Percentual das Grandes Áreas na UEFS



Fonte: Elaboração própria, (2024).

As grandes áreas de conhecimento são importantes para a organização do sistema de ensino e pesquisa no Brasil e ajudam a definir as diretrizes curriculares para cursos de graduação e pós-graduação, permitindo melhor estruturação do conhecimento científico e coordenação eficaz das atividades acadêmicas e de pesquisa, além de orientar a alocação de recursos, bolsas e financiamento para projetos de pesquisa. São usadas como critério para avaliação de desempenho acadêmico e produção científica pelas principais agências de fomento e avaliação no país, como CNPq, CAPES e FINEP. (CAPES, Sobre as áreas de avaliação; CAPES, Tabela de áreas de conhecimento/avaliação).

De acordo com a CAPES (Brasil, 2022), a organização feita das áreas do conhecimento disponibiliza uma hierarquização em quatro níveis (Grande Área; Área; Subárea; Especialidade) e determina 9 grandes áreas do conhecimento, mais amplas, nas quais se distribuem as 49 áreas de avaliação, mais específicas dentro de cada área do conhecimento. Para a CAPES (2022) “estas áreas de avaliação, por sua vez, agrupam áreas básicas (ou áreas do conhecimento), subdivididas em subáreas e especialidades”. As 9 grandes áreas de conhecimento são: 1. Ciências Agrárias; 2. Ciências Biológicas; 3. Ciências da Saúde; 4. Ciências Humanas; 5. Ciências Sociais Aplicadas; 6. Linguística, Letras e Artes; 7. Ciências Exatas e da Terra; 8. Engenharias; e 9. Multidisciplinar. A área “Interdisciplinar” fica dentro da grande área “Multidisciplinar”. Cada uma das áreas determinadas é subdividida em diferentes subáreas, permitindo uma avaliação mais específica dos programas de pós-graduação e pesquisa (CAPES, sobre as áreas de avaliação; CAPES, Tabela de Áreas de Conhecimento/Avaliação).

Nas análises realizadas, por vezes relacionamos professores da UEFS aos seus departamentos da universidade ou a sua grande área/área de conhecimento, dependendo da informação relacionada aos nós. A rede de coautoria tem informações relacionadas ao departamento do professor, da mesma forma para os professores dos grupos de pesquisa. Porém, em relação às redes semânticas, as relações são relacionadas aos grupos de pesquisa, sua grande área ou área que consta no cadastro do CNPq.

3.4.3 Seleção, Coleta e Extração dos dados

Para avaliar a pesquisa e sua disseminação utilizamos duas fontes:

1. Dados selecionados dos currículos lattes de todos os professores que estavam atuando na UEFS, coletados entre junho e outubro de 2021;
2. Dados dos grupos de pesquisa cadastrados no CNPq, suas temáticas e professores pesquisadores da UEFS (CNPQ, Diretório dos Grupos de Pesquisa no Brasil). Tais dados foram coletados em outubro de 2023.

A escolha inicial foi a de trabalhar com duas fontes de informações, as quais apresentaremos no Quadro 5.

Quadro 5. Seleção, coleta e extração dos dados

FONTE	DADOS EXTRAÍDOS	COMO FOI FEITA A EXTRAÇÃO
1. Currículo Lattes	<p>1. Dados Gerais: Áreas do Conhecimento. Nome Completo. Orcid². Nome em Citações Bibliográficas. Texto do Resumo. Endereço Profissional. Nome da Instituição ou Empresa. Grande área do conhecimento. Subárea do conhecimento. Formação acadêmica. Titulação. Universidade de formação. Idioma.</p> <p>2. Referente aos artigos e livros/capítulos de livros):</p> <p>Título. Idioma. Ano. País de Publicação. Autores. Nome completo do autor. Nome para citação. Palavras-chave. Grande área do conhecimento. Subárea do conhecimento.</p>	Script em Python para extração de dados dos currículos lattes e montagem das planilhas com as informações coletadas. (Apêndice - Script 01).
2. Grupos de Pesquisa da IES cadastrados no CNPq	Nomes dos grupos/projetos. Pesquisadores que fazem parte dos grupos e Linhas de Pesquisa.	Seleção manual no site CNPq.

Fonte: Elaboração própria, (2023).

O Currículo Lattes é uma das fontes mais completas de informação sobre acadêmicos no Brasil, e apesar de ser preenchido pelo pesquisador, no caso da nossa pesquisa, todos são professores da instituição, e devem estar comprometidos em declarar apenas a verdade, sob risco de sofrer sanções inerentes ao cargo público. Para complementar as informações sobre colaboração/cooperação científica, adicionamos dados sobre os grupos de pesquisa cadastrados no CNPq, para melhor descrever o panorama de pesquisa da universidade.

3.4.3.1 Aspectos éticos da seleção dos dados

Quando começamos a coletar os dados em 2019, a proposta era que, após identificação dos pesquisadores importantes nas redes, deveríamos entrevistá-los, para tentar conhecer melhor o professor, suas linhas de pesquisas e afinidades. Com o surgimento da pandemia do vírus SARS-COV-2, a pesquisa ficou parada e entendemos que não seria viável realizar a entrevista e aplicar o questionário com os professores, além disso, os currículos baixados em 2019 já estavam defasados. Chegamos a solicitar aprovação para o Comitê de Ética da UEFS, ainda em 2019, mas com as mudanças de plano por conta da pandemia, optamos por modificar o projeto de pesquisa, renunciando às entrevistas e do questionário e usando, inicialmente, as informações dos currículos lattes. Com a decisão sobre a inviabilidade de aplicação de questionário e realização de entrevista, optamos pela busca de mais informações do próprio currículo lattes dos professores e acrescentaríamos outros dados relacionados a grupos de pesquisa da instituição.

Em outubro de 2021, baixamos novamente os currículos lattes dos professores, entendendo que os dados anteriores poderiam estar ultrapassados, e que tal defasagem de tempo poderia empobrecer a pesquisa. Além disso, com a desistência dos questionários, não seria mais necessário a permissão do Comitê de Ética e Pesquisa - CEP, já que não há obrigatoriedade de submissão ao Comitê, sendo que tais dados estão disponíveis ao público. Ainda assim, os nomes ou dados específicos que possam levar ao reconhecimento dos professores não foram mostrados, pois compreendemos que os professores não podem, de forma alguma, ser expostos a qualquer incômodo, ainda que tenhamos que perder alguma informação ou deixar de divulgar algum dado.

A necessidade de obter aprovação de um comitê de ética para o uso de dados de

domínio público vai da natureza da pesquisa e das políticas da instituição à qual o pesquisador está afiliado. Geralmente, o uso de dados de domínio público não requer revisão de comitê de ética, uma vez que esses dados estão disponíveis ao público em geral e não envolvem questões de privacidade, confidencialidade ou riscos significativos para os participantes da pesquisa. Pontuamos aqui algumas questões referentes ao uso de dados de domínio públicos que seguimos à risca:

1. Questões éticas: Caso os dados de domínio público fossem usados em um contexto que levanta questões éticas, como o uso de dados sensíveis ou controversos, poderia ser apropriado buscar orientação de um comitê de ética, o que não acontece no nosso trabalho, onde os dados dos pesquisadores não foram revelados, e todas as informações dos pesquisadores foram codificadas. Portanto, vão aparecer os códigos que representam os professores/pesquisadores, códigos dos artigos e livros/capítulos de livros publicados, e códigos de grupos de pesquisa.
2. Combinação de dados: Caso houvesse intenção de combinar dados de domínio público com outros conjuntos de dados ou informações que não são de domínio público, seria necessário obter aprovação ética para o uso dessa combinação de dados. Na nossa pesquisa, todos os dados utilizados são de domínio público.
3. Políticas da instituição: Na Universidade Federal da Bahia - UFBA, instituição responsável pelo doutorado a qual estamos ligados por este projeto, o Comitê de Ética em Pesquisa/Instituto de Ciências da Saúde CEP/ICS - UFBA é o responsável por essa avaliação. No Brasil, o responsável pelos aspectos éticos que envolvem seres humanos é o sistema CEP/CONEP, ligado ao Conselho Nacional de Saúde - CNS. Esse sistema é dividido em duas partes; a primeira é a Comissão Nacional de Ética e Pesquisa - CONEP, que é a instância máxima de avaliação ética de pesquisa. A segunda é formada pelos Comitês de Ética e Pesquisa - CEPs (Brasil, Conselho Nacional de Saúde). A CEP/ICS, ressalta a questão de número 24 respondida pelo CONEP.

24) Minha pesquisa é uma revisão sistemática com ou sem metanálise ou minha pesquisa utiliza base de dados públicos. Devo dar entrada no CEP? De acordo com o documento “Perguntas e Respostas” elaborado pelo CONEP que pode ser encontrado no site http://conselho.saude.gov.br/web_comissoes/conep/index.html as pesquisas envolvendo apenas dados de domínio público que não identifiquem os participantes da pesquisa, ou apenas revisão bibliográfica, sem envolvimento

de seres humanos, não necessitam aprovação por parte do Sistema CEP/CONEP.

Entretanto, deve-se salientar que, se houver a possibilidade de identificação do sujeito da pesquisa, a pesquisa deverá ser protocolada junto ao CEP. O pesquisador também deve considerar a possibilidade de solicitação de aprovação por CEP por parte de revista científica na submissão de artigo para publicação. Nesse caso, o pesquisador deverá entrar em contato com o CEP previamente para obter maiores esclarecimentos (UFBA, Comitê de Ética e Pesquisa).

4. Uso de dados de maneira ética: mesmo quando se utiliza dados de domínio público, é fundamental garantir que a pesquisa seja conduzida de maneira ética, respeitando princípios éticos gerais, como a honestidade na análise dos dados e a devida atribuição aos criadores originais dos dados. Afirmamos que o nosso trabalho mantém o respeito pela fidelidade dos dados dos pesquisadores, salvando o princípio ético em relação a essas informações, mesmo que haja algum tipo de prejuízo para o resultado das análises.

Entramos em contato com a CEP da UEFS, que nos confirmou que a Resolução CNS 510/16 assegura os projetos que não necessitam passar pelo Comitê. A Resolução CNS 510/2016, do Conselho Nacional de Saúde (CNS), estabelece diretrizes éticas para pesquisas em Ciências Humanas e Sociais no Brasil. Ela foi criada para complementar a Resolução CNS 466/2012, que regula pesquisas envolvendo seres humanos, mas tem um enfoque mais biomédico. Em seu Artigo 1º, §1º a resolução considera pesquisa com seres humanos aquelas que envolvem interação direta com participantes ou uso de informações identificáveis. No seu artigo 1º, §2º a resolução afirma que: “Estão dispensadas da apreciação ética pelo sistema CEP/CONEP as pesquisas que utilizam informações de acesso público, conforme a legislação vigente, e as que utilizam bases de dados cujas informações são agregadas, sem possibilidade de identificação individual”.

Como tanto os dados dos currículos lattes, como os grupos de pesquisa disponível no site do CNPq, são de acesso público (conforme a Lei de Acesso à Informação 12.527/2011) e não há interação direta com os participantes (entrevistas, questionários, observação direta), esta pesquisa dispensa a permissão do CEP.

3.4.3.2 Currículo Lattes

A Plataforma Lattes foi criada pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq, em 1999, e é um sistema que disponibiliza currículos de estudantes e pesquisadores de forma virtual e gratuita. Qualquer pessoa pode cadastrar seu currículo, apesar deste sistema ser mais utilizado por pessoas com envolvimento com a vida acadêmica. Nesse espaço, os pesquisadores podem registrar todas as informações e dados da sua vida acadêmica e/ou profissional. Inicialmente o lattes era apenas utilizado no âmbito do Ministério da Ciência e Tecnologia, mas tem cumprido a missão de disseminar as informações relacionadas às produções intelectuais e atividades do ensino, tornando-se importante via de divulgação de tais atividades/informações, mais do que isso, representa a principal fonte de divulgação de pesquisadores e de suas realizações científicas. Hoje a Plataforma é fonte não apenas para estratégias de planejamento e ação como também para políticas de diferentes órgãos do governo, na área de ciência, tecnologia e ação (Brasil, 2023).

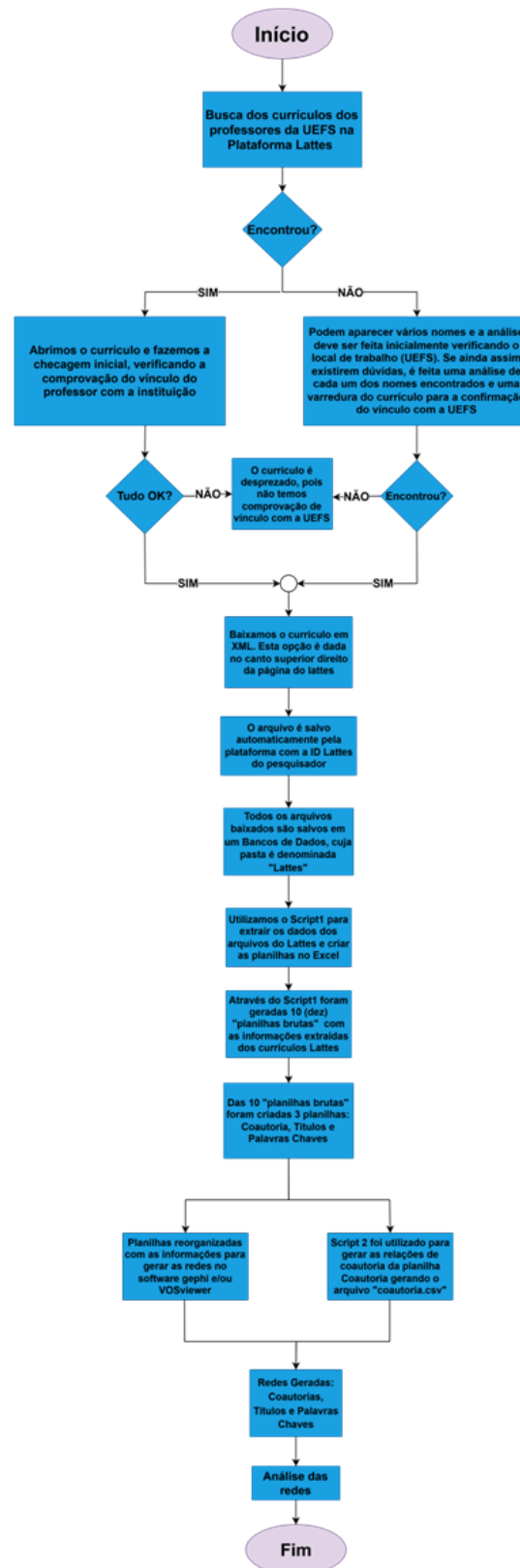
Os dados do lattes são “autorreferidos”, ou seja, declarados pelo dono do currículo, e precisamos reconhecer que existe a possibilidade de alterações, informações errôneas, ou a projeção de uma imagem mais favorável sobre si mesmo. Contudo, todos os currículos, cujos dados foram utilizados neste trabalho, são de professores da UEFS, que passaram por concurso público (efetivos) ou seleção pública (contratados por determinado período). Em ambos os casos, uma banca realizou uma investigação do currículo e documentação para o ingresso na instituição, mas ainda assim, não temos como assegurar 100% (cem por cento) de confiabilidade em dados que são inseridos pelos próprios professores.

Apesar dos dados na Plataforma Lattes serem de domínio público, para consulta de toda e qualquer pessoa, o procedimento de baixar currículo da plataforma lattes, um por um, é bastante trabalhoso. Principalmente com a inserção, em 2015, do sistema “CAPTCHA”, cujo objetivo é de não sobrecarregar os servidores e dificultar a extração automatizada de vários currículos (Matias, 2017).

Para extração dos dados dos currículos lattes, a principal dúvida foi qual seria a melhor e mais rápida maneira de extrair os dados para montagem de planilhas com informações para as Redes. Em busca de extratores que servissem em específico para o nosso problema, não encontramos nenhum satisfatório, pois a extração do nosso trabalho é muito específica, de forma que selecionamos todos os itens do currículo lattes que poderiam servir

para todas as redes que poderiam ser importantes para uma análise da pesquisa científica da UEFS. Diante da dificuldade em encontrar um programa gratuito que servisse para extração dos dados dos currículos nos moldes pretendidos, decidimos utilizar um programa desenvolvido em Python para extrair os dados específicos do nosso Banco de Dados, conforme script 01 (APÊNDICE B). Os currículos salvos após extração do script têm formato XML (Dados semiestruturados em marcadores ou tags). Todo o processo realizado para obtenção das informações dos currículos lattes está representado a seguir, na Figura 4.

Figura 4. Processo de captura e extração de dados dos Currículos Lattes



Fonte: Elaboração própria, (2022).

O número de professores da UEFS em atividade em 2021 eram 936, todos eles foram procurados na plataforma lattes e tiveram seus currículos baixados em XML e salvos automaticamente com o número do ID Lattes atribuído pela própria plataforma em um Banco de Dados “Lattes”. Esse processo foi finalizado em dezembro de 2021.

Para coleta dos dados dos currículos lattes dos professores da IES, percorremos os seguintes passos, também representados na Figura 4:

1. Iniciamos o processo de busca dos nomes dos professores da UEFS no site (<https://lattes.cnpq.br/>) (CNPq, Lattes), da seguinte forma:

Busca pelo nome do professor => Seleção do nome escolhido (caso tenha mais de um, a análise é feita buscando informações iniciais na identificação da página, muitas vezes consta a instituição principal de trabalho e/ou cidade que reside);

2. Ao encontrar o pesquisador, abrimos o currículo e checamos o vínculo com a Instituição;
3. Confirmado o nome, baixamos o currículo em formato XML;
4. Salvamos o arquivo (currículo) numa pasta no google drive, a qual denominamos “Lattes”. Quando as informações não eram suficientes para identificar o professor/pesquisador (alguns casos encontramos muitos nomes muito parecidos, alguns sem identificação do local de trabalho), abrimos o currículo de cada um deles até confirmar qual deles seria professor da UEFS. Seguimos o mesmo procedimento, até encontrar o professor e salvar o seu currículo.

Com o arquivo contendo os currículos salvos em XML na pasta de nome “Lattes” no Google Drive, seguimos os passos para obtenção de novas planilhas:

1. Identificamos (via script) as informações que devem ser capturadas do currículo lattes de cada pesquisador (Ver Quadro 5: Seleção, coleta e extração dos dados);

2. O script 01 (APÊNDICE B) fez a varredura do currículo, selecionando as informações que foram solicitadas. É feita a leitura dos arquivos zipados em XML (baixados de currículos lattes da plataforma através do site da plataforma Lattes) (CNPq, Lattes). Os arquivos são nomeados (salvos) na pasta “Lattes” no google Drive com seu número (ID lattes). Esse número é atribuído pelo sistema lattes mesmo, no momento do salvamento no formato XML.

Com os dados capturados dos currículos lattes dos professores da UEFS, foram geradas 10 planilhas, que, para facilitar a importação e exportação para o Excel e softwares utilizados, foram salvas em csv (comma-separated-values), ou seja, no formato de valores separados por vírgula. À medida que os dados eram tratados, modificamos a extensão para “.xls”, no intuito de facilitar as alterações e organização dos dados. Descrevemos os nomes de cada Tabela, na sua forma bruta, do jeito que os dados foram capturados pelo script inicialmente. E o conteúdo de cada uma das 10 (dez) planilhas encontradas está descrito no Quadro seguinte.

Quadro 6. Planilhas geradas via script

PLANILHA GERADA PELO SCRIPT	O QUE SIGNIFICA
table_pessoa	Tabela com os dados pessoais
table_artigos	Tabela com os títulos dos artigos
table_coautorias	Tabela com todos os coautores
table_artigo_coautoria	Tabela de relação artigo coautoria
table_pessoa_coautor	Tabela de relação de autor e coautoria
table_artigo_keyword	Tabela de relação artigo e palavras-chave
table_keyword	Tabela com palavras-chave
table_art_sub_area_conh	Tabela de relação de artigo com subáreas
table_area_conh	Tabela de área do conhecimento
table_sub_area_conh	Tabela de subárea do conhecimento

Fonte: Elaboração própria, (2023).

Focando na qualidade das publicações que seriam utilizadas na nossa pesquisa, optamos por selecionar do currículo lattes, no que se diz respeito às publicações, apenas os artigos publicados em periódicos, além de livros e capítulos de livros. A decisão levou em conta também que a modalidade artigo de eventos não garante que o trabalho foi publicado e divulgado, já que vai depender da forma de apresentação do trabalho exigida pelo evento. Alguns eventos solicitam apenas apresentação de pôsteres, ou vídeo, outros publicam apenas

um resumo do trabalho nos anais do evento, o que não garante a divulgação do trabalho completo, condição garantida nos periódicos e livros.

Dos 936 professores ativos em 2021, baixamos em formato XML o total de 913 (novecentos e treze). Desses 23 currículos que faltaram, 8 (oito) deles não foram encontrados na plataforma lattes. Atribuímos o fato a que, alguns professores, principalmente os mais antigos na instituição, não aderiram à plataforma, por não sentirem necessidade ou ter alguma dificuldade no seu preenchimento. Os outros 15 (quinze) não conseguimos informações suficientes no currículo, que garantisse que fossem realmente professores da instituição, alguns tinham apenas o resumo escrito, sem nenhuma identificação que confirmasse o vínculo com a UEFS. Outros preencheram muito superficialmente, com pouquíssimos dados, insuficientes para a pesquisa. Isso foi verificado no momento de baixar o currículo em XML, ou seja, quando abríamos o currículo, fazíamos uma leitura inicial uma verificação rápida de cada currículo, no intuito de confirmar se realmente seria o pesquisador que consta na lista de professor da universidade. No total, esses professores representam apenas 2,46% dos currículos, concluímos que é um índice baixo e não afetaria o nosso resultado. Os demais nomes tiveram os dados extraídos do lattes e aparecem na planilha bruta, o que não garante que todos irão aparecer em todas as redes, pois, como alguns não tinham publicações nos moldes que escolhemos para a nossa pesquisa, então o script captura apenas os dados encontrados e cria as “planilhas brutas” preenchendo as informações que constam no currículo. Porém, caso o professor não tenha as publicações nos moldes determinados, este vértice não irá aparecer na rede de coautoria, por exemplo. Isso pôde ser verificado ao fazermos uma busca dos currículos dos pesquisadores, onde confirmamos que nenhum dos professores que “sumiu” de alguma das planilhas, tinha as publicações selecionadas, ratificando o funcionamento do script 01.

As planilhas que foram encontradas estão definidas no Quadro 7. Para auxiliar o trabalho de manipulação dos dados e preservar os dados dos professores, criamos códigos (via script) para algumas informações que apareceriam nas planilhas, conforme o Quadro a seguir:

Quadro 7. Informações sobre os códigos das planilhas dos Currículos Lattes

INFORMAÇÕES QUE CONSTAM NA PLANILHAS	O QUE É?	QUAL A INTENÇÃO DA INFORMAÇÃO?
	Código criado para representar o autor/coautor. Representamos por 5	

Código do Coautor	(cinco) números, seguidos das letras “auc”: 00000auc	Preservar a identidade do autor
ID Artigos/livros	Número gerado para representação dos artigos ou livros/capítulos; são cinco números em ordem crescente da captura do script e mais a identificação do trabalho: 00000ar (artigo) e 00000cl (capítulo de livro ou livro)	Preservar a identificação dos trabalhos
Tipo de Publicação	Identificar o tipo de publicação, se for artigo o código terá as letras (art), se for livro ou capítulo de livro (cl)	Representar na rede os trabalhos de acordo com o seu tipo, para visualização diferenciada dos trabalhos (artigo/ livros/capítulos de livro).
ID Key	Número que identifica o artigo, o código terá 4 (quatro) números e duas letras kw, representando a “key word”, assim: 0000kw	Código para identificar as palavras-chave

Fonte: Elaboração própria, (2022).

Os códigos foram criados basicamente por dois motivos: 1. No intuito de facilitar o manuseio dos dados e 2. Preservar os dados dos pesquisadores, sem a possibilidade de identificação deles.

3.4.3.3 Grupos de Pesquisa UEFS/CNPq

Outra fonte de dados para análise de pesquisa na UEFS foi o Diretório dos Grupos de Pesquisa no Brasil (DGP), seus dados são públicos e é uma iniciativa do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), com a principal função organizar, catalogar e divulgar informações sobre grupos de pesquisa atuantes em instituições brasileiras, principalmente nas áreas de ciência, tecnologia e inovação (CNPQ, Diretório dos Grupos de Pesquisa no Brasil - O que é).

O cadastro no diretório é feito pela instituição interessada, que deve estar cadastrada no Diretório de Instituições (DI) do CNPq e atender a um dos requisitos previstos no art. 5º. da Portaria CNPq 1.513, de 27 de outubro de 2023. Tais requisitos estão relacionados a condições da instituição, como; possuir curso stricto sensu, bolsa de produtividade, professor Doutor com dedicação exclusiva, produção científica ou tecnológica considerada relevante pelo CNPq e Bolsa de Iniciação Científica (PIBIC) (CNPq, Diretório dos Grupos de Pesquisa

no Brasil - Quem pode participar, s.d.).

Os grupos selecionados para o nosso estudo são formados por pesquisadores da UEFS, e grupos/projetos de pesquisa da referida instituição ou de outras, mas que tenham ao menos 1 (um) professor da instituição. Assim:

1. Buscamos os grupos de pesquisa da UEFS cadastrados no CNPq através da plataforma lattes, no link “Diretório dos grupos de pesquisa no Brasil” (DGP). Os dados são referentes aos grupos de pesquisa cadastrados CNPq. Na página do diretório dos grupos selecionamos a opção “Termo de Busca” digitamos “UEFS”. Em “Consultar por” digitamos “Grupos”. Como resultado obtivemos todos os grupos que têm pesquisadores/professores da UEFS, sejam com grupos da própria instituição, seja por grupos de outras instituições, mas que possua pesquisadores/professores da UEFS entre os pesquisadores.
2. Organizamos as informações obtidas em uma planilha chamada “Grupo de Pesquisa”, com as seguintes informações: 1. O nome do grupo de pesquisa cadastrado no CNPq; 2. Nome do professor da UEFS participante do grupo; 3. Grande Área de Pesquisa; 4. Linhas de Pesquisa.

A Figura 5 mostra a tela inicial do DGP:

Figura 5. Tela inicial de consulta do Diretórios dos Grupos de Pesquisa no Brasil (DGP)



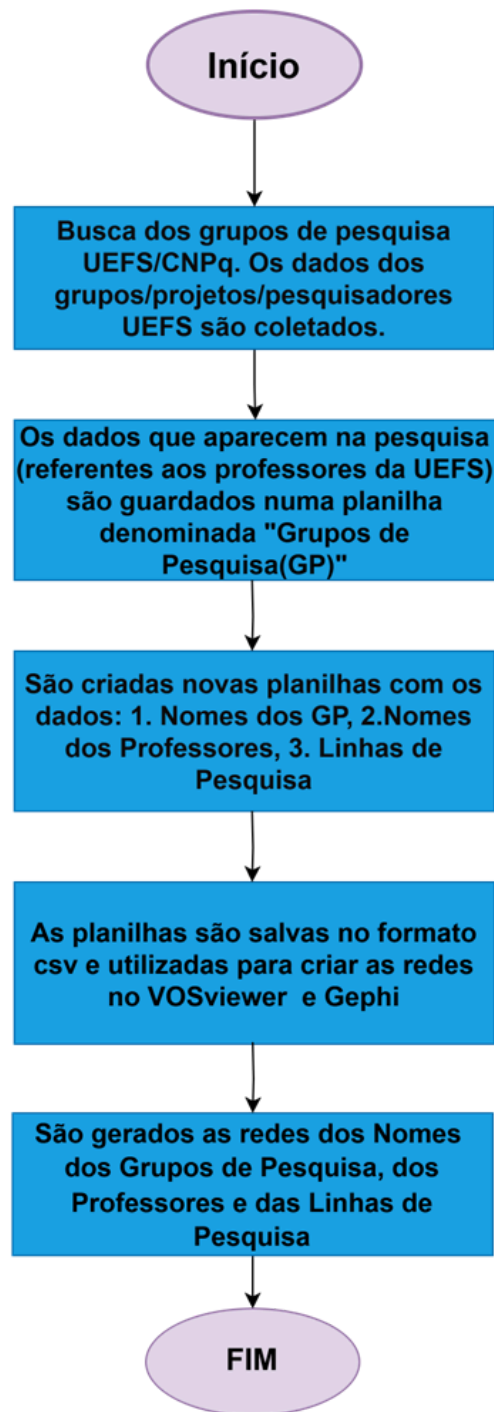
Fonte: CNPq, Diretório dos Grupos de Pesquisa no Brasil - Consulta Parametrizada, s.d).

Os dados dos grupos de pesquisa foram conferidos para verificar o vínculo do pesquisador com a UEFS. Essa checagem é importante porque precisamos garantir os mesmos

códigos de professores definidos na planilha de coautoria. Os grupos/núcleos foram identificados com as letras “gp” (grupo de pesquisa da UEFS) ou “gpe” (grupo de pesquisa externo) enumerando de 01 a 40, representando o número de grupos/núcleos encontrados relacionados a pesquisadores da UEFS, 13 grupos internos e 27 externos.

O processo de captura das informações sobre os grupos/projetos de pesquisa da UEFS, vinculados ao CNPq, é representado pela Figura 6, a seguir:

Figura 6. Processo de captura e extração de dados dos Grupos de Pesquisa da UEFS/CNPq



Fonte: Elaboração própria, (2023).

O Quadro 8 ajuda o entendimento do formato da “planilha bruta” dos grupos de pesquisa, obtida com os dados do diretório dos grupos de pesquisa:

Quadro 8. Planilha dos Grupos de Pesquisa UEFS/CNPq

TÓPICOS DA “PLANILHA BRUTA” DOS GRUPOS DE PESQUISA UEFS/CNPq	INFORMAÇÃO LIGADA AO TÓPICO	INTENÇÃO
Nome do Grupo de Pesquisa	O nome do grupo de pesquisa da UEFS/CNPq ou de outra instituição cujo grupo tenha algum professor/pesquisador da UEFS	Utilizar esta informação para identificar as temáticas/palavras utilizadas nos projetos de pesquisa
Código para o projeto de pesquisa	Introduzir um código formado por “gp” (grupo de pesquisa), mais 2 (dois) números (gp00). Caso o grupo não seja da UEFS, mas tenha algum professor/pesquisador, o grupo foi identificado como “gpe”(grupo de pesquisa externo)	Como os códigos irão aparecer nas redes, os nomes dos professores e dos grupos serão preservados
Departamento/Grande Área do professor ou Área	Departamento e grande área/área na qual está inserida o pesquisador de influência na rede	Identificar a participação dos departamentos/grandes áreas/áreas da UEFS que aparecem nas redes
Professor/Pesquisador da UEFS	Nome do professor/pesquisador que faz parte do projeto	O nome não irá aparecer na rede, mas receberá o mesmo código da planilha de coautoria caso tenha seu nome cadastrado, caso contrário, receberá um novo código
Código do Professor/Pesquisador	Código que identifica o professor da UEFS, que será o mesmo utilizado nas redes referentes aos currículos lattes. Se o professor não tiver código, um código será criado, nos mesmos moldes; cinco números seguidos das letras “auc”	Preservar a identidade dos professores/pesquisadores

Fonte: Elaboração própria, (2023).

As redes geradas pelos dados dos projetos de pesquisa irão identificar a relação entre os professores, grandes áreas ou áreas e/ou departamentos, nomes dos grupos e as linhas de pesquisa dos grupos. Com esses dados e as métricas utilizadas podemos identificar as relações das redes e comparar com outras redes geradas pelos dados dos currículos lattes².

3.4.4 Reprocessamento e Transformação dos dados

3.4.4.1 Currículo Lattes

As planilhas salvas através dos dados dos currículos lattes, após utilização do script 01, apresentaram muitos ruídos, além de informações repetidas que deveriam ser limpas e organizadas. Com estes dados obtivemos as tabelas brutas com captura dos dados via script 01, porém muitos dados estavam truncados, faltando letras, símbolos da linguagem XML, informações repetidas. Então fizemos uma varredura para deixar as informações mais enxutas e organizadas. Essa limpeza leva algum tempo, pois das informações iniciais dos 913 professores tínhamos planilhas de até 89.486 linhas para verificar, contendo, em cada linha da planilha, os nomes não só dos professores da instituição, como também de todos os coautores que escreveram com os professores. Verificamos na conversão para as planilhas, que muitos professores utilizam formas diferentes de apresentar o seu nome no currículo lattes, ou seja, para cada artigo/livro publicado, o (a) professor (a) cita seu nome de uma forma diferente, modificam as iniciais ou, no caso de professoras, muitas mudaram seus nomes adicionando o sobrenome dos cônjuges após o casamento, ou ainda, utilizam citações diferentes, sendo a mesma pessoa, o que é natural, pois geralmente escrevemos com outras pessoas e a cada artigo publicado, pode-se mudar a forma de citação. Essa checagem foi feita analisando as

² Os grupos de pesquisa identificados são os que têm atividades cadastradas no CNPq. Existem na UEFS vários outros grupos de pesquisa relacionados a outras formas de incentivos à pesquisa, além de projetos, núcleos e grupos de pesquisa e extensão que não estão cadastrados pelo CNPq, mas que contabilizam carga horária complementar para estes professores e estudantes, chamadas Atividades Complementares - ACCs. A Lei de Diretrizes e Bases da Educação Nacional (LDB), nº 9.394/1996, no seu artigo 47, menciona a necessidade de incluir atividades acadêmicas complementares nos currículos dos cursos de graduação. Cada universidade tem autonomia para definir quais atividades serão consideradas como complementares, levando em conta suas características, objetivos institucionais e o perfil do curso. Essas atividades podem incluir estágios, participação em eventos acadêmicos, projetos de extensão, atividades de pesquisa etc. Na UEFS, a resolução CONSEPE 54/2001 regulamenta as atividades complementares para graduação na instituição. Outras resoluções complementam as regras referente a cada um dos cursos, de acordo com o que estabelece a Lei Federal nº 9.394/96 e o Parecer nº 776/97 da Câmara de Ensino Superior do Conselho Nacional de Educação - CNE (UEFS, Atividades Complementares; Brasil, 1996).

citações que constam em seus currículos, ou seja; identificamos os dados dos nomes “parecidos” e, com a confirmação de se tratar da mesma pessoa (através do lattes) identificamos as várias citações do mesmo autor. Com a confirmação de se tratar da mesma pessoa, escolhemos uma das citações a ser utilizada em todas as produções, alterando o código do coautor para um só (já que se tratava da mesma pessoa). Com isso reduzimos as planilhas em número de linhas. Em relação ao número de professores/pesquisadores da UEFS, tínhamos começado na Coleta e Extração de Dados com o total de 913 currículos de professores/pesquisadores da UEFS, porém, desse total, apenas 833 tinham dados relacionados a artigos completos publicados e/ou livros ou capítulos de livros, para serem utilizados para as redes de coautoria, títulos e palavras chaves.

Outro fato importante a ser descrito é que verificamos professores que participam de grupos de pesquisa nacionais e internacionais, com um número muito grande de pesquisadores, todos descritos nos trabalhos (artigos). Dessa forma, separamos os seus artigos que tinham um número muito alto de autores, para isso, resolvemos utilizar uma planilha de trabalho com artigos de até 50 autores. Assim, os artigos que tinham mais de 50 autores foram retirados das planilhas utilizadas para traçar as redes e foram alocadas em uma planilha em separado a qual chamada “ArtG”. Dessa forma, o número de autores desses artigos G variam de 51 a 1.925 autores, apresentando uma média de 347,85 autores por artigo. Foram retirados da rede um total de 54 artigos “G” (todos são artigos, nenhum livro ou capítulo de livro), sendo que em 53 deles tem um mesmo pesquisador da UEFS, dentre outros. Observamos que, em alguns artigos “G”, de um a seis professores da UEFS também publicaram os artigos juntos, ou seja, de um a seis professores da UEFS provavelmente fazem parte dos mesmos grupos de pesquisa que produziram alguns dos 54 artigos gigantes. Esses artigos, além de conter um número enorme de autores, ainda apareceram várias vezes nas planilhas, pois cada professor da UEFS tem o artigo no seu Lattes e o script 01 capturou o mesmo artigo/capítulo de livro em cada um desses currículos, nomeando como se fossem diferentes. Fizemos essa correção; identificamos as repetições, através dos títulos que apareciam repetidos e retiramos das planilhas “Coautoria”, “Títulos” e “Palavras-Chave”.

Essa repetição também aparecia em vários outros artigos e/ou livros ou capítulos de livro, onde muitos professores da UEFS publicaram com outros colegas da universidade, e os artigos aparecem o mesmo número de vezes, afetando o resultado na rede, por isso fizemos novamente uma limpeza das repetições, utilizando o Excel.

A planilha de “Títulos” dos artigos iniciais tinha um número total de 15.261 linhas

(títulos). O primeiro passo foi excluir as repetições, título a título, os que são iguais, posteriormente, ano da publicação, tipo de publicação e idioma. Os títulos que apresentavam “repetição” identificamos para comparar aos dados dos currículos e, confirmada a repetição, deletamos os artigos repetidos para otimizar a rede apresentada. Foram retirados também os títulos dos Artigos Gigantes (ArtG) que não fariam mais parte da Tabela de Coautoria. Isso afetou também os dados de coautoria e tivemos que retirar da planilha coautoria todos os artigos repetidos. Com essas alterações e retiradas dos artigos “Gigantes” o número de títulos da planilha “Títulos” chegou a 13.287 títulos distintos e os coautores da planilha “Coautoria” caiu de 89.486 para 49.815 coautores.

Após o processo ETL, as planilhas foram otimizadas, de forma que todas as informações necessárias ficassem em apenas três. Para chegar a este número, selecionamos as informações necessárias para as redes de coautoria, títulos e palavras chaves e acrescentamos dados de várias planilhas em uma só, reduzindo assim o número total de planilhas. As três novas planilhas reprocessadas e tratadas, derivadas dos dados dos currículos lattes receberam os nomes de: 1. Coautoria, 2. Títulos, 3. Palavras-chave. Ressaltamos que todas as planilhas brutas foram guardadas, caso haja necessidade da utilização de algum dado.

3.4.4.2 Grupos de Pesquisa UEFS/CNPq

Foram identificados 40 grupos de pesquisa cadastrados no CNPq, que tem algum professor da UEFS como participante. Construímos a planilha, de forma manual, para que pudesse suprir as necessidades exigidas pela pesquisa, não havendo necessidade de limpeza de dados. As informações dos grupos do CNPq são concisas, diferentemente dos currículos lattes que possuem muitas informações, necessitando do uso de script. Apenas os dados referentes à área/departamento (ao qual o docente pertence) e a confirmação de que eram docentes da UEFS foram retirados da planilha de coautoria para complementar a planilha a partir da qual encontraremos as redes dos nomes dos grupos, dos nomes dos professores/pesquisadores e das linhas de pesquisa de cada um dos grupos de pesquisa.

Mais uma vez analisando o objetivo desta pesquisa, apesar de termos feito uma análise descritiva inicial de todos os grupos, optamos por focar nos grupos internos (que são 13 do total dos 40), pois o número de professores da UEFS nos grupos externos é muito pequeno,

com um ou, no máximo, dois professores da UEFS presentes, o que não iria refletir na conexão interna desses pesquisadores.

3.4.5 Mineração e Interpretação dos Dados

Após a aplicação da última etapa ETL, onde os dados foram organizados no formato adequado para utilização dos softwares, encontramos duas redes sociais: a Rede Mãe (RM) de coautoria, com todos as conexões da rede de pesquisa da UEFS, e a Rede Filha (RF) de coautoria, gerada com os dados dos docentes da UEFS, retirados os nós representados como “EXTERNOS”, então esses dados foram analisados e interpretados. A RF foi obtida, depois de observarmos a necessidade de análise da colaboração dos professores da UEFS.

Além das redes de coautoria, obtivemos as redes semânticas de títulos e palavras-chaves com os dados dos currículos. Nas redes maiores (as de coautoria), utilizamos o software Gephi (Apêndice C) para visualização das redes e análise de redes sociais. A partir daí, foram aplicados algoritmos, por meio dos quais analisamos as redes, encontramos os nós de destaque nas centralidades para identificar os nós estratégicos e fizemos a avaliação da topologia das redes. Nas redes semânticas (títulos e palavras-chave), inicialmente utilizamos o software VOSviewer (Apêndice D), que tem sua visualização focada nos *clusters* (comunidade formada por nós com alta conectividade) da rede, depois, utilizamos o software Gephi para aplicação dos algoritmos e cálculo das métricas. Além das estatísticas e análise da posição dos nós influentes, as redes semânticas foram analisadas quanto ao seu conteúdo e contexto.

Das informações dos grupos de pesquisa da UEFS cadastrados no CNPq foram geradas três redes: uma social, representando os professores que fazem parte dos grupos de pesquisa (com mesmos códigos das redes de autoria do lattes), outras duas semânticas: uma com os nomes dos grupos de pesquisa, outra com as linhas de pesquisa destes grupos. Os passos para chegar aos resultados foram os mesmos seguidos nas respectivas redes do currículo lattes (social e semântica).

3.4.5.1 Teoria dos Grafos como Metodologia de Análise

Para analisar o processo de pesquisa e colaboração científica da UEFS recorremos à metodologia das redes, que descreve abordagens e técnicas utilizadas para o estudo e análise das estruturas da rede que representam sistemas complexos. O nosso modelo também representa um sistema complexo, e para esta construção utilizamos as ferramentas de rede e métodos estatísticos para descrever as relações de pesquisa e colaboração encontradas nas redes obtidas dos dados coletados (Rosen, 2010).

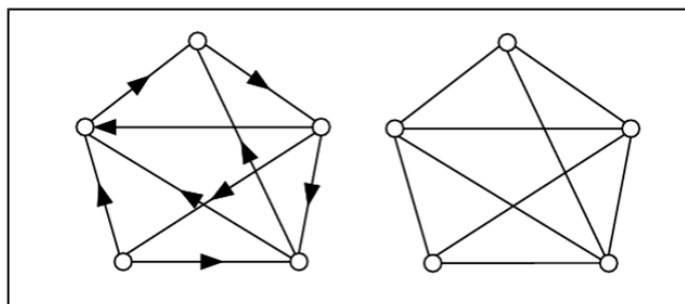
Quando falamos em Ciência das Redes, falamos de conjuntos de objetos conectados, cujas conexões representam as relações entre estes objetos. Para entender estes componentes, iremos defini-los melhor. As redes são construídas como um grafo, ferramenta fundamental na teoria dos grafos. Falamos um pouco sobre a origem dos grafos quando descrevemos o problema das pontes de Königsberg. Agora iremos definir esse componente matemático. Um grafo $G = (V, A)$ consiste em um conjunto não vazio de vértices (ou nós), V e um conjunto não vazio de arestas, A (Rosen, 2010). Essa notação permite a representação de várias situações da vida real. Para formar as redes propostas na nossa pesquisa, devemos identificar os nós e as arestas de cada uma delas. Mas antes, algumas definições de grafos são importantes de enfatizar, para facilitar o entendimento do conceito.

Quando um vértice tem uma aresta conectada a ele mesmo, dizemos que temos um laço. Se um grafo não tem mais de uma aresta, nem laços, dizemos que é um grafo simples. Se um grafo tem arestas múltiplas, chamamos de multi grafo (grafo desenhado por Euler no problema das sete pontes). Outra definição importante é a de grafos orientados, que são grafos cujas arestas têm direção (formando uma seta) cuja direção é determinada pelo vértice de partida para o vértice de chegada (v_1, v_2). Se as arestas não tem direção, os grafos são chamados de não orientados. Podemos direcionar ou não as arestas, dependendo do problema a ser modelado. Por exemplo, numa rota aérea, considerando dois aeroportos distintos como os nós do grafo, caso exista a rota do nó A para o nó B e não exista a rota do B para A, precisamos representar com uma aresta direcionada de A para B, pois a aresta não direcionada estabelece a conexão nas duas vias. Em uma rede de influência, costumamos utilizar grafo direcionado para representar de quem é a influência sobre quem (se o pesquisador A influencia o pesquisador B, logo a direção da aresta será de A para B), o que estabelece um

direcionamento na relação.

Numa rede de coautoria, os autores se relacionam entre si, ou seja, se A publicou com B, então B publicou com A, por isso não é necessário direcionar essas arestas, porém, estes autores podem publicar mais de uma vez (diferentes trabalhos), o que estará representado na rede por arestas múltiplas. No caso das redes propostas nesta pesquisa: de Coautoria, Títulos, Palavras-chave, Nomes dos grupos, Pesquisadores/Grupos e Linhas de Pesquisa, também foram gerados grafos não orientados, pois todos os vértices, sejam eles pesquisadores ou palavras, a direção não importa e sim a conexão. A Figura 7 mostra as representações de grafos orientados e não orientados.

Figura 7. Grafo orientado X Grafo não orientado

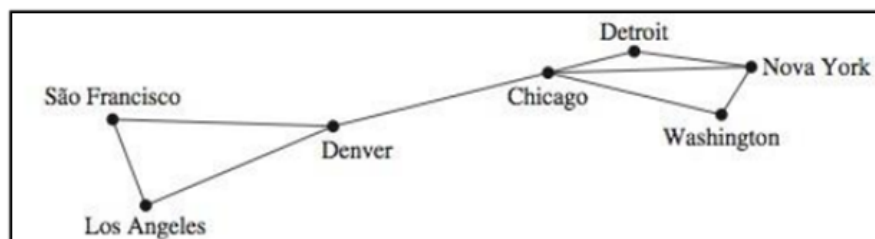


Fonte: Edson Prestes, [s.d.]

A abordagem de rede tem características da interdisciplinaridade e por isso representa tão bem os sistemas complexos, assim, inúmeras relações podem ser representadas como um grafo, seus vértices e arestas. Destas relações formadas pelas conexões, emergem características que podem ser identificadas através das métricas estatísticas. As próximas Figuras representam diferentes situações de grafos, onde todas elas interligam sete cidades dos Estados Unidos que são os centros de dados (vértices) e links de comunicação (arestas). A Figura 8 mostra um grafo simples (sem laços e sem arestas múltiplas). O grau de um vértice é o número de arestas que chegam ou saem desse vértice. Caso o grafo seja orientado, esse grafo terá o grau de entrada e o grau de saída. Na Figura 8, por exemplo, os vértices São Francisco e Los Angeles têm grau 2, Denver tem grau 3, Chicago tem grau 4. Um vértice de grau 0 (zero) é chamado de vértice isolado, pois não tem ligação com nenhum outro vértice. Na Figura 9 temos um grafo não orientado, porém não é um grafo simples, pois tem arestas múltiplas, que neste caso são necessárias para representar os links múltiplos de conexões. A

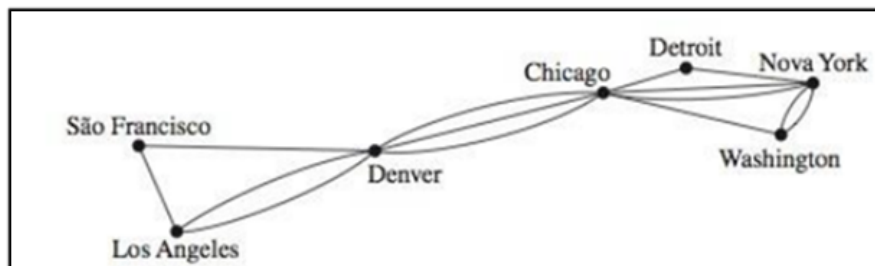
Figura 10 representa a mesma rede, porém, foram acrescentados links de diagnóstico, onde cada nó pode fazer o diagnóstico da sua rede (testes a serem feitos no próprio nó). Para isso, foram adicionados laços (arestas que ligam o mesmo nó) em cada um dos nós da rede. A Figura 11 representa um grafo com a necessidade de direcionar a conexão dessas redes, já que muitas vezes apenas uma direção de conexão importa. Para esta representação utilizamos flechas (arestas orientadas) que representam o sentido da conexão (grafos orientados) (Rosen, 2010).

Figura 8. Uma rede de computadores representada por um grafo simples não orientado



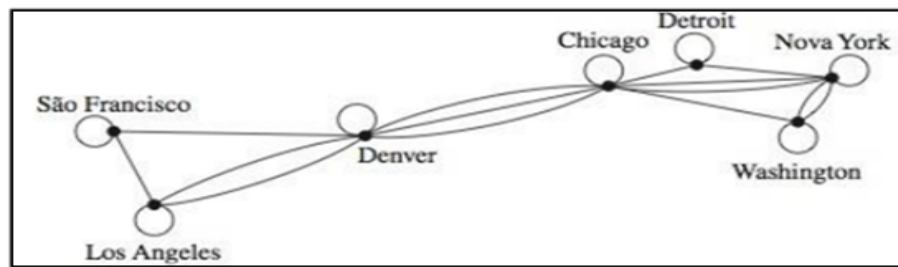
Fonte: (Rosen, 2010, p. 590).

Figura 9. Uma rede de computadores representada por um grafo não orientado com arestas múltiplas e sem laços (Multigrafo)



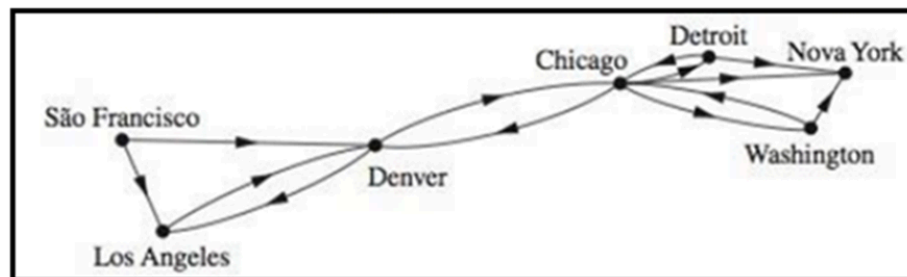
Fonte: (Rosen, 2010, p. 590).

Figura 10. Uma rede de computadores, com laços e arestas múltiplas, representada por grafo não orientado



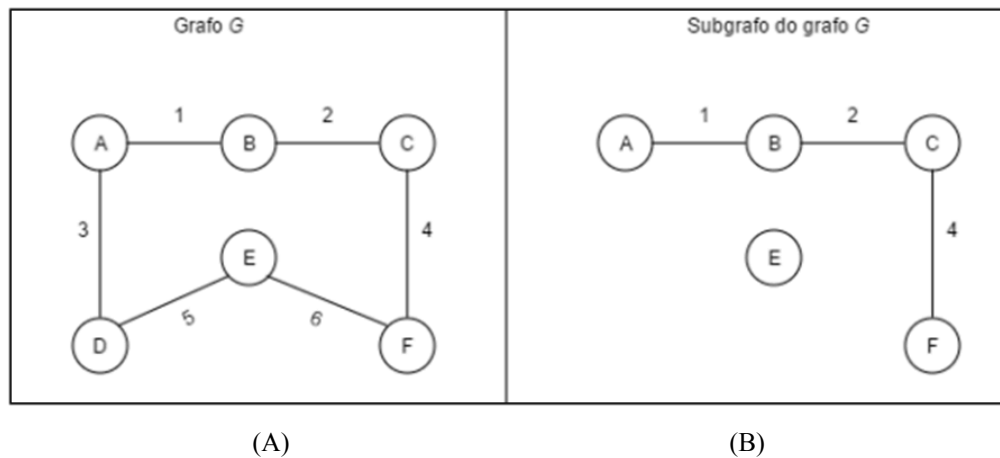
Fonte: (Rosen, 2010, p. 591).

Figura 11. Uma rede de computadores representada por um grafo orientado com arestas múltiplas e sem laços



Fonte: Rosen (2010, p. 591).

Outro conceito importante, antes das análises dos grafos, é o subgrafo. Um subgrafo de um grafo é uma parte de um grafo, ou seja, ele tem um conjunto de vértices e arestas que são subconjuntos do grafo original. A Figura 12 (B) representa um subgrafo do grafo da Figura 12 (A).

Figura 12. Grafo Ponderado G (A) e Sub grafo Ponderado do Grafo G (B)

Fonte: Filho (2017, p. 16).

Um caminho em um grafo é representado pelos vértices visitados, onde v_0, \dots, v_k descreve o caminho de v_0 a v_k . O comprimento de um caminho é determinado contando-se o número de arestas do caminho, se a aresta for utilizada mais de uma vez, ela deve ser contada o número de vezes que for utilizada. Se o caminho começa e termina num mesmo nó, chamamos de ciclo. Um grafo sem ciclo é chamado de acíclico (Rosen, 2010).

A importância de um grafo são as relações existentes nesse grafo, as conexões formadas entre os nós e arestas. Dessa forma, dois grafos podem ser representados por Figuras diferentes, porém, têm as mesmas conexões. Chamamos isso de isomorfismo, ou seja, dois grafos são ditos isomorfos se pudermos encontrar uma correspondência entre nós e arestas dos grafos.

Os grafos são chamados de ponderados quando suas arestas possuem um valor (peso) e são frequentemente utilizados para calcular menores distâncias, custos, tempo etc. Dois problemas bem conhecidos foram resolvidos por algoritmos com a lógica de grafos; o problema do carteiro chinês (PCC), que consiste em buscar o menor caminho. O problema do caixeiro viajante, onde o caixeiro deve visitar n cidades exatamente uma vez voltando para a cidade de partida (mesmo vértice, formando um ciclo) e percorrendo uma distância mínima. Na prática, muitos algoritmos são utilizados para o cálculo de menor caminho, um dos mais conhecidos é o algoritmo de Dijkstra.

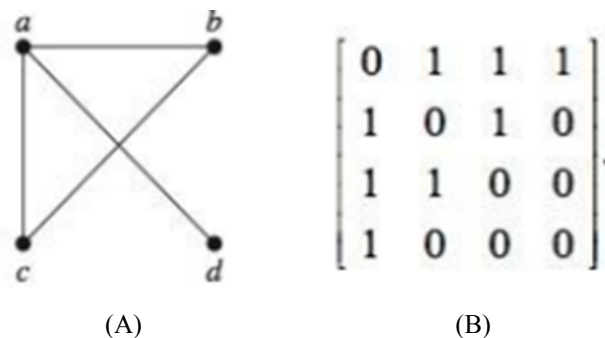
Para representar as conexões entre os nós dos grafos, são muitas vezes utilizadas matrizes. A matriz mais utilizada para essa representação é a Matriz de Adjacência. A Matriz

de Adjacência A do grafo G é formada por zeros (0s) se os vértices v_i e v_j não forem adjacentes e uns (1s), se os vértices v_i e v_j forem adjacentes (Rosen, 2010, p. 612). Dessa forma:

$$a_{ij} = \begin{cases} 0, & \text{se não existe aresta entre } v_i \text{ e } v_j \\ 1, & \text{se existe uma aresta entre } v_i \text{ e } v_j \end{cases}$$

A Figura 13 (A) e (B) representa um grafo e a sua matriz adjacência. Considerando grafos de arestas múltiplas e laços, pode-se ter valores maiores que 1. No caso de grafos orientados, devemos ter a matriz de adjacência de entrada e de saída. Em relação à Figura 13 (A) e (B) têm as matrizes adjacências:

Figura 13. Grafo não orientado (A) e sua Matriz Adjacência (B)



Fonte: Rosen (2010, p. 612-613).

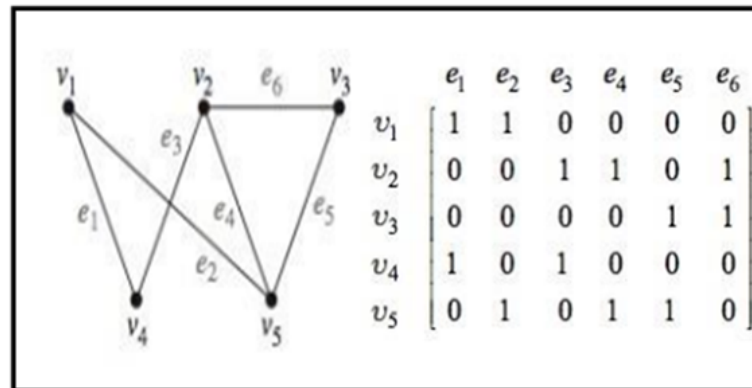
Observe que a representação matricial de um grafo não orientado é sempre simétrica, ou seja, os elementos acima da diagonal principal são iguais aos elementos abaixo da diagonal principal, que funciona como um eixo de simetria refletindo os valores acima e abaixo do eixo.

Uma outra forma de representação de grafos é a Matriz de Incidência. Seja um grafo G com $\{v_1, v_2, \dots, v_n\}$ vértices e $\{a_1, a_2, \dots, a_m\}$ arestas, representando, respectivamente, as colunas e as linhas da matriz de incidência I . Assim:

$$I_{ij} = \begin{cases} 1, & \text{quando existe a aresta } a_{ij} \text{ para o vértice } v_i \\ 0, & \text{caso contrário} \end{cases}$$

A Figura 14 representa um grafo e sua matriz de incidência:

Figura 14. Grafo não orientado e a sua matriz de incidência



Fonte: Rosen, (2010, p. 614).

Da mesma forma que a matriz de adjacência, a matriz de incidência terá valores diferentes de 1, caso o grafo tenha arestas múltiplas e/ou laços.

Há ainda como representar as conexões de um grafo sem arestas múltiplas listando todas as arestas do grafo. Ou ainda, utilizando lista de adjacência, onde, através de uma tabela, são listados todos os vértices do grafo e seus respectivos vértices adjacentes. Porém, as matrizes de adjacências são mais frequentemente utilizadas por possuírem um formato naturalmente simples e usual.

3.4.5.2 Métricas Utilizadas para Análise das Redes Propostas (Grafos não Orientados e não Ponderados)

Utilizamos algumas métricas para análises das redes pesquisadas. Descrevemos cada uma das medidas para análise nas redes propostas nos Quadros 9, 10, 11 e 12.

Quadro 9. Medidas Estatísticas para grafos não orientados (Visão Geral da Rede)

MEDIDA	DESCRIÇÃO	FÓRMULA
<p>Grau Médio</p> <p>(gr_m)</p>	<p>Encontra a média do número de arestas (ou conexões) que cada nó na rede possui, ou seja, a média do número de vizinhos dos nós na rede. Podemos utilizá-lo para entender e comparar diferentes redes em termos de conectividade (Newman, 2003).</p>	<p>É calculado multiplicando o número total de arestas por 2 e dividindo-se pelo número total de nós da rede.</p> $gr_m = \frac{2A}{N}$ <p>Onde:</p> <p>A é o número total de arestas</p> <p>N é o número total de nós</p> <p>g_m é o Grau Médio</p> <p>(Adaptado de Rosen, 2010, p. 598)</p>

<p>Diâmetro da Rede (D)</p>	<p>O diâmetro é definido como a maior distância geodésica (maior distância) entre quaisquer dois nós do grafo, ou seja, é a maior quantidade mínima de arestas que precisa ser percorrida para ir de um nó a outro em toda a rede. É uma métrica global que pode ser útil para entender a eficiência da comunicação ou a propagação de informações através da rede. Esta medida dá uma ideia de quão "grande" ou "espalhada" é a rede em termos de conexões entre os nós (Newman, 2001).</p>	<p>Para determinar o diâmetro de uma rede, é necessário calcular as distâncias geodésicas, que correspondem aos caminhos mais curtos entre todos os pares de nós. O diâmetro é então definido como a maior dessas distâncias (NEWMAN, 2003). O algoritmo de Ulrik Brandes, baseado em busca em largura (<i>Breadth-First Search – BFS</i>), é utilizado para calcular a centralidade de intermediação, e como parte de seu funcionamento, determina as distâncias geodésicas de um nó para todos os outros, podendo ser utilizado também para inferir o diâmetro da rede (BRANDES, 2001).</p> $D = \max_{u,v \in V} d(u,v)$ <p>Onde:</p> <p>D é o diâmetro da rede</p> <p>V é o conjunto de todos os nós da rede</p> <p>$d(u,v)$ é a distância entre o nó u e o nó v.</p>
<p>Densidade do Grafo (d)</p>	<p>A densidade de um grafo mede o quão próximo ele está de ser um grafo completo. Um grafo completo contém todas as arestas possíveis entre seus nós, e, portanto, sua densidade é igual a 1. A densidade é calculada como a razão entre o número de arestas existentes e o número máximo possível de arestas, variando entre 0 (nenhuma conexão) e 1 (totalmente conectado) (NEWMAN, 2010; WASSERMAN; FAUST, 1994).</p>	<p>A densidade de uma rede não direcionada é dada por:</p> $d = \frac{2A}{N(N-1)}$ <p>Onde:</p> <p>d é a densidade</p>

		<p>A é o número total de arestas da rede</p> <p>N é o número total de nós da rede</p> $\frac{N(N-1)}{2}$ <p>Dá o número máximo possível de arestas, onde cada nó (N) está conectado a todos os outros (N-1).</p> <p>$N(N-1) \Rightarrow$ Princípio Multiplicativo que dá o máximo possível de conexões.</p>
Componentes Conectados (Cc)	<p>Dizer que os componentes são conectados significa que o cálculo levará em consideração se há algum caminho entre os nós, não necessariamente uma aresta direta entre eles, ou seja, a ligação entre um nó <i>a</i> e um nó <i>b</i> pode ser através de um outro nó, não precisa haver uma aresta diretamente entre <i>a</i> e <i>b</i> (Rosen, 2010, p. 626). A estatística de "Componentes Conectados" é usada para identificar e analisar sub-redes dentro de uma rede maior. Um componente conectado é um subgrafo no qual qualquer par de nós está conectado por um caminho.</p>	<p>O cálculo dos componentes conectados é encontrado utilizando um algoritmo de Busca em Profundidade (Depth-First Search, DFS). O Algoritmo de Robert Tarjan explora cada vizinho de um nó antes de voltar e explorar os vizinhos de vizinhos, e assim por diante. Após explorar todos os nós, o contador de componentes dará o número total de componentes conectados na rede. Cada grupo de nós marcados durante uma busca representa um componente conectado (TARJAN, 1972).</p>

Fonte: Elaboração própria, (2024).

Quadro 10. Medidas Estatísticas para grafos não orientados (Detecção de Comunidades)

MEDIDA	DESCRIÇÃO	FÓRMULA
Modularidade (M)	<p>A modularidade é uma métrica que avalia a densidade das conexões dentro das comunidades comparada com a densidade das conexões entre diferentes comunidades. Mede a qualidade da divisão da rede em comunidades. Seu valor varia entre -0.5 e 1, porém, nas redes reais, o comum é encontrarmos valores entre 0 e 1. Valores positivos próximos de 1 indicam fortes divisões em comunidades. Tal medida é encontrada por algoritmos de detecção de comunidade. A opção “aleatório” é uma escolha do tipo de algoritmo, e, nesse caso, produz uma decomposição melhor, mas aumenta o tempo de processamento (Newman & Girvan, 2004).</p>	<p>Newman & Girvan (2004) é um dos principais trabalhos que introduz a métrica de modularidade. O artigo estudou algoritmos de estrutura de rede e define um método para detectar a estrutura de comunidade em redes complexas, bem como uma métrica para avaliar a qualidade dessa estrutura. A formalização encontrada por Newman e Girvan é baseada na comparação da densidade de arestas dentro e entre comunidades e dada por (fórmula adaptada pelos autores):</p> $M = \frac{1}{2a} \sum_{ij} \left(A_{ij} - \frac{g_i g_j}{2a} \right) f(c_i, c_j)$ <p>Onde:</p> <p>M é a Modularidade</p> <p>a é o número total de arestas da rede</p> <p>i,j são os nós da rede</p> <p>A_{ij} é o elemento da matriz adjacência da rede. Será igual a 1, se houver aresta entre os nós i e j e 0, caso contrário</p> <p>g_i, g_j São os graus dos nós i e j, respectivamente</p> <p>$f(c_i, c_j)$ que retorna ao valor 1 se os nós pertencem à mesma comunidade ($c_i=c_j$) e 0 caso contrário.</p>

Fonte: Elaboração própria, (2024).

Quadro 11. Medidas Estatísticas para grafos não orientados (Visão Geral dos Nós)

MEDIDA	DESCRIÇÃO	FÓRMULA
<p align="center">Coefficiente de Aglomeração (Coefficiente de <i>Clustering</i> Médio)</p> <p align="center">(CC_m)</p>	<p>Mede a tendência dos nós de uma rede se agruparem em <i>clusters</i> ou "aglomerados". Em outras palavras, ele quantifica o grau de coesão local na rede, indicando a probabilidade de que os vizinhos de um nó também sejam vizinhos entre si. Um coeficiente de aglomeração alto indica que há muitos triângulos na rede, ou seja, os vizinhos dos nós tendem a ser vizinhos uns dos outros. Isso sugere uma estrutura altamente coesa com muitos <i>clusters</i> locais. O coeficiente de aglomeração baixo indica que há poucos triângulos na rede, isso sugere uma estrutura mais esparsa e menos coesa. Duncan J. Watts e Steven H. Strogatz introduziram o coeficiente de <i>clustering</i> em seu artigo seminal sobre "small-world networks," que explora a tendência dos nós de formar <i>clusters</i> (Watts & Strogatz, 1998).</p>	<p>O coeficiente de aglomeração de um nó é definido como a proporção entre o número de triângulos (ciclos de comprimento 3) que passam por esse nó e o número total de tríades (conjuntos de três nós que são conectados por pelo menos duas arestas) possíveis com esse nó. Formalmente, o coeficiente de aglomeração é dado por:</p> $CC_v = \frac{2 \cdot (\text{número de triângulos através de } v)}{g_v(g_v - 1)}$ <p>Onde:</p> <p>CC_v é o Coeficiente de Aglomeração</p> <p>g_v é o grau do nó v</p> <p>$\frac{g_v(g_v - 1)}{2}$ Dá o número de pares de vizinhos que v pode ter (número de tríades possíveis)</p> <p>Assim, o Coeficiente de Aglomeração será:</p> $CC_m = \frac{1}{n} \sum_{v=1}^n CC_v$ <p>Onde:</p>

		CC_m é o Coeficiente de Aglomeração (ou de <i>Clustering</i> Médio) n é o número total de nós da rede
--	--	--

Fonte: Elaboração própria, (2024)

Quadro 12. Medidas Estatísticas para grafos não orientados (Visão Geral das Arestas)

MEDIDA	DESCRIÇÃO	FÓRMULA
<div>Comprimento Médio de Caminho</div> <div>(L)</div>	<p>É a média das distâncias mais curtas entre todos os pares de nós na rede. Watts e Steven Strogatz (1998) descreveram redes que combinam alta <i>clusterização</i> (como redes regulares) com curtos caminhos médios (como redes aleatórias), criando um modelo intermediário — chamado de “rede small-world”.</p>	<p>Calcula a distância mais curta entre todos os pares de nós na rede. Após calcular as distâncias, o algoritmo computa a média dessas distâncias, resultando no comprimento do caminho médio. O diâmetro da rede, que também pode ser calculado, é a maior distância mais curta entre qualquer par de nós na rede. Ele representa o caminho mais longo em termos de número de arestas que precisa ser percorrido entre dois nós quaisquer. Assim:</p> <div>$L = \frac{1}{n(n-1)} \sum_{i \neq j} d(i,j)$</div> <p>Onde:</p> <p>L é o Comprimento Médio</p> <p>n é o número total de nós da rede</p> <p>d(i,j) é a distância mais curta entre os nós i e j.</p> <p>Assim como o diâmetro da rede, o cálculo é feito a partir das distâncias mais curtas entre os nós, sendo que o comprimento médio do caminho oferece uma visão geral da conectividade da rede.</p>

Fonte: Elaboração própria, (2024).

3.4.5.3 Análise Topológica da Rede

Sabendo que as redes modeladas seguem as redes reais, algumas análises podem ser feitas para caracterizar o comportamento dessas redes. Para isso, usamos topologias para atribuir as características às redes complexas; *small world* (mundo pequeno), aleatórias (Watts & Strogatz, 1998) e Livre de Escala (BARABÁSI, 2003). As redes aleatórias “simulam” um sistema aleatório, onde os nós são conectados aos vértices de forma aleatória. Tal rede foi criada no software Gephi, com o mesmo número de nós e densidade aproximada a da rede a ser analisada, mediante uma probabilidade p para o nível de aleatoriedade. Assim, cada nó é conectado aos seus vizinhos mais próximos, em uma estrutura circular, então algumas arestas são reconectadas de forma aleatória com probabilidade p , que controla o quanto a rede é transformada de regular para aleatória (Watts & Strogatz, 1998). Redes mundo pequeno modelam o efeito mundo pequeno, assim, a distância média entre os nós é pequena, o que as torna eficiente para comunicação da rede. Já as redes livres de escala, tem como característica muitos nós conectados a poucos, e poucos nós ligados a muitos outros nós, o que faz com que novos nós da rede se conectem a nós com alto grau (muitas conexões) (Barabási & Albert, 1999). Tais redes obedecem a lei de potência e, com uma análise de distribuição de graus, frequência e gráfico de dispersão, podemos analisar esse comportamento.

As comparações foram feitas a partir de algumas das métricas definidas no Quadros 10 (distribuição de graus), Quadro 11 (coeficiente de aglomeração) e Quadro 12 (comprimento médio). Fizemos a análise topológica das redes e identificamos as características tipo mundo pequeno, seguindo o modelo de Watts e Strogatz (1998) e avaliamos as características de uma rede Livre de Escala (Barabási & Albert, 1999). A rede pode ser caracterizada como mundo pequeno se o comprimento médio do caminho (L) da rede é comparável ao de uma rede aleatória de mesmo número de nós e arestas e o coeficiente de aglomeração (CC_m) da rede é significativamente maior do que o de uma rede aleatória (gerada no software Gephi).

$$CC_{m_{rede}} \geq C_{aleatória} \quad e \quad L_{rede} \approx L_{aleatória}$$

A rede é Livre de Escala se a distribuição de grau segue uma distribuição de lei de potência $P(k) \approx k^{-\gamma}$, onde k é o grau e γ (gama) é o expoente da lei de potência. (Barabási & Albert, 1999). Newton (2005) sugere plotar o gráfico Log-Log da distribuição de graus e sua

frequência, encontrar a reta de regressão aproximada, e o ajuste do gráfico a uma reta decrescente sugere uma lei de potência.

Na mesma equação da reta linear, encontramos o valor de R^2 (coeficiente de determinação), que indica a proporção da variação dos dados. Num contexto de regressão, o R^2 representa o quão os dados se ajustam ao modelo da reta de regressão linear, variando de 0 a 1.

3.4.5.4 Análise das Centralidades da Rede

Além da análise topológica, fizemos um estudo relacionado à centralidade dos nós da rede para identificar o papel dos nós e suas relações na rede. Para essa análise, utilizamos três das medidas de centralidade do Laboratório de Dados do Gephi; através dos cálculos das Centralidade de Grau (CGr), Centralidade de Proximidade (CP) e Centralidade de Intermediação (CI) (Freeman, 1978, 1979; Wasserman & Faust, 1994; Girvan & Newman, 2004). As centralidades são ferramentas fundamentais para a análise de redes, fornecendo informações sobre o papel e a importância dos nós e suas relações na rede. O Quadro 13 mostra a importância de cada medida de centralidade para o estudo das redes sociais e semânticas.

Quadro 13. Métricas utilizadas para análise das redes

MÉTRICA	POR QUÊ?
Centralidade de Grau	Mede o número de conexões diretas de um nó, útil para identificar nós com muitas conexões na rede.
Centralidade de Proximidade	Mede o quão perto o nó está dos outros nós, com base na soma das distâncias geodésicas (os caminhos mais curtos) entre ele e todos os outros nós.
Centralidade de Intermediação	Avalia quantas vezes um nó funciona como ponte entre outros nós, sendo útil para identificar os nós que conectam diferentes <i>clusters</i> .

Fonte: Elaboração própria, baseado nos conceitos de (Freeman, 1978, 1979; Wasserman & Faust, 1994; Barabási, 2016), (2024).

3.4.6 Validação e Conhecimento

Para validar o nosso modelo e obter os resultados necessários para construção do conhecimento pretendido, seguimos todos os passos descritos neste percurso utilizando os dados relacionados à pesquisa científica dos professores da Universidade Estadual de Feira de Santana (UEFS). Tal estudo está detalhado no próximo tópico deste trabalho.

4. ESTUDO DAS REDES DA UNIVERSIDADE ESTADUAL DE FEIRA DE SANTANA

Para construção do nosso modelo de análise de pesquisa e colaboração científica na UEFS, definimos cada uma das redes de pesquisa propostas para serem interpretadas e validadas. Mas por que utilizar redes complexas? A representação de situações reais por redes é cada vez mais utilizada. Castells (2006) explica em seu livro “Sociedade em Rede” como as redes moldam as dinâmicas sociais, econômicas e culturais no contexto das tecnologias da informação e da globalização defendendo que o poder não estaria apenas na posição hierárquica de um indivíduo ou organização, mas na capacidade de influenciar e criar conexões dentro das redes (Leydesdorff, 2011).

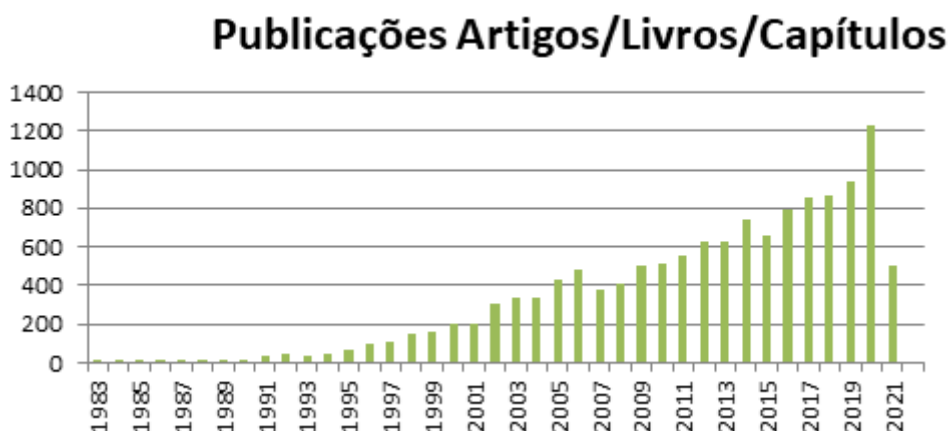
A análise de colaboração e disseminação partiu das interpretações de redes sociais e semânticas, buscando compreender essas conexões num contexto multi e interdisciplinar da comunidade científica da UEFS (Leydesdorff, 2011; Fortunato, 2016; Recuero, 2017; Wasserman & Faust, 1994; Guedes, 2010; Watts & Strogatz, 1998; Barabási & Albert, 1999; Freeman, 1979; Newman, 2001; Newman & Girvan, 2004; Fazenda, 1994; Japiassu, 1976; Pombo 2006). A representação das redes e seus dados qualitativos e quantitativos devem responder aos objetivos desta pesquisa, que são: 1. Mapear a rede de pesquisadores da UEFS para compreender o processo de colaboração entre os professores/pesquisadores; 2. Propor possibilidades de aumentar a colaboração, diante da posição dos pesquisadores nas redes e/ou grupos de pesquisas; 3. Compreender como os professores e suas temáticas se posicionam e se conectam nas redes de pesquisas da UEFS; 4. Identificar (por códigos) os atores das comunidades que colaboram na rede, e seus departamentos na UEFS; 5. Identificar temáticas de destaque na rede e/ou que são comuns a professores de grupos distintos.

Para as análises qualitativas utilizamos uma abordagem interpretativa indutiva e iterativa para compreensão do contexto acadêmico em relação às redes (Chamaz, 2009), além do conjunto de técnicas de análise de conteúdo para interpretação e categorização temática das redes semânticas; títulos/palavras-chaves/nomes dos grupos/ linhas de pesquisa (Bardin, 2011). Para a modelagem utilizamos a teoria dos grafos, por meio dos atores das redes (nós) e suas conexões (arestas) e suas relações. Iremos interpretar e avaliar as redes científicas no contexto social e colaborativo (Freeman, 1979; Newman, 2011; Watts & Strogatz, 1998; Hanneman *et al*, 2005). Numa abordagem qualitativa e quantitativa estrutural, utilizamos a

análise de redes sociais e complexas para entendimento dos atores, suas colaborações, nós estratégicos e estruturas das redes de pesquisa relacionados às métricas, informações quantitativas, estatísticas e padrões sociais e semânticos (Hanneman *et al*, 2005; Recuero, 2017). A primeira fonte de informações das redes foram os dados coletados dos currículos lattes dos professores, que foram identificados por pertencerem ao quadro de docentes da UEFS. Dentre as informações retiradas dos currículos lattes desses professores estão as publicações dos artigos completos, livros ou capítulos de livros. Destas informações geramos as seguintes redes: de coautoria, títulos e palavras-chave.

A quantidade de publicações do quadro docente está representada pelo gráfico da Figura 15, onde é possível observar um constante crescimento no decorrer dos anos, que atingiu um número máximo de publicações no ano de 2020. O ano da primeira publicação dos currículos foi em 1983, e como os currículos foram baixados no fim do ano de 2021, podemos ter publicações de 1983 até 2021 (incompleto). Em março de 2020 aconteceu a pandemia da COVID-19, o aumento pode ter sido fruto desse período, quando os professores ficaram mais em casa, sem aulas ou com aulas virtuais, colaborando para estudos e publicações de trabalhos. Essa tendência de aumento de publicações no Brasil no período da pandemia foi confirmada pela SciELO, que registrou aumento de 23% na submissão de artigos no ano de 2020 e afirmou que a produção científica sobre COVID-19 afetou o equilíbrio da geração de conhecimento. Foram identificados 500 mil estudos sobre a pandemia, em revistas e repositórios (Revista Fapesp, 2022).

Figura 15. Quantidade de trabalhos publicados na UEFS por ano



Fonte: Elaboração própria, (2024).

Os dados dos grupos de pesquisa cadastrados no CNPq foram coletados em outubro de 2023 para gerar a rede de professores dos grupos de pesquisa, a rede dos nomes dos grupos e a rede das linhas de pesquisa. As datas de coleta variam entre as fontes de dados, pois a ideia de utilizar esses grupos surgiu algum tempo depois da obtenção dos dados do Lattes. Porém, o processo de cadastramento dos grupos é bem diferente dos currículos, pois os grupos são cadastrados muitos anos antes e se perpetuam por muito tempo, sendo alimentados por novos trabalhos. Dessa forma, exceto um dos grupos de pesquisa tem data de cadastro em 2022 (período totalmente fora dos dados dos currículos), outro de 2021 (período parcialmente incluído nos currículos), todos os demais (84,62%) foram cadastrados anteriormente a esta data, percentual considerável em termos de comparação com os dados da rede de coautoria, visto que, grande parte das publicações de 2021 também foi considerada, aumentando este percentual.

4.1 REDES SOCIAIS QUE FAZEM PARTE DO MODELO

Redes sociais podem ser entendidas como estruturas compostas por atores; pessoas, grupos ou entidades conectadas por relações (arestas). Através da ARS (análise de redes sociais), buscamos compreender as interações e analisar as características que emergem dessas conexões (Wasserman & Faust, 1994; Freeman, 1979). Nosso modelo propõe a análise e interpretação das seguintes redes sociais: rede de coautoria mãe, rede de coautoria filha (ambas com os dados dos currículos lattes) e rede de professores dos grupos de pesquisa cadastrados no CNPq. Ambas as redes (RM e RF) foram geradas a partir de planilhas encontradas através do script 02 (Apêndice E).

4.1.1 Rede de Coautoria Mãe (RM)

No intuito de mapear a rede de professores/pesquisadores da UEFS, que cumpriram a condição inicial de publicação (de artigos completos, livros e/ou capítulos de livros), e para

compreender o processo de colaboração entre eles, a primeira rede gerada para análise é uma rede de coautoria, onde o nó representa o autor, neste caso o professor, e a aresta representa a conexão entre dois professores, caso eles compartilhem a autoria de uma publicação. (Barabási, 2016; Borgatti; 2018; Wasserman & Faust; 1994). Inicialmente analisaremos a rede de todos os nós (professores/pesquisadores) que publicaram com professores da UEFS, a qual chamamos de Rede Mãe (RM).

A Rede “Mãe” (RM) foi obtida após a organização dos dados de professores da UEFS que publicaram juntos (artigos, livros ou capítulos de livros). Analisamos as redes obtidas pelas informações da planilha coautoria. Na Figura 16 é exibido um “recorte” da planilha “Coautoria”. Os números da primeira coluna representam a numeração dada para cada linha. Os números que faltam na sequência foram retirados por formarem repetições de mesmas publicações ou retirada dos “artigos gigantes” que não entraram na planilha da RM por terem mais de 50 autores. A segunda coluna (id_artigo) enumera cada uma das publicações. A terceira coluna (id_coautor) dá um código para cada autor. A quarta e quinta coluna tiveram seu conteúdo subtraído nesta Figura pois identificam, respectivamente, os nomes na forma de citação de cada autor e seu nome completo. A sexta coluna identifica se o autor pertence ao quadro de professores da UEFS ou se é um membro “EXTERNO” (de fora do quadro de professores da UEFS). A sétima coluna identifica o departamento ao qual o professor pertence (caso esteja no quadro de professores da UEFS), caso contrário, identificamos como “EXTERNO”. A oitava e última coluna apresenta o ano da publicação. Essa é a planilha “Coautoria RM”, da qual foram extraídos os dados para montar a planilha “.csv” para utilização no software Gephi. As relações de coautoria foram obtidas por meio do script 02 (Apêndice E).

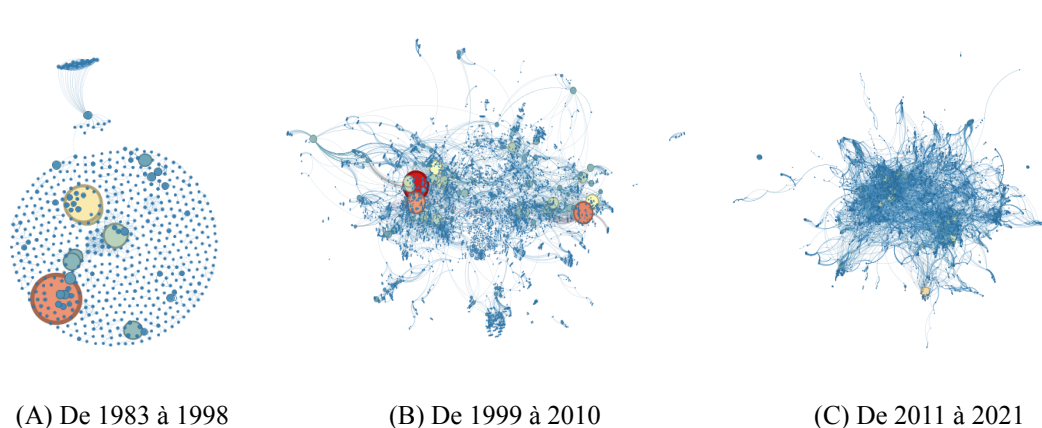
Figura 16. Recorte da planilha de coautoria

Coautoria	id_artigo	id_coautor	nome/citação	Nome Completo	MEMBRO	DEPTO	ANO
1	00001ar	0001auc	SILVA JR, E.	Edgar Silva Junior	UEFS	DTEC	2012
2	00001ar	0002auc	CARRIO, G. A.	Gilberto Arantes Carrão	EXTERNO	EXTERNO	2012
4	00002ar	0001auc	SILVA JR, E.	Edgar Silva Junior	UEFS	DTEC	2018
5	00002ar	0002auc	CARRIO, G. A.	Gilberto Arantes Carrão	EXTERNO	EXTERNO	2018
3	00002ar	0003auc	OSTA SANTOS, RAFAEL	Rafael Costa Santos	UEFS	DEXA	2018
6	00003ar	0001auc	SILVA JR, E.	Edgar Silva Junior	UEFS	DTEC	2005
7	00003ar	0002auc	CARRIO, G. A.	Gilberto Arantes Carrão	EXTERNO	EXTERNO	2005
8	00004ar	0008auc	ANDRADE, C. B. de.	Carla Borges de Andrade	UEFS	DSAU	2012
14	00006ar	0008auc	ANDRADE, C. B. de.	Carla Borges de Andrade	UEFS	DSAU	2013
15	00006ar	0013auc	RODRIGUES, A. B.	Aline Brito Rodrigues	EXTERNO	EXTERNO	2013
16	00006ar	0014auc	PEREIRA, L. L.	LEIAN LIMA PEREIRA	EXTERNO	EXTERNO	2013
17	00007cl	0008auc	ANDRADE, C. B. de.	Carla Borges de Andrade	UEFS	DSAU	2020

Fonte: Elaboração própria, (2024).

Com 16.482 nós e 124.861, a rede tem 66,75% de pesquisadores de outras instituições. Esses pesquisadores publicaram com os 33,25% dos professores/pesquisadores da Universidade Estadual de Feira de Santana (UEFS). Os números mostram a grande quantidade de conexões com pesquisadores externos, que não pertencem ao quadro de professores da instituição. A Figura 17 representa a evolução da RM entre 1983 até 2021, onde a primeira rede (A) mostra as conexões iniciais da UEFS, é nela que aparecem os maiores vértices, indicando que os autores com mais conexões estão nesse período, ou seja, são professores que estão na UEFS desde o início da rede de pesquisa e continuaram publicando. A segunda rede (B) ainda apresenta professores em destaque, também indicando professores com mais tempo na instituição e que continuaram publicando nos anos subsequentes. Por último, a (C) mostra uma rede mais densa, porém, sem nós de grande destaque, indicando que a maioria dos autores começaram a publicar nesse período e por isso têm graus mais próximos.

Figura 17. As conexões da Rede “Mãe” (RM) nos intervalos de tempo A, B e C

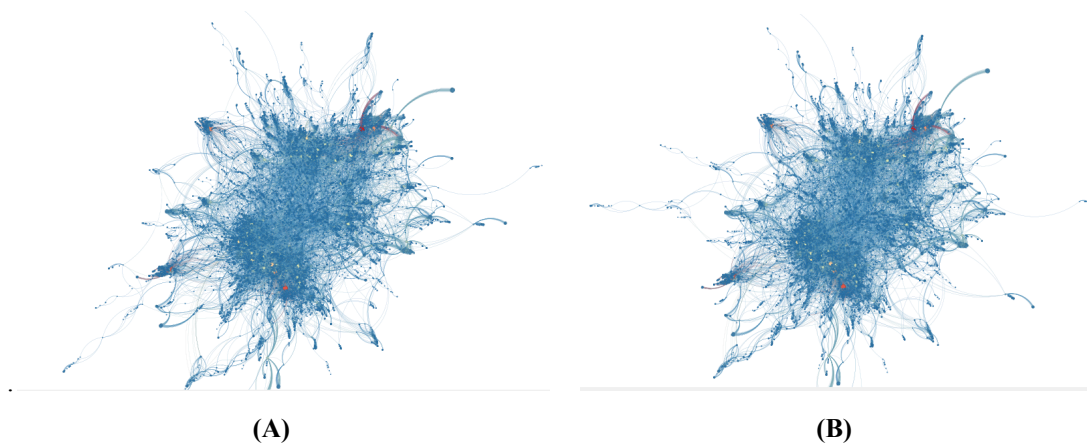


Fonte: Elaboração própria, (2024).

Apresentaremos os dados da RM junto com os dados da componente gigante (CG). A CG de uma rede representa a maior componente conectada da rede, ou seja, um subgrafo da rede em que todos os nós estão diretamente ou indiretamente conectados entre si por meio de caminhos (arestas). A componente gigante da rede de coautoria tem um total de 15.753 nós (95,58% da RM), e 122.557 arestas, ou seja, 98,15% em relação à RM, indicando que a rede está bem conectada, já que a grande maioria dos nós e arestas da componente gigante está na

rede original (com apenas 4,42% dos nós estão fora da CG). As redes geradas são apresentadas na Figura 18 (A) e (B), representam a RM e sua CG.

Figura 18. (A) Rede “Mãe” (RM) e (B) Sua Componente Gigante (CGRM)



Fonte: Elaboração própria com auxílio do software Gephi, (2024).

Para análise da RM, encontramos as estatísticas (conforme descrições dos Quadros 10, 11, 12 e 13), que estão representadas na Tabela 1.

Tabela 1. Estatísticas da Rede “Mãe” (RM), sua Componente Gigante (CGRM)

ESTATÍSTICAS	RM	CG
Nós	16.482	15.753 (95,58%)
Arestas	124.861	122.557 (98,15%)
Grau Médio	15,151	15,56
Diâmetro da Rede	14	14
Densidade do Grafo	0,001	0,001
Componentes Conectados	159	1
Modularidade	0,869	0,864
Coefficiente de Aglomeração	0,856	0,855
Comprimento Médio do Caminho	4,769	4,769

Fonte: Elaboração própria, (2024).

O número médio de conexões por professor/pesquisador é bem próximo para as duas redes (em torno de 15), indica o número médio de conexões de cada professor/pesquisador. O diâmetro igual a 14 representa a maior distância geodésica (o maior caminho mais curto) entre dois nós no grafo, assim, o par de autores mais "distante" na rede conectada está separado por 14 arestas. A densidade muito baixa indica uma rede esparsa, onde a maioria dos possíveis pares de nós não estão conectados, fato comum em redes de coautoria, onde as conexões tendem a se formar em pequenos grupos ou comunidades. Os 159 componentes conectados, mostram que 158 desses grupos conectados devem ser muito pequenos, já que o maior deles, a CG, contém um percentual muito alto da RM. O coeficiente de aglomeração alto é uma característica de redes mundo pequeno, indicando que os vizinhos de um nó tendem a ser conectados entre si, refletindo em colaborações intensas de pequenos grupos de pesquisa (*clusters*). O comprimento médio do caminho relativamente baixo (quase 5), combinado com o alto coeficiente de aglomeração, reforça a propriedade de mundo pequeno, isso significa que a rede é eficiente em conectar pesquisadores. A modularidade alta indica que a rede está organizada em comunidades bem definidas.

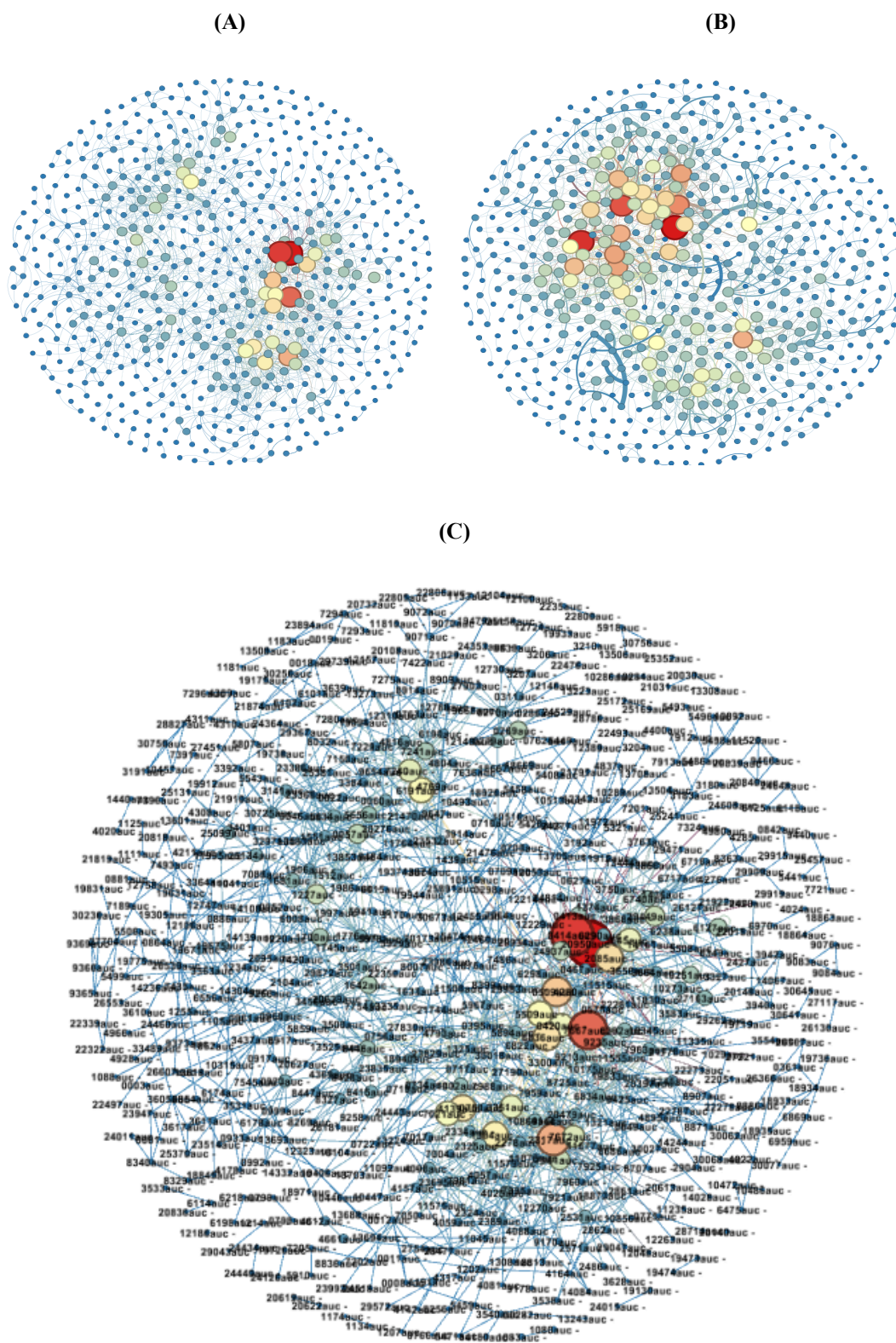
Entendemos que existe colaboração entre os pesquisadores, dando um bom indicativo de disseminação do conhecimento na RM, entre os pesquisadores da UEFS e fora dela. Porém, para entender as conexões que acontecem dentro UEFS, geramos uma nova rede, de coautoria, agora apenas com as conexões dos professores da UEFS: a Rede de coautoria Filha (RF), da qual faremos análises das métricas, da topologia e das centralidades (medidas que dão características aos nós) para identificar melhor as relações da rede e responder aos objetivos desta pesquisa.

4.1.2 Rede de Coautoria Filha (RF)

Na rede filha (RF) só aparecem os professores da UEFS, dessa forma, os nós representam os professores da instituição e a aresta entre dois professores vai existir, se eles compartilham uma mesma publicação. Os dados da RM foram organizados, e os autores que não fazem parte do quadro de professores da UEFS foram retirados desta planilha. Para atingir o objetivo 1, 2 e 4 desta pesquisa, vamos responder alguns objetivos específicos: 1.

Mapear a rede de coautoria (RF) para compreender o processo de colaboração entre os nós da rede; 2. Identificar os atores das comunidades que colaboram na rede e seus departamentos na UEFS; 4. Propor possibilidades de aumentar a colaboração em relação à posição dos nós da rede RF; As redes mapeadas (seguindo com o objetivo 1) estão representadas na Figura 19 (A), (B) e (C), e através delas e das suas métricas, topologia e centralidades, observamos como acontecem as interações entre esses coautores (professores) para compreender o comportamento da pesquisa na UEFS e verificar características dos nós que possam proporcionar mais multi/interdisciplinaridade, colaboração e disseminação do conhecimento.

Figura 19. (A) Rede de Coautoria “Filha” (RF); (B) Sua Componente Gigante (CGRF); (C) RF com os rótulos



Fonte: Elaboração própria com auxílio do software Gephi, (2024).

A Tabela 2 mostra as estatísticas encontradas na Rede Filha (RF) e na sua componente gigante CGRF.

Tabela 2. Estatísticas da Rede Filha (RF) e sua Componente Gigante (CGRF)

ESTATÍSTICAS	RF	CGRF
Nós	833	675 (81,03%)
Arestas	4.833	4.813 (99,59%)
Grau Médio	11,604	14,261
Diâmetro da Rede	14	14
Densidade do Grafo	0,014	0,021
Componentes Conectados	256	1
Modularidade	0,767	0,768
Coeficiente de Aglomeração	0,444	0,439
Comprimento Médio do Caminho	4,675	4,675

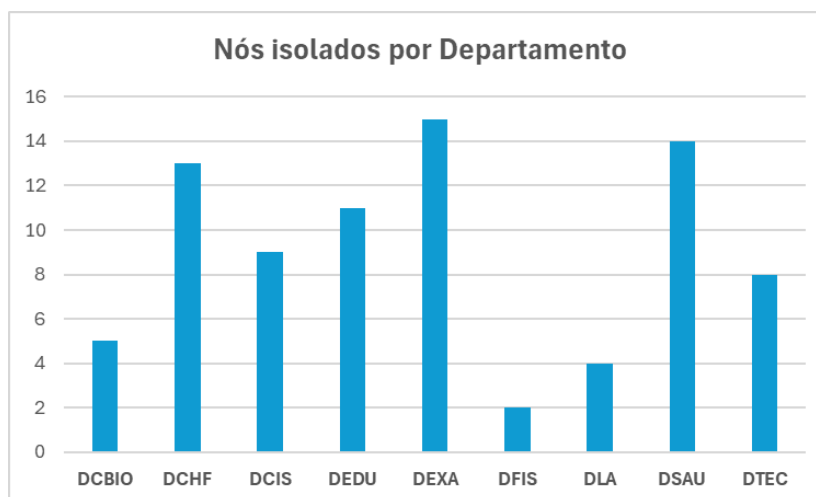
Fonte: Elaboração própria, (2024).

A RF tem 833 nós (professores) e 4.833 arestas (coautorias). A Componente Gigante da Rede Filha (CGRF) tem 81,03% dos nós da RF e manteve 99,59% das arestas da Rede Filha (RF), revelando que, mesmo com a perda de aproximadamente 19% dos nós, a CGRF só perdeu 0,41% das arestas, resultando num aumento do grau médio da RF de 11,064 conexões por nó da CGRF para 14,261 conexões por nó, mostrando mais conectividade que a RF. Outro destaque para a RF é a queda no valor de grau médio em relação à RM, o que indica que, com retirada dos nós denominados “externos” a rede fragmentou-se, desconectando uma quantidade considerável de autores. Uma atenção especial deve ser dada aos nós isolados da RF, que representam muitos professores que publicam sozinhos. A RF possui 81 nós isolados (9,72%). A inserção desses nós na rede de colaboração é importante para manter uma rede mais conectada e mais colaborativa.

A Figura 20 mostra os departamentos aos quais pertencem os professores que publicaram sozinhos, onde maior número de professores faz parte dos departamentos de

Ciências Exatas, seguido por Ciências da Saúde e Ciências Humanas.

Figura 20. Nós isolados por Departamento da UEFS



Fonte: Elaboração própria, (2024).

Encontramos o mesmo valor do diâmetro nas duas (RF e CGRF), indicando que o CGRF cobre a maioria da rede em termos de conectividade. A densidade da rede RF é baixa, e indica que apenas 1,4% das conexões possíveis entre autores estão presentes e, embora haja colaboração entre os autores, a rede ainda é bastante esparsa, com muitas relações potenciais que não estão sendo aproveitadas. Esse valor é um pouco menor para CGRF, e aumenta um pouco para a componente gigante (2,1%). A RF tem 256 componentes conectados, ou seja, 256 subconjuntos/subgrafos de professores que estão conectados entre si, mas não com os outros subconjuntos, este número aumentou bastante em relação à RM, confirmando que a retirada de alguns autores externos fragmentou a RF.

A modularidade é alta para as duas redes, e sugere uma estrutura de comunidade bem definida, o que significa que os autores tendem a colaborar em grupos distintos. Em relação ao coeficiente de aglomeração, que mede a probabilidade de que dois autores que colaboram com um terceiro autor também colaborem entre si, encontramos valores em torno de 0,44 nas duas redes, indicando que há uma quantidade razoável de colaboração entre colegas professores. Avaliando o comprimento médio do caminho como a distância média entre dois nós (autores) na rede, observamos que um valor em torno de 4,7 nas duas redes representa que, em média, qualquer professor pode ser alcançado por outro em cerca de cinco conexões.

O valor idêntico em RF e CGRF reflete uma boa conectividade.

Para verificar se a rede atende aos critérios estabelecidos para redes mundo pequeno, geramos a rede aleatória do Gephi e comparamos os valores para identificar características de rede mundo pequeno. Além do coeficiente de aglomeração médio (CCm), analisamos os valores dos comprimentos médios do caminho comparando os valores da rede com os valores da RF com os valores da rede aleatória (Watts & Strogatz, 1998) (Tabela 3). Segundo Newman (2010), a rede aleatória fornece uma referência útil para comparação. A rede social é considerada de topologia mundo pequeno, se as distâncias típicas forem comparáveis àquelas em um gráfico aleatório (Newman, 2000; Watts & Strogatz, 1998).

Tabela 3. Métricas da Rede Aleatória (RA)

ESTATÍSTICAS	RA
Coeficiente de Aglomeração	0,025
Comprimento Médio do Caminho	2,591

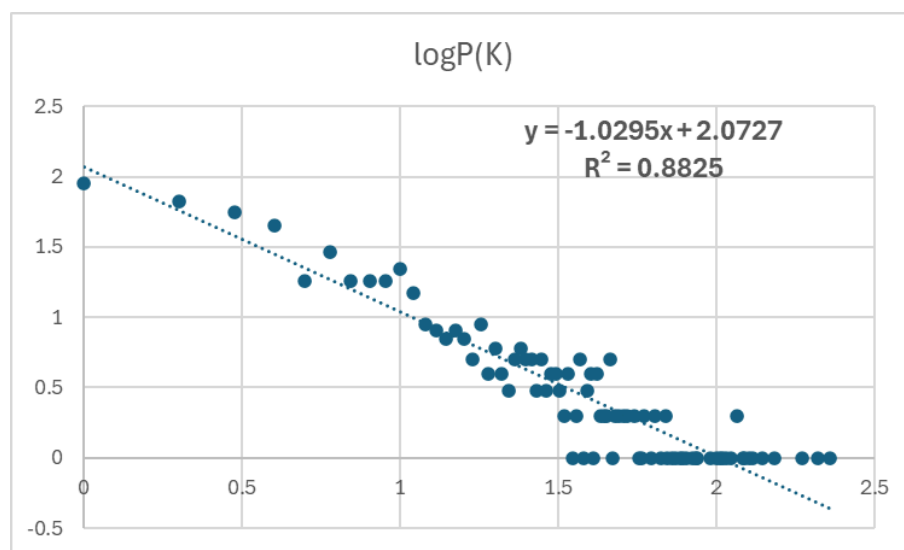
Fonte: Elaboração própria, (2024).

Para a RF é válido que $CCm_{rede} = 0,439 \geq CCm_{aleatória} = 0,025$, ou seja, o coeficiente de aglomeração da rede é muito maior que o da rede aleatória, indicando que a rede tem uma estrutura mais coesa e localmente agrupada (característica de redes mundo pequeno). Os comprimentos médios $L_{rede} = 4,675 \approx L = 2,591$ têm valores considerados próximos. A rede possui características de mundo pequeno, comum em redes de coautoria científica. (Watts & Strogatz, 1998). Newman (2001) destaca resultados que evidenciam as redes do tipo “mundo pequeno”, para redes de coautoria, onde dois coautores estão separados por uma distância pequena e a rede apresenta alta aglomeração. Assim, a estrutura de rede “mundo pequeno” está diretamente associada a um ambiente colaborativo e eficiente, onde os nós conseguem se conectar rapidamente a outros nós da rede.

Para identificar se a RF atende aos critérios da lei de potência (característica de rede livre de escala), utilizamos a distribuição de graus, que foi organizada em uma planilha Excel juntamente com as suas frequências. Calculamos os \log_{10} , para cada um dos graus e das suas frequências. Obtivemos o gráfico Log-Log, junto com a reta de regressão linear e o R^2 (que determina o quão próximo está a equação do comportamento da função). Quando R^2 tem um

valor próximo de 1, sugere uma lei de potência, e podemos inferir que a rede tem característica do fenômeno livre de escala. O ajuste gráfico à reta decrescente é um bom indicativo para caracterizar que a distribuição segue a lei de potência (Newman, 2005). Os valores encontrados estão no gráfico da Figura 21, a seguir:

Figura 21. Gráfico Log-Log da Rede de Cautoria Filha (RF)



Fonte: Elaboração própria, (2024).

O valor de $\gamma = 1,0295$ é próximo de 1, indicando que a rede possui uma alta concentração de nós com muitos graus (Barabási & Albert, 1999; Newman, 2003). O coeficiente de determinação (R^2), que mede a qualidade do ajuste do modelo de regressão aos dados da RF é 0,8464 e significa que há uma forte correlação (de 84,64%), indicando que as distâncias variam logaritmicamente com o número de professores nesta rede de pesquisa e que há tendência a um comportamento livre de escala. Isso significa que a RF tem poucos coautores que se destacam pela alta colaboração e muitos coautores com pouca colaboração.

A rede RF apresenta características de uma rede livre de escala e mundo pequeno, sendo assim, a rede possui *hubs* (professores) importantes para disseminação de conhecimento (livre de escala). Ao mesmo tempo que a propagação de informações ocorre de forma rápida e todos estão “próximos” dentro desta rede científica (mundo pequeno). No entanto, as conexões não estão bem distribuídas entre os professores, e a saída de um *hub* pode fragmentar a rede de pesquisa.

Para atingir os objetivos da pesquisa, caminhamos para identificar padrões significativos de colaboração. Para isso, fizemos análises com auxílio das estatísticas, identificamos os autores que se destacam pelos seus valores de medidas maiores (por isso são nós mais influentes). Para identificar tais atores que fazem a diferença na rede de coautoria buscaremos a análise de centralidade. Assim, selecionamos as seguintes métricas: 1. Centralidade de Grau (CGr); 2. Centralidade de Proximidade (CP); 3. Centralidade de Intermediação (CI) (Freeman, 1978; Freeman, 1979; Newman, 2001; Borgatti, 2006 Wasserman & Faust, 1994). Essas medidas devem descrever melhor as relações que acontecem entre os departamentos e verificar como estão dialogando na rede.

Para analisar os nós mais centrais quanto ao número de conexões, proximidade e intermediação, selecionamos os nós 10% mais (10%+) nas três centralidades, ou seja, os primeiros 62 (sessenta e dois) nós mais “centrais da rede” em cada uma das centralidades listadas no Quadro 13. As células coloridas na Tabela 4 representam (com seus pares) os nós que aparecem mais de uma vez entre o total de 186 nós. Ou seja, como são 62 nós que se destacam em cada centralidade, o número total de nós nas três centralidades é 186.

Tabela 4. Os nós 10% mais nas Centralidade de Grau (CGr), Centralidade de Proximidade (CP) e Centralidade de Intermediação (CI)

	Código Autor	CGr	Código Autor	CP	Código Autor	CI
1	20950auc	228	29919auc	0.170	0509auc	20910.148
2	0414auc	209	8340auc	0.170	7240auc	15578.823
3	7667auc	187	22476auc	0.170	2085auc	13556.281
4	2317auc	153	25370auc	0.170	8446auc	10860.430
5	0509auc	140	7205auc	0.168	6834auc	10774.673
6	2085auc	130	0854auc	0.168	20950auc	9486.309
7	6836auc	127	18845auc	0.168	2317auc	9015.242
8	4139auc	122	3206auc	0.167	5929auc	6992.085
9	1304auc	121	0881auc	0.167	6191auc	6500.424
10	5509auc	116	22497auc	0.165	1302auc	6395.796
11	6191auc	116	0001auc	0.163	7667auc	6303.826
12	7021auc	111	9072auc	0.161	3501auc	6004.637
13	0420auc	107	0019auc	0.161	0420auc	5964.324
14	3645auc	105	22805auc	0.161	5634auc	5868.921
15	3351auc	103	9070auc	0.161	3437auc	5832.944
16	7240auc	101	1133auc	0.161	13444auc	5827.311
17	7672auc	96	14180auc	0.158	3351auc	5455.844
18	7961auc	87	0361auc	0.157	2325auc	5452.824
19	29449auc	87	12724auc	0.157	3363auc	5401.628
20	1227auc	85	24353auc	0.157	24937auc	5395.185
21	0057auc	81	26367auc	0.157	3456auc	5377.634
22	24937auc	80	6475auc	0.157	1436auc	5309.404
23	0269auc	77	1207auc	0.157	30062auc	5300.000
24	6127auc	74	29572auc	0.157	0414auc	5269.715
25	1642auc	72	30645auc	0.155	1776auc	5142.248

26	7241auc	71	7912auc	0.155	6102auc	5049.885
27	10251auc	70	1125auc	0.152	1304auc	4931.069
28	1700auc	70	21819auc	0.152	7021auc	4892.985
29	0413auc	67	9369auc	0.150	1642auc	4809.954
30	6740auc	64	9365auc	0.150	6836auc	4672.946
31	8446auc	64	22322auc	0.149	3645auc	4636.538
32	5656auc	63	19831auc	0.149	13705auc	4595.633
33	27163auc	59	7391auc	0.148	1439auc	4529.673
34	1302auc	59	5498auc	0.148	2531auc	4412.757
35	1681auc	58	5496auc	0.147	0763auc	4140.399
36	6292auc	57	10092auc	0.147	4280auc	4110.980
37	0279auc	55	5493auc	0.147	1700auc	3850.182
38	2325auc	55	7296auc	0.146	27704auc	3801.396
39	9235auc	52	8329auc	0.146	0270auc	3732.419
40	6834auc	52	0842auc	0.146	0018auc	3731.000
41	10869auc	51	24011auc	0.145	5509auc	3541.051
42	4025auc	50	5910auc	0.144	5894auc	3495.745
43	9213auc	49	3210auc	0.143	5656auc	3476.132
44	2531auc	49	7294auc	0.139	3874auc	3406.916
45	25099auc	48	22806auc	0.139	1551auc	3373.664
46	25134auc	48	18934auc	0.136	0467auc	3314.627
47	5634auc	48	10472auc	0.136	26539auc	3313.258
48	0923auc	47	9366auc	0.130	7241auc	3294.120
49	8415auc	46	30236auc	0.130	7420auc	3239.847
50	7963auc	46	22339auc	0.130	23532auc	3233.670
51	6290auc	46	1183auc	0.122	1325auc	3228.619

52	5894auc	46	7293auc	0.122	25241auc	3209.000
53	1776auc	46	20732auc	0.122	0019auc	3204.000
54	10273auc	45	18933auc	0.120	12747auc	3187.141
55	0060auc	45	18935auc	0.120	4139auc	3050.624
56	3501auc	45	10486auc	0.119	1906auc	3032.878
57	20479auc	43	1181auc	0.109	0012auc	3017.630
58	4816auc	43	23894auc	0.109	0279auc	3009.090
59	1374auc	42	12186auc	0.000	20276auc	2975.502
60	3141auc	42	13508auc	0.000	6008auc	2775.013
61	4098auc	42	20140auc	0.000	7988auc	2715.469
62	4793auc	42	25352auc	0.000	7294auc	2682.000

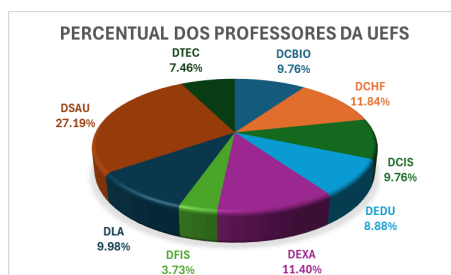
Fonte: Elaboração própria, (2024).

Professores com alta centralidade de grau podem ser vistos como principais colaboradores, pois são mais produtivos pelo número de conexões na rede, tendo assim, muitas publicações em coautoria, o que facilita no acesso aos colegas da rede. Alta centralidade de proximidade indica acesso rápido aos outros colegas da rede de coautoria, facilitando a disseminação de informações. Os que possuem alta centralidade de intermediação, atuam como intermediário entre os nós, conectando áreas ou grupos distintos.

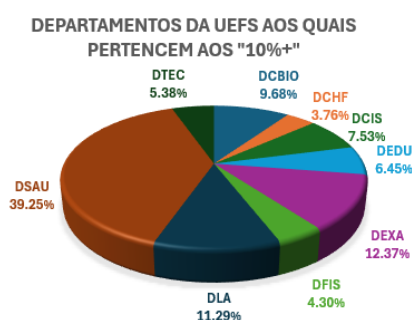
Fizemos um comparativo entre o total de professores UEFS por departamento e a sua representatividade entre os 186 professores que se destacam entre os “10%+”. Em relação ao quantitativo total ativo do corpo docente da UEFS, o departamento que possui o maior número de professores é o de Saúde (DSAU), seguido dos departamento de Ciências Humanas (DCHF), o departamento de Exatas (DEXA), o departamento de Linguística, Letras e Artes (DLA), o departamento de Ciências Sociais Aplicadas (DCIS), o departamento de Ciências Biológicas (DCBIO), o departamento de Educação (DEDU), o departamento de

Tecnologia (DTEC) e por último, o departamento de Física (DFIS) [Figura 22 (A) e (B)]. Relacionando a quantidade de professores por departamento e as posições destes professores entre os “10%+”, verificamos que, em geral, mantêm-se a proporcionalidade. Os destaques são: primeiro DSAU (com percentual de representatividade maior entre os [Figura 22 (A)], mantendo o primeiro lugar em número de professores na UEFS e entre os “10%+”. Em seguida o DEXA, que também manteve a proporção em relação ao número de professores, mostrando um leve avanço entre os, assim como o DLA e o DCBIO. O próximo no ranking é o DCIS que também apresentou valores equivalentes, com um leve decréscimo percentual entre os “10%+”, assim como DEDU e DTEC. O DFIS aparece com percentual maior nos “10%+” do que na distribuição do número total de professores. Por último aparece o DCHF, cujo número de professores é o segundo, mas fica em último lugar em relação ao “10%+” [Figura 22 (B)], ou seja, apesar de ter o 2º maior número de professores da universidade, o departamento tem o menor número de professores entre os “10%+”.

Figura 22. Representatividade dos professores na UEFS x “10%+” nas três Centralidades



(A) Representatividade dos departamentos ao qual pertencem os professores da UEFS



(B) Representatividade dos departamentos dos professores que estão entre os “10%+”

Fonte: Elaboração própria, (2025).

Para atingir o objetivo 2 (propor possibilidades de aumentar a colaboração, diante da posição dos pesquisadores na rede e/ou grupos de pesquisa) e o objetivo 4 (identificar os atores das comunidades que colaboram na rede e seus departamentos), analisamos o papel dos nós na RF para identificar os mais influentes na rede de acordo com as centralidades encontradas. Do total dos 186 (cento e oitenta e seis) professores do “10%+”, identificamos os que estão em duas das centralidades ao mesmo tempo (eles aparecem coloridos na Tabela 4). As características destes nós indicam a sua importância na rede: são um total de 33 professores, sendo que apenas dois desses nós estão na centralidade de grau e de proximidade, todos os outros 31 estão, ao mesmo tempo, na centralidade de grau (muitas conexões com outros professores, tornando-os torna facilitadores na disseminação das informações) e na centralidade de intermediação (indicando a característica de conectar diferentes grupos na rede, facilitando a conexão entre diferentes departamentos e consequentemente, proporcionando a possibilidade de colaborar com grupos distintos).

Os 31 professores que aparecem, ao mesmo tempo, nas centralidades de grau e intermediação são: 0279auc, 0414auc, 0420auc, 0509auc, 1302auc, 1304auc, 1642auc, 1700auc, 1776auc, 2085auc, 20950auc, 2317auc, 2325auc, 24937auc, 2531auc, 3351auc, 3501auc, 3645auc, 4139auc, 5509auc, 5634auc, 5656auc, 5894auc, 6191auc, 6834auc, 6836auc, 7021auc, 7240auc, 7241auc, 7667auc e 8446auc. Além destes, os nós 0019auc e 7294auc, aparecem em destaque na centralidade de grau e de proximidade. Esses professores representam departamentos/áreas mais ativas e estratégicas, cujo trabalho colaborativo pode auxiliar na gestão das relações da rede, propondo temáticas e conexões que reforcem o incentivo à colaboração entre diferentes grandes áreas (representadas pelos departamentos). Para Recuero (2017), os atores centrais (com centralidade de grau e de intermediação) têm mais capacidade de influenciar o fluxo de informação da rede.

Entre os 31 professores que se destacam nas duas centralidades ao mesmo tempo (em número de conexões e intermediação), 64,5% pertencem aos departamentos de saúde, 16,2% são do departamento de exatas, 12,9% do departamento de ciências biológicas, 3,2% do departamento de linguística, letras e artes e 3,2% do departamento de física. Os dois professores que aparecem ao mesmo tempo com destaque em número de conexões e proximidade são do departamento de linguística, letras e artes, apresentando maior facilidade de se conectarem a diferentes partes da rede de pesquisa. Na análise do posicionamento de cada professor e seu departamento, deve-se levar em conta as características de cada um e os seus respectivos papéis na rede.

Entre os 9 departamentos da instituição, os 5 que mencionamos acima têm professores que se destacam entre os em duas centralidades diferentes, tornando-os influentes na rede de pesquisa, porém, precisamos entender o papel dos professores dos demais departamentos. Um deles é o departamento de educação, que tem 12 professores que aparecem entre os “10%+”, 7 estão na centralidade de proximidade e 5 na de intermediação, ou seja, o departamento de educação (DEDU), que é um departamento cujos professores atuam de outros departamentos da UEFS, colaborando com disciplinas de educação, apresentam nós que partilham publicações com professores de outros departamentos, acessam outros grupos de estudo e conseguem conectar grupos distintos, mas precisam se destacar no número de conexões, por exemplo, se conectando a nós com alta centralidade de grau de outros grupos, já que eles têm alta intermediação. Os professores com influência na intermediação e proximidade podem colaborar nas conexões da rede para aumentar o número de conexões e colaboração.

O departamento de Ciências Sociais Aplicadas (DCIS) tem 14 professores entre os “10%+”, apenas um aparece na intermediação, os outros 13 estão na centralidade de proximidade, mostrando facilidade em chegar em outros grupos da rede, mas tem apenas 1 professor entre os da intermediação, e não conseguem intermediar estas relações, também não têm muitas conexões na rede. Estes professores estão próximos de outros nós importantes, porém têm poucas conexões diretas e exercem pouca influência como intermediários nas relações da rede. Os destaques da centralidade de proximidade têm potencial para aumentar o número de conexões interdisciplinares, engajando outros nós na rede.

O departamento de Ciências Humanas e Filosofia (DCHF), tem o menor número de professores no “10%+”: total de 7. Todos eles aparecem em destaque na centralidade de proximidade, indicando participação em outros grupos, porém sem estabelecer maiores conexões entre eles. O departamento de ciências humanas e filosofia tem um potencial muito grande para colaborações, pois estuda o ser humano e suas conexões sociais, culturais e históricas, podendo utilizar temáticas comuns a todos os outros departamentos. Para maior coesão à rede, deve ser dada uma atenção especial ao número de coautorias (grau) e às conexões com grupos de outros departamentos, funcionando como pontes (intermediação). Como não existem nós de maior destaque nas conexões, cabe um trabalho da gestão da UEFS, provocando e incentivando trabalhos mais interdisciplinares, com propostas temáticas inserindo professores de todos os departamentos neste contexto.

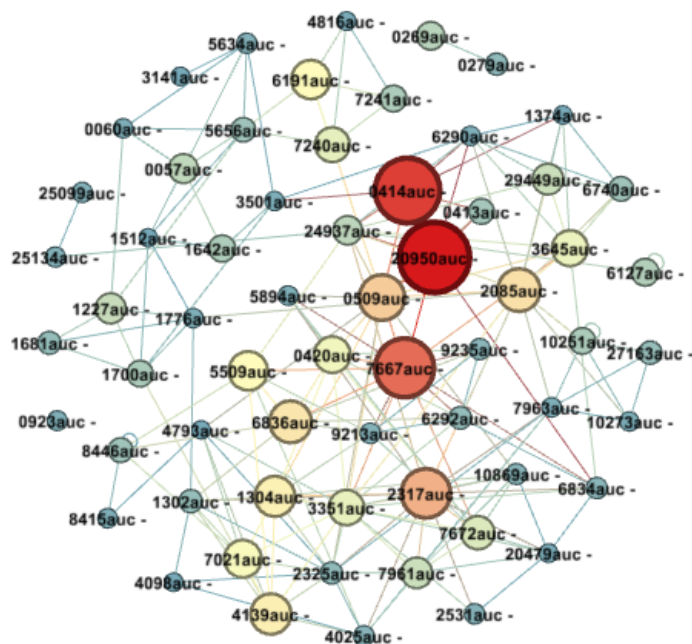
O departamento de tecnologia (DTEC) tem 10 professores entre os “10%+”, apenas 1 se destaca no número de conexões (centralidade de grau). Cinco deles aparecem na

centralidade de proximidade e 4 na centralidade de intermediação. Os números balanceados mostram que o departamento de tecnologia consegue algum destaque em acessar e conectar grupos distintos da rede, porém precisam melhorar o grau das conexões. Eles desempenham um papel importante como mediador e estão bem posicionados para acessar outros nós com rapidez, contudo, têm poucas conexões diretas. Os nós destacados em intermediação e proximidade podem trabalhar para aumentar estas conexões.

Os professores do departamento de física (DFIS) aparecem 8 vezes no total entre os “10%+”, este valor fica acima da proporção de representatividade dos professores deste departamento na UEFS. Além disso, o DFIS possui representante em todas as centralidades entre os “10%+”. Depois aparecem os professores do departamento de ciências biológicas (DCBIO), com 19, departamento de linguística, letras e artes (DLA), com 21, e o departamento de ciências exatas (DEXA), com 23, e o departamento de saúde (DSAU) lidera com 73 professores. A centralidade de proximidade é a única em que o departamento de saúde não tem maior número de professores. Esta é a centralidade mais interdepartamental de todas da rede de pesquisa da UEFS, e indica professores que podem alcançar rapidamente qualquer outro na rede, tornando-os mais influentes em termos de alcance direto. A formação de conexões através da centralidade de intermediação indica professores que conectam áreas diferentes atuando como facilitadores da disseminação de informações e conhecimento.

A Figura 23 mostra a rede de coautoria, na qual aplicamos o filtro de 10% (que são os 10%+ em relação à centralidade de grau). Assim, a rede apresenta os nós com mais arestas da RF, trazendo os 62 professores (10%) com mais conexões (coautorias). Borgatti (2018) identifica tais atores como os principais colaboradores numa rede de coautoria. Estes professores têm facilidade em disseminar informações na rede por conta do maior número de conexões, por isso são influentes. Desses 62 professores com mais conexões que aparecem na Figura 23, 38 são do departamento de saúde, 9 do departamento de ciências exatas, 8 do departamento de biologia, 4 do departamento de física, 2 de linguística, letras e artes e 1 do departamento de tecnologia. Os departamentos de ciências humanas e filosofia, ciências sociais e de educação não aparecem entre esses 62 professores com mais conexões (grau) da rede de coautoria.

Figura 23. Rede com filtro de 10% dos nós com maior grau da rede de coautoria RF



Fonte: Elaboração própria, (2024).

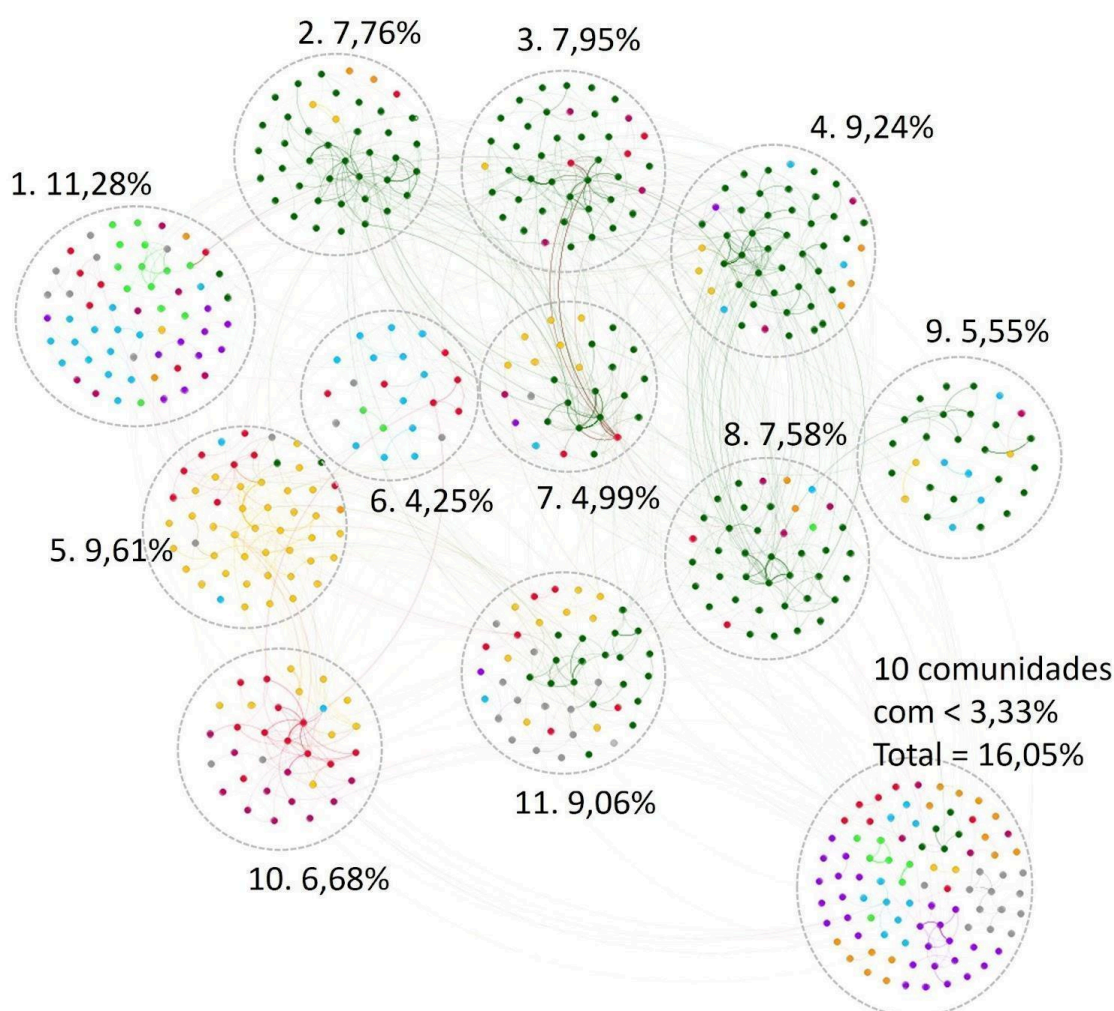
Em uma rede de pesquisa científica, para aumentar a colaboração, o ideal é identificar e integrar as contribuições de cada professor. Professores com alta centralidade de grau podem ajudar a eficiência no fluxo de informações pelo número de arestas, enquanto os com alta centralidade de intermediação garantem a coesão e conectividade da rede. Caso a rede perca conexões importantes, professores com alta intermediação podem ser fundamentais para reorganizar a rede, garantir sua funcionalidade e manter os resultados de pesquisa da rede. (Wasserman & Faust, 1994). Os professores que se destacam com alta centralidade de proximidade conseguem acessar outros nós com um pequeno número de conexões, por isso têm facilidade em cooperar com grupos distintos. Todos eles juntos formam a espinha dorsal da colaboração e inovação científica, e definir os papéis de cada ator na rede pode impulsionar a pesquisa científica.

Para complementar os objetivo 1, de mapear e compreender o processo de colaboração e o objetivo 4, de identificar os atores das comunidades que colaboram na rede e seus departamentos na UEFS, recorreremos à representação da RF em comunidades, utilizando o algoritmo de modularidade para identificar a divisão da rede em comunidades ou subgrupos densamente conectados (Guedes, 2010; Newman & Girvan, 2004). Tal identificação é

especialmente útil para entender a estrutura interna de uma rede de coautoria e suas colaborações, ajudando a investigar o trabalho multi/interdisciplinar nos grupos formados na rede.

Na Figura 24, estão representadas as comunidades da rede RF, onde cada grupo da rede é formado pelos nós que estão nos círculos, para destacar que são comunidades distintas. As 10 comunidades com menores quantidades de vértices (menor que 3,33%) foram agrupadas em apenas um círculo. Os nós são pintados em cores distintas para representar departamentos distintos (Quadro 14).

Figura 24. Divisão em comunidades com nós (professores) de departamentos distintos



Fonte: Elaboração própria, (2025).

Quadro 14. Representação de cada departamento da UEFS por cores

Departamento	Cores	
DCBIO	FFCC00	
DCHF	CC0066	
DCIS	FF9900	
DEDU	00CCFF	
DEXA	FF0033	
DFIS	33FF33	
DLA	9900FF	
DSAU	6600	
DTEC	999999	

Fonte: Elaboração própria, (2025).

Essa forma de representação, conforme vemos na Figura 24, permitiu a visualização dos nós que representam professores de diferentes departamentos que publicam e colaboram em uma mesma comunidade, o que sugere troca de conhecimento entre grandes áreas distintas, indicando multi/interdisciplinaridade e disseminação de conhecimento na RF (via publicações). As conexões entre comunidades permitem o compartilhamento de informações, que são fundamentais para a multi/interdisciplinaridade e disseminação de conhecimento. (Fortunato, 2016; Bertalanffy, 1968; Pombo, 2004; Fazenda, 2003). Índícios de multi e interdisciplinaridade aparecem na Figura 24, com fortes conexões dentro de cada comunidade e entre comunidades distintas (Pombo, 2004).

A modularidade da rede de coautoria filha é alta, aproximadamente 0,8, e já havíamos alertado sobre fortes divisões em comunidades que emergiram das métricas da rede. As conexões mais intensas da RF são dos nós verde escuro, representando professores do departamento de saúde, que também têm maior representatividade em número de comunidades. Na comunidade 1 podemos verificar a presença de todos os departamentos da UEFS (todas as cores), indicando colaboração e troca entre os membros. O departamento de saúde (em verde escuro) só não apresenta membros em duas delas, na comunidade 6 e na 10. Os professores do departamento de ciências biológicas (em amarelo) também aparecem em todas, exceto na comunidade 6 e na 8.

Todas as comunidades têm, no mínimo, 3 professores de diferentes departamentos,

sugerindo um trabalho colaborativo. Em quase todas é visível uma quantidade maior de professores de um determinado departamento, apenas na comunidade 1 e no “grupão” (que contém as 10 comunidades com número de vértices menor que 3,33%) podemos visualizar uma quantidade proporcional de professores de departamentos diferentes. No contexto de redes de coautoria ou de pesquisa, comunidades podem ser grupos de pesquisadores que frequentemente colaboram em projetos ou publicações.

Da rede de coautoria filha (RF) emergem muitas informações importantes para entender o comportamento dos professores em relação às publicações. A análise completa da rede de coautoria mãe não é necessária, porém, a RM oferece alguns pontos importantes para a análise da RF, pois deu a dimensão da rede completa e as suas conexões com pesquisadores externos.

O Quadro 15 destaca os tópicos mais importantes para análise da RF.

Quadro 15. Propostas/pontos a serem avaliados na rede de professores dos grupos de pesquisa da UEFS/CNPq de autoria filha (RF)

Grau Médio	O valor teve um decréscimo considerável com a retirada dos nós externos (RM), tornando a rede mais fragmentada. A RF precisa ser fortalecida com o aumento de publicações de coautoria entre professores da UEFS.
Diâmetro	O valor do diâmetro igual a 14, que é a maior quantidade mínima de arestas a serem percorridas de um nó a outro na rede. Mais conexões (publicações) em coautoria podem propiciar redução deste caminho.
Densidade	O valor de densidade máxima é 1 e representa a conexão de cada nó com todos os outros nós da rede. A RF com densidade de 0,014 revela a conexão de 1,4%, muito baixa para potencial de conectividade da RF. Mais conexões podem aumentar esse valor. Segundo Silva (2006) a ciência funciona bem quando a comunidade de pesquisadores é densamente conectada.
Componentes Conectados	A rede tem uma quantidade muito grande de pequenos componentes conectados, muitos deles resultado da fragmentação da rede causada pela retirada dos nós externos. Pequenas conexões através de publicações com os grupos e nós isolados podem deixar a rede mais conectada e colaborativa. A fragmentação da rede é contrária ao modo multi/interdisciplinar, pois representa a fragmentação do conhecimento produzido na rede (Fazenda, 2004).
Característica de rede Livre de Escala	Revela alguns poucos professores com muitas publicações de coautoria e muitos professores com poucas publicações. A retirada de um nó <i>hub</i> (com muitas conexões) causará grande fragmentação na rede. A multi/interdisciplinaridade deve buscar igualdade entre as publicações dos departamentos na rede. Conectar os nós isolados e componentes desconectados pode melhorar essa média. Para maior colaboração da rede de pesquisa, devemos buscar as características de rede pequeno mundo, com professores mais próximos.
Centralidade de Grau/Proximidade e Intermediação	Análise das centralidades em cada um dos departamentos de acordo com os “10%+”, buscando uma proporcionalidade entre o número de professores por departamento e sua representatividade entre os “10%+”. Departamentos com percentuais muito abaixo devem ser incentivados a publicar mais (maior grau), interagir mais com outros grupos e departamentos (proximidade e intermediação). Essas centralidades podem indicar multi/interdisciplinaridade na rede e disseminação do conhecimento na rede de pesquisa.

Comunidades	A divisão em comunidades mostrou boa conexão tanto entre os professores das comunidades, como entre professores de comunidades distintas, apesar disso, essas conexões ainda podem ser melhoradas. Essa representação também mostrou multi e interdisciplinaridade nas comunidades, porém, ainda é possível visualizar comunidades mais homogêneas, onde se pode incentivar o trabalho mais interdisciplinar.
--------------------	---

Fonte: Elaboração própria, (2025).

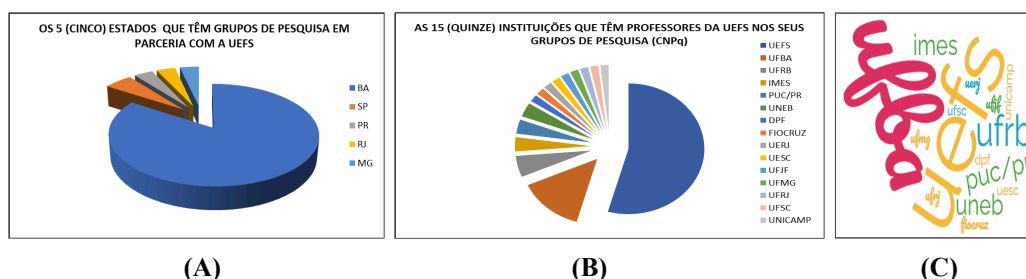
Para complementar os dados das redes de coautoria, analisaremos uma outra rede social da UEFS, formada pelos pesquisadores dos grupos de pesquisa cadastrados do CNPq.

4.1.3 A Rede Social Formada por Professores dos Grupos de Pesquisa da UEFS Cadastrados no CNPq

A segunda rede social é formada pelos professores participantes dos grupos de pesquisa da UEFS. Do total de 40 (quarenta) grupos cadastrados, 13 são da UEFS, os demais são de outras instituições, porém, têm algum professor da UEFS cadastrado como membro do grupo.

A Figura 25 mostra: (A) estados a que pertencem os grupos de pesquisa externos, que têm ao menos um professor da UEFS como participante e (B) (C) respectivas instituições a que pertencem os grupos externos (em gráfico e em nuvem de palavras).

Figura 25. Instituições parceiras da UEFS em grupos de pesquisa CNPq



- (A) Os Estados que têm Instituições com grupos de pesquisa em parceria com a UEFS
 (B) As Instituições que têm professores da UEFS nos seus grupos de pesquisa
 (C) Nuvem de palavras com as siglas das Instituições envolvidas nos Grupos de Pesquisa externos

Fonte: Elaboração própria, (2025).

Em relação aos grupos cadastrados, percebemos que dos 27 grupos externos (que são de outras instituições, mas tem algum professor da UEFS) o número de professores da UEFS variou de 1 a 2 participantes, o que não mostraria um resultado da colaboração entre os pares da UEFS. Para ter uma visão mais clara de como acontece a interação/colaboração entre os pesquisadores da UEFS, optamos por focar nos 13 grupos de pesquisa da UEFS, para

entender como eles funcionam e seus resultados de colaboração na instituição.

Através da Figura 26, podemos verificar que apenas 5 grandes áreas do conhecimento são contempladas pelos grupos de pesquisa da UEFS. As Ciências da Saúde lideram o número de grupos de pesquisa. Em segundo aparece a grande área Multidisciplinar, seguida de Ciências Exatas e Ciências Biológicas empatadas em terceiro e por último “Ciências Humanas”. Pelos dados do Apêndice, os grupos multidisciplinares são fruto dos cursos Lato Sensu e Stricto Sensu da UEFS e, por sua natureza interdisciplinar, devem ter professores de outras áreas, indicando colaboração e multi/interdisciplinaridade. A Figura 26 ilustra as grandes áreas a que pertencem os grupos de pesquisa da UEFS.

Figura 26. Grupos de Pesquisa da UEFS dividido por “Grandes Áreas” do Conhecimento

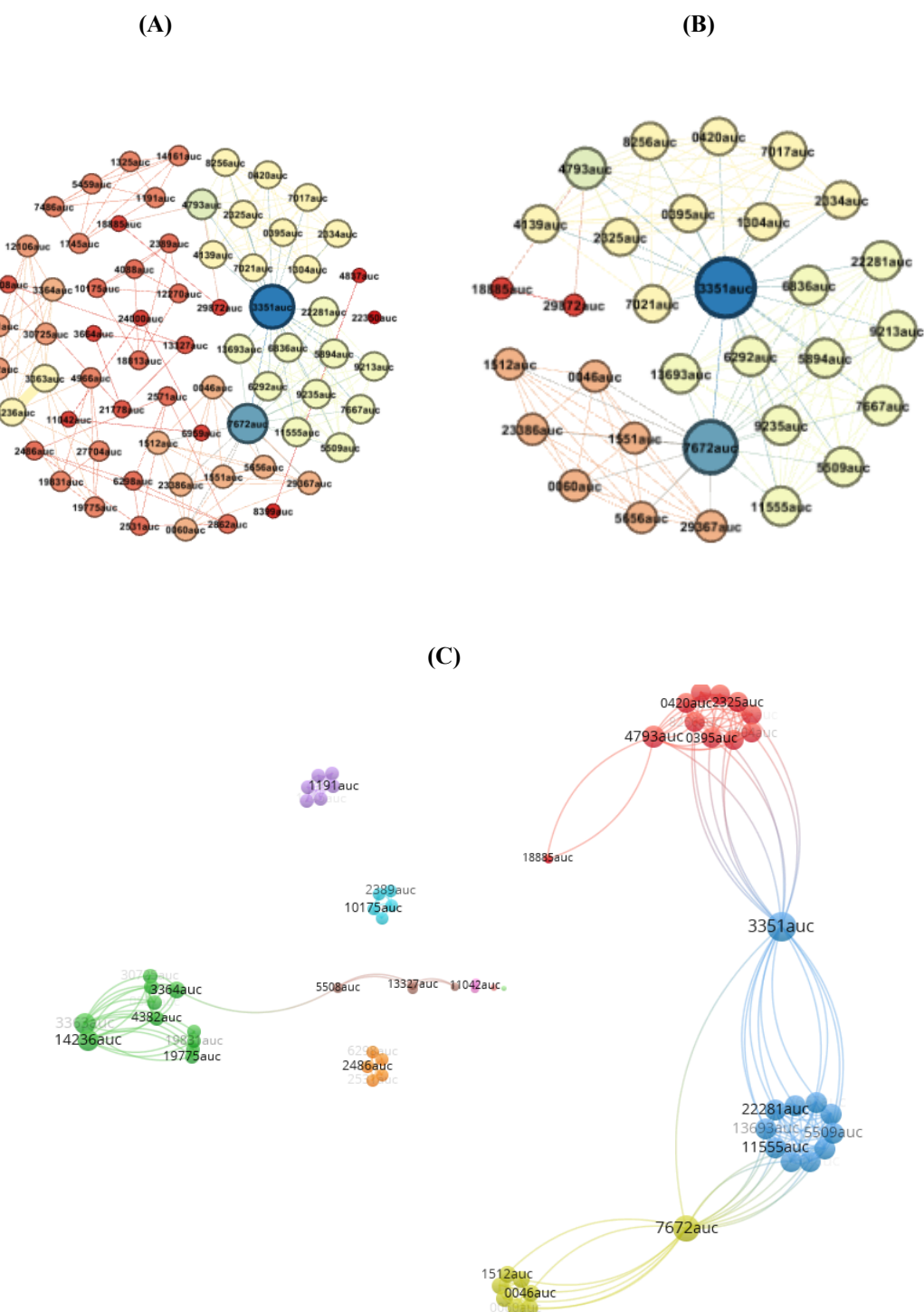


Fonte: Elaboração própria, (2024).

Para complementar as informações da rede filha e ratificar os objetivos 1 (Mapear redes de pesquisas e compreender o processo de colaboração entre os professores), 2 (Propor possibilidades de aumentar a colaboração, diante da posição dos pesquisadores na rede e/ou grupos de pesquisa) e 4 (identificar os atores das comunidades que colaboram na rede e seus departamentos) utilizaremos a rede social formada pelos professores participantes dos grupos de pesquisa, onde os nós são os professores da UEFS, e dois nós têm uma aresta se participam do mesmo grupo. Os professores representados nesta rede têm os mesmos códigos que foram utilizados na rede de coautoria filha (RF), assim podemos cruzar e /ou acrescentar novas informações à RF.

O total de professores da UEFS que participam dos grupos cadastrados no CNPq na instituição é 68, onde 55,88% deles são do departamento de saúde. As Figuras 27 (A) e (B) mostram a rede com as relações entre os pesquisadores e sua componente gigante. No *site* do CNPq consta a data de início de cada grupo. Os grupos de pesquisa da UEFS em atuação (em 22 de outubro de 2023) tiveram as seguintes datas de cadastramento no CNPq: 1998 (o mais antigo) da Área de Saúde, seguidos de 2000, 2004, 2005, 2009, 2011, 2015, 2016, 2018 (dois grupos), 2019, 2021 e 2022 (o mais recente). A rede visualizada no VOSviewer está representada pela Figura 27 (C) e apresenta 68 termos (professores) e 11 *clusters*. O tamanho de cada nó indica a quantidade de conexões de cada professor com os grupos de pesquisa e a divisão em cores no VOSviewer utiliza um algoritmo baseado na modularidade para identificar e diferenciar os *clusters* ou comunidades dentro da rede (Figura 27).

Figura 27. Rede de professores participantes de grupos de pesquisa cadastrados no CNPq



- (A) Rede de professores dos grupos de pesquisa da UEFS cadastrados no CNPq (*software gephi*)
 (B) A Componente gigante da rede de pesquisadores dos grupos de pesquisa da UEFS
 (C) Rede de professores dos grupos de pesquisa da UEFS cadastrados no CNPq (*software VOSviewer*)

Fonte: Elaboração própria, (2025).

Para entender se há relação entre os professores que participam dos grupos de pesquisa com a rede de autoria filha (RF), fizemos um comparativo entre os nós que se destacam na centralidade de grau nesta rede de pesquisadores. Com os códigos dos 68 professores dos grupos de pesquisa, fizemos uma relação com os professores da rede de coautoria. A comparação foi feita utilizando as duas planilhas (coautoria RF e professores dos grupos de pesquisa). Cruzando essas informações, obtivemos todas as intersecções dos elementos das duas redes, e encontramos 25 professores dos grupos de pesquisa que estão entre os “10%+” da rede de coautoria. Observamos que 12 desses professores dos grupos de pesquisa, aparecem em duas das centralidades ao mesmo tempo na rede de coautoria (grau e intermediação), ou seja, como se repetem na Tabela “10%+”, eles aparecem 37 ($25 + 12 = 37$) vezes nessa tabela. Dessa forma, 54,41% dos professores que fazem parte desses grupos de pesquisa, têm seus nomes entre os “10%+” na rede de coautoria. Um número considerável, se avaliarmos que este percentual se refere a apenas 13 grupos, dentre muitos outros existentes dentro da UEFS, com outras formas de cadastro, que não é via CNPq. Esse percentual mostra a influência desses grupos de pesquisa na rede e sua importância para as conexões nas publicações que aparecem na rede de coautoria.

O maior grupo, o gp09 (em vermelho) é interdisciplinar da grande área “multidisciplinar”, e tem como líderes, professores do departamento de saúde. Este grupo absorve o grupo gp05, que juntos formam um *cluster* só. O grupo gp05 é formado por professores do departamento de exatas, e a junção mostra que esses professores trabalham de forma colaborativa, conectando diferentes áreas de conhecimento, ajudando a responder o objetivo 3 desta pesquisa (compreender como os professores e suas temáticas se posicionam e se conectam nas redes de pesquisa), e o objetivo 5 (identificar temáticas de destaque na rede e/ou que são comuns a professores de grupos distintos).

A visualização da Figura 27 (C) mostra as comunidades formadas pela proximidade dos nós, assim podemos verificar, por exemplo, que o nó 3351auc (departamento de saúde) é destaque no número de conexões e conecta os dois outros maiores grupos da rede, um da área de saúde (gp07) e outro, o gp09 (em vermelho), interdisciplinar. Este nó também aparece na rede de coautoria filha (RF), entre os “10%+”, como destaque na centralidade de grau e na centralidade de intermediação, assim, e sua relevância na rede é confirmada pela rede dos grupos de pesquisa [Figura 27 (A) e (B)].

Outro nó que liga dois grupos é o 7672auc, também do departamento de saúde. O nó conecta o gp06 (das ciências biológicas, em verde claro) e o gp07 (de saúde, em azul). Este nó

também é um dos que aparece entre os “10%+”, como destaque na centralidade de grau e de intermediação. O gp07, apesar de ser da área de saúde, possui professores de três departamentos distintos: saúde, letras e artes e exatas, indicando um trabalho colaborativo de diferentes áreas. (Fazenda, 2003; Pombo, 2004). Os dois maiores grupos da rede (gp07 e gp09) formam o maior componente conectado da rede de pesquisadores dos grupos de pesquisa, conectando 4 grupos distintos: gp05 (exatas), gp06 (biológicas), gp07 (saúde), gp09 (interdisciplinar). Todos os quatro grupos têm professores de departamentos diferentes como participantes, representando a importância da troca de informações entre diferentes departamentos em grupos de pesquisa diferentes.

Pela Figura 27 (C), os grupos desconectados da rede dos grupos de pesquisa são três: o gp08, o gp11 e o gp12. Dois deles, o gp11 (de saúde, em azul claro, ao centro) e o gp12 (de saúde, em laranja), têm apenas professores do departamento de saúde em sua composição e em número muito pequeno. O grupo interdisciplinar (gp08, em lilás), é liderado por professores do departamento de saúde e tem poucos integrantes. Apesar de possuir professores dos departamentos de educação, saúde e biológicas, o grupo não interage com os demais.

O nó 5508auc (do departamento de saúde), conecta o grupo verde (gp02) do departamento de educação a outro grupo pequeno, o gp04 (marrom) do departamento de saúde, novamente indicando intermediação entre grandes áreas distintas.

Os dados das redes de pesquisadores dos grupos de pesquisa, juntamente com a rede de coautoria, ofereceram valiosas informações para o entendimento da comunidade científica da UEFS, suas colaborações e temáticas. As próximas redes semânticas devem colaborar ainda mais para o entendimento da rede, mais especificamente em relação a multi/interdisciplinaridade do grupo. O Quadro 16 apresenta propostas de análises referente à rede dos professores dos grupos de pesquisa.

Quadro 16. Propostas/pontos a serem avaliados da rede de professores dos grupos de pesquisa da UEFS/CNPq

Grandes Áreas	<p>Apenas 5 das 9 grandes áreas têm grupos cadastrados pelo CNPq. Reavaliar os programas Lato e Stricto Sensu, das áreas não contempladas, as quais poderiam cadastrar grupos de pesquisa, tornando a rede mais conectada.</p> <p>Somente 6 departamentos da UEFS estão representados nesta rede. É válido o incentivo à inserção dos demais departamentos na rede de pesquisa (Ciências Sociais, Ciências Humanas e Física).</p>
Multi/interdisciplinaridade	<p>Avaliar os membros dos grupos interdisciplinares, de forma que contemplem professores de diferentes grandes áreas para conectar diferentes grupos. Tais professores podem ser os destaques nas centralidades de proximidade ou intermediação entre os “10%+”. Os grupos mais isolados [Figura 27 (C)]: gp08 (em lilás), o gp11 (azul claro, ao centro), o gp12 (em laranja) devem ser consultados quanto à conexão com professores de outros grupos da rede de pesquisa, identificando nós que possam colaborar com diferentes grupos para troca de informações e maior conexão da rede de pesquisa.</p>

Fonte: Elaboração própria, (2025).

4. 2 REDES SEMÂNTICAS

As redes semânticas são originadas na associação de palavras, conceitos ou frases com base em seus significados ou relações no contexto de um corpus textual, onde os nós representam as palavras, conceitos ou termos, as arestas indicam conexões semânticas, que podem ser baseadas em coocorrência, sinônimos, ou outro relacionamento linguístico (Fadigas, 2011). Tais redes são utilizadas em análise de redes para identificar padrões de significado dentro de textos, mapear relações entre conceitos centrais de um campo ou tema e identificar a estrutura de um conhecimento ou vocabulário específico. Dessa forma, podem oferecer mais argumentos para responder o objetivo 3: Compreender como os professores e suas temáticas se posicionam e se conectam nas redes de pesquisa da UEFS e o objetivo 5: Identificar temáticas de destaque na rede e/ou que são comuns a pesquisadores de grupos

distintos.

Os atores (nós) numa rede semântica são as palavras (ou junção delas) selecionadas por seus significados intrínsecos. Os relacionamentos (arestas) acontecem entre as palavras de um título, palavras-chave ou qualquer outra conexão de palavras (Newman, 2010; Fadigas, 2011). Em relação aos dados dos currículos lattes, duas redes semânticas serão geradas; a de títulos e a de palavras-chave. Em relação à rede dos grupos de pesquisa, serão geradas as redes dos nomes dos grupos e das linhas de pesquisa.

É importante ressaltar que a análise das redes semânticas tem o intuito de trazer mais informações, que devem ser acrescentadas às outras redes propostas, para responder aos objetivos relacionados a temáticas, colaboração e multi/interdisciplinaridade. Leydesdorff (2001) ressalta como as redes de comunicação científica (incluindo redes de palavras-chave, títulos e coautorias) revelam padrões de colaboração e organização do conhecimento, que são frequentemente associados a abordagens interdisciplinares.

Nesta pesquisa, as palavras que emergem nas redes foram analisadas de forma contextualizada, pois, grande parte destes termos podem ter diferentes significados dependendo do departamento/grande área/área considerada, mudando o sentido semântico. Enquanto outros termos são sinônimos e, assim, podem ser compartilhados por outras áreas do conhecimento distintas, sob diferentes perspectivas. Essa avaliação faz parte das etapas de tratamento, inferência e interpretação, onde é feita a extração de significados, tendências ou implicações da unidade de análise; a unidade de registro, que é a menor unidade analisada, neste caso, a palavra ou expressão (Bardin, 2011).

As redes semânticas foram encontradas através do VOSviewer. O próprio software retira todos os acentos, símbolos, cedilhas e transforma letras maiúsculas em minúsculas. Após a limpeza dos dados, as redes foram encontradas no VOSviewer e salvas em “.gml” para rodar no software Gephi e calcular as demais estatísticas.

4.2.1 Redes de Títulos dos Artigo/Livros/Capítulos de Livros

As redes de títulos podem identificar os temas centrais de um conjunto de publicações. Os nós da rede de títulos desta pesquisa são as palavras (ou conjunto de palavras) encontradas nos títulos dos artigos/livros/capítulos de livros. Segundo Leydesdorff (2011), esse tipo de

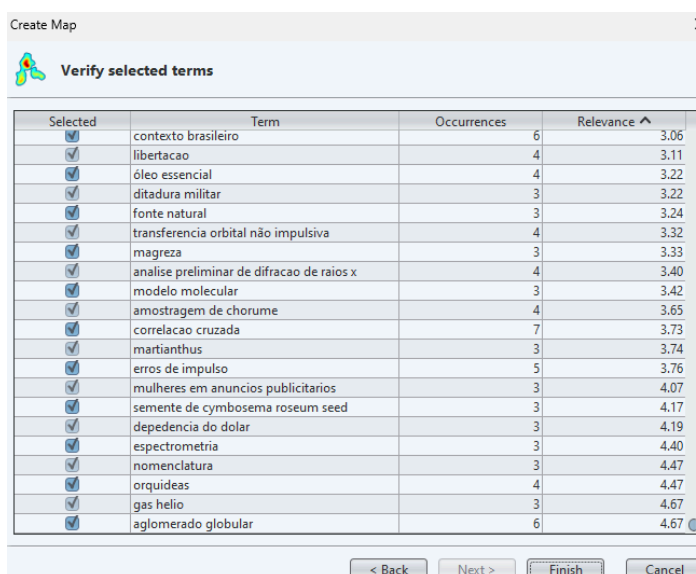
rede pode identificar tendências temáticas em áreas específicas e suas conexões. Nessa rede, foram consideradas as palavras que aparecem três ou mais vezes nos títulos das publicações.

As arestas conectam as palavras no mesmo título ou as mesmas palavras que aparecem em títulos diferentes, assim, cada título forma uma clique (grafo completo, onde todos os nós estão conectados) (Fadigas, 2011). Os dados foram organizados de forma a transformar o conteúdo qualitativo e quantitativo dessas palavras possam revelar padrões e comportamentos da rede. Para obtenção de uma rede que tivesse caracterização temática representativa, foram seguidos os passos (Bardin, 2011):

1. Pré-análise: Organização, sistematização e seleção dos títulos da planilha “Títulos”:
 - Verificamos e retiramos as duplicidades dos títulos na planilha “Título” para evitar contaminação nos resultados;
 - Traduzimos os títulos em idiomas diferentes do português. Para isso utilizamos o DeepL Tradutor.
2. Exploração: Codificação. Fragmentação e Regras:
 - Ao utilizar o arquivo referente aos títulos no software VOSviewer, selecionamos mínimo de ocorrências igual a três para identificar as palavras dos títulos que aparecem, no mínimo, três vezes.
 - Após a visualização das palavras destacadas (Figura 28), e observando a necessidade de alterações, fizemos um tesouro, que funciona como um dicionário de sinônimos, com substituições/exclusões/correções necessárias para tornar a rede mais clara e concisa.
3. Tratamento. Inferência e Interpretação
 - Encontramos as redes de títulos e seus *clusters* para identificar as comunidades formadas. Visualmente foi feita uma análise das unidades de registro, verificando a importância contextual para optar pela rede com mais possibilidades de apresentar as características que respondessem aos objetivos específicos;
 - Encontradas as métricas, topologia e centralidades da rede de títulos, fizemos a seleção da menor unidade de registro (palavras ou expressões) para extração de significado, contextualização e interpretação dos resultados com as informações obtidas (Charmaz, 2009).

Quando o arquivo de títulos é selecionado, antes da inserção do tesauro, o software apresenta um quadro de palavras, com todos os termos da rede (Figura 28), número de ocorrências de cada uma e o índice de relevância de cada palavra (Manual VOSviewer, p. 44). Assim, podemos visualizar as palavras que fariam parte da rede e realizar as alterações necessárias. Nesse momento podemos acrescentar o tesauro, com as mudanças para otimização da rede e/ou retirar palavras descontextualizadas.

Figura 28. Fragmento da Tabela de Palavras referente a Rede Títulos no Software VOSviewer



Selected	Term	Occurrences	Relevance ^
<input checked="" type="checkbox"/>	contexto brasileiro	6	3.06
<input checked="" type="checkbox"/>	libertacao	4	3.11
<input checked="" type="checkbox"/>	óleo essencial	4	3.22
<input checked="" type="checkbox"/>	ditadura militar	3	3.22
<input checked="" type="checkbox"/>	fonte natural	3	3.24
<input checked="" type="checkbox"/>	transferencia orbital não impulsiva	4	3.32
<input checked="" type="checkbox"/>	magreza	3	3.33
<input checked="" type="checkbox"/>	analise preliminar de difracao de raios x	4	3.40
<input checked="" type="checkbox"/>	modelo molecular	3	3.42
<input checked="" type="checkbox"/>	amostragem de chorume	4	3.65
<input checked="" type="checkbox"/>	correlacao cruzada	7	3.73
<input checked="" type="checkbox"/>	martianthus	3	3.74
<input checked="" type="checkbox"/>	erros de impulso	5	3.76
<input checked="" type="checkbox"/>	mulheres em anuncios publicitarios	3	4.07
<input checked="" type="checkbox"/>	semente de cymbosema roseum seed	3	4.17
<input checked="" type="checkbox"/>	depedencia do dolar	3	4.19
<input checked="" type="checkbox"/>	espectrometria	3	4.40
<input checked="" type="checkbox"/>	nomenclatura	3	4.47
<input checked="" type="checkbox"/>	orquideas	4	4.47
<input checked="" type="checkbox"/>	gas helio	3	4.67
<input checked="" type="checkbox"/>	aglomerado globular	6	4.67

Fonte: Elaboração própria, (2024).

O tesauro anexado ao projeto do VOSviewer foi feito no Excel e depois salvo em formato .txt, possui um vocabulário para padronização da rede, onde retiramos e/ou alteramos palavras da rede (Figura 29). Palavras iguais que estão com representações diferentes, ou que podem ser substituídas para melhor compreensão semântica, ou até mesmo retiradas por não fornecerem sentido semântico à rede.

Figura 29. Fragmento do Tesauro (.txt) anexado ao projeto ‘Títulos’ no Software VOSviewer

Arquivo	Editar	Exibir
Label	Replace by	
apocynaceae	asclepiadoideae	apocynaceae
araceae	araceae	
bioma caatinga	caatinga	
bioma caatinga semiarido	caatinga	
bioma semiarido	caatinga	
bioma semiarido da caatinga	caatinga	
brasil e	brasil	
brasil	brasil	
casos clinico	caso clinico	
catole	chapada diamantina bahia	
celulas	t humanas tipo celulas t	
chapada diamantina bahia	chapada diamantina bahia	
chapada diamantina bahia	brasil chapada diamantina bahia	
checklist		
cirurgiaodentista	cirurgiao dentista	
class		
concepco		
consideracoes	sobre	
contexto da	contexto	
contribuico		
coord		
corpo e corpo		
da chapada diamantina bahia	brasil	chapada diamantina bahia
diabetes mellitus tipo	diabetes	
diagnosticos	diagnostico	
diamantina bahia	brasil chapada diamantina bahia	
diptera psychodidae	diptera	
diptera psychodidae	diptera	
diptera psychodidae	diptera	
durante		
em feira	feira de santana bahia	
ensino de	ensino	
estudos		
feira de santana ba	feira de santana bahia	
feira de santana bahia	feira de santana bahia	
feira de santana bahia	brasil feira de santana bahia	
feira de santanaba	feira de santana bahia	

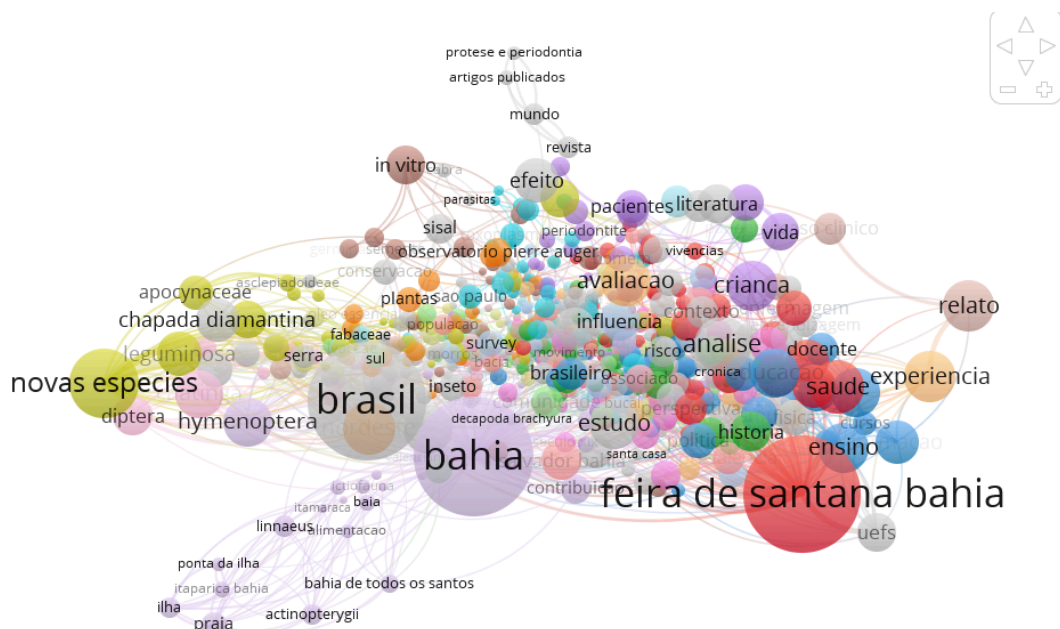
Fonte: Elaboração própria (2024)

A planilha, depois de organizada, formada pelos títulos dos trabalhos publicados (artigos/livros/capítulos de livros), apresentou 13.289 títulos. Copiamos o arquivo do Excel em formato “.txt” para utilizar no VOSviewer, onde foram encontrados 33.776 termos (palavras). Após a inserção do tesauro, o total de termos ficou em 33.044. Selecionamos o número de ocorrência igual a 3 (mínimo de vezes que o termo deve aparecer nos títulos das publicações para ser utilizado na rede). Com essa limitação, foram encontrados 1.083 termos. A Figura 30 mostra a rede encontrada.

O VOSviewer apresenta a rede considerando a coocorrência das palavras, assim pode aparecer na rede de títulos, a junção de palavras (expressões), pois o algoritmo identifica que a junção aparece muitas vezes nas publicações e considera como uma unidade de registro. As dez palavras (ou junções de palavras), da rede de títulos que aparecem com os maiores graus na rede são: “brasil” (com grau 354), “bahia” (334), “feira de santana” (215), “análise” (128), “novas espécies” (113), “nordeste” (109), “estudo” (95), “ensino” e “professores” (ambos com grau 92). A rede de títulos encontrada tem 55 *clusters* (grupos de palavras de títulos com conexões mais densas). Os nós maiores mostrados na Figura 30, representam as palavras que

aparecem mais nas publicações, e as cores determinam os *clusters* (grupos) da rede. (Manual VOSviewer 1.6.20, 2023, p. 9).

Figura 30. Rede de Títulos do software VOSviewer com o mínimo de 3 ocorrências da palavra



Fonte: Elaboração própria (2024)

Com o objetivo de retirar da rede os termos mais genéricos e menos conectados, analisamos a componente gigante da rede (mais conectada), e aplicamos no VOSviewer uma relevância de 60% (60% dos termos mais relevantes para a rede). A rede encontrada tem 512 termos e 26 *clusters*.

A nova rede está representada na Figura 31, nesta rede as palavras “brasil”, “bahia” e “feira de santana bahia” não aparecem mais, pois o algoritmo considera que são palavras genéricas. Estas palavras, vistas como genéricas pelo algoritmo, não atribuem valor semântico à rede, assim, não identificam características multi/interdisciplinares, de colaboração e nem disseminação entre os grupos.

A rede apresenta alguns *clusters* com graus menores e mais distantes do seu centro, que são relacionados a grandes áreas específicas como: as ciências biológicas ou linguística, letras e artes. Já os termos relacionados à saúde estão mais coesos, possuem nós maiores e estão localizados mais ao centro da rede.

Pelas métricas podemos descrever o comportamento das palavras dos títulos na componente gigante. O grau médio, aproximadamente igual a 4 representa o número médio de conexões das palavras dos títulos. O diâmetro, medindo a maior distância entre duas palavras quaisquer é de 15, o valor é alto comparado ao número de nós e a conectividade da componente gigante. A densidade baixa indica uma rede pouco coesa e modularidade alta revela estruturas de comunidades de palavras na rede. O coeficiente de aglomeração é de quase 0,36 revelando uma probabilidade de conexão de 36% de que os vizinhos de uma palavra também sejam vizinhas entre si. O comprimento médio do caminho indica comprimento mínimo de quase seis “passos” entre quaisquer dois termos da rede, indicando um comportamento mundo pequeno.

A Tabela 6 traz as estatísticas da rede aleatória para análise topológica.

Tabela 6. Estatísticas Rede Aleatória (RA)

Estatísticas	Ra (Com Relevância De 60%)
Coeficiente de Aglomeração (CCm)	0,049
Comprimento Médio do Caminho (L)	2,611

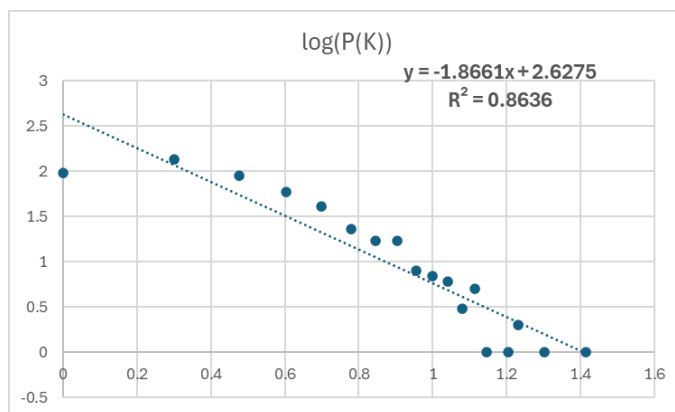
Fonte: Elaboração própria, (2025).

Observamos que $CCm_{rede} = 0,355 \geq CCm_{aleatória} = 0,049 \Rightarrow$ A rede tem coeficiente de aglomeração maior que a rede de aleatória, o que indica tendência a formar “grupos”, característica de redes mundo pequeno. E os valores do comprimento médio $L_{rede} = 5,783 \approx L_{aleatória} = 2,611 \Rightarrow$ são consideravelmente próximos, aceitáveis para considerar uma rede de mundo pequeno (Watts & Strogatz, 1998). A rede de títulos tem forte organização temática local, mas as palavras mais distantes (ou comunidades de títulos) ainda estão interligadas, pois fazem parte da componente gigante. Para a análise semântica, isso sugere que diferentes tópicos podem ser acessados rapidamente, mas possuem subtemas bem definidos.

O valor do R^2 de 86,36% indica um modelo linear ajustado (Figura 32), sugerindo uma topologia livre de escala, com poucas palavras se conectando a muitas outras e muitas palavras se conectando a poucas palavras. Certas palavras dos títulos desempenham um papel crucial para manter a rede conectada. A rede é robusta a falhas aleatórias, mas vulnerável à

remoção desses *hubs* cuja retirada pode desconectar a rede (Barabási & Albert, 1999; Newman, 2003).

Figura 32. Gráfico de Dispersão, Reta de regressão, Equação Linear e o R^2



Fonte: Elaboração própria, (2025).

Para entender como acontecem as relações das palavras dos títulos (nós) da rede, analisaremos, a partir das medidas de CGr - Centralidade de Grau, que mede o número de conexões diretas de cada nó, CP - Centralidade de Proximidade, que mede quão perto um nó está de todos os outros na rede e a CI - Centralidade de Intermediação, que aponta os nós conectores da rede.

Os termos que se destacam na rede estão na Tabela 7, onde, de cada medida de centralidade, selecionamos 20 termos, encontrando o total de 60 termos nas três centralidades, repetidos ou não, representando 10% do total das palavras. A escolha das palavras em número menor que os nós da rede de coautoria se deve ao fato de que um percentual maior traria muitas palavras sem maior importância na rede, assim, quanto mais palavras, mais termos sem importância apareceriam. Os 20 termos selecionados estão representados como as “20 mais” (“20+”) palavras destacadas da rede de títulos nas três centralidades.

Tabela 7. As “20+” dos Títulos nas três Centralidades

	Palavra	CCg	Palavra2	CP	Palavra3	CI
1	in vitro	26	in vitro	0.274	in vitro	30519.348
2	tratamento	20	sisal	0.265	tratamento	17482.007
3	apocynaceae	17	tratamento	0.259	apocynaceae	8817.993
4	linnaeus	17	temperatura	0.254	linnaeus	5924.665
5	praia	16	extrato	0.250	praia	4287.735
6	sisal	14	desenvolvimento	0.247	sisal	12725.591
7	actinopterygii	13	semente	0.242	actinopterygii	870.779
8	dna	13	light	0.241	dna	10177.820
9	individual	13	dna	0.241	individual	6532.944
10	laser	13	germinacao	0.241	laser	7372.455
11	revisao sistematica	13	laser	0.240	revisao sistematica	5169.353
12	temperatura	13	cabra	0.239	temperatura	11048.914
13	concentracao	12	associacoes	0.237	concentracao	5601.488
14	ilha	12	cultivo	0.237	ilha	101.331
15	crescimento	11	crescimento	0.237	crescimento	10176.145
16	discurso	11	toxoplasmose	0.236	discurso	5723.943
17	metanalise	11	apocynaceae	0.236	metanalise	4846.501
18	metodo	11	sempreviva	0.235	metodo	9376.616
19	semente	11	annona	0.235	semente	3864.368
20	toxoplasmose	11	parasitas	0.234	toxoplasmose	8957.366

Fonte: Elaboração própria, (2024).

Do total de 60 termos, 10 se destacam nas três centralidades, são elas: “in vitro”, “tratamento”, “dna”, “sisal”, “apocynaceae”, “crescimento”, “sementes”, “sisal”, “temperatura” e “toxoplasmose”. Estas palavras são importantes porque aparecem muito nas redes nas três centralidades. Porém, é necessário avaliar o sentido semântico dessas palavras para verificar se cabe a utilização em títulos de publicações de áreas/grande áreas/departamentos distintos. Por exemplo, temos palavras específicas da área de botânica, como “apocynaceae”, que é uma planta. Podemos ter várias publicações, da mesma área, utilizando essa palavra nos seus títulos, pois dificilmente artigos de áreas distintas vão utilizar esta palavra, para verificar indícios de multi/interdisciplinaridade, precisamos analisar as relações da palavra na rede.

Outro exemplo é a palavra “crescimento”, ela pode estar relacionada ao crescimento de uma planta, ou o crescimento demográfico, ou de uma criança, e não terá sentido semântico, assim como a palavra “tratamento”. Nenhuma delas pode indicar interdisciplinaridade pelo simples fato de se destacarem na rede, mesmo inseridas em um domínio bem delimitado, como biologia, saúde ou áreas correlatas. Isso indica que, embora as

palavras estejam conectadas na rede de títulos, a interdisciplinaridade pode não ser tão evidente ou robusta, por isso devemos aprofundar a análise para obtenção de outros indícios. Apenas as medidas de centralidades não levaram à indícios de multi/interdisciplinaridade, então optamos por analisar alguns *clusters* e verificar sua posição na rede. A palavra “tratamento” [(Figura 33 (A))] está conectada a outros grupos de títulos na rede, como sisal, laser e dinâmica molecular. Porém, não podemos considerar indício de interdisciplinaridade, pois a palavra “tratamento” pode estar sendo utilizada em contextos diferentes, como; “o tratamento do sisal”, um “tratamento a laser” ou um “tratamento através do método de dinâmica molecular”. Ou seja, não podemos afirmar que a palavra “tratamento” conecta grandes áreas/áreas ou departamentos, pois ela pode estar sendo utilizada em diferentes contextos.

Vamos analisar a palavra “apocynaceae”, que é o nome científico de uma planta. Uma palavra que não tem como ser utilizada em diferentes contextos, ou seja, quando aparece esta palavra, não há dúvidas de que se trata de um tipo específico de planta. A Figura 33 (B) mostra o nó “apocynaceae” (em azul) e cujas palavras indicam ser da grande área das ciências biológicas. A palavra está conectada a apenas duas cores distintas (que pertencem a *clusters* diferentes), porém, as palavras que pertencem ao *cluster* em amarelo, como germinação e biogeografia, também pertencem às ciências biológicas. A palavra “in vitro” (na cor cinza) é uma palavra que pode ser utilizada tanto pelas ciências biológicas quanto pelas ciências da saúde, não podemos dessa forma, afirmar conexões multi ou interdisciplinares.

Por último, analisaremos a junção de palavras “baixo peso ao nascer”, que forma um nó no *cluster* em azul que pertence à grande área das ciências da saúde, mais especificamente, da área de enfermagem, e que está conectada a dois *clusters* distintos: um deles, em amarelo, também da grande área da saúde, porém da área odontologia. A palavra “estudo piloto” também pertence à área odontologia, já que está conectada a outras palavras ligadas a esta área, como a palavra “terapia periodontal”. Nesse caso, verificamos conexões entre áreas distintas, dentro da mesma grande área da saúde.

Os títulos são mais extensos que as palavras-chave, e contêm textos mais explicativos, diferente das palavras-chave que têm que ser mais diretas e concisas. Por isso, a rede de títulos fornece uma visão mais detalhada das áreas específicas de estudo, enquanto a próxima rede (de palavras-chave) pode ajudar a identificar tendências amplas e transversais, oferecendo maior clareza temática. Assim, as palavras encontradas dos títulos devem ser analisadas relacionando-se às conexões do nó e aos *clusters* aos quais pertencem, sempre de forma contextualizada.

Para pontos de melhorias na rede de títulos, apresentamos um só quadro juntando as informações das redes de títulos e de palavras chaves.

4.2.2 Redes de Palavras-Chave dos Artigos/Livros/Capítulos de Livros

Buscando mais indícios que possam complementar as informações obtidas, geramos mais uma rede semântica. Esta rede foi formada com as palavras-chave dos artigos/livros/capítulos de livro retirados dos currículos lattes dos professores da UEFS. As palavras-chave que aparecem, no mínimo três vezes, formam os nós, e uma aresta é criada entre dois nós (palavras-chave) quando essas palavras aparecem juntas em uma mesma publicação ou, se a mesma palavra-chave aparece em outras publicações (sempre considerando o mínimo de ocorrência igual a três). Esta rede ajuda a identificar como conceitos estão interligados em um conjunto de publicações, e podem revelar padrões semânticos e potenciais relações interdisciplinares, por isso, pode ser utilizada para complementar respostas dos objetivos específicos propostos na rede de títulos (Charmaz, 2009).

Com o intuito de dar característica temática aos termos (Bardin 2011), seguimos os passos bem parecidos com os da rede de títulos:

1. Pré-análise: Organização, sistematização e seleção dos títulos da planilha “PalavrasChave”:

- As palavras da planilha “PalavrasChave” que estavam em outro idioma foram traduzidas para o português, para isso, utilizamos o DeepL Tradutor.

2. Exploração: Codificação. Fragmentação e Regras:

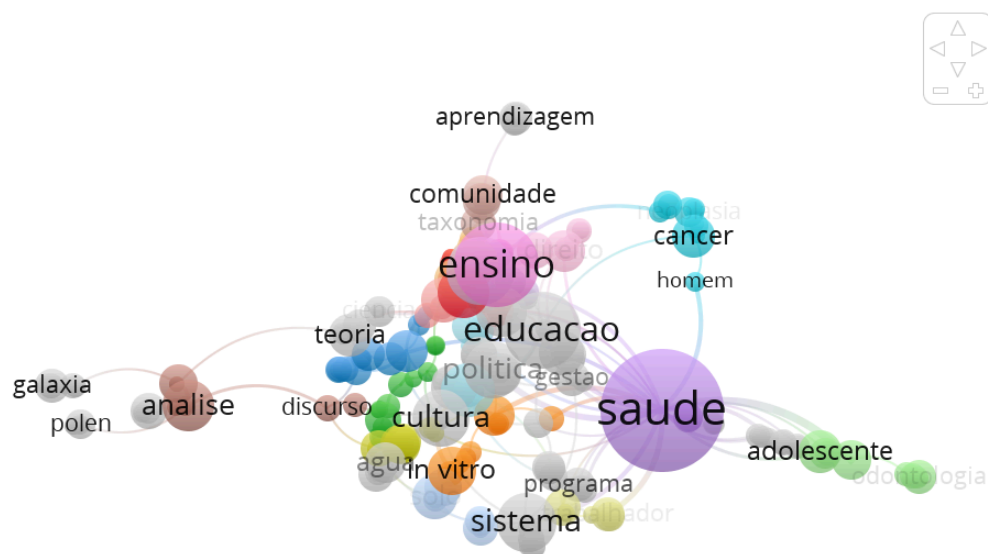
- Inserimos o arquivo PalavrasChaves.txt no VOSviewer e selecionamos o mínimo de ocorrência das palavras igual a três.
- Após visualizar as palavras destacadas e observando a necessidade de alterações, fizemos um tesauro, que funcionou como um dicionário de sinônimos, com substituições/exclusões/correções necessárias para tornar a rede mais clara e concisa. Tal arquivo foi acrescentado ao projeto PalavrasChave;

3. Tratamento. Inferência e Interpretação

- Encontramos as redes de palavras e seus *clusters* para identificar as comunidades formadas. Visualmente foi feita uma análise das unidades de registro, verificando a importância contextual para optar pela rede com mais possibilidades de apresentar características que respondessem aos objetivos específicos;
- Encontradas as métricas da rede, topologia e centralidades da rede de títulos, fizemos a seleção da menor unidade de registro (palavras ou expressões) para extração de significado e interpretação dos resultados com as informações obtidas (Charmaz, 2009).

O total de termos encontrados na rede de palavras-chave foi 8.441, com a condição de 3 ocorrências da palavra. Aplicamos o tesauro e, com a relevância de 60%, foram encontrados 421 termos e sua componente gigante tem 286 palavras conectadas e 38 *clusters*. Optamos mais uma vez pela análise da componente gigante, por ser uma rede mais conectada, com as palavras-chave mais relevantes (coocorrência de 60%), retirando os termos mais genéricos na sua versão conectada (Figura 34).

Figura 34. Componente Gigante da Rede de Palavras-Chave dos artigos/livros/capítulos de livros dos professores/pesquisadores da UEFS com relevância de 60%



Fonte: Elaboração própria, (2024).

As palavras-chave que aparecem nas redes da Figura 34 são as de maior número de conexões da rede, as mais utilizadas pelos professores nas suas publicações. Para análise quanto à posição destas palavras-chave na rede, analisamos o ranking das 20 palavras-chave que aparecem mais nas medidas de centralidade de grau, de proximidade e de intermediação (“20+”), conforme Tabela 8:

Tabela 8. As “20+” Palavras-Chave nas três Centralidades

	Palavra-Chave 1	CGr	Palavra-Chave2	CP	Palavra-Chave3	CI
1	saude	41	saude	0.306	saude	16395.462
2	ensino	31	ensino	0.298	ensino	12462.938
3	historia	17	tecnologia	0.286	sistema	4922.745
4	enfermagem	11	literatura	0.276	historia	4726.754
5	educacao	10	educacao	0.275	teoria	4527.122
6	sistema	10	direito	0.272	analise	4500.900
7	brasil	10	educacao fisica	0.265	rede	4148.670
8	literatura	10	escola	0.264	educacao fisica	3868.503
9	trabalho	10	sistema	0.264	tecnologia	3771.334
10	direito	9	professor	0.264	discurso	3737.360
11	cultura	8	cultura	0.263	identidade	3610.793
12	portugues	8	filosofia	0.261	cultura	3561.361
13	bahia	8	gestao	0.260	direito	3544.960
14	in vitro	7	politica	0.257	literatura	3415.740
15	politica	7	distancia	0.257	enfermagem	3235.233
16	campo	7	historia	0.256	trabalho	3196.645
17	analise	7	trabalho	0.255	educacao	3159.133
18	gestao	7	comunicacao	0.252	agua	2854.444
19	agua	7	politicas publicas	0.252	qualidade	2487.819
20	cuidado	6	cuidado	0.251	genero	2447.410

Fonte: Elaboração própria, (2025).

Estão destacadas em cores, as palavras que emergem nas três ou em duas centralidades, porém, no caso das redes de palavras-chave, muitas palavras que têm relação podem ter significados ou contextos distintos, dependendo da área a que está se referindo. Assim, é necessário que seja feita uma análise para verificar se a palavra pode ser utilizada em publicações com temas que podem ser construídos entre diferentes departamentos. As seguintes palavras aparecem nas três centralidades, sendo palavras influentes na rede de palavras-chave: “saude”; “ensino”; “cultura”, “direito”, “educacao”, “historia” e “literatura”. Em duas centralidades aparecem as seguintes palavras: “agua”, “analise”, “cuidado”, “educacao fisica”, “enfermagem”, “gestao”, “politica” e “sistema”.

Algumas destas palavras influentes não podem ser consideradas parte de um contexto interdisciplinar pois, por si só, não se definem. Por exemplo: “direito”, “analise”, “gestão” ou “sistema”. Elas dependem de uma outra palavra que lhes dê o contexto, o que pode ser identificado através das análises das conexões dos *clusters*.

As palavras “saude” e “ensino” aparecem nos primeiros lugares e nas três centralidades, sugerindo que são conceitos centrais e aparecem com frequência como palavras-chave das publicações. O tema saúde é considerado transversal pelos parâmetros curriculares (MEC, 1997) e justifica sua posição de destaque nas palavras-chave das

publicações. A palavra ensino pode estar sendo utilizado de forma colaborativa e interdisciplinar, mas a recomendação é uma análise mais precisa nas conexões deste nó e verificar o contexto. Segundo Thiesen (2008), ao se pensar a interdisciplinaridade como um movimento articulador no processo educativo, ele atenta que pode não ser suficientemente eficaz uma avaliação por palavras-chave.

Por outro lado, há algumas palavras que por si só já determinam o seu contexto. A palavra “educação física” não pode ser retirada de contexto, então ela pode indicar um trabalho colaborativo entre grupos diferentes. A palavra “educação física” aparece nas centralidades de proximidade e intermediação, ou seja, ela não é destaque na quantidade de conexões, e sim nas conexões com grupos distintos, seja como parceria (proximidade) seja como intermediadora (intermediação). Os indícios são que a palavra “educação física” não é amplamente utilizada como termo principal de palavras-chave, mas têm um papel de ligação entre diferentes grupos, podendo ser um tema intermediário que permeia múltiplos contextos, sem ser o foco principal. Para identificar melhor a rede de palavras-chave, encontramos as principais métricas da rede (Tabela 9).

Tabela 9. Estatísticas da Componente Gigante da Rede de Palavras-chave dos artigos/livros/capítulos de livro

ESTATÍSTICAS	CG DA REDE DE PALAVRAS
Nós	286
Arestas	381
Grau Médio	2,664
Diâmetro da Rede	14
Densidade do Grafo	0,009
Componentes Conectados	1
Modularidade	0,733
Coeficiente de Aglomeração	0,17
Comprimento Médio do Caminho	5,184

Fonte: Elaboração própria, (2024).

A rede é esparsa indicando pouca integração entre os temas, onde cada palavra está conectada a aproximadamente 2,7 outras palavras (menor que a coocorrência) já que estamos

analisando a componente gigante. O diâmetro de 14 é alto pelo tamanho da rede sugerindo palavras distantes. A densidade é baixa indicando uma rede pouco conectada, composta por termos específicos. Porém, como estamos falando da componente gigante, todas as palavras estão conectadas direta ou indiretamente, e não temos redes isoladas, ideais para análise de centralidade e observação de interdisciplinaridade. A modularidade alta (0,733) indica divisão em comunidades de palavras-chave bem definidas, que podem ser melhor identificadas pelos *clusters* da rede. O coeficiente de aglomeração médio de 0,17 é baixo, e indica que palavras-chave vizinhas não estão fortemente conectadas entre si. A rede está organizada em grupos temáticos bem definidos, revelando áreas específicas de pesquisa.

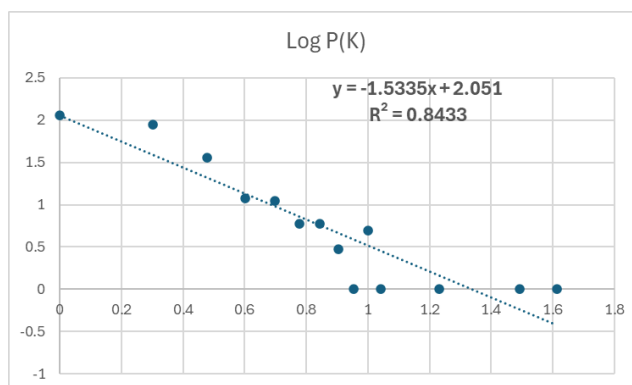
As estatísticas da rede aleatória para análise topológica são apresentadas pela Tabela 10.

Tabela 10. Estatísticas da Rede Aleatória (RA)

ESTATÍSTICAS	RA
Coeficiente de Aglomeração Médio (CCm)	0,013
Comprimento Médio do Caminho (L)	5,56

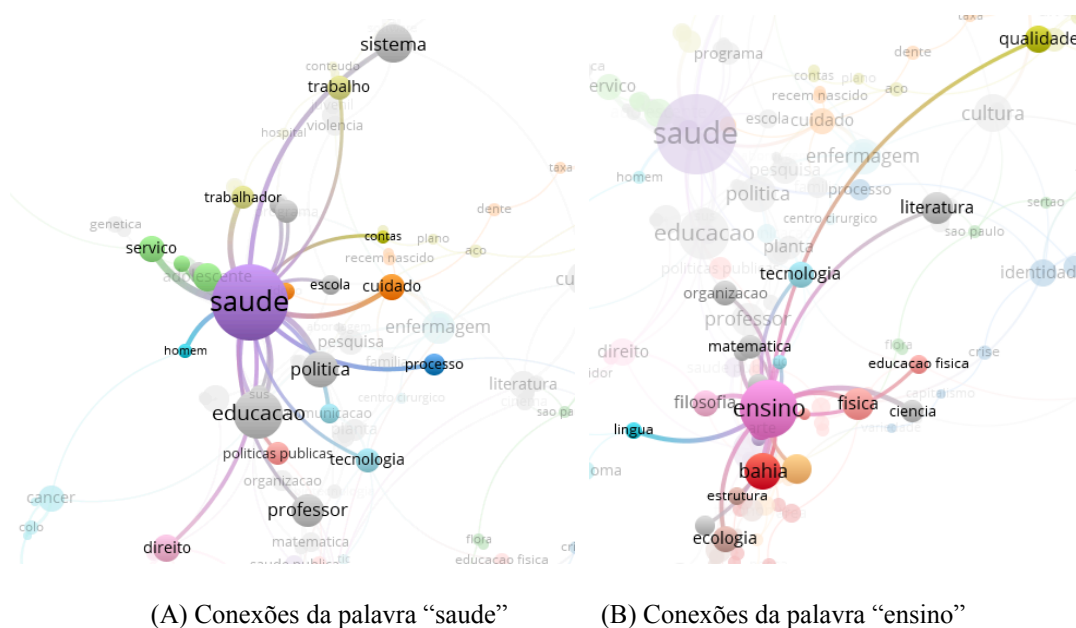
Fonte: Elaboração própria, (2025).

Para a análise topológica da rede, o $CCm_{rede} = 0,17 \geq CCm_{aleatória} = 0,013 \Rightarrow$ indica que apesar dos valores baixos, a rede tem coeficiente de aglomeração maior que a rede aleatória com característica de rede de mundo pequeno. O valor do comprimento médio $L_{rede} = 5,184 \approx L_{aleatória} = 5,56 \Rightarrow$ indica valores muito próximos, assim, a rede tem características de mundo pequeno. O gráfico representado pela Figura 35 mostra que a rede também tem características de uma rede livre de escala, pois 84,33% para o R^2 em um gráfico de dispersão na análise de rede livre de escala, significa um grau de ajuste do modelo de lei de potência aos dados observados da rede e reflete a proporção de variação nos dados que é explicada pelo modelo ajustado. Uma rede onde as palavras estão próximas, mas tem poucas palavras com muitas conexões e muitas com poucas conexões.

Figura 35. Gráfico de Dispersão. Reta de regressão, Equação Linear e R^2 

Fonte: Elaboração própria, (2025).

As Figuras 36 (A) e (B) mostram as conexões de duas das palavras-chave.

Figura 36. Conexões nós “saude” e “ensino”

Fonte: Elaboração própria, (2025).

A palavra "saude" na [Figura 36 (A)], está conectada a muitos outros *clusters* e palavras, como "politica", "trabalhador", "educacao", "escola", "homem", "professor", entre outras, atuando como um nó influente (seu grau a define como destaque na rede). Ela está

conectada a termos que podem fazer parte de diferentes áreas do conhecimento, grandes áreas ou departamentos. As interações das ciências humanas e saúde, por exemplo, indicam que o significado da palavra “saúde”, transcende o conceito de uma só área ou disciplina. Essas conexões indicam a existência de um diálogo multi/interdisciplinar, essencial para compreender a complexidade do tema da saúde (Pombo, 2004; Morin, 2006) confirmando as centralidades das “20+”.

A segunda palavra, “ensino”, está relacionada às ciências humanas, e representa um leque de opções de inserções multi/interdisciplinares. Entre os diferentes *clusters*/palavras relacionadas, pode-se trabalhar ensino em “ecologia”, “literatura”, “ciência” ou “matemática”, com recortes diferentes, mas complementares e relacionados ao sentido da palavra “ensino”. Os *clusters* diretamente conectados à palavra [Figura 36 (B)] ratificam os indícios de uma posição interdisciplinar desta palavra, que em todos os contextos, terá uma o foco relacionado à educação.

O caráter multi/interdisciplinar possível de ser construído com as palavras-chave respondem os objetivos específicos 3: Compreender como os professores e suas temáticas se posicionam e se conectam nas redes de pesquisa da UEFS, e: Identificar temáticas de destaque na rede e/ou que são comuns a pesquisadores de grupos distintos.

A rede de palavras-chave apresenta palavras que frequentemente conectam áreas distintas, como “saúde”, “educação” e “política”, que são mais abrangentes e transversais. Na rede de palavras-chave, termos como “educação física” frequentemente possuem um significado mais direto e autoexplicativo. As palavras-chave que emergem nas redes mostram maior potencial para interdisciplinaridade em relação às palavras dos títulos, já que possuem temas mais gerais e conectores entre diferentes áreas de pesquisa. Observamos que, tanto na rede de títulos como na rede de palavras-chave, para buscar multi/interdisciplinaridade, devemos procurar os termos mais específicos, que não possam ser confundidos quando se modifica o contexto.

Como proposta dos pontos sensíveis das redes de títulos e palavras chaves, apresentamos o Quadro 17.

Quadro 17. Propostas/pontos a serem avaliados das Redes de Títulos e Palavras-chave

Grau Médio	O grau médio é quase 4 para a rede de títulos e quase 3 para a rede de palavras chave. Utilizar mais palavras-chave do núcleo da rede deve aumentar essas conexões. Esse aumento deve oferecer possibilidades de outras áreas discutirem assuntos que estejam relacionados, além de permitir trabalhos conjuntos entre departamentos/áreas distintas.
Diâmetro	Ambas as redes têm valores altos, respectivamente, 15 e 16. Redes esparsas indicando palavras dos títulos e palavras-chave distantes, periféricas. Se a intenção for publicações mais homogêneas nas temáticas e heterogêneas entre departamentos e/ou grandes áreas, utilização de palavras mais próximas do núcleo poderá criar conexões tornando as palavras mais conectadas e aproximando áreas distintas.
Densidade	Valores baixos, nas duas redes indicando apenas 0,7% e 0,9% das conexões possíveis entre as palavras dos títulos e as palavras-chave. Mais conexões às palavras mais próximas ao núcleo da rede podem tornar a rede mais densa.
Modularidade	Nas duas redes, modularidade alta com densidade baixa nas duas redes semânticas indica que, apesar de esparsa, as redes têm <i>clusters</i> fortes com grupos temáticos específicos.
Característica Rede Livre de Escala	Tendência a ter poucas palavras com muitas conexões e poucas palavras com muitas conexões. Avaliar a inserção de títulos e palavras-chave que estejam mais próximas ao núcleo da rede.
Centralidade de Grau/Proximidade e Intermediação	Avaliar a posição das palavras relevantes da rede, seu sentido semântico, contextualização e <i>clusters</i> para investigar possíveis temas multi/interdisciplinares.

Fonte: Elaboração própria, (2025).

4.2.3 Redes Semânticas dos Grupos de Pesquisa

São duas as redes semânticas geradas a partir dos dados dos grupos de pesquisa e como são poucos grupos de pesquisa da UEFS cadastrados no CNPq, são elas: 1. Rede de nomes dos grupos de pesquisa; 2. Redes das linhas de pesquisa dos grupos de pesquisa. A quantidade de dados é pequena, por isso optamos por uma análise voltada para a visualização destas redes, a análise de conteúdo e do contexto no qual estão inseridas (Bardin, 2011; Charmaz, 2009). Para isso utilizamos os softwares VOSviewer e Gephi. Tais redes devem complementar as respostas aos objetivos 3 e 5; que estão relacionados à compreensão e

identificação das temáticas da rede de pesquisa da UEFS. O tratamento temático dos dados segue os mesmos passos das demais redes semânticas da UEFS relacionadas aos currículos lattes (Bardin. 2011).

As redes semânticas dos grupos de pesquisa foram formadas pelos dados coletados de forma manual no site do CNPq e salvos no Excel. Seguindo os seguintes passos: (Bardin, 2011):

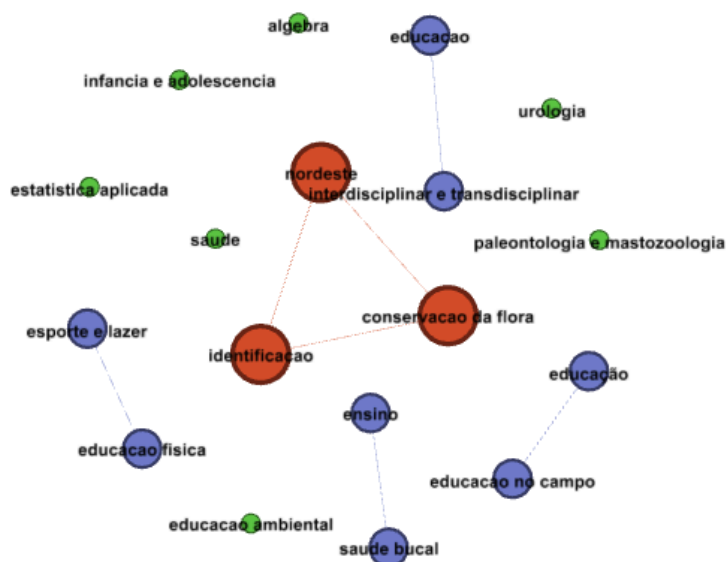
1. Pré-análise: Organização, sistematização e seleção dos nomes dos grupos e suas linhas de pesquisa.
 - Os nomes dos grupos e as linhas de pesquisa de cada grupo foram coletados dos grupos de pesquisa da UEFS via site do CNPq;
2. Exploração: Codificação. Fragmentação e Regras:
 - Não utilizamos coocorrência nem relevância, já que tínhamos um número pequeno de termos;
 - Para a rede dos nomes dos grupos não foi necessário tesauro para limpeza e organização dos dados. Porém, utilizamos o tesauro para a rede de linhas de pesquisa;
3. Interpretação. Comparação e extração de significados:
 - As redes geradas foram analisadas de forma qualitativa; quanto ao seu contexto, visualização e significado.

4.2.3.1 Redes dos Nomes dos Grupos de Pesquisa da UEFS cadastrados no CNPq

Para criar a rede, selecionamos todos os nomes dos 13 grupos de pesquisa da UEFS. Os nós representam os nomes dos grupos, e uma aresta é criada entre a(s) palavra(s) do mesmo grupo ou, se a mesma palavra aparece em outros grupos. A rede criada com os nomes dos grupos de pesquisa da UEFS, tem apenas 18 termos, e mostra os pequenos subgrafos formados pelos nomes destes grupos. Lembrando que podemos ter mais de uma palavra representando apenas um nó, pois o software identifica conexões de palavras com sentido semântico na rede e a coocorrência, caso apareçam sempre juntas. A maioria das palavras estão soltas e o maior grupo de pesquisa tem uma clique, formada pelas palavras; “nordeste”, “conservação da flora” e “identificação”, referente a um grupo das ciências biológicas. Estas

palavras formam parte do nome de um dos grupos, onde todos os nós estão conectados entre si, ou seja, formam um grafo completo dentro da rede. Um tópico para observação é que o grupo que aparece na rede de nomes, das ciências biológicas, não se destacou nas outras redes. Neste caso, o mais provável é que o nome do grupo seja extenso, diferente dos demais grupos que têm nomes pequenos e autoexplicativos. Além disso, podemos destacar a relação entre as palavras “interdisciplinar e transdisciplinar” com a palavra “educação”, cujo próprio nome do grupo já identifica seu perfil inter/transdisciplinar. As palavras “saúde bucal” e “ensino”, já indicam indícios de multi/interdisciplinaridade, ao mesmo tempo que as palavras “educação” e “ensino”, ratificam o potencial multi/inter dessas conexões.

Figura 37. Rede dos Nomes dos Grupos de Pesquisa da UEFS cadastrados no CNPq



Fonte: Elaboração própria (2024)

No nosso estudo de caso, a rede de nomes de grupo de pesquisa agrega muito pouco para as análises. As poucas conexões existentes não traduzem nenhuma outra informação e a maioria aparecem como nós isolados. Por outro lado, pode servir como análise para outros trabalhos ou para outras instituições que que utilizem o modelo e possam ter um maior número de dados. Assim, decidimos manter a análise desta rede, como parte do processo de

distintas relacionadas à educação, odontologia, educação física e até engenharia de computação. Isso acontece porque são poucos grupos (13), então há maior proporcionalidade entre eles, e a rede parece mais balanceada em sua representatividade. No nosso estudo de caso, tal rede deverá servir para direcionar os novos grupos da UEFS em seus estudos, diversificando essas linhas e enriquecendo a rede de pesquisa. Poderá servir também para que gestores possam direcionar linhas e grupos de pesquisa complementares à rede.

Todas as redes e análises encontradas servirão como base para expandir e otimizar a rede de pesquisa e colaboração na UEFS. Este trabalho será direcionado, inicialmente, aos gestores e setores responsáveis da instituição, principalmente às pós-graduações e grupos de pesquisa, para que possam conhecer o panorama de pesquisa da instituição e melhor trabalhar suas publicações, estudos, trabalhos e grupos de pesquisa. Ressaltamos que este trabalho poderá ser utilizado para outras instituições que tenham objetivos parecidos aos desta pesquisa. Para isso, descrevemos o nosso modelo de análise de redes de pesquisa e colaboração científica.

5. MODELO DE ANÁLISE DE REDES DE PESQUISA E COLABORAÇÃO CIENTÍFICA

A partir dos passos seguidos durante o processo do estudo de caso realizado na UEFS fizemos um fluxograma (Figura 39), onde esquematizamos o modelo de análise de pesquisa proposto, que parte do objetivo geral da pesquisa e dos seus objetivos específicos.

O objetivo principal desta pesquisa é desenvolver um modelo de análise de cooperação e colaboração multi e interdisciplinar entre professores utilizando redes de pesquisa (redes de colaboração científica e redes semânticas) para compreender as dinâmicas de colaboração e identificar padrões, estruturas e relações que possam influenciar a produção científica e o fortalecimento da rede multi e interdisciplinar e de disseminação de conhecimento na instituição. Após a identificação do objetivo geral, o modelo é criado para responder a cada um dos objetivos específicos. No nosso estudo de caso, as redes, trabalhadas em conjunto, devem responder os seguintes objetivos:

1. Mapear a rede de professores da UEFS para compreender o processo de colaboração entre eles: através da rede de coautoria e dos professores dos grupos de pesquisa, observamos as comunidades formadas por professores de departamentos distintos nos *clusters* e a colaboração deles. As estatísticas de grau, proximidade, intermediação, componente gigante, comprimento médio do caminho, densidade, modularidade e topologia da rede também mostraram indícios de colaboração e conectividade.

2. Propor possibilidades de aumentar a colaboração, diante da posição dos professores nas redes e/ou grupos de pesquisas: através da rede de coautoria, de professores dos grupos de pesquisa, títulos e palavras-chave, emergiram indícios de colaboração nos *clusters* que se conectam na rede. As análises das medidas de centralidade forneceram informações sobre a posição dos nós e sua influência nas redes, identificando os destaques que podem auxiliar na colaboração. A caracterização estrutural também deu indícios do funcionamento da rede e possíveis pontos fracos.

3. Compreender como os professores e suas temáticas se posicionam e se conectam nas redes de pesquisas da UEFS: as redes de títulos e palavras-chave permitiram a identificação de temáticas predominantes e suas relações, enquanto a rede de coautoria ajudou a posicionar os professores no contexto dos seus grupos de pesquisa. As centralidades foram importantes para responder este objetivo. Palavras que se destacam nas centralidades, após

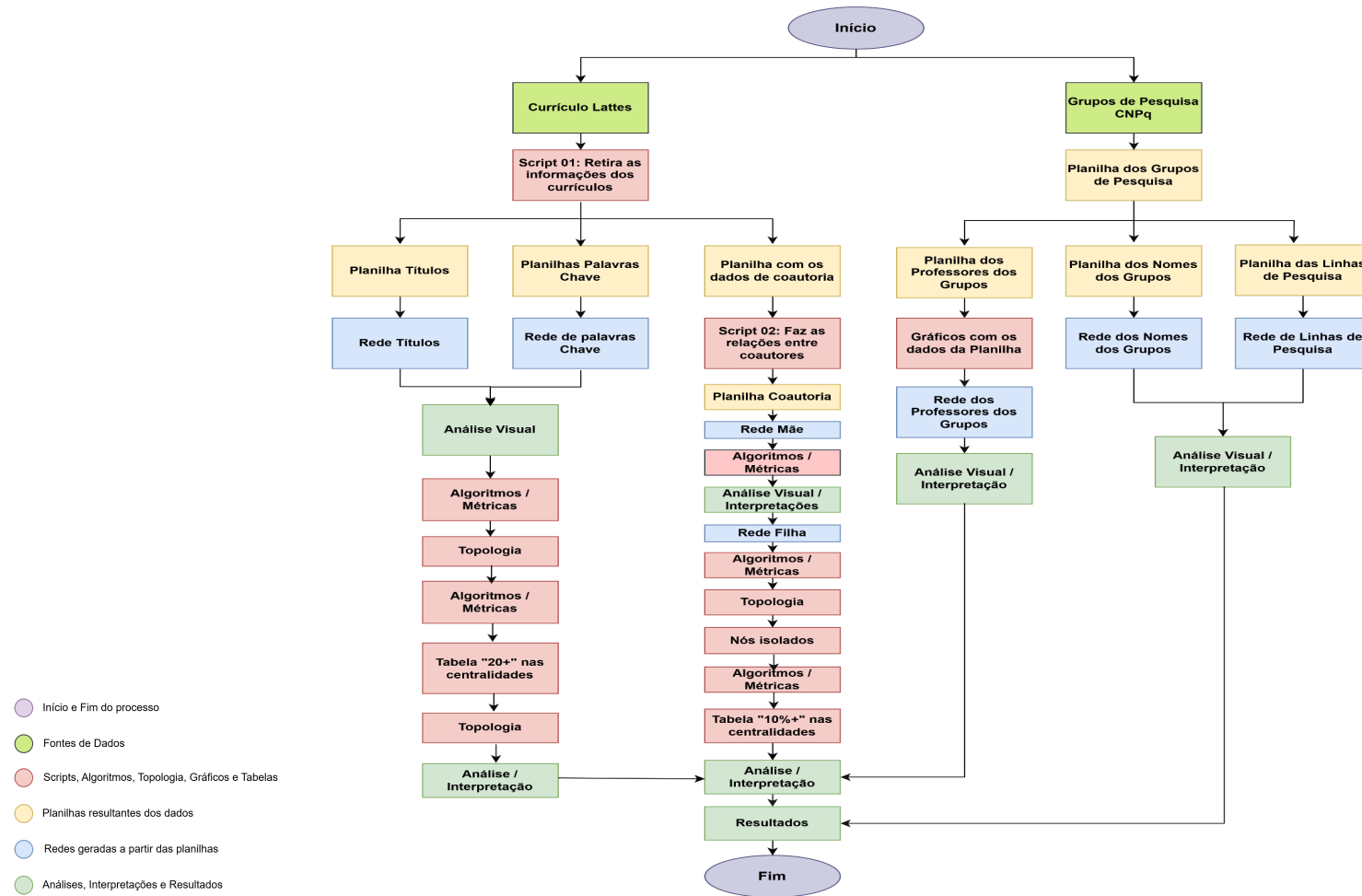
análise de significado e contexto, indicaram (ou não) conexão entre grupos distintos. As redes de nomes dos grupos e linhas de pesquisa, mostraram os tópicos discutidos nos grupos de pesquisa.

4. Identificar os atores das comunidades que colaboram na rede, e seus departamentos/grandes áreas/áreas da UEFS: pela rede de coautoria, idem objetivo 1, através da análise de modularidade, identificamos as comunidades de departamentos distintos. Pelas redes de professores dos grupos de pesquisa, destacamos os professores que mediaram conexões entre grupos distintos na rede.

5. Identificar temáticas de destaque na rede e/ou que são comuns a pesquisadores de grupos distintos: as redes semânticas através das medidas de centralidade identificamos palavras mais frequentes e suas conexões contextualizadas. Através da rede de professores dos grupos e de coautoria também identificamos indícios de conexões entre grupos/departamentos/grandes áreas/áreas distintas, sugerindo colaboração e multi/interdisciplinaridade.

A Figura 39 descreve as interações entre as redes que analisamos durante todo o processo.

Figura 39. Modelo de análise de Redes de Pesquisa e Colaboração da UEFS



Fonte: Elaboração própria, (2025)

O fluxograma descreve o processo de construção e análise de redes acadêmicas a partir de dados extraídos de Currículos Lattes e da base de Grupos de Pesquisa do CNPq. A proposta metodológica combina as fases clássicas de ETL (Extração, Transformação e Carga) com as etapas do processo KDD (Knowledge Discovery in Databases), aplicado ao contexto de ciência de redes e análise institucional acadêmica.

A primeira etapa envolve a obtenção das informações dos currículos lattes dos professores da instituição e dos seus grupos de pesquisa cadastrados no CNPq. Essas fontes foram essenciais para compor os elementos que formam as redes de coautoria, rede de título e palavras-chave. Os dados dos currículos lattes passaram pelo script 01, de pré processamento, para organizar e estruturar as informações em planilhas. O script 02 foi utilizado para relacionar os professores que publicaram juntos, formando uma planilha específica para a rede de coautoria. Outras planilhas foram geradas diretamente ou por algumas transformações/organização e limpeza dos dados: planilhas de títulos, palavras-chave (dos currículos lattes) e a planilha dos grupos de pesquisa.

Com as planilhas organizadas partimos para a carga/transformação dos dados, e então foram geradas as seguintes redes (em amarelo na Figura 39): as redes de coautoria (rede mãe e rede filha), de títulos, de palavras-chave, dos professores dos grupos, dos nomes desses grupos e suas linhas de pesquisa. As redes foram geradas através dos softwares gephi e/ou VOSviewer, através dos quais foram feitas as análises quantitativas e visuais.

A mineração dos dados foi feita através da aplicação dos algoritmos e métricas para avaliação das características das redes. Para isso, utilizamos: o número de nós, número de arestas, grau médio, diâmetro da rede, densidade do grafo, componentes conectados, modularidade, coeficiente de aglomeração e comprimento médio do caminho. Com essas medidas encontramos as três centralidades (grau, intermediação, proximidade), além da topologias (mundo pequeno e livre de escala). As tabelas das centralidades encontradas representam 10% dos nós mais importantes das redes de coautoria (10%+) e as 20 principais palavras ou junção de palavras que aparecem nas redes de títulos e de palavras-chave (20+).

A base conceitual da pesquisa está fundamentada na Teoria das Redes, que entende os fenômenos sociais, científicos e informacionais como sistemas compostos por nós (professores/pesquisadores) e arestas (relações, coautorias, coocorrências). A Análise de

Redes Sociais (ARS) foi adotada para mensurar e interpretar as estruturas de colaboração científica, especialmente por meio das redes de coautoria e redes derivadas dos grupos de pesquisa institucionais. A interpretação dos resultados foi orientada pela Teoria Fundamentada nos Dados (TFD), especialmente na abordagem construtivista de Charmaz (2009) e conforme sistematizado por Tarozzi (2011). Nas redes de coautoria pudemos identificar agrupamentos estruturais (detectados por modularidade) correspondem a agrupamentos temáticos, fortalecendo a interpretação institucional dos dados. As redes semânticas, geradas a partir dos títulos dos artigos e palavras-chave, foram analisadas com base em técnicas de análise de conteúdo (Bardin, 2011) identificando frequência e relevância das palavras e possibilidades de multi/interdisciplinaridade. Dessa forma, foi feita a interpretação e avaliação dos resultados de forma quantitativa e qualitativa, compondo sobre a estrutura e intensidade da colaboração científica na instituição, sua organização, as relações multi e interdisciplinares de suas produções e grupos, bem como dos atores principais e sua representatividade na instituição. As análises visuais funcionaram como complementação e fortalecimento das interpretações obtidas das redes geradas.

Todo esse processo descrito e representado pelo fluxograma, articula de forma sistemática a extração, organização e análise dos dados acadêmicos coletados. A aplicação combinada das fases ETL e KDD possibilita não apenas o mapeamento da produção científica, mas também a identificação de dinâmicas institucionais e redes de colaboração. A metodologia é flexível e replicável, podendo ser aplicada a outras instituições ou contextos de avaliação científica.

6. CONTRIBUIÇÕES E PESQUISAS FUTURAS

O modelo será de especial importância para a UEFS, pois deve contribuir para a análise do processo de pesquisa e oferecer cenários de possibilidades para tornar essas interações mais produtivas e multi/interdisciplinares, possibilitando maior integração dos atores da rede e fluidez das informações no processo de construção do conhecimento mais colaborativo. A implementação de um modelo de análise, deve identificar parcerias na UEFS, reduzir a fragmentação e fortalecer a rede de pesquisa da instituição. A otimização das redes de pesquisa deve focar em estratégias que aumentem a conectividade e a interação multi/interdisciplinar. Medidas como o incentivo à colaboração em publicações, a criação de eventos interdepartamentais e a formação de grupos de pesquisa multi e interdisciplinares podem gerar resultados mais expressivos e diversificados, impactando positivamente a produção científica e a formação acadêmica na UEFS. A proposta de um grupo de pesquisa multi/interdisciplinar para um estudo específico destas redes de pesquisa da UEFS é a intenção inicial para conseguir colocar em prática este trabalho.

Como trabalhos futuros pretendemos acrescentar um script para identificar os dados repetidos da planilha, facilitando o trabalho manual, em relação às planilhas da rede de coautoria e de títulos. Pensamos em disponibilizar a ferramenta para implementar o modelo para utilização da UEFS e/ou outras instituições que possam ter interesse na sua implementação. Ainda como planos futuros, pensamos em disponibilizar um instrumento de atualização dos dados do lattes, assim, para captar novas inserções do lattes, não haverá mais a necessidade de baixar todos os dados novamente para realizar novas análises, apenas atualizar o sistema e fazer os novos estudos. Esse recurso é muito importante para os próximos passos, porque, após a análise inicial, a universidade poderá acompanhar as mudanças nas redes, os pontos positivos e negativos para melhoria da rede de pesquisa da comunidade científica.

Além da aplicação no modelo na UEFS, a proposta pode ser de relevância para implementação similar em sistemas governamentais responsáveis por gerenciar e disseminar informações de pesquisa científica no Brasil. Podemos citar aqui o BrCris, do qual falamos no início desta pesquisa. O “Sistema de Informação de Pesquisa Corrente acrônimo CRIS” foi lançado com o intuito de disponibilizar informações sobre instituições de pesquisa. Porém, um modelo de análise como o que propomos neste trabalho, pode ser de grande auxílio nas

análises de pesquisas nas instituições, disponibilizando ferramentas que possam unificar a forma de avaliação da pesquisa nas instituições de ensino superior do Brasil.

7. LIMITAÇÕES DO ESTUDO

Identificamos algumas limitações encontradas no decorrer da pesquisa, que podemos aqui destacar. Uma delas é que, apesar dos dados serem públicos, o sigilo desses dados coletados da plataforma lattes pode limitar a divulgação direta dos sujeitos que fazem parte dos resultados. Da mesma forma que temos o ponto positivo ético de não expor os dados desses pesquisadores, tais informações poderiam auxiliar, de forma mais incisiva, nas mudanças que podem ser feitas para obtenção de melhores resultados na divulgação e produção científica.

Outro limitante identificado durante a pesquisa foi a dificuldade quanto à limpeza dos dados após a extração. São muitas informações que se repetem, onde professores cadastram a “citação” dos seus nomes, de várias formas diferentes, dificultando o processo, pois a maioria não identifica as citações no seu próprio lattes. A pesquisa pode trazer resultados imediatos para a UEFS, porém, a inserção de scripts para facilitar esse trabalho repetitivo é importante para os próximos trabalhos.

8. CONSIDERAÇÕES FINAIS

O objetivo da nossa pesquisa é desenvolver um modelo de análise de cooperação e colaboração multi e interdisciplinar. Utilizamos como estudo de caso, as redes de pesquisa (redes de colaboração científica e redes semânticas) da UEFS para compreender as dinâmicas de colaboração e identificar padrões, estruturas e relações que possam influenciar a produção científica e o fortalecimento da rede multi/interdisciplinar e de disseminação de conhecimento na instituição.

Os dados utilizados para as análises tiveram duas fontes, a primeira, os currículos lattes dos professores da UEFS. A plataforma Lattes foi criada pelo CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) e disponibiliza currículos a qualquer pessoa que deseje, de forma gratuita e virtual. As Instituições de Ensino Superior utilizam a plataforma para guardar os seus dados no currículo lattes e sua utilização é comum na comunidade científica do Brasil. Os currículos dos professores da UEFS foram baixados em “xml” e os dados que poderiam ser utilizados para análise foram capturados pelo script 01 e transformados em planilhas brutas, depois organizadas de forma a serem utilizadas para gerar as redes sociais e redes semânticas. Com estes dados foram geradas as redes e coautoria mãe (RM), a rede de coautoria filha (RF), a rede de títulos e a rede de palavras-chave.

As universidades possuem muitos meios de fazer pesquisa, incentivadas pelas próprias universidades ou pelas agências de apoio e fomento à pesquisa. Porém, esses dados nem sempre estão disponíveis ao público, o que dificulta o acesso a essas informações. Como segunda fonte de dados optamos por uma análise com as informações dos grupos de pesquisa da UEFS que fossem cadastrados no CNPq, já que esses dados são públicos e de fácil acesso. Através dos dados dos grupos de pesquisa, as informações foram selecionadas de forma manual e criamos as redes de professores da UEFS que participavam dos grupos, a rede dos nomes e a rede de linhas de pesquisa.

O modelo proposto foi mapeado partindo da teoria dos grafos, de forma a descrever as relações de colaboração e pesquisa. Todas as redes foram geradas no VOSviewer e/ou gephi, ambos são softwares livres utilizados para visualização das redes e cálculos das estatísticas a serem analisadas dentro do contexto de análise de colaboração, multi e interdisciplinaridade e disseminação do conhecimento.

Através da Análise de Redes Sociais (ARS) e complexas, uma análise interpretativa,

indutiva e iterativa, análise de conteúdo e contextualização das redes semânticas, analisamos quais atores podem interferir nas redes e a relevância de parcerias estratégicas, buscando responder os seguintes objetivos específicos: mapear as redes de pesquisa para compreender o processo de colaboração entre os professores; propor possibilidades de aumentar a colaboração diante da posição desses autores na rede; compreender como os professores e suas temáticas se posicionam e se conectam nas redes; identificar os atores das comunidades que colaboram na rede e seus respectivos setores e identificar temáticas de destaque na rede e/ou que são comuns a pesquisadores de grupos distintos.

A multi/interdisciplinaridade e a disseminação do conhecimento não dependem apenas de indivíduos ou grupos isolados, mas da conexão entre todos os componentes da rede acadêmica e as redes “falam” isso de várias formas diferentes, por isso são um excelente ponto de partida para análise e intervenção, para promover e fortalecer esses vínculos, um passo fundamental para transformar a pesquisa em um instrumento democrático e poderoso de inovação e desenvolvimento social.

Porém, não é tão simples “ouvir” o que elas insistem em gritar! A utilização de uma área interdisciplinar (Ciência das Redes), no trabalho de um curso da área de análise cognitiva (Doutorado em Difusão do Conhecimento) também de natureza interdisciplinar, revela o caráter complexo de descrever o processo cognitivo de um modelo que responda ao objetivo de uma pesquisa. Descrever esse processo em meio a tantas complexidades esbarra nas limitações humanas, primeiro no seu entendimento e depois na sua descrição.

Foi o que tentamos fazer nesta pesquisa, e esperamos ter minimamente contribuído para a descrição do processo de pesquisa da instituição. Desde o entendimento de como o conhecimento é distribuído na rede de pesquisa e pesquisadores, como ele flui na rede, a interpretação das interações que emergem das redes mapeadas, a identificação de atores e temáticas que participam do processo, e até os que estão fora dele. A visualização de pequenos indícios de multi e interdisciplinaridade e as possíveis soluções para aumentar a colaboração e a disseminação do conhecimento, as propostas de melhorias nas colaborações e o fomento à multi/interdisciplinaridade, exigem uma interpretação cuidadosa de como os pesquisadores acessam, utilizam e compartilham informações, conectando dados das redes a decisões estratégicas.

A contribuição à UEFS foi dada e representa muito para nós como comunidade científica. A rede de pesquisa da UEFS só tem a ganhar com este trabalho colaborativo.

REFERÊNCIAS

ABBAS, G. **A review of “Random graphs and complex networks” by Hofstad**. *Complex Adapt Syst Model* 5, 11 (2017). Disponível em: <https://doi.org/10.1186/s40294-017-0050-2> Acesso em: 25 set. 2022.

ALBAGLI, Sarita. **Ciência Aberta em questão**. In: Seminário Internacional Ciência Aberta, Questões Abertas. Rio de Janeiro, 2014. Trabalho apresentado. Rio de Janeiro: Liinc; IBICT; OKF; Unirio, 2014. Disponível em: <http://www.cienciaaberta.net/encontro2014/> Acesso em: 25 set. 2022.

BARABÁSI, A. L.; ALBERT, R. Emergence of scaling in random networks. **Science**, v. 286, p. 509-512, 1999.

BARABÁSI, A. L. **Linked: How everything is connected to everything else and what it means for business, science and everyday life**. New York: Basic Books, 2002.

BARABÁSI, Albert-László. **Linked: How Everything Is Connected to Everything Else and What It Means for Business, Science, and Everyday Life**. Plume, 2003.

BARABÁSI, Albert-László. **Network Science**. Cambridge University Press, 2016. Disponível em: <https://networksciencebook.com> Acesso em: 12 jan. 2024.

BARDIN, Laurence. **Análise de conteúdo**. São Paulo: Edições 70, 2011.

BERTALANFFY, Ludwig Von. **General System Theory: Foundation, Development, Applications**. George Braziller, 1968.

Boletim Anual OCTI. Centro de Gestão e Estudos Estratégicos. Brasília: Centro de Gestão e Estudos Estratégicos. Ano 3 - junho de 2023. Disponível em: [CGEE_OCTI_Boletim_Anual_do_OCTI_2022.pdf](#) Acesso em: 8 dez. 2023.

BONACICH, P. Factoring and weighting approaches to status scores and clique identification. **Journal of Mathematical Sociology**, 2(1), 113-120, 1972.

BORNMANN, L.; LEYDESDORFF, L. Scientometrics in a changing research landscape. **EMBO Reports**, 15(12), 1228-1232. 2014. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC4264924/pdf/embr0015-1228.pdf> Acesso em: 28 fev. 2024.

BRANDES, Ulrik. A Faster Algorithm for Betweenness Centrality. **Journal of Mathematical Sociology**, 25(2), 163–177, 2001.

BRASIL. **Lei de Diretrizes e Bases da Educação Nacional**, LDB. 9394/1996. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l9394.htm Acesso em: 16 out. 2021.

BRASIL. CAPES. 2022. Disponível em :
<https://www.gov.br/capes/pt-br/aceso-a-informacao/acoes-e-programas/avaliacao/sobre-a-avaliacao/areas-avaliacao/sobre-as-areas-de-avaliacao/sobre-as-areas-de-avaliacao> Acesso em: 9 jun. 2022.

BRASIL. Ministério da Ciência, Tecnologia e Inovação – MCTI, 2023. **CGEE apresenta panorama da produção científica no Brasil e no mundo.** Disponível em:
<https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/noticias/2023/06/cgee-apresenta-panorama-da-producao-cientifica-no-brasil-e-no-mundo> Acesso em 8 nov. 2023.

BRASIL. **Plataforma Lattes.** 2023. Disponível em:
<https://www.gov.br/cnpq/pt-br/aceso-a-informacao/acoes-e-programas/plataforma-lattes>. Acesso em 14 set. 2023.

BRASIL. **Ministério da Educação.** Disponível em: <https://www.gov.br/mec/pt-br> Acesso em 4 fev. 2021.

BRASIL. MEC. **CEB nº 3, de 26 de junho de 1998.** 2018. Disponível em:
http://portal.mec.gov.br/cne/arquivos/pdf/rceb03_98.pdf Acesso em: 20 set. 2023.

BRASIL. **Exame Nacional de Desempenho dos estudantes - ENADE.** Disponível em:
<https://enade.inep.gov.br/enade/> Acesso em: 25 fev. 2023.

BRASIL. **Conselho Nacional de Saúde, Ministério da Saúde.** Disponível em:
<https://conselho.saude.gov.br/comissoes-cns/conep#:~:text=O%20Sistema%20CEP%2FConep%20%C3%A9,dispostas%20em%20todo%20territ%C3%B3rio%20brasileiro> Acesso em: 12 jul. 2021.

BRASIL, Gov.br - **Educação e Pesquisa.** Disponível em:
<https://www.gov.br/pt-br/servicos/cadastrar-grupos-de-pesquisa-na-plataforma-lattes> Acesso em: 16 ago. 2024.

BRASIL. **Sistema Nacional de Avaliação da Educação Superior - SINAES.** Disponível em:
<http://portal.mec.gov.br/component/content/270-programas-e-acoes-1921564125/sinaes-2075672111/12303-sistema-nacional-de-avaliacao-da-educacao-superior-sinaes> Acesso em: 28 fev. 2023.

BRASIL. **Plataforma Sucupira.** Disponível em: <https://sucupira.capes.gov.br/sucupira/> Acesso em: 27 nov. 2022.

BRASIL. **O que é o gov.br?** Disponível em: <https://www.gov.br/pt-br/sobre> Acesso em: 29 out. 2022.

BrCris. **Ecossistema de Informação de Pesquisa Científica Brasileira.** Disponível em:
<https://brcris.ibict.br/> Acesso em: 12 jul. 2024.

BORTOLIERO, S. O papel das universidades na promoção da cultura científica: formando jornalistas científicos e divulgadores da ciência. In PORTO, CM., org. **Difusão e cultura científica: alguns recortes** [online]. Salvador: EDUFBA, 2009. p. 45-73. ISBN

978-85-2320-912-4. Available from SciELO Books.

BURT, R.S. **Structural holes**: the social structure of competition. Cambridge: Harvard University Press, 1992.

CAMILO, C. O.; SILVA, J. C. **Mineração de dados**: Conceitos, tarefas, métodos e ferramentas. UFG - Universidade Federal de Goiás. 2009. Disponível em: http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_001-09.pdf Acesso em: 14 abr. 2024.

CAPES. **Sobre as áreas de avaliação**. Disponível em: <https://www.gov.br/capes/pt-br/aceso-a-informacao/acoes-e-programas/avaliacao/sobre-a-avaliacao/areas-avaliacao/sobre-as-areas-de-avaliacao/sobre-as-areas-de-avaliacao> Acesso em: 14 out. 2022.

CAPES. **Tabela de áreas de conhecimento/avaliação**. Disponível em: <https://www.gov.br/capes/pt-br/aceso-a-informacao/acoes-e-programas/avaliacao/instrumentos/documentos-de-apoio-1/Tabela-de-areas-de-conhecimento-avaliacao> Acesso em: 14 out. 2022.

CARTA CAMPINAS. **Verdade 06**: as universidades públicas são responsáveis por 90% das pesquisas no Brasil. 06 jul. 2019. Disponível em <https://cartacampinas.com.br/2019/07/verdade-6-as-universidades-publicas-sao-responsaveis-por-90-das-pesquisas-no-brasil/> Acesso: 06 set. 2023.

CASTELLS, Manuel. **The Network Society**: A Cross-Cultural Perspective. Edward Elgar Publishing, 2004.

CASTELLS, Manuel. **A Sociedade em Rede**, Volume I, 9ª. Edição. São Paulo: Editora Paz e Terra, 2006. Tradução: Roneide Venancio Majer com a colaboração de Klauss Brandini Gerhardt.

CHARMAZ, K. **A construção da teoria fundamentada**: guia prático para análise qualitativa. Tradução: Joice Elias Costa. Porto Alegre, Artmed. 2009.

CNPq. **Currículo Lattes**. (s.n). Disponível em: <https://lattes.cnpq.br/> Acesso em: 21 de set. 2021.

CNPq. **Diretório dos Grupos de Pesquisa (DGP) - Consulta Parametrizada**. (s.n). Disponível em: https://dgp.cnpq.br/dgp/faces/consulta/consulta_parametrizada.jsf. Acesso em: 30 out. 2023.

CNPq. **Diretório dos Grupos de Pesquisa (DGP) - Lattes**. (s.n). Disponível em: https://dgp.cnpq.br/dgp/faces/consulta/consulta_parametrizada.jsf Acesso em: 30 out. 2023.

CNPq. **Diretório dos Grupos de Pesquisa (DGP) - O que é**. (s.n). Disponível em: <https://lattes.cnpq.br/web/dgp/o-que-e> Acesso em: 30 out. 2023.

CNPq. **Diretório dos Grupos de Pesquisa (DGP) - Quem pode participar**. (s.n).

Disponível em: <https://lattes.cnpq.br/web/dgp/quem-pode-participar> Acesso em: 30 out. 2023.

CONSELHO NACIONAL DE EDUCAÇÃO. **Resolução CNE/CES n.º 3, de 2 de julho de 2007**. Institui Diretrizes Curriculares Nacionais para o curso de graduação. Diário Oficial da União: Brasília, DF, 2007.

CREESE, A; DANIELS, H.; NORWICH, B. **Teacher Support Teams in Primary and Secondary Schools**. London: Fulton, 1997.

Cursos de Graduação, Oferta Regular. Disponível em: <https://feiradegraduacao.uefs.br/cursos-de-graduacao/> Acesso em: 20 jan. 2024.

DENNETT, D. C.. **A consciência explicada**. Porto Alegre: Bookman, 2004.

DIAS, T. M. R. *et al.* BrCris: plataforma para integração, análises e visualização de dados técnicos-científicos. **Revista Informação & Informação**, Londrina, v. 27, n. 3, p. 622–638, jul./set. 2022. Disponível em: https://doi.org/10.1007/978-3-030-02511-3_25 Acesso em: 27 mar. 2023.

EDUCAMAISBRASIL, 2019a. **Por que a universidade pública é mais concorrida?** Disponível em: <https://www.educamaisbrasil.com.br/educacao/noticias/por-que-a-universidade-publica-e-mais-concorrida> Acesso em 18 nov. 2021.

EDUCAMAISBRASIL, 2019b. **Parte da linguística que estuda o significado das palavras**. Disponível em: <https://www.educamaisbrasil.com.br/enem/lingua-portuguesa/semantica> Acesso em: 18 nov. 2021.

ERDŐS, P., & RÉNYI, A. **On the Evolution of Random Graphs**. Publications of the Mathematical Institute of the Hungarian Academy of Sciences, 5, 17-61. 1960. Disponível em: <https://snap.stanford.edu/class/cs224w-readings/erdos60random.pdf> Acesso em: 04 jun. 2024.

FADIGAS, Inácio Souza. **Produção e difusão do conhecimento em educação matemática sob perspectiva das redes sociais e complexas**. 2011. 200p. Tese (Doutorado Multiinstitucional e Multidisciplinar de Difusão do Conhecimento), Faculdade de Educação, UFBA, Salvador, 2011.

FAVARÃO, N. R. L.; ARAÚJO, C. S. A. Importância da Interdisciplinaridade no Ensino Superior. **EDUCERE** - Revista da Educação, p. 103-115, vol. 4, n.2, jul./dez., 2004.

FAYYAD, U. M.; PLATETSKY-SHAPIO, G.; SMYTH, P. **From Data Mining to Knowledge Discovery: An Overview**. Knowledge Discovery and Data Mining, Menlo Park: AAAI Press, 1996.

FAZENDA, I. C. A. **Interdisciplinaridade: história, teoria e pesquisa**. Campinas, SP: Papirus, 1994.

FILHO, José da Silva. **Criação de uma nova ferramenta informativa sobre Teoria dos Grafos**. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação), Centro Universitário Luterano de Palmas, Palmas: CEULP/ULBRA, 2017.

FISHER, Ronald A. **Statistical Methods for Research Workers**. Oliver and Boyd, 1925.

FREEMAN, L. C.. **Centrality in social networks: Conceptual clarification**. Social Networks, 1(3), 215–239, 1978.

FREEMAN, L. C.. **Centrality in social networks: Conceptual clarification**. Social Networks, 1(3), 215–239, 1979. DOI: 10.1016/0378-8733(78)90021-7

FREEMAN, Linton C. **The Development of Social Network Analysis: A Study in the Sociology of Science**. Empirical Press, 2004.

GEPHI 0.10.1 (Site Oficial). **Gephi Makes graphs handy**. Disponível em: <https://gephi.org/users/> Acesso em: 18 fev. 2024.

GEPHI 0.10.1 (Software). Disponível em: <https://gephi.org/users/> Acesso em: 18 fev. 2024.

GEPHI 0.10.1 (Tutorial). Disponível em: https://gephi.org/tutorials/gephi-tutorial-quick_start.pdf Acesso em: 14 mar. 2024.

GERSTING, Judith L. **Fundamentos Matemáticos para a Ciência da Computação**. 3.ed., Rio de Janeiro: LTC Editora, 1995.

GIL, Antônio Carlos. **Métodos e técnicas de pesquisa social**. 6.ed. São Paulo: Atlas, 2008.

GIMENEZ, Ana Maria Nunes; BONACELLI, Maria Beatriz Machado. **Repensando o Papel da Universidade no Século XXI: Demandas e Desafios**. Revista Tecnologia e Sociedade. v.9, n.18, 2013.

GLUCKMAN, Max. **Custom and Conflict in Africa**. Oxford; Blackwell, 1955.

GOMES, Caio Cesar Piffero. **O papel social da universidade**. XIV Colóquio Internacional de Gestão Universitária – CIGU. A Gestão do Conhecimento e os Novos Modelos de Universidade. Florianópolis. Dez de 2014. Disponível em: <https://repositorio.ufsc.br/xmlui/bitstream/handle/123456789/131807/2014-175.pdf> Acesso em 20 jan. 2021.

GRAY, B.. **Collaborating: Finding Common Ground for Multiparty Problems**. Jossey-Bass, 1989.

INGOLD, T. **Chega de etnografia!** A educação da atenção como propósito da antropologia. Educação, [S. l.], v. 39, n. 3, p. 404–411, 2016. DOI: 10.15448/1981-2582.2016.3.21690. Disponível em: <https://revistaseletronicas.pucrs.br/ojs/index.php/faced/article/view/21690> Acesso em: 4 fev. 2023.

JAPIASSU, Hilton. **Interdisciplinaridade e patologia do saber**. Rio de Janeiro: Imago, 1976.

Jornal da USP. **10 mitos sobre a universidade pública no Brasil**. 28 jun. 2019. Disponível em: <https://jornal.usp.br/universidade/10-mitos-sobre-a-universidade-publica-no-brasil/> Acesso em: 8 abr. 2023.

KIMBALL, R.; CASERTA, J. **The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data**. Wiley, 2004.

KLEINBERG, Jon M. **Authoritative Sources in a Hyperlinked Environment**. 1998. Disponível em: <https://cse.msu.edu/~cse960/Papers/LinkAnalysis/auth.pdf> Acesso em: 04 mar. 2023.

KRIPPENDORF, K. **Content analysis: An introduction to its methodology**. Beverly Hills, CA: Sage, 1980.

LAKATOS, Eva Maria; MARCONI, Marina de Andrade. **Fundamentos de Metodologia Científica**. 8. ed., São Paulo: Atlas, 2017.

LATOUR, Bruno. **Reagregando o Social: uma introdução à Teoria do Ator-Rede**. Tradução: Gilson César Cardoso de Sousa. Editora EDUFBA-EDUSC. Salvador- Bauru. 2012.

LUIZ, Isabela das Chagas. E-book - **VOSviewer: Tutorial para iniciantes**. Fundação Universidade do Estado de Santa Catarina, 2022. Disponível em: https://www.udesc.br/arquivos/cct/documentos/E_book_VOSviewer_17181006477198_10724.pdf Acesso em: 22 out. 2024.

ECK, N. J. V.; Waltman, L. **Manual do VOSviewer 1.6.20**. Universiteit London, 2023. Disponível em: https://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.20.pdf Acesso em: 12 mar. 2024.

MARTINS, D. L.; FERREIRA, S. M. S. P. Mapeamento e avaliação da produção científica da Universidade de São Paulo com foco na estrutura e dinâmica de suas redes de colaboração científica. **Liinc em Revista**, v. 9, n. 1, p. 181-195, 2013.

MATIAS, Mesailde de Souza Oliveira; AMARAL, Roniberto Morato do; MATIAS, Paulo. **Proxy customizado para acesso ao web service da Plataforma Lattes**. 2017. Disponível em: https://www.researchgate.net/profile/Paulo-Matias-4/publication/317348747_Proxy_customizado_para_acesso_ao_web_service_da_Plataforma_Lattes/links/59356ab70f7e9beee7da87e3/Proxy-customizado-para-acesso-ao-web-service-da-Plataforma-Lattes.pdf Acesso em: 20 jun. 2020.

MELO, José Damião de. **Modelagem de redes de cooperação científica na educação profissional e tecnológica**. 2022. 135p. Tese (Doutorado em Difusão do Conhecimento) - Faculdade de Educação, Universidade do Estado da Bahia. Salvador, 2022.

MILGRAM, S. The small world problem. **Psychology Today**, 1(1), p. 61-67, 1967.

MORIN, Edgar. **Introduction to Complex Thought**. Seuil, 1990.

MORIN, Edgar. **A cabeça bem-feita: repensar a reforma, reformar o pensamento**. Tradução Eloá Jacobina - 8a. ed. - Rio de Janeiro: Bertrand Brasil, 2003.

NASCIMENTO, J. B. *et al.* A teoria fundamentada em dados aplicada ao campo da educação superior. **Research, Society and Development**, v. 10, n.5, 2021.

NASCIMENTO, A. G. **Altmetria para bibliotecários: guia prático de métricas alternativas para avaliação da produção científica**. Porto Alegre: Revolução eBook, 2016. Disponível em: <http://revolucaoebook.com.br/ebook/altmetria-para-bibliotecarios-isbn-9788569333821/>. Acesso em: 5 fev. 2021.

NEWMAN, M. E. J., & GIRVAN, M. (2004). Finding and evaluating community structure in networks. **Physical Review E**, 69(2), 026113. DOI: 10.1103/PhysRevE.69.026113. Disponível em: https://www.cse.cuhk.edu.hk/~cslui/CMSC5734/newman_community_struct_networks_phys_rev.pdf (cuhk.edu.hk). Acesso em: 2 jun. 2024.

NEWMAN, M. E. J. Scientific collaboration networks. I. Network construction and fundamental results. **Physical Review E**, v.98, n.2, p.404-409, 16 jan. 2001.

NEWMAN, M. E. J. Scientific collaboration networks. I. Network construction and fundamental results. **Physical Review E**, v.64, p. 016131, 28 jun. 2001a.

NEWMAN, M. E. J. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. **Physical Review E**, v. 64, p. 016132, 28 jun. 2001b.

NEWMAN, M. E. J. (2003). The structure and function of complex networks. **SIAM Review**, 45(2), 167–256, v. 45, Iss.2 (2003). DOI: 10.1137/S003614450342480.

NEWMAN, M. E. J. **Networks: An Introduction**. Oxford: Oxford University Press, 2010.

ORTEGA, Fábio Silva; BRANDÃO, Carlos da Fonseca. A história da pós-graduação no Brasil e a construção do espaço acadêmico científico da educação. **Revista Educação em Foco**, ano 23, n.39 - jan/abr. 2020 - p.249-269. Belo Horizonte - Minas Gerais.

PETRAGLIA, I. C. **Interdisciplinaridade**. São Paulo. Pioneira. 1993.

PIMENTA, S. G.; ANASTASIOU, L. G. C.. **Docência no ensino superior**. Coleção Docência em Formação, 2. ed. - São Paulo: Cortez, 2005.

POMBO, O. **Interdisciplinaridade: Ambição e Fracasso**. Lisboa: Fundação Calouste Gulbenkian. 2004.

POMBO, O. Interdisciplinaridade e integração dos saberes. **Liinc em Revista**, [S. l.], v. 1, n. 1, 2006. DOI:10.18617/liinc.v1i1.186. Disponível em: <https://revista.ibict.br/liinc/article/view/3082>. Acesso em: 27 jun. 2023.

PRIGOL, Edna Liz; BEHRENS, Marilda Aparecida. Teoria Fundamentada: metodologia aplicada na pesquisa em educação. **Revista Educação e Realidade** 44 (3), 2019. Disponível em: <https://www.scielo.br/j/edreal/a/rDDFsHvBCQyWKpthzPjMGzk/> Acesso em: 6 fev. 2022.

RECUERO, Raquel. **Introdução à análise de redes sociais**. Coleção Ciberultura. Salvador: EDUFBA, 2017.

Revista Pesquisa Fapesp. **Produção científica sobre Covid-19 afeta o equilíbrio da geração de conhecimento**. Edição 318, agosto de 2022. Disponível em: <https://revistapesquisafapesp/producao-cientifica-sobre-covid-19-afeta-o-equilibrio-da-geracao-o-de-conhecimento/>. Acesso em: 17 jan. 2025.

RIBEIRO, Flávia Nascimento. Edgar Morin, o pensamento complexo e a educação. **Pró-Discente**: Caderno de Prod. Acad.-Cient. Progr. Pós-Grad. Educ., Vitória, v. 17, n. 2, jul./dez. 2011. Disponível em: <https://periodicos.ufes.br/prodiscente/issue/view/404> Acesso em: 2 out. 2022.

ROSEN, Kenneth H. **Matemática Discreta e Suas Aplicações**. Tradução técnica: Helena Castro, João Guilherme Giudice - 6ª. ed. - Porto Alegre: AMGH, 2010.

RUSSO, Luis. Quem sobreviverá? Fundamentos da Sociometria, da psicoterapia de grupo e do sociodrama. **Revista Brasileira de Psicodrama**. vol.18 - no.2 - São Paulo: 2010.

SANTOS, Francisco Coelho dos; CYPRIANO, Cristina Petersen. Redes Sociais, Redes de Sociabilidade. **Revista brasileira de Ciências Sociais**. 29 (85). Junho de 2014. Disponível em <<https://www.scielo.br/j/rbcsoc/a/k5ykGdRVvtzwfCq9Twh6ZGq/?lang=pt>> Acesso em: 23 fev. 2023.

SEVERINO, Antônio Joaquim. **Metodologia do Trabalho Científico**. 1ª ed. Eletrônica. São Paulo: Cortez Editora, 2014. Disponível em: https://www.ufrb.edu.br/ccaaab/images/AEPE/Divulga%C3%A7%C3%A3o/LIVROS/Metodologia_do_Trabalho_Cient%C3%ADfico_-_1%C2%AA_Edi%C3%A7%C3%A3o_-_Antonio_Joaquim_Severino_-_2014.pdf Acesso em: 26 mar. 2023.

SILVA, Franklin Leopoldo e. Teoria do conhecimento. In: CHAUÍ, Marilena *et al.* **Primeira Filosofia. Lições Introdutórias**. 2 ed., p.175-195. São Paulo: Editora Brasiliense, 1987.

SILVA, Larissa Paula Ferreira de. **Interdisciplinaridade, transdisciplinaridade e multidisciplinaridade**: reflexões teóricas sobre a prática escolar. 2018. 26 f. Trabalho de Conclusão de Curso (Graduação em Língua Portuguesa) - Centro de Ciências Humanas e Exatas, Universidade Estadual da Paraíba, Monteiro. Disponível em: <https://dspace.bc.uepb.edu.br/jspui/bitstream/123456789/17572/1/PDF%20-%20Larissa%20P aula%20Ferreira%20de%20Sousa.pdf> Acesso em: 25 set. 2023.

SILVA, Wagner Rodrigues. Construção da Interdisciplinaridade no Espaço Complexo de Ensino e Pesquisa. **Revista Eletrônica Outros Temas**. Cadernos de Pesquisa. v.41 n.143 maio/ago, 2011. Disponível em: <https://www.scielo.br/j/cp/a/GYGJGyQhgStnPsTMNQ48bZb/?format=pdf&lang=pt> Acesso em: 6 fev. 2022.

Software Gephi 0.10.1. Disponível em: <https://gephi.github.io/> Acesso: 4 mai. 2021.

SOUZA, Jansen Cruz de. **Utilizando Títulos de Artigos Científicos na Construção de Redes Semânticas para Caracterizar Áreas de Pesquisa**. 2015. 82 p. Dissertação (Mestrado em Informática) - Centro de Informática, Universidade Federal da Paraíba, João Pessoa 2015. Disponível em: <https://repositorio.ufpb.br/jspui/bitstream/tede/7858/2/arquivototal.pdf>. Acesso em: 14 de out. de 2023.

TARJAN, R. E. Depth-first search and linear graph algorithms. **SIAM Journal on Computing**, v. 1, n. 2, p. 146–160, 1972.

TAROZZI, Massimiliano. **O que é Grounded Theory**: metodologia de pesquisa e de teoria fundamentada nos dados. Tradução: Carmem Lussi. Petrópolis, RJ. Vozes, 2011.

TEIXEIRA, E. F. B. Emergência da inter e da transdisciplinaridade na universidade. In: AUDY, J. L. N.; MOROSINI, M. C. (Org.). **Inovação e interdisciplinaridade na universidade**. Porto Alegre: EDIPUCRS, 2007. p. 58-80.

TREVISOL, J. V; GARMUS, R. (2021). O princípio da autonomia na universidade brasileira: sentidos em disputa. **Revista Brasileira de Política e Administração da Educação** - Periódico científico editado pela ANPAE, 37(1), 307–326. Disponível em: <https://doi.org/10.21573/vol37n12021.105947> Acesso em: 12 set. 2023.

TUCKMAN, B. W.. The Impact of Knowledge Dissemination on Educational Practice. **The Journal of Educational Research**, 80(3), 146–150, 1987.

UEFS, 2020. **Cursos**. Disponível em: <https://www.uefs.br/modules/conteudo/conteudo.php?conteudo=16>. Acesso em: 17 out. 2021.

UEFS. **Atividades Complementares**. Disponível em: <https://drive.google.com/file/d/1SV9Qh9cW9Qbo3fWcjgMhtmj52Hc--Nx0/view> Acesso em: 9 fev. 2023.

UEFS, **Cursos de Graduação**. Disponível em: <https://feiradegraduacao.uefs.br/cursos-de-graduacao/> Acesso em: 17 out. 2024.

UEFS. **Cursos Lato Sensu**. Disponível em: <https://www.uefs.br/modules/conteudo/conteudo.php?conteudo=10> Acesso em: 17 out. 2024.

UEFS. **Cursos Stricto Sensu**. Disponível em:

<https://www.uefs.br/modules/conteudo/conteudo.php?conteudo=11> Acesso em: 17 out. 2024.

UEFS. **Departamentos**. Disponível em:

<https://www.uefs.br/modules/conteudo/conteudo.php?conteudo=18> Acesso em: 6 nov. 2021.

UEFS. **Nossa História**. Disponível em:

<https://www.uefs.br/modules/conteudo/conteudo.php?conteudo=12> Acesso em: 12 maio 2022.

UEFS. **Uefs é melhor estadual baiana no ranking QS das universidades latinoamericanas**. Disponível em:

<https://servidores.rhbahia.ba.gov.br/noticias/2023-09-14/uefs-e-melhor-estadual-baiana-no-ranking-q-s-das-universidades-latinoamericanas> Acesso em: 2 out. 2023.

UFBA. **Comitê de Ética em Pesquisa**. Instituto de Ciências da Saúde - UFBA. Disponível em: <http://cep.ics.ufba.br/duvidas.html#content> Acesso em: 10 jan. 2022.

UFES. **Gephi** - Um software de manipulação e visualização de grafos. Oficina Gephi - Mapeando e analisando a vida das redes sociais - Labic (Laboratório de Estudos sobre Imagem e Cibercultura - Departamento de Comunicação Social. (Apostila) [s.d].

WASSERMAN, Stanley; FAUST, Katherine. **Social Network Analysis: Methods and Applications**. Cambridge University Press, 1994.

WATTS, Duncan J.; STROGATZ, Steven H. Collective dynamics of 'small-world' networks. **Nature**, v.393, n. 6684, p. 440-442, 1998.

Watts, D. J.. **Small Worlds: The Dynamics of Networks between Order and Randomness**. Princeton University Press. 1999.

WERNECK, Vera Rudge. Sobre o processo de construção do conhecimento: o papel do ensino e da pesquisa. **Revista Ensaio: Avaliação e Políticas Públicas em Educação**. 14 (51). Jun. 2006. Disponível em: <https://www.scielo.br/j/ensaio/a/yy5rBTwpjnh4mq7QWcFDwN/#> Acesso em: 12 dez. 2021.

APÊNDICE A – CURSOS OFERECIDOS PELA UEFS/GRANDE ÁREA A QUE PERTENCE

CURSOS DE GRADUAÇÃO	GRANDE ÁREA
Bacharelado em Administração	Ciências Sociais Aplicadas
Bacharelado em Agronomia	Ciências Agrárias
Bacharelado em Ciências Biológicas	Ciências Biológicas
Licenciatura em Ciências Biológicas	Ciências Biológicas
Bacharelado em Ciências Contábeis	Ciências Sociais Aplicadas
Bacharelado em Ciências Econômicas	Ciências Sociais Aplicadas
Bacharelado em Direito	Ciências Sociais Aplicadas
Bacharelado em Enfermagem	Ciências da Saúde
Bacharelado em Engenharia Civil	Engenharias
Bacharelado em Engenharia de Computação	Engenharias
Bacharelado em Engenharia de Alimentos	Engenharias
Bacharelado em Farmácia	Ciências da Saúde
Licenciatura em Filosofia	Ciências Humanas
Bacharelado em Filosofia	Ciências Humanas
Licenciatura em Física	Ciências Exatas e da Terra
Bacharelado em Física	Ciências Exatas e da Terra
Licenciatura em Geografia	Ciências Humanas
Bacharelado em Geografia	Ciências Humanas
Licenciatura em História	Ciências Humanas
Bacharelado em Medicina	Ciências da Saúde
Bacharelado em Odontologia	Ciências da Saúde
Bacharelado em Psicologia	Ciências Humanas
Licenciatura em Educação Física	Ciências da Saúde
Licenciatura em Letras - Inglês	Linguística, Letras e Artes
Licenciatura em Letras - Português e Espanhol	Linguística, Letras e Artes
Licenciatura em Letras - Língua Portuguesa	Linguística, Letras e Artes

Licenciatura em Letras - Português e Francês	Linguística, Letras e Artes
Licenciatura em Matemática	Ciências Exatas e da Terra
Licenciatura em Música	Linguística, Letras e Artes
Licenciatura em Pedagogia	Ciências Humanas
Licenciatura em Química	Ciências Exatas e da Terra
CURSOS LATO SENSU	GRANDE ÁREA
Especialização em Saúde - Modalidade a Distância (PNAP/UAB)	Ciências da Saúde
Especialização em Biologia Celular	Ciências Biológicas
Especialização em Contabilidade Gerencial com ênfase em Controladoria	Ciências Sociais Aplicadas
Especialização em Desenho	Linguística, Letras e Artes
Especialização em Direito e Prática Jurídica	Ciências Sociais Aplicadas
Especialização em Educação Inclusiva	Ciências Humanas
Especialização em Educação na Cultura Digital	Ciências Humanas
Especialização em Engenharia Estrutural	Engenharias
Especialização em Filosofia Contemporânea	Ciências Humanas
Especialização em Gestão de Projetos	Ciências Humanas
Especialização em Gestão Pública - Modalidade a Distância - UAB	Ciências Humanas
Especialização em Gestão Universitária	Ciências Sociais Aplicadas
Especialização em História da Bahia	Ciências Sociais Aplicadas
Especialização em Linguística e Ensino-Aprendizagem da Língua Portuguesa	Linguística, Letras e Artes
Especialização em Matemática	Ciências Exatas e da Terra
Especialização em Música e Contemporaneidade - Modalidade a Distância - UAB	Linguística, Letras e Artes
Programa de Residência Multidisciplinar em Atenção à Urgência e Emergência	Multidisciplinar
Residências Multiprofissionais em Saúde da Família	Ciências da Saúde
CURSOS STRICTO SENSU	GRANDE ÁREA
Mestrado Profissional em Enfermagem	Ciências da Saúde
Mestrado Profissional em Saúde Coletiva	Ciências da Saúde

Programa de Pós-Graduação em Astronomia (Mestrado Profissional)	Ciências Exatas e da Terra
Programa de Pós-Graduação em Biotecnologia	Multidisciplinar
Programa de Pós-Graduação em Botânica	Ciências Biológicas
Programa de Pós-Graduação em Ciência da Computação	Ciências Sociais Aplicadas
Programa de Pós-Graduação em Ciências Farmacêuticas	Ciências da Saúde
Programa de Pós-Graduação em Desenho, Cultura e Interatividade	Linguística, Letras e Artes
Programa de Pós-Graduação em Ecologia e Evolução	Ciências Biológicas
Programa de Pós-Graduação em Educação	Ciências Humanas
Programa de Pós-Graduação em Engenharia Civil e Ambiental	Engenharias
Programa de Pós-Graduação em Ensino de Matemática	Ciências Exatas e da Terra
Programa de Pós-Graduação em Estudos Linguísticos	Linguística, Letras e Artes
Programa de Pós-Graduação em Estudos Literários	Linguística, Letras e Artes
Programa de Pós-Graduação em História	Ciências Humanas
Programa de Pós-Graduação em Modelagem em Ciências da Terra e do Ambiente	Multidisciplinar
Programa de Pós-Graduação em Planejamento Territorial (Mestrado Profissional)	Ciências Sociais Aplicadas
Programa de Pós-Graduação em Recursos Genéticos Vegetais	Ciências Biológicas
Programa de Pós-Graduação em Saúde Coletiva	Ciências da Saúde
Programa de Pós-Graduação em Gestão, Organizações e Sociedade	Ciências Sociais Aplicadas
Programa de Pós-Graduação em Direito	Ciências Sociais Aplicadas
PROFCIAMB - Programa de Pós-Graduação em Rede Nacional para o Ensino das Ciências Ambientais (Em Rede)	Multidisciplinar
PROFMAT - Mestrado Profissional em Matemática (Em Rede)	Ciências Exatas e da Terra
PROFIS - Mestrado Nacional Profissional em Ensino de Física (Em Rede)	Ciências Exatas e da Terra
PROFBIO - Mestrado em Ensino de Biologia em Rede Nacional (Em Rede)	Ciências Biológicas
Mestrado e Doutorado em Ensino, Filosofia e História da Ciência (Interinstitucional UEFS/UFBA)	Ciências Humanas
Doutorado em Difusão do Conhecimento (Interinstitucional UEFS/UFBA/IFBA/UNEB/LNCC/SENAI CIMATEC)	Multidisciplinar

Fonte: Elaboração própria, baseado nos dados disponíveis em (UEFS, Cursos de Graduação - Oferta Regular); (UEFS, Cursos Lato Sensu); (UEFS, Cursos Stricto Sensu); (CAPES, Tabelas de Áreas de Conhecimento/Avaliação) (2024).

APÊNDICE B – SCRIPT 01

RETIRA OS DADOS (ANTERIORMENTE DEFINIDOS) DOS CURRÍCULOS LATTES DOS PROFESSORES QUE FORAM BAIXADOS EM XML, CRIA PLANILHAS COM OS DADOS EXTRAÍDOS.

```
# import xml.etree.ElementTree as ET
```

```
from bs4 import BeautifulSoup as BS
```

```
import pandas as pd
```

```
import urllib.request
```

```
import lxml
```

```
import json
```

```
# import xmldict
```

```
import glob
```

```
import zipfile as zip
```

```
import os , shutil, sys
```

```
import csv
```

```
path = os.getcwd()
```

```
folderzip = os.path.join(path, 'Lattes')
```

```
print(folderzip)
```

```
all_zip = glob.glob(folderzip + '/*.zip')
```

```

def levarDicttoList(mydict):

    #tempC = ['@NOME-COMPLETO',
"@NOME-EM-CITACOES-BIBLIOGRAFICAS", 'AREAS-DE-ATUACAO',
'ATUACOES-PROFISSIONAIS',

    #
'FORMACAO-ACADEMICA-TITULACAO','UNIVERSIDADE-FORMACAO','IDIOMAS'
,'RESUMO-CV']

    tempC =
'@NOME-COMPLETO;@NOME-EM-CITACOES-BIBLIOGRAFICAS;AREAS-DE-ATU
ACAO;ATUACOES-PROFISSIONAIS;'

    tempC = tempC +
'FORMACAO-ACADEMICA-TITULACAO;UNIVERSIDADE-FORMACAO;IDIOMAS;R
ESUMO-CV\n'


    nn = len(mydict['@NOME-COMPLETO'])

    MYLIST = list()

    MYLIST.append(tempC)

    print(MYLIST)

    #for ii in range(nn):

    #    rows = list()

    #    rows.append(str(mydict['@NOME-COMPLETO'][ii]))

    #

    rows.append(str(mydict['@NOME-EM-CITACOES-BIBLIOGRAFICAS'][ii]))

    #    rows.append(str(mydict['AREAS-DE-ATUACAO'][ii]))

```

```

#         rows.append(str(mydict['ATUACOES-PROFISSIONAIS'][ii]))

#
rows.append(str(mydict['FORMACAO-ACADEMICA-TITULACAO'][ii]))

#         rows.append(str(mydict['UNIVERSIDADE-FORMACAO'][ii]))

#         rows.append(str(mydict['IDIOMAS'][ii]))

#         rows.append(str(mydict['RESUMO-CV'][ii]))

#         #print(rows)

#         MYLIST.append(rows)

for ii in range(nn):

    rows = str(mydict['@NOME-COMPLETO'][ii]) + ";"

    rows = rows +

str(mydict['@NOME-EM-CITACOES-BIBLIOGRAFICAS'][ii]) + ";"

    rows = rows + str(mydict['AREAS-DE-ATUACAO'][ii]) + ";"

    rows = rows + str(mydict['ATUACOES-PROFISSIONAIS'][ii]) + ";"

    rows = rows + str(mydict['FORMACAO-ACADEMICA-TITULACAO'][ii])

+ ";"

    rows = rows + str(mydict['UNIVERSIDADE-FORMACAO'][ii]) + ";"

    rows = rows + str(mydict['IDIOMAS'][ii]) + ";"

    rows = rows + str(mydict['RESUMO-CV'][ii]) + '\n'

# print(rows)

MYLIST.append(rows)

```



```
#print(MYLIST)
```

```
with open("salida.csv", "w") as fichero:
```

```
    writer = csv.writer(fichero,delimiter='\n')
```

```
    writer.writerow(MYLIST)
```

```
def set_row_table_pessoa (df_tmp, idLattes, nome_p, instituicao, area_act, resumo, orcid):
```

```
    dict_tmp = {
```

```
        'id_lattes': idLattes,
```

```
        'nome': nome_p,
```

```
        'nome_instituicao': instituicao,
```

```
        'area_atuacao': area_act,
```

```
        'resumo_lattes': resumo,
```

```
        'orcid_id': orcid
```

```
    }
```

```
    df_tmp = df_tmp.append(dict_tmp, ignore_index=True)
```

```
    return df_tmp
```

```
def set_row_table_pessoa_coaut (df_tmp, id_pessoa, idLattes, id_coaut, nome_coaut):
```

```
    dict_tmp = {
```

```
        'id_pessoa_coautor': id_pessoa,
```

```
        'id_lattes': idLattes,
```

```
        'id_coautor': id_coaut,
```

```

        'nome': nome_coaut
    }

    df_tmp = df_tmp.append(dict_tmp, ignore_index=True)

    return df_tmp

def set_row_table_coautorias(df_tmp, id_coaut, nome_cita, nome_comp):

    dict_tmp = {

        'id_coautor': id_coaut,

        'nome_citacao': nome_cita,

        'nome_completo': nome_comp

    }

    df_tmp = df_tmp.append(dict_tmp, ignore_index=True)

    return df_tmp

def set_row_table_artigos(df_tmp, id_art, idLattes, titulo, ano, pais, idioma, tipo):

    dict_tmp = {

        'id_artigo': id_art,

        'id_lattes': idLattes,

        'titulo_pub': titulo,

        'ano_pub': ano,

        'pais_pub': pais,

        'idioma': idioma,

        'tipo_pub': tipo

```

```
}
```

```
df_tmp = df_tmp.append(dict_tmp, ignore_index=True)
```

```
return df_tmp
```

```
def set_row_table_artigo_coautoria(df_tmp, id_art_coa, id_art, id_coaut, idLattes):
```

```
    dict_tmp = {
```

```
        'id_art_coautor': id_art_coa,
```

```
        'id_artigo': id_art,
```

```
        'id_coautor': id_coaut,
```

```
        'id_lattes': idLattes
```

```
    }
```

```
    df_tmp = df_tmp.append(dict_tmp, ignore_index=True)
```

```
    return df_tmp
```

```
def set_row_table_artigo_keyword(df_tmp, id_art_kk, id_art, id_key):
```

```
    dict_tmp = {
```

```
        'id_art_keyword': id_art_kk,
```

```
        'id_artigo': id_art,
```

```
        'id_key': id_key
```

```
    }
```

```
    df_tmp = df_tmp.append(dict_tmp, ignore_index=True)
```

```
    return df_tmp
```

```
def set_row_table_keyword(df_tmp, id_key, word):
```

```

dict_tmp = {

    'id_key': id_key,

    'palavra': word

}

df_tmp = df_tmp.append(dict_tmp, ignore_index=True)

return df_tmp

def set_row_table_art_sub_area_conh(df_tmp, id_art_arC, id_art, id_ar, id_sub_ar):

    dict_tmp = {

        'id_art_areaC': id_art_arC,

        'id_artigo': id_art,

        'id_areaC': id_ar,

        'id_sub_areaC': id_sub_ar

    }

    df_tmp = df_tmp.append(dict_tmp, ignore_index=True)

    return df_tmp

def set_row_table_area_conh(df_tmp, id_ar, nome_ar_conh):

    dict_tmp = {

        'id_areaC': id_ar,

        'nome_area_conh': nome_ar_conh

    }

    df_tmp = df_tmp.append(dict_tmp, ignore_index=True)

```

```

        return df_tmp

def set_row_table_sub_area_conh(df_tmp, id_sub_arC, nome_sub_arC, id_ar):

    dict_tmp = {

        'id_sub_areaC': id_sub_arC,

        'nome_sub_area_conh': nome_sub_arC,

        'id_areaC': id_ar

    }

    df_tmp = df_tmp.append(dict_tmp, ignore_index=True)

    return df_tmp

def set_id_art (numb, sufixo):

    n_text = "

    resto = 5 - len(str(numb))

    if resto > 0:

        for ii in range(resto):

            n_text += '0'

    n_text += str(numb) + sufixo

    return n_text

def set_id_autor (numb):

    n_text = "

    resto = 4 - len(str(numb))

    if resto > 0:

```

```

        for ii in range(resto):

            n_text += '0'

        n_text += str(numb) + 'auc'

    return n_text

def set_id_keyword (numb):

    n_text = ""

    resto = 4 - len(str(numb))

    if resto > 0:

        for ii in range(resto):

            n_text += '0'

        n_text += str(numb) + 'kw'

    return n_text

def research_coautor (df_tmp, nomcoaut):

    ls_coautor = df_tmp.loc[df_tmp['nome_citacao'] == nomcoaut]

    if len(ls_coautor.index.values) == 0:

        return False, ""

    else:

        return True, ls_coautor.id_coautor.values[0]

def research_keywords (df_tmp, kkeyword):

    ls_kkeyword = df_tmp.loc[df_tmp['palavra'] == kkeyword]

    if len(ls_kkeyword.index.values) == 0:

```

```

        return False, "

    else:

        return True, ls_kkeyword.id_key.values[0]

def research_area_Conhecimento (df_tmp, grande_area):

    ls_areaC = df_tmp.loc[df_tmp['nome_area_conh'] == grande_area]

    if len(ls_areaC.index.values) == 0:

        return False, "

    else:

        return True, int(ls_areaC.id_areaC.values[0])

def research_sub_area_Conhecimento (df_tmp, grande_sub_area):

    ls_areaC = df_tmp.loc[df_tmp['nome_sub_area_conh'] == grande_sub_area]

    if len(ls_areaC.index.values) == 0:

        return False, "

    else:

        return True, int(ls_areaC.id_sub_areaC.values[0])

# DataFrame para a Tabela pessoa

table_pessoa = pd.DataFrame(columns=('id_lattes', 'nome',
'nome_instituicao','area_atuacao','resumo_lattes','orcid_id'))

# Dataframe para a tabela pessoa e coautoria

table_pessoa_coaut = pd.DataFrame(columns=('id_pessoa_coautor', 'id_lattes',
'id_coautor','nome'))

# DataFrame para a Tabela coautorias

```

```
table_coautorias = pd.DataFrame(columns=('id_coautor', 'nome_citacao', 'nome_completo'))
```

```
# Data frame para a Tabela artigos
```

```
table_artigos = pd.DataFrame(columns=('id_artigo', 'id_lattes',  
'titulo_pub','ano_pub','pais_pub','idioma','tipo_pub'))
```

```
# Data frame para a Tabela relação artigo coautoria
```

```
table_artigo_coautoria =  
pd.DataFrame(columns=('id_art_coautor','id_artigo','id_coautor','id_lattes'))
```

```
# Dataframe que relaciona as palavras claves ao artigo
```

```
table_artigo_keyword = pd.DataFrame(columns=('id_art_keyword','id_artigo','id_key'))
```

```
# Dataframe para salvar todas as palavras claves
```

```
table_keyword = pd.DataFrame(columns=('id_key','palavra'))
```

```
# Datafram para realcionar artigos a sub areas de conhecimento
```

```
table_art_sub_area_conh = pd.DataFrame(columns=('id_art_areaC',  
'id_artigo','id_areaC','id_sub_areaC'))
```

```
# dataframe que salva a lista de area de conhecimento
```

```
table_area_conh = pd.DataFrame(columns=('id_areaC','nome_area_conh'))
```

```
# dataframe que salca a sub area de conhecimento
```

```
table_sub_area_conh = pd.DataFrame(columns=('id_sub_areaC','nome_sub_area_conh',  
'id_areaC'))
```

```
id_art_areaC = 1
```

```
id_areaC = 1
```

```
id_sub_areaC = 1
```

```
id_art_coautor = 1
```



```
id_pessoa_coautor = 1
```

```
id_lattes_p = "
```

```
id_artig = "
```

```
num_art = 0
```

```
num_coaut = 0
```

```
id_pessoa_couat = 1
```

```
num_keyword = 1
```

```
id_keywords = "
```

```
id_art_keyword = 1
```

```
id_areaC = 1
```

```
id_sub_areaC = 1
```

```
id_art_areaC = 1
```

```
for zfileN in all_zip:
```

```
    # print("zip extraido \n", zfileN)
```

```
    id_lattes_p = str(zfileN.split('/')[-1][:4])
```

```
    print("id_lattes = ", id_lattes_p)
```

```
    zfile = zip.ZipFile(zfileN, 'r')
```

```
    for nameXML in zfile.namelist(): #
```

```
        replicas = 0
```

```
        #root = getroot('example.xml')
```

```

fXML = zfile.read(nameXML)

# print(fXML)

soup = BS(fXML,'xml')

# curriculo = soup.find_all('CURRICULO-VITAE')

dados_gerais = soup.find('DADOS-GERAIS')

# curriculo = soup.find('AREAS-DO-CONHECIMENTO')

# result = list(curriculo.children)

# print(result[0])

# print(result[1])

atributo_nome = dados_gerais.get('NOME-COMPLETO')

print('NOME-COMPLETO = ', atributo_nome)

orcid_ids = dados_gerais.get('ORCID-ID')

print('ORCID-ID = ',orcid_ids)

nome_citacoes =
dados_gerais.get('NOME-EM-CITACOES-BIBLIOGRAFICAS')

print("nomes de citaciones bibliograficas = ", nome_citacoes)

print(" ")

try:

    resumo_CV = soup.find('RESUMO-CV')

    resumene = resumo_CV.get('TEXTO-RESUMO-CV-RH')

    print("RESUMO = ", resumene)

    print(" ")

```

```

except:

    resumen = "

try:

    endereco_prof = soup.find('ENDERECO-PROFISSIONAL')

    endereco =
endereco_prof.get('NOME-INSTITUICAO-EMPRESA')

    print('endereço profissional ', endereco)

    print(" ")

except :

    endereco = 'Universidade Estadual de Feira de Santana'

try:

    area_atuacao = soup.find('AREA-DE-ATUACAO')

    profissao =
area_atuacao.get('NOME-GRANDE-AREA-DO-CONHECIMENTO')

    print("Area de Conhecimento = ", profissao)

    print(" ")

except:

    profissao = "

    # 'id_lattes', 'nome',
'nome_instituicao','area_atuacao','resumo_lattes','orcid_id'

    table_pessoa = set_row_table_pessoa(table_pessoa, id_lattes_p,
atributo_nome, endereco, profissao, resumene, orcid_ids)

```

```

ls_art_pub = soup.find('ARTIGOS-PUBLICADOS')

if ls_art_pub != None:

    for art in ls_art_pub:

        tipo_pub = 'artigo'

        print('tipo de publicação == ', tipo_pub)

        dados_art =
art.find('DADOS-BASICOS-DO-ARTIGO')

        try:

            titulo_art = dados_art.get('TITULO-DO-ARTIGO')

            print(" TITULO-DO-ARTIGO = ", titulo_art)

            idioma_art = dados_art.get('IDIOMA')

            print(" IDIOMA = ", idioma_art)

            ano_art = dados_art.get('ANO-DO-ARTIGO')

            print(" ANO-DO-ARTIGO = ", ano_art)

            pais_art = dados_art.get('PAIS-DE-PUBLICACAO')

            print('PAIS-DE-PUBLICACAO = ', pais_art)

            num_art += 1

            id_artig = set_id_art (num_art, 'ar')

# 'id_artigo', 'id_lattes','titulo_pub','ano_pub','pais_pub','idioma','tipo_pub'

table_artigos = set_row_table_artigos(table_artigos,id_artig, id_lattes_p, titulo_art, ano_art,
pais_art, idioma_art, tipo_pub)

```

```

autores = art.find_all('AUTORES')

for cc, aut in enumerate(autores):

    nome_aut = aut.get('NOME-COMPLETO-DO-AUTOR')

    print('NOME-COMPLETO-DO-AUTOR = ',
nome_aut)

    nome_cita =
aut.get('NOME-PARA-CITACAO')

    print("NOME-PARA-CITACAO = ", nome_cita)


# verificando a existencia do nome citação

coaut_exist, id_coautor = research_coautor(table_coautorias, nome_cita)

    if coaut_exist == False:

        num_coaut += 1

        id_coautor = set_id_autor (num_coaut)


# 'id_coautor', 'nome_citacao', 'nome_completo'

table_coautorias = set_row_table_coautorias(table_coautorias,id_coautor, nome_cita,
nome_aut)

# 'id_pessoa_coautor', 'id_lattes', 'id_coautor','nome'

table_pessoa_coaut = set_row_table_pessoa_coaut(table_pessoa_coaut,

id_pessoa_couat,id_lattes_p, id_coautor, nome_aut)

id_pessoa_couat += 1

```

```

# 'id_art_coautor','id_artigo','id_coautor','id_lattes'

table_artigo_coautoria =
set_row_table_artigo_coautoria(table_artigo_coautoria,id_art_coautor, id_artig, id_coautor,
id_lattes_p)

id_art_coautor += 1

palavras_chaves = art.find_all('PALAVRAS-CHAVE')

# print("PALAVRAS-CHAVE : ",
palavras_chaves)

# ls_palavras = []

for pal_ls in palavras_chaves:

    # print(pal_ls)

    for ii in range(1, 7):

        # print(ii)

        pal_chave = pal_ls.get('PALAVRA-CHAVE-' + str(ii))

        if pal_chave != None and pal_chave != "":

            print('PALAVRA-CHAVE-' + str(ii) , pal_chave)

# verificando a existencia do nome citação

kkeywords_exist, id_keywords = research_keywords(table_keyword, pal_chave)

```

```
if kkeywords_exist == False:
```

```
    id_keywords = set_id_keyword(num_keyword)
```

```
    num_keyword += 1
```

```
    # 'id_key','palavra'
```

```
    table_keyword = set_row_table_keyword(table_keyword, id_keywords, pal_chave)
```

```
    # ls_palavras.append(pal_chave)
```

```
    print("    ", pal_chave)
```

```
#
```

```
    'id_art_keyword','id_artigo','id_key'
```

```
    table_artigo_keyword = set_row_table_artigo_keyword(table_artigo_keyword,
```

```
    id_art_keyword, id_artig, id_keywords)
```

```
    id_art_keyword += 1
```

```
area_conhec = art.find_all('AREAS-DO-CONHECIMENTO')
```

```
    # ls_area = []
```

```
    # ls_subarea = []
```

```
    for area in area_conhec:
```

```
        area_grande =
```

```
        area.get('NOME-GRANDE-AREA-DO-CONHECIMENTO')
```

```

        print('NOME-GRANDE-AREA-DO-CONHECIMENTO = ',
area_grande )

        if area_grande != None and area_grande != "":

            areaC_exist, id_areaC = research_area_Conhecimento(table_area_conh,
area_grande)

            if areaC_exist == False:

                # ls_area.append(area_grande)


            # 'id_areaC','nome_area_conh'

            table_area_conh = set_row_table_area_conh(table_area_conh,id_areaC,
area_grande)

            id_areaC += 1

            subarea_grande = area.get('NOME-DA-SUB-AREA-DO-CONHECIMENTO')

            print('NOME-DA-SUB-AREA-DO-CONHECIMENTO = ', subarea_grande )


            if subarea_grande != None and subarea_grande != "":

                # ls_subarea.append(subarea_grande)


            subareaC_exist, id_sub_areaC = research_sub_area_Conhecimento(table_sub_area_conh,
subarea_grande)

```



```

if subareaC_exist == False:

# 'id_sub_areaC','nome_sub_area_conh','id_areaC'table_sub_area_conh =
set_row_table_sub_area_conh(table_sub_area_conh,id_sub_areaC, subarea_grande,
id_areaC)

        id_sub_areaC += 1

# 'id_art_areaC', 'id_artigo','id_areaC','id_sub_areaC'

table_art_sub_area_conh =
set_row_table_art_sub_area_conh(table_art_sub_area_conh,id_art_areaC, id_artig, id_areaC,
id_sub_areaC)

id_art_areaC += 1

except:

print("POR ALGUM ERRO NAO SALVOU")

ls_cap_livro_pub = soup.find('CAPITULOS-DE-LIVROS-PUBLICADOS')

        if ls_cap_livro_pub != None:

                for cap_livro in ls_cap_livro_pub:

                        tipo_pub = 'capitulo_livro'

                        print('tipo de publicação == ', tipo_pub)

                        dados_cap =
cap_livro.find('DADOS-BASICOS-DO-CAPITULO')

                        try:

                                titulo_cap =
dados_cap.get('TITULO-DO-CAPITULO-DO-LIVRO')

                                print(" TITULO-DO-CAPITULO-DO-LIVRO = ",
titulo_cap)

```

```

ano_pub = dados_cap.get('ANO')

print("Ano da publicação ", ano_pub)

idioma_art = dados_cap.get('IDIOMA')

print(" IDIOMA = ", idioma_art)

pais_art = dados_cap.get('PAIS-DE-PUBLICACAO')

print('PAIS-DE-PUBLICACAO = ', pais_art)

num_art += 1

id_artig = set_id_art (num_art, 'cl')

# 'id_artigo', 'id_lattes', 'titulo_pub', 'ano_pub', 'pais_pub', 'idioma', 'tipo_pub'

table_artigos = set_row_table_artigos(table_artigos, id_artig, id_lattes_p, titulo_cap, ano_pub,
pais_art, idioma_art, tipo_pub)

autores = cap_livro.find_all('AUTORES')

for cc, aut in enumerate(autores):

    nome_aut = aut.get('NOME-COMPLETO-DO-AUTOR')

    print('NOME-COMPLETO-DO-AUTOR = ', nome_aut)

    nome_cita = aut.get('NOME-PARA-CITACAO')

    print("NOME-PARA-CITACAO = ", nome_cita)

    # verificando a existencia do nome citação

coaut_exist, id_coautor = research_coautor(table_coautorias, nome_cita)

if coaut_exist == False:

    num_coaut += 1

```

```

id_coautor = set_id_autor (num_coaut)

# 'id_coautor', 'nome_citacao', 'nome_completo'

table_coautorias = set_row_table_coautorias(table_coautorias,

id_coautor, nome_cita, nome_aut)

# 'id_pessoa_coautor', 'id_lattes', 'id_coautor', 'nome'

table_pessoa_coaut = set_row_table_pessoa_coaut(table_pessoa_coaut,
id_pessoa_coaut, id_lattes_p, id_coautor, nome_aut)

id_pessoa_coaut += 1

# 'id_art_coautor', 'id_artigo', 'id_coautor', 'id_lattes'

table_artigo_coautoria = set_row_table_artigo_coautoria(table_artigo_coautoria,

id_art_coautor, id_artig, id_coautor, id_lattes_p)

id_art_coautor += 1

palavras_chaves = cap_livro.find_all('PALAVRAS-CHAVE')

print("PALAVRAS-CHAVE : ")

ls_palavras = []

for pal_ls in palavras_chaves:

    for ii in range(1, 7):

        pal_chave = pal_ls.get('PALAVRA-CHAVE-' + str(ii))

```

```

        if pal_chave != " and pal_chave != None:

# ls_palavras.append(pal_chave)

# print(" {}    {} ".format(ii, pal_chave))

    print('PALAVRA-CHAVE-' + str(ii) , pal_chave)


# verificando a existencia do nome citação

keywords_exist, id_keywords = research_keywords(table_keyword, pal_chave)

if keywords_exist == False:

id_keywords = set_id_keyword(num_keyword)

num_keyword += 1


# 'id_key','palavra'

table_keyword = set_row_table_keyword(table_keyword, id_keywords, pal_chave)

# ls_palavras.append(pal_chave)

print("    ", pal_chave)

# 'id_art_keyword','id_artigo','id_key'

table_artigo_keyword = set_row_table_artigo_keyword(table_artigo_keyword,

id_art_keyword, id_artig, id_keywords)

id_art_keyword += 1

```

```

# area_conhec = cap_livro.find_all('AREAS-DO-CONHECIMENTO-2')

# print(area_conhec)

# areas_con =

ls_area = []

ls_subarea = []

    for cc in range(1, 4):

        try:

            areas_con = cap_livro.find('AREA-DO-CONHECIMENTO-' + str(cc))

# print(areas_con)

            area_grande = areas_con.get('NOME-GRANDE-AREA-DO-CONHECIMENTO')

# ls_area.append(area_grande)

            print('NOME-GRANDE-AREA-DO-CONHECIMENTO = ', area_grande )

            if area_grande != None and area_grande != "":

                areaC_exist, id_areaC = research_area_Conhecimento(table_area_conh, area_grande)

                if areaC_exist == False:

                    ls_area.append(area_grande)


# 'id_areaC','nome_area_conh'

            table_area_conh = set_row_table_area_conh(table_area_conh,id_areaC, area_grande)

            id_areaC += 1

```

```

subarea_grande = areas_con.get('NOME-DA-SUB-AREA-DO-CONHECIMENTO')

# ls_subarea.append(subarea_grande)

print('NOME-DA-SUB-AREA-DO-CONHECIMENTO = ', subarea_grande )

if subarea_grande != None and subarea_grande != "":

# ls_subarea.append(subarea_grande)

subareaC_exist, id_sub_areaC = research_sub_area_Conhecimento(table_sub_area_conh,
subarea_grande)

if subareaC_exist == False:

# 'id_sub_areaC','nome_sub_area_conh', 'id_areaC'

table_sub_area_conh = set_row_table_sub_area_conh(table_sub_area_conh,

id_sub_areaC, subarea_grande, id_areaC)

id_sub_areaC += 1

# 'id_art_areaC', 'id_artigo','id_areaC','id_sub_areaC' table_art_sub_area_conh =
set_row_table_art_sub_area_conh(table_art_sub_area_conh,id_art_areaC, id_artig, id_areaC,
id_sub_areaC)

id_art_areaC += 1

except :

        break

except:

        print("ALGO DEU ERRADO ! ")

```

```

#

print("Tabela Dados pessoais ", table_pessoa.shape)

print("Tabela de artigos ", table_artigos.shape)

print("Tabela de coautorias ", table_coautorias.shape)

print("Tabela relação artigo coautoria ", table_artigo_coautoria.shape)

print("Tabela relação autor e coautorias ", table_pessoa_coaut.shape)

print("Tabela Dados pessoais ", table_artigo_keyword.shape)

print("Tabela de palavras chaves ", table_keyword.shape)

print("Tabela de relação de artigos com sub areas de conhecimento ",
table_art_sub_area_conh.shape)

print("Tabela area de conhecimento ", table_area_conh.shape)

print("Tabela de sub area de conhecimentos ", table_sub_area_conh.shape)


#####Salvando todas as Tabelas #####

print('SALVANDO TODOS AS TABELAS EM \n =====> '+ path)

name_csv = 'table_dados_pessoais_autor.csv'

folder_save = os.path.join(path, name_csv)

print(folder_save)

table_pessoa.to_csv(folder_save, index= False, encoding='utf-8')


name_csv = 'table_dados_artigos.csv'

folder_save = os.path.join(path, name_csv)

```

```
table_artigos.to_csv(folder_save, index= False, encoding='utf-8')
```

```
name_csv = 'table_dados_coautorias.csv'
```

```
folder_save = os.path.join(path, name_csv)
```

```
table_coautorias.to_csv(folder_save, index= False, encoding='utf-8')
```

```
name_csv = 'table_relacao_artigos_coautor.csv'
```

```
folder_save = os.path.join(path, name_csv)
```

```
table_artigo_coautoria.to_csv(folder_save, index= False, encoding='utf-8')
```

```
name_csv = 'table_relacao_autor_coautoria.csv'
```

```
folder_save = os.path.join(path, name_csv)
```

```
table_pessoa_coaut.to_csv(folder_save, index= False, encoding='utf-8')
```

```
name_csv = 'table_relacao_artigos_palavras_chaves.csv'
```

```
folder_save = os.path.join(path, name_csv)
```

```
table_artigo_keyword.to_csv(folder_save, index= False, encoding='utf-8')
```

```
name_csv = 'table_dados_palavras_chaves.csv'
```

```
folder_save = os.path.join(path, name_csv)
```

```
table_keyword.to_csv(folder_save, index= False, encoding='utf-8')
```



```
if table_sub_area_conh.shape[0] > 0:
```

```
    name_csv = 'table_sub_area_conhecimento.csv'
```

```
    folder_save = os.path.join(path, name_csv)
```

```
    table_sub_area_conh.to_csv(folder_save, index= False, encoding='utf-8')
```

```
    name_csv = 'table_area_conhecimento.csv'
```

```
    folder_save = os.path.join(path, name_csv)
```

```
    table_area_conh.to_csv(folder_save, index= False, encoding='utf-8')
```

```
    name_csv = 'table_relacao_artigo_subarea_conhecimento.csv'
```

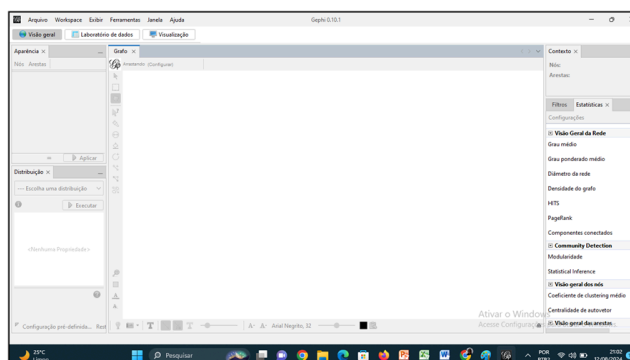
```
    folder_save = os.path.join(path, name_csv)
```

```
    table_art_sub_area_conh.to_csv(folder_save, index= False, encoding='utf-8')
```

APÊNDICE C – SOFTWARE GEPHI

A seguir, a Figura C1 mostra a tela do Gephi 0.10.1 com a rede inicial de coautoria. Software que foi utilizado na visualização e análise das redes de pesquisa.

Figura C1. Tela do Gephi versão 0.10.1



Fonte: Elaboração própria, (2024).

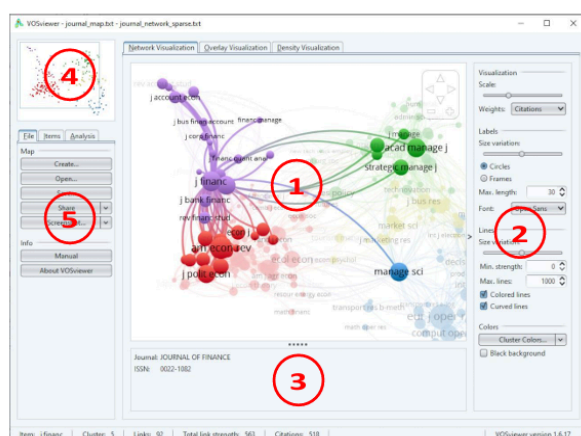
O Gephi 0.10.1 dá algumas opções na forma de apresentação da rede, algumas já estão no programa e outras podem ser acrescentadas usando *plugins*. Acima da tela tem três tópicos importantes do software: 1. Visão Geral, onde a rede aparece; 2. O Laboratório de Dados, onde temos acesso a todos os dados que compõem o grafo e através dele é possível adicionar, editar, excluir, importar e exportar dataset em csv. E por último, o tópico 3. Visualização, que representa a área de exportação do projeto. Do lado direito da tela estão dispostas a aba Filtros (que podem ser aplicados na rede) e as estatísticas, que podem ser executadas para encontrar as métricas da rede. As estatísticas estão divididas em cinco partes: 1. Visão geral da rede; 2. Community Detection; 3. Visão geral dos nós; 4. Visão geral das arestas, e 5. Dinâmico.

As métricas fornecem uma compreensão abrangente da estrutura e dinâmica de redes sociais. Usando o Gephi versão 0.10.1 podemos visualizar e analisar essas métricas de forma interativa, permitindo uma exploração das propriedades e características das redes estudadas. A escolha do Gephi se deve ao fato de ser um software gratuito, ter uma interface agradável, e várias opções de formas de apresentação, facilitando sua manipulação, visualização e entendimento da estrutura e dinâmica das redes. Além de tudo isso, o Gephi oferece o cálculo de medidas estatísticas satisfatórias para análise das redes propostas.

APÊNDICE D – SOFTWARE VOSVIEWER

Além do Gephi, o software gratuito VOSviewer 1.6.20 auxiliou na visualização das redes. A rede apresentada é definida por cores diferentes representando a força dos *clusters* da rede (termos com maior conexão), ou seja, as redes são criadas com base na frequência de termos, coocorrência e agrupamentos. A Figura D1 mostra a tela inicial do software (Manual VOSviewer 1.6.20, p. 5).

Figura D1. Janela principal do VOSviewer



LEGENDA DAS FUNÇÕES:

- (1) Painel Principal
- (2) Painel de Opções
- (3) Painel de Informações
- (4) Painel de Visão Geral
- (5) Painel de Ação

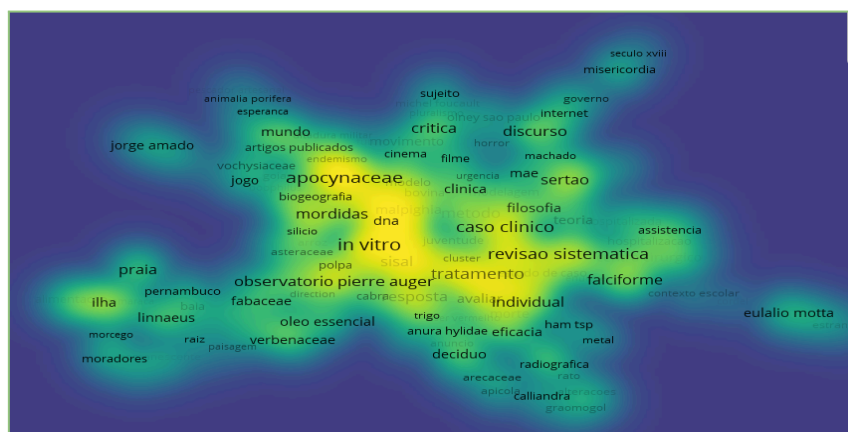
Fonte: Manual VOSviewer 1.6.20, p. 8.

O algoritmo do VOSviewer permite a atribuição de um peso de relevância $R(t)$ a cada termo, que é combinado com a frequência e o contexto, priorizando termos mais significativos e deixando em segundo plano os termos mais genéricos, assim, termos mais genéricos, mesmo com alta frequência $f(t)$, terão menor valor de especificidade $s(t)$, levando em conta a seguinte fórmula: $R(t) = f(t) \cdot s(t)$, sendo $R(t)$ a relevância da palavra, $f(t)$ a frequência absoluta da palavra e $s(t)$ o valor da especificidade de cada palavra. A escolha padrão mostra 60% das palavras mais relevantes. É importante destacar que as redes criadas no VOSviewer apresentam, por padrão, o máximo de 1.000 arestas. Dessa forma, mesmo tendo muitos dados, só os mais representativos aparecerão na rede. A rede apresentada pode ter duas formas: a “Network Visualizacion” que é a forma padrão, onde os nós e arestas

aparecem no espaço bidimensional e as cores representam os clusters; e a “Density Visualizacion”, que não apresenta arestas e mostra cores mais intensas quanto maior for a densidade do termo (Manual VOSviewer 1.6.20, p. 7). As redes exportadas do VOSviewer foram rodadas no Gephi para complementar as análises visuais com métricas estruturais. Essa integração permitiu uma compreensão mais profunda das relações entre os termos e autores, combinando a abordagem estatística do Gephi com a visualização intuitiva do VOSviewer.

A partir de todos esses dados e métricas, fizemos uma análise para entender o comportamento das redes, seus nós e suas conexões (arestas) em cada uma das redes.

Figura D2. Density Visualizacion



Fonte: Elaboração própria (2025)

APÊNDICE E – SCRIPT 02

Organiza os dados das planilhas em linhas (coautores) para facilitar a utilização dos dados nos softwares.

```
# Bibliotecas necessárias para o código abaixo
```

```
library(tidyverse)
```

```
library(dplyr)
```

```
library(tidyr)
```

```
library(stringr)
```

```
# Informando o diretório onde os arquivos estão localizados
```

```
# Substitua <Defina aqui seu diretório com os arquivos> pelo caminho completo
```

```
setwd("<Defina aqui seu diretório com os arquivos>")
```

```
# Esse comando acima pode ser substituído no RStudio clicando em:
```

```
# Session -> Set Working Directory -> Choose Directory
```

```
# Escolha o diretório padrão onde estão os arquivos desse código
```

```
# Em seguida clique em Choose
```

```
# Lendo o arquivo base
```

```
dados <- read_csv("base.csv")
```

```
# Agrupando por artigo com a combinação dos pares de autores
```

```
resultado <- dados %>%
```

```

group_by(artigo,ano) %>%

summarise(autores = paste(autor, collapse = ", ")) %>%

rowwise() %>%

mutate(

  pares = list(if (length(str_split(autores, ", ")[1]) >= 2) {

    combn(str_split(autores, ", ")[1], 2, simplify = FALSE)

  } else {

    NULL

  })

) %>%

unnest(pares) %>%

mutate(par = sapply(pares, function(x) paste(x, collapse = ", "))) %>%

select(artigo, ano, par) %>%

arrange(artigo)

# Gerando o arquivo "resultado_pares.csv" com o resultado da combinação

# Gravando no diretório definido no comando setwd (linha 10)

write.csv(resultado, "resultado_pares.csv", row.names = FALSE)

# Analisando as Duplicidades

aresta <- read_csv("arestas.csv")

base_quantidade <- (aresta %>%

```

```
select(Source,Target,Artigo,Ano) %>%  
  
group_by(Source,Target,Artigo,Ano) %>%  
  
summarise(Total=n()) %>%  
  
filter(Total > 1) %>%  
  
arrange(desc(Total))  
  
)  
  
# Gerando o arquivo "total_duplicidades.csv" com o resutlado das duplicidades  
  
# Considere duplicidade a sumarização de registros com total maior que 1  
  
# Gravando no diretório definido no comando setwd (linha 10)  
  
write.csv(base_quantidade, "total_duplicidades.csv", row.names = FALSE)
```