



UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
DEPARTAMENTO DE ESTATÍSTICA E INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA APLICADA

NÉFI DE MEDEIROS FERNANDES

SUPORTE À ELABORAÇÃO DE MATRIZES
CURRICULARES E DISCIPLINAS USANDO
INTELIGÊNCIA ARTIFICIAL GENERATIVA E
GRAFOS

RECIFE – PE

2025

NÉFI DE MEDEIROS FERNANDES

**SUPORTE À ELABORAÇÃO DE MATRIZES
CURRICULARES E DISCIPLINAS USANDO
INTELIGÊNCIA ARTIFICIAL GENERATIVA E
GRAFOS**

Dissertação submetida à Coordenação do
Programa de Pós-Graduação em Informática
Aplicada do Departamento de Estatística
e Informática - DEINFO - Universidade
Federal Rural de Pernambuco, como parte
dos requisitos necessários para obtenção do
grau de Mestre.

ORIENTADOR: Gabriel Alves de Albuquerque Júnior

RECIFE – PE

2025

Dados Internacionais de Catalogação na Publicação
Sistema Integrado de Bibliotecas da UFRPE
Gerada automaticamente, mediante os dados fornecidos pelo(a)
autor(a)

F363s Fernandes, Néfi de Medeiros.
 Suporte à elaboração de matrizes curriculares e
disciplinas usando inteligência artificial generativa e grafos
/ Néfi de Medeiros Fernandes. – Recife, 2025.
 147 f.; il.

Orientador(a): Gabriel Alves de Albuquerque Junior.

Dissertação (Mestrado) – Universidade Federal Rural de
Pernambuco, Programa de Pós-Graduação em Informática
Aplicada, Recife, BR-PE, 2025.

Inclui referências e apêndice(s).

1. Educação superior. 2. Design curricular. 3. Inteligência
artificial generativa. 4. Teoria dos grafos I. Junior, Gabriel
Alves de Albuquerque, orient. II. Título

CDD 004

NÉFI DE MEDEIROS FERNANDES

SUPORTE À ELABORAÇÃO DE MATRIZES CURRICULARES E DISCIPLINAS USANDO INTELIGÊNCIA ARTIFICIAL GENERATIVA E GRAFOS

Dissertação submetida à Coordenação do
Programa de Pós-Graduação em Informática
Aplicada do Departamento de Estatística
e Informática - DEINFO - Universidade
Federal Rural de Pernambuco, como parte
dos requisitos necessários para obtenção do
grau de Mestre.

Aprovada em: 29 de Agosto de 2025.

BANCA EXAMINADORA

Gabriel Alves de Albuquerque Júnior (Orientador)
Universidade Federal Rural de Pernambuco
Departamento de Estatística e Informática

Andre Camara Alves do Nascimento
Universidade Federal Rural de Pernambuco
Departamento de Computação

Emanuel Marques Queiroga
Instituto Federal de Educação, Ciência e Tecnologia
Sul-rio-grandense

Dedico este trabalho à minha esposa Rebeca, e aos meus filhos Edgar e Cecília, por todo o amor, companheirismo e paciência durante essa jornada; acima de tudo, dedico ao Néfi de seis anos de idade que sonhou em ser cientista.

Agradecimentos

A conclusão desta jornada não seria possível sem o apoio, a paciência e a inspiração de muitas pessoas a quem sou profundamente grato.

Primeiramente e acima de tudo, agradeço à minha esposa, Rebeca, e aos meus filhos, Edgar e Cecília. Vocês foram meu alicerce e minha maior motivação. Agradeço por toda a paciência, compreensão e amor que me deram a força necessária para seguir e finalizar mais esta importante etapa. Aos meus amigos Adalberto, Everton, Lucas, Rold, Sérgio e Thiago, obrigado pelo apoio e por tornarem a vida mais leve. Um agradecimento especial a meus sogros, Valquíria e Hector, por todo apoio e incentivo durante essas e outras empreitadas.

Agradeço especialmente ao meu orientador, Professor Gabriel Alves, por todos os conselhos, pela orientação precisa e pela paciência fundamental ao longo de todo este processo. Sua clareza e sinceridade foram indispensáveis para a evolução deste trabalho e da minha evolução como pesquisador e como pessoa. Estendo meus agradecimentos à Professora Ceça pela orientação e pelo incentivo nos estágios iniciais desta pesquisa.

Um agradecimento especial aos meus líderes na Prime Control, Daniel Mizobuchi e Ana Paula Rocha, por me incentivarem a continuar com o mestrado e, principalmente, pela flexibilidade e pelo apoio para conciliar os momentos de aula e pesquisa com o trabalho. Sem o incentivo inicial de vocês, esta jornada não teria começado.

“N3o existe verdade abstrata. A verdade 3
sempre concreta.”

(Vladimir Lenin, 1904)

Resumo

A elaboração de matrizes curriculares para cursos de graduação é um processo complexo, de baixa escalabilidade e de alto impacto para a qualidade do ensino superior. Esta dissertação teve como objetivo propor, desenvolver e validar um método original para dar suporte a este processo, combinando a capacidade de síntese da Inteligência Artificial com a validação do especialista humano. O método proposto integra técnicas de Inteligência Artificial Generativa (IAGen), Processamento de Linguagem Natural (PLN) e Análise de Grafos, sob a orientação de um paradigma *Human-in-the-Loop* (HITL). A abordagem utiliza a vetorização de textos e a similaridade de cosseno para identificar e fundir disciplinas com conteúdos sobrepostos utilizando grafos para localizar os cursos similares entre si e usando a IAGen para gerar novas disciplinas a partir dos grupos formados; uma arquitetura de Geração Aumentada por Recuperação (RAG) para orientar o modelo de IAGen na seleção de disciplinas alinhadas a documentos normativos; e métricas de centralidade de grafo para analisar a estrutura curricular e atribuir papéis pedagógicos aos seus componentes a partir das métricas de centralidade. A validação do método foi realizada através de dois estudos de caso completos, nos domínios de Ciência de Dados e Administração, para testar sua eficácia e generalização. Os resultados demonstram que os currículos gerados possuem alta coerência semântica com matrizes de referência nacionais e internacionais. A análise de aderência, validada por testes de hipótese, comprovou um alinhamento com as diretrizes normativas estatisticamente indistinguível do de cursos consolidados. A análise estrutural revelou uma organização de núcleo-periferia coerente em ambos os domínios, validando a classificação dos papéis pedagógicos. O trabalho contribui com um método original de ponta a ponta, reproduzível e validado, que demonstra o potencial de utilizar a Inteligência Artificial Generativa com o Análise de Grafos para suporte no design curricular de forma eficiente.

Palavras-chave: Educação Superior. Design Curricular. Inteligência Artificial Generativa. Teoria dos Grafos.

Abstract

The development of curriculum matrices for undergraduate programs is a complex, low-scalability process with a high impact on the quality of higher education. This dissertation aimed to propose, develop, and validate an original method to support this process, combining the synthesis capabilities of Artificial Intelligence with validation by human experts. The proposed method integrates techniques from Generative Artificial Intelligence (GenAI), Natural Language Processing (NLP), and Graph Analysis, under the guidance of a Human-in-the-Loop (HITL) paradigm.

The approach uses text vectorization and cosine similarity to identify and merge courses with overlapping content, employing graphs to locate similar programs and using GenAI to generate new courses based on the resulting clusters. It also incorporates a Retrieval-Augmented Generation (RAG) architecture to guide the GenAI model in selecting courses aligned with normative documents, and uses graph centrality metrics to analyze the curriculum structure and assign pedagogical roles to its components based on those metrics.

The method was validated through two complete case studies in the fields of Data Science and Business Administration to test its effectiveness and generalizability. The results show that the generated curricula exhibit high semantic coherence with national and international reference matrices. The adherence analysis, validated by hypothesis testing, demonstrated alignment with normative guidelines that is statistically indistinguishable from that of established programs. Structural analysis revealed a consistent core-periphery organization in both domains, validating the classification of pedagogical roles.

This work contributes an original, end-to-end, reproducible, and validated method that demonstrates the potential of combining Generative Artificial Intelligence and Graph Analysis to efficiently support curriculum design.

Keywords: Higher Education. Curriculum Design. Generative Artificial Intelligence. Graph Theory.

Lista de Figuras

Figura 1 – Exemplo do processo de RAG.	33
Figura 2 – Exemplo de valores de similaridade.	38
Figura 3 – Comparação esquemática da ordem de exploração da BFS e da DFS em um mesmo grafo.	42
Figura 4 – Representação da função dos nós no grafo.	44
Figura 5 – Representação do paradigma Human-In-The-Loop.	45
Figura 6 – Fluxograma representando as etapas do método proposto.	62
Figura 7 – Diagrama do processo de criação da base de conhecimento a partir do documento de referência.	73
Figura 8 – Fluxo da Fase 1: avaliação individual da aderência de cada disciplina, utilizando a base de conhecimento para contextualizar a decisão da IAGen.	74
Figura 9 – Fluxo da Fase 2: seleção consolidada da matriz curricular final, utilizando uma consulta específica e a lista de disciplinas pré-aprovadas.	75
Figura 10 – Distribuição das similaridades de cosseno entre as <i>keywords</i> geradas e as ementas originais.	88
Figura 11 – Distribuição da média de similaridade de cosseno entre as disciplinas geradas e seus respectivos grupos de origem.	91
Figura 12 – Grafo do currículo de Ciência de Dados em <i>layout</i> hierárquico, ilustrando o fluxo pedagógico proposto.	107
Figura 13 – Grafo do currículo de Ciência de Dados em <i>layout</i> direcionado por força, revelando a estrutura de núcleo-periferia.	108
Figura 14 – Grafo do currículo de Administração em <i>layout</i> direcionado por força, revelando a estrutura de núcleo-periferia e a segregação dos papéis pedagógicos.	114
Figura 15 – Grafo do currículo de Administração em <i>layout</i> hierárquico, ilustrando o fluxo pedagógico proposto.	114

Lista de tabelas

Tabela 1 – Comparativo entre stemização e lematização	36
Tabela 3 – Comparativo entre modelos de vetorização de texto.	37
Tabela 4 – Métricas de Centralidade: definições e interpretação pedagógica	43
Tabela 5 – Protocolo de Validação Quantitativa no Fluxo Metodológico.	86
Tabela 6 – Análise de Sensibilidade da Estrutura do Grafo com Diferentes Limiares de Similaridade.	90
Tabela 7 – Dados Originais das Disciplinas da Comunidade G35.	92
Tabela 8 – Dados Originais das Disciplinas da Comunidade G41.	94
Tabela 9 – Dados Originais das Disciplinas da Comunidade G46.	96
Tabela 10 – Distribuição da Origem das Disciplinas na Matriz Curricular Proposta. .	99
Tabela 11 – Análise Consolidada de Similaridade de Conteúdo com Matrizes Nacionais e Internacionais.	100
Tabela 13 – Resultados dos Testes de Hipótese para Alinhamento com as Diretrizes da SBC.	102
Tabela 14 – Teste de Hipótese (Mann-Whitney U) para Coerência com Matrizes Internacionais.	103
Tabela 15 – Classificação Final das Disciplinas da Matriz Proposta por Papel Pedagógico.	105
Tabela 15 – (Continuação) Classificação Final das Disciplinas da Matriz Proposta por Papel Pedagógico.	106
Tabela 16 – Análise Comparativa de Similaridade para o Estudo de Caso de Administração.	109
Tabela 17 – Análise Comparativa de Similaridade para o Estudo de Caso de Administração.	110
Tabela 18 – Resultados dos Testes de Hipótese para Alinhamento com as Diretrizes de Administração.	110
Tabela 20 – Classificação Final das Disciplinas da Matriz de Administração por Papel Pedagógico.	112
Tabela 20 – (Continuação) Classificação Final das Disciplinas da Matriz de Administração.	113

Tabela 21 – Visão Geral dos Principais Resultados de Validação do Método.	115
Tabela 22 – Exemplo de disciplinas utilizadas do curso de Administração da UFPE.	128
Tabela 22 – (Continuação) Disciplinas obrigatórias utilizadas do curso de Administração da UFPE.	129
Tabela 23 – Exemplo de disciplinas utilizadas do curso de Ciências Biológicas da UFPE.	130
Tabela 23 – (Continuação) Disciplinas obrigatórias utilizadas do curso de Ciências Biológicas da UFPE.	131
Tabela 24 – Exemplo de disciplinas utilizadas do curso de Ciências da Computação da UFPE.	132
Tabela 24 – (Continuação) Disciplinas obrigatórias utilizadas do curso de Ciências da Computação da UFPE.	133
Tabela 25 – Exemplo de disciplinas utilizadas do curso de Engenharia da Computação da UFPE.	133
Tabela 25 – (Continuação) Disciplinas obrigatórias utilizadas do curso de Engenharia da Computação da UFPE.	134
Tabela 26 – Exemplo de disciplinas utilizadas do curso de Sistemas de Informação da UFPE.	135
Tabela 26 – (Continuação) Disciplinas obrigatórias utilizadas do curso de Sistemas de Informação da UFPE.	136
Tabela 27 – Exemplo de disciplinas utilizadas do curso de Ciência de Dados da UFC.	137
Tabela 27 – (Continuação) Disciplinas obrigatórias utilizadas do curso de Ciência de Dados da UFC.	138
Tabela 27 – (Continuação) Disciplinas obrigatórias utilizadas do curso de Ciência de Dados da UFC.	139
Tabela 28 – Exemplo de disciplinas utilizadas do curso de Ciência de Dados da UFPB.	139
Tabela 28 – (Continuação) Disciplinas obrigatórias utilizadas do curso de Ciência de Dados da UFPB.	140
Tabela 28 – (Continuação) Disciplinas obrigatórias utilizadas do curso de Ciência de Dados da UFPB.	141
Tabela 29 – Métricas de Centralidade Calculadas para as Disciplinas da Matriz de Ciência de Dados (T=0,6957).	142

Tabela 29 – (Continuação) Métricas de Centralidade Calculadas para Ciência de
Dados. 143

Tabela 29 – (Continuação) Métricas de Centralidade Calculadas para Ciência de
Dados. 144

Tabela 30 – Métricas de Centralidade Calculadas para as Disciplinas da Matriz de
Administração (T=0,6798). 144

Tabela 30 – (Continuação) Métricas de Centralidade Calculadas para Administração. 145

Tabela 30 – (Continuação) Métricas de Centralidade Calculadas para Administração. 146

Lista de Algoritmos

1	BFS(G, s)	41
2	DFS(G, u)	41
3	Formação de Grupos de Equivalência por Componentes Conectados	69

Lista de Siglas

API	<i>Application Programming Interface</i>
BFS	Busca em Largura
BERT	<i>Bidirectional Encoder Representations from Transformers</i>
CoT	<i>Chain-of-Thought</i>
CFA	Conselho Federal de Administração
DFS	Busca em Profundidade
GenAI	<i>Generative Artificial Intelligence</i>
HITL	<i>Human-in-the-Loop</i>
IAGen	Inteligência Artificial Generativa
LLM	<i>Large Language Model</i>
MoE	<i>Mixture-of-Experts</i>
MEC	Ministério da Educação
NDE	Núcleos Docentes Estruturantes
NLP	<i>Natural Language Processing</i>
PLN	Processamento de Linguagem Natural
RAG	<i>Retrieval-Augmented Generation</i>
RL	<i>Reinforcement Learning</i>
RTF	<i>Role, Task, Format</i>
SBC	Sociedade Brasileira de Computação
SFT	<i>Supervised Fine-Tuning</i>
TF-IDF	<i>Term Frequency–Inverse Document Frequency</i>
UFC	Universidade Federal do Ceará
UFPE	Universidade Federal de Pernambuco
UFPB	Universidade Federal da Paraíba

Sumário

1	Introdução	20
1.1	Justificativa	21
1.2	Problema de Pesquisa	22
1.3	Objetivos	23
1.4	Delimitação do Tema	24
2	Referencial Teórico	26
2.1	Teoria Curricular	26
2.1.1	Conectivismo	27
2.2	Inteligência Artificial Generativa	29
2.2.1	Grandes Modelos de Linguagem	29
2.2.1.1	Modelo de Síntese: <i>deepseek-chat</i>	30
2.2.1.2	Modelo de Deliberação: <i>deepseek-reasoner</i>	31
2.2.2	Engenharia de <i>Prompt</i>	31
2.2.3	<i>Retrieval-Augmented Generation</i> (RAG)	32
2.3	Processamento de Linguagem Natural	34
2.3.1	Pré-processamento Textual	34
2.3.2	Vetorização de Textos e Similaridade de Cosseno	36
2.4	Teoria dos Grafos	38
2.4.1	Análise de Modularidade	39
2.4.2	Algoritmos de Busca em Grafos	40
2.4.2.1	Busca em Largura (BFS)	40
2.4.2.2	Busca em Profundidade (DFS)	41
2.4.3	Complexidade e Aplicação Metodológica	42
2.4.4	Métricas de Centralidade como Indicadores Estruturais	42
2.4.5	Definições Formais e Interpretação Pedagógica	43
2.5	Paradigma Human-in-the-Loop (HITL)	44
3	Revisão da Literatura	47
3.1	Considerações Finais	58
4	Método e Ferramentas	60
4.1	Dados Utilizados	61

4.1.1	Coleta e Composição do Corpus	62
4.1.2	Pré-processamento e Padronização	63
4.2	Extração de Palavras-chave para Enriquecimento Semântico	63
4.2.1	Engenharia de Prompt para Extração	64
4.2.2	Protocolo de Validação e Refinamento	65
4.2.2.1	Ciclo de Refinamento Qualitativo (HITL).	65
4.2.2.2	Validação Quantitativa.	65
4.3	Construção da Matriz de Similaridade Semântica	65
4.3.1	Protocolo de Validação e Calibração do Limiar	67
4.3.2	Ciclo de Refinamento Qualitativo (HITL).	67
4.3.3	Suporte à Decisão Quantitativo.	67
4.4	Criação de Novas Disciplinas por Fusão Semântica	68
4.4.1	Identificação de Pares e Calibração do Limiar de Similaridade	68
4.4.2	Formação de Grupos de Equivalência por Componentes Conectados	69
4.4.3	Geração de Conteúdo via IAGen	70
4.4.4	Protocolo de Validação e Refinamento	70
4.4.4.1	Ciclo de Refinamento Qualitativo (HITL).	71
4.4.4.2	Validação Quantitativa.	71
4.4.5	Integração ao Corpus Expandido	72
4.5	Seleção de Disciplinas Guiada por Documento de Referência	72
4.5.1	Preparação da Base de Conhecimento	72
4.5.2	Fase 1: Filtragem Individual por Aderência	73
4.5.3	Fase 2: Seleção Final da Matriz Curricular	74
4.5.4	Protocolo de Validação e Refinamento da Seleção	75
4.5.4.1	Seleção do Teste de Hipótese Apropriado	77
4.6	Estruturação Curricular e Análise de Papéis Pedagógicos via Grafos	77
4.6.1	Construção do Grafo Curricular e Calibração do Limiar	78
4.6.1.1	Análise de Centralidade para Definição de Papéis Pedagógicos	78
4.6.1.2	Classificação das Disciplinas em Eixos Formativos	79
4.6.2	Protocolo de Validação e Refinamento da Estrutura Curricular	81
4.6.2.1	Ciclo de Refinamento Qualitativo (HITL).	81

4.6.2.2	Validação Quantitativa e Estrutural.	82
4.6.3	Ambiente Experimental e Ferramentas	82
4.7	Limitações da Metodologia	84
4.8	Considerações sobre o Método Proposto	85
5	Resultados e Análise Experimental	87
5.1	Estudo de Caso I: Validação da Extração de Keywords	87
5.2	Estudo de Caso II: Validação da Criação de Novas Disciplinas por Fusão	89
5.2.0.1	Calibração do Limiar de Agrupamento via Análise de Sensibilidade	89
5.2.1	Validação da Coesão Semântica das Disciplinas Geradas	90
5.2.2	Análise Qualitativa: Exemplos de Fusão Semântica na Prática	91
5.2.2.1	Exemplo de Integração Vertical: Comunidade G35	92
5.2.2.2	Exemplo de Consolidação Horizontal: Comunidade G41	93
5.2.2.3	Exemplo de Síntese de Tópicos Avançados: Comunidade G46	95
5.3	Estudo de Caso III: Criação da Matriz Curricular para Ciência de Dados	97
5.3.1	Composição da Matriz Curricular Proposta	98
5.3.2	Análise de Coerência com Matrizes de Referência	99
5.3.3	Análise de Aderência e Coerência por Testes de Hipótese	101
5.3.3.1	Aderência às Diretrizes Nacionais (SBC)	102
5.3.3.2	Validação da Coerência em Contexto Global	102
5.3.4	Análise Estrutural e Validação do Grafo Curricular	103
5.3.4.1	Construção e Classificação do Grafo.	104
5.3.4.2	Análise Visual da Estrutura Curricular.	106
5.4	Estudo de Caso IV: Criação da Matriz Curricular para Administração	108
5.4.1	Análise de Coerência com Matrizes de Referência	109
5.4.2	Análise de Aderência às Diretrizes Normativas (MEC/CFA)	109
5.4.3	Análise Estrutural e Validação do Currículo de Administração	111
5.4.3.1	Construção e Classificação do Grafo.	111
5.4.3.2	Análise Visual da Estrutura Curricular.	113
5.5	Discussão Consolidada dos Resultados	115

5.5.1	Resposta à Pergunta de Pesquisa 1: Geração e Validação de Componentes Curriculares Coesos	115
5.5.2	Resposta à Pergunta de Pesquisa 2: Desempenho na Elaboração de Matrizes Curriculares Alinhadas e Interdisciplinares	116
6	Considerações Finais	118
	Referências	122
	APÊNDICES	127
	APÊNDICE A Matrizes Curriculares Utilizadas na Pesquisa	128
A.1	Matriz Curricular do Curso de Bacharelado em Administração - Universidade Federal de Pernambuco (UFPE)	128
A.2	Matriz Curricular do Curso de Bacharelado em Ciências Biológicas - Universidade Federal de Pernambuco (UFPE)	130
A.3	Matriz Curricular do Curso de Bacharelado em Ciências da Computação - Universidade Federal de Pernambuco (UFPE)	131
A.4	Matriz Curricular do Curso de Bacharelado em Engenharia da Computação - Universidade Federal de Pernambuco (UFPE)	133
A.5	Matriz Curricular do Curso de Bacharelado em Sistema da Informação - Universidade Federal de Pernambuco (UFPE)	134
A.6	Matriz Curricular do Curso de Bacharelado em Ciência de Dados - Universidade Federal do Ceará (UFC)	136
A.7	Matriz Curricular do Curso de Bacharelado em Ciência de Dados - Universidade Federal da Paraíba (UFPB)	139
	APÊNDICE B Métricas de Centralidade Detalhadas	142
B.1	Estudo de Caso: Ciência de Dados	142
B.2	Estudo de Caso: Administração	144

1 Introdução

O design de matrizes curriculares no ensino superior é amplamente reconhecido como um processo complexo, frequentemente marcado por subjetividade e baixa escalabilidade (BIANCHESSI, 2021). Instituições educacionais enfrentam crescente pressão para criar currículos que sejam, simultaneamente, robustos, atualizáveis e interdisciplinares (YANG et al., 2024). Tradicionalmente, esse processo envolve comitês de especialistas cuja deliberação, embora fundamental, tende a ser lenta e difícil de replicar em larga escala.

Nesse cenário, ferramentas de Inteligência Artificial (IA), especialmente no campo do Processamento de Linguagem Natural (PLN) e da IA Generativa (IAGen), surgem como alternativas promissoras. Essas tecnologias oferecem capacidades significativas para análise de grandes volumes textuais e geração de conteúdo coerente, possibilitando novas formas de apoio ao planejamento educacional (MIAO et al., 2023). Apesar desse potencial, a ausência de marcos regulatórios consolidados levanta preocupações éticas, sobretudo quanto à privacidade de dados e à validação pedagógica das soluções propostas.

A relevância de se abordar este problema reside nas consequências diretas de um design curricular ineficiente. Matrizes curriculares que não evoluem com a velocidade necessária podem resultar na formação de egressos com competências defasadas em relação às demandas do mercado de trabalho e da sociedade. Além disso, a dificuldade em realizar análises curriculares comparativas e em larga escala pode levar à perpetuação de redundâncias, lacunas de conhecimento e baixa interdisciplinaridade, impactando negativamente a qualidade e a atratividade dos cursos oferecidos pelas instituições de ensino.

A integração entre IAGen e a Teoria dos Grafos revela-se promissora, uma vez que permite tanto o processamento semântico do conteúdo disciplinar quanto a modelagem estrutural do currículo como uma rede complexa de conhecimento (STAVRINIDES; ZUEV, 2023a). Pesquisas recentes evidenciam o valor analítico de redes curriculares, utilizando métricas de centralidade para identificar disciplinas com papel estratégico dentro da estrutura curricular (YANG et al., 2024). No entanto, a implementação integrada dessas tecnologias ainda é incipiente. A construção de grafos de conhecimento, por exemplo, é frequentemente manual e dispendiosa (AYTEKIN et al., 2024), o que reforça a necessidade de *frameworks* automatizados, validados e escaláveis.

Emerge, assim, uma lacuna clara na literatura: a ausência de um *frameworks* de ponta a ponta que não apenas aplique técnicas de IA de forma isolada, mas que as integre em um fluxo de trabalho coeso, desde a análise de fontes heterogêneas até a geração e estruturação de uma matriz final. Mais importante, falta um modelo que formalize a colaboração com o especialista humano, não como um passo final, mas como um componente de supervisão e validação contínua, garantindo que a eficiência computacional seja sempre guiada pela expertise pedagógica.

Diante desse panorama, esta dissertação propõe e valida um método de apoio à decisão que combina as técnicas de IAGen e Análise de Grafos. O objetivo é apoiar educadores e planejadores curriculares na construção e estruturação de matrizes para cursos de graduação, promovendo maior eficiência, coerência e inovação no design curricular.

1.1 Justificativa

A relevância desta pesquisa fundamenta-se na importância estratégica do design curricular para o ensino superior. A matriz curricular é um dos instrumentos que define a qualidade e a pertinência da formação acadêmica. Currículos desalinhados com os avanços científicos e as demandas socioprofissionais podem resultar em egressos com competências defasadas, impactando negativamente a empregabilidade e a capacidade de inovação. A importância de se resolver o problema da elaboração curricular reside, portanto, na necessidade de se criar percursos formativos que sejam, ao mesmo tempo, sólidos em seus fundamentos e ágeis em sua capacidade de adaptação.

Contudo, as abordagens atuais para enfrentar este desafio apresentam limitações significativas. O processo tradicional, baseado em deliberação humana, é de baixa escalabilidade e enfrenta dificuldades para processar a crescente complexidade de informações de forma ágil. Por outro lado, as soluções computacionais existentes tendem a operar de forma isolada e incompleta. A aplicação de IAGen de forma isolada pode agilizar a criação de conteúdo, mas carrega riscos de alucinações factuais e superficialidade. Adicionalmente, a implementação de qualquer solução baseada em dados no contexto brasileiro enfrenta barreiras organizacionais, conforme diagnosticado por [Schneider et al. \(2023\)](#), que apontam o baixo nível de maturidade na governança de dados e na adoção de *Learning Analytics* na maioria das Instituições de Ensino Superior (IES) do país.

A justificativa para a presente dissertação emerge diretamente da insuficiência das abordagens isoladas frente a este cenário complexo. Não basta apenas gerar conteúdo com IAGen, nem apenas analisar uma estrutura com Grafos. A contribuição necessária é um método integrado que utilize a essas tecnologias e consiga superar os desafios do contexto real. Nesta pesquisa, a IAGen é ancorada por uma arquitetura RAG para direcionar ao alinhamento normativo. A Análise de Grafos é usada tanto para guiar a geração de novas disciplinas quanto para validar a estrutura final. O direcionamento e controle de todo este processo pelo paradigma Human-in-the-Loop não é um mero adicional, mas o componente que garante a responsabilidade pedagógica em um cenário de baixa maturidade de dados, como o descrito por [Scheneider et al. \(2023\)](#).

Finalmente, o trabalho se justifica por sua relevância social e pela busca de equidade digital. Conforme destacado por [Freitas et al. \(2023\)](#), a aplicabilidade de soluções de IA em cenários com infraestrutura de conectividade limitada é um desafio crítico para a inclusão no Brasil. Embora o método proposto em sua forma atual dependa de APIs, ele foi projetado com componentes (como a execução local de modelos de *embedding*) e um protocolo de validação que representam um passo em direção a um sistema mais robusto e adaptável a diferentes realidades institucionais. Portanto, este trabalho se justifica pela necessidade de uma solução que seja, ao mesmo tempo, computacionalmente eficiente, pedagogicamente robusta e consciente dos desafios de implementação no contexto educacional brasileiro.

1.2 Problema de Pesquisa

A literatura aponta para uma lacuna metodológica significativa no processo de design curricular. Métodos tradicionais garantem qualidade, mas carecem de escalabilidade. Já as abordagens computacionais existentes atuam de forma isolada: ferramentas baseadas em IAGen limitam-se a gerar componentes curriculares sem considerar sua integração estrutural, enquanto métodos com grafos frequentemente se restringem a pré-requisitos formais ou exigem construção manual ([AYTEKIN et al., 2024](#)).

Dessa forma, destaca-se a ausência de um método que combine a síntese textual da IAGen com a análise estrutural da Teoria dos Grafos em um fluxo de trabalho completo. Além disso, falta incorporar sistematicamente a colaboração do especialista humano por

meio do modelo HITL.

As questões de pesquisa que norteiam este trabalho são:

1. De que forma a combinação de IAGen e similaridade semântica permite gerar e validar novos componentes curriculares coesos a partir da fusão de conteúdos de fontes heterogêneas?

Esta pergunta investiga a viabilidade de se utilizar a IA para a criação dos “blocos de construção” de um currículo. Sua importância reside na necessidade de se ter componentes individuais de alta qualidade antes de se montar a estrutura completa. A presente pesquisa pretende responder a esta questão ao propor um método de fusão semântica e validá-lo quantitativamente por meio de uma métrica de coesão, que mede a fidelidade do componente gerado em relação às suas fontes.

2. Qual o desempenho da utilização de IAGen e Análise de Grafos para elaborar matrizes curriculares interdisciplinares alinhadas a referenciais formativos?

Esta segunda questão aborda o desafio macro: a montagem do artefato final. Sua relevância está em avaliar se o método integrado é capaz de produzir não apenas componentes individuais, mas um currículo completo que seja, ao mesmo tempo, interdisciplinar, coerente e normativamente válido. A dissertação responde a esta questão através de dois estudos de caso, nos quais o desempenho do método é medido por meio de análises de similaridade comparativa e testes de hipótese (para o alinhamento) e pela análise da estrutura do grafo curricular gerado (para a elaboração e organização).

1.3 Objetivos

Propor, desenvolver e validar um método, governado pelo paradigma Human-in-the-Loop (HITL), que integra Inteligência Artificial Generativa e Análise de Grafos para dar suporte à elaboração e estruturação de matrizes curriculares para cursos de graduação. O objetivo proposto será alcançado através dos Objetivos Específicos apresentados abaixo:

1. Desenvolver um método para enriquecimento e representação semântica de componentes curriculares utilizando técnicas de PLN e IAGen.
2. Projetar e avaliar um processo de fusão de conteúdo para a geração de novas

disciplinas, a partir de agrupamentos semanticamente similares.

3. Implementar um sistema baseado em *Retrieval-Augmented Generation* (RAG) para selecionar disciplinas alinhadas a referenciais normativos (LEWIS et al., 2020).
4. Aplicar a Teoria dos Grafos e métricas de centralidade para analisar a estrutura da matriz curricular gerada (STAVRINIDES; ZUEV, 2023a).
5. Validar o método completo em dois domínios distintos: Ciência de Dados e Administração.

1.4 Delimitação do Tema

Esta pesquisa foca no problema do design e da estruturação de matrizes curriculares em um nível institucional. O método desenvolvido é, portanto, concebido como uma ferramenta de suporte à decisão para planejadores curriculares, como coordenadores de curso, membros de Núcleos Docentes Estruturantes (NDE) e comitês pedagógicos. O objetivo central é aumentar a capacidade analítica desses especialistas, fornecendo informações sobre as relações semânticas e estruturais entre os componentes curriculares, que são difíceis de serem percebidos em uma análise puramente manual.

É igualmente importante definir as fronteiras desta pesquisa em relação a problemas adjacentes na literatura de IA na Educação. O método aqui proposto não foca na criação de trilhas de aprendizagem personalizadas ou no agendamento de disciplinas (*course scheduling*) para estudantes, abordagens investigadas por trabalhos como o de (VASILEIOU; YEOH, 2024). Tais abordagens partem da premissa de um currículo já existente e consolidado para otimizar o percurso individual do aluno. A presente pesquisa, em contrapartida, aborda o problema anterior e mais fundamental: como gerar e estruturar o próprio “mapa” curricular sobre o qual essas trilhas individuais poderão, futuramente, ser traçadas.

Por fim, a abordagem se diferencia de propostas de automação total ao adotar, como princípio de design, o paradigma Human-in-the-Loop (HITL). Conforme enfatizado por Mosqueira-Rey et al. (2022) e outros, a aplicação de IA em domínios de alto impacto, como a educação, exige mecanismos de supervisão e validação. Portanto, o objetivo desta tese não é criar um sistema autônomo que substitua o especialista, mas sim um sistema de suporte aos especialistas. O método foi projetado para que a capacidade de processamento

em escala da IA seja constantemente guiada e validada pela expertise pedagógica humana, mitigando os riscos de vieses e garantindo a qualidade e a relevância do produto final.

2 Referencial Teórico

A elaboração de matrizes curriculares no ensino superior é um processo complexo, com desafios conhecidos de subjetividade, escalabilidade e dificuldade de atualização. Esta pesquisa aborda esses desafios por meio de uma metodologia que integra teorias pedagógicas e técnicas computacionais. O presente capítulo tem como objetivo delinear os fundamentos teóricos que sustentam o método proposto para o design curricular.

A metodologia se apoia em quatro pilares conceituais. O primeiro é a teoria do Conectivismo, que oferece uma base pedagógica para a modelagem do currículo como uma rede de conhecimentos. O segundo pilar abrange os fundamentos da Inteligência Artificial Generativa (IAGen), incluindo Grandes Modelos de Linguagem (do inglês *Large Language Models - LLMs*), Engenharia de *Prompt* e a arquitetura de Geração Aumentada por Recuperação (do inglês *Retrieval Augmented Generation - RAG*). O terceiro pilar foca em técnicas de Processamento de Linguagem Natural (PLN) para a vetorização de textos e cálculo de similaridade, e na Teoria dos Grafos para a modelagem estrutural e análise da interconectividade das disciplinas.

Finalmente, toda a abordagem é governada pelo paradigma do Human-in-the-Loop (HITL). Este conceito é apresentado como um componente metodológico e ético essencial para garantir a validação, a qualidade pedagógica e a responsabilidade do processo. Ao detalhar cada um desses fundamentos, este capítulo fornece o arcabouço conceitual necessário para a compreensão do método, sua implementação e a interpretação dos resultados.

2.1 Teoria Curricular

O campo da teoria curricular transcende a visão reducionista do currículo como uma mera lista de disciplinas ou conteúdos a serem transmitidos. Em uma acepção mais ampla e academicamente robusta, a teoria curricular se ocupa da investigação sobre os processos de seleção, organização, e justificação do conhecimento que constitui a experiência educacional (PINAR, 2019). Trata-se de um campo de estudo que questiona não apenas o que se ensina, mas por que se ensina determinado conhecimento em detrimento de outros, e quais as implicações sociais, políticas e subjetivas dessas escolhas.

Autores como (LOPES, 2013) destacam a complexidade e a hibridização do campo, que se afasta de respostas definitivas para se constituir como uma “conversa complicada” (PINAR, 2019). Essa conversa envolve a articulação entre o conhecimento acadêmico e a experiência vivida, buscando compreender como o currículo molda identidades e visões de mundo. No contexto contemporâneo, marcado pela aceleração da produção de conhecimento e pela onipresença das tecnologias digitais, essa conversa se torna ainda mais crítica, demandando novos referenciais teóricos capazes de dialogar com a natureza fluida e conectada do saber. É nesse cenário que o Conectivismo emerge como uma lente pedagógica particularmente relevante para fundamentar a presente pesquisa.

2.1.1 Conectivismo

O Conectivismo é uma teoria da aprendizagem desenvolvida para a era digital, proposta fundamentalmente por George Siemens e Stephen Downes. A teoria parte da premissa de que as abordagens clássicas — como o Behaviorismo, o Cognitivismo e o Construtivismo — se mostram insuficientes por centrarem a aprendizagem no indivíduo, como um processo que ocorre “dentro” da mente. O Conectivismo, em contrapartida, desloca o foco para a rede: a aprendizagem é o processo de criar, manter e navegar por conexões entre diferentes fontes de informação ou “nós” de conhecimento (SIEMENS, 2025).

Segundo (SIEMENS, 2025), o conhecimento em si é distribuído através de uma rede de conexões e, portanto, a aprendizagem consiste na habilidade de construir e atravessar essas redes. Essa perspectiva se sustenta em alguns princípios fundamentais que dialogam diretamente com a proposta metodológica desta dissertação:

1. A aprendizagem reside na diversidade de opiniões e conexões: O conhecimento não é um artefato único e monolítico, mas emerge da articulação de múltiplos pontos de vista. Este princípio justifica a construção de um corpus de dados heterogêneo, que agrega disciplinas de diferentes cursos e instituições, como matéria-prima para a criação de uma matriz curricular mais rica e interdisciplinar.
2. A aprendizagem é um processo de conectar nós ou fontes de informação: Este é o pilar central que fundamenta a modelagem do currículo como um grafo. Nesta pesquisa, as disciplinas são tratadas como nós de conhecimento, e as relações de similaridade

semântica entre elas, calculadas por meio de técnicas de PLN, representam as conexões. A estrutura curricular, portanto, não é vista como uma sequência linear, mas como uma rede topológica que pode ser analisada e otimizada.

3. A aprendizagem pode residir em artefatos não-humanos: O Conectivismo reconhece que o conhecimento pode ser armazenado e manipulado por tecnologia, como bancos de dados ou, no caso deste trabalho, sistemas de Inteligência Artificial (DOWNES, 2008). Este princípio oferece o embasamento pedagógico para o uso de LLMs como ferramentas legítimas no processo de análise, síntese e geração de conteúdo curricular. A IA atua como um nó especializado na rede, capaz de processar conexões em uma escala inviável para um especialista humano.
4. A capacidade de saber é mais crítica do que o que se sabe atualmente: Em um mundo onde o conhecimento se torna obsoleto rapidamente, a habilidade de encontrar, avaliar e conectar novas informações é mais valiosa do que o domínio de um corpo de conhecimento estático. Um método que permite a reconfiguração ágil e a atualização de matrizes curriculares, como o proposto aqui, alinha-se diretamente a essa demanda por adaptabilidade contínua.

A metodologia desenvolvida nesta tese operacionaliza diretamente os princípios do Conectivismo. A teoria sustenta que o conhecimento é uma rede; o método, por sua vez, modela a matriz curricular como um grafo, uma representação matemática de uma rede, onde as disciplinas são os nós e a similaridade semântica entre elas forma as conexões. O Conectivismo afirma que a aprendizagem é o processo de navegar e entender essa rede; a metodologia, portanto, utiliza a Análise de Grafos e as métricas de centralidade para decodificar a estrutura dessa rede, identificando quais disciplinas são centrais, quais atuam como “pontes” e quais são periféricas. Finalmente, a teoria valoriza a diversidade de fontes de informação; o método manifesta isso ao utilizar um corpus heterogêneo de matrizes curriculares e ao empregar a Inteligência Artificial Generativa para fundir esses diferentes pontos da rede em novos componentes curriculares coesos.

Dessa forma, o Conectivismo não é apenas um pano de fundo teórico para esta dissertação, mas seu alicerce pedagógico. Ele oferece um modelo conceitual que legitima e dá sentido à aplicação das técnicas computacionais ao design curricular. A modelagem em grafo se torna uma manifestação da estrutura em rede do conhecimento em conjunto com a análise de centralidade se torna uma forma de entender a importância de cada nó nessa

rede e a fusão de disciplinas por IAGen se torna um processo conectivista de criar novo conhecimento a partir da síntese de fontes diversas. A metodologia, portanto, pode ser compreendida como uma implementação prática de uma teoria da aprendizagem concebida para a complexidade do século XXI.

2.2 Inteligência Artificial Generativa

A IAGen representa um paradigma tecnológico que transcende a mera automação de tarefas, capacitando sistemas computacionais a gerar conteúdo original e complexo, como textos, imagens e códigos. No contexto desta pesquisa, a IAGen não é empregada como uma ferramenta monolítica, mas como um ecossistema de tecnologias cuidadosamente orquestradas para enfrentar o desafio multifacetado do design curricular. Esta abordagem sistêmica se apoia em três componentes interdependentes que, juntos, formam a espinha dorsal do método proposto. O primeiro pilar são os LLMs, que funcionam como o motor computacional, fornecendo a capacidade bruta de processamento semântico e geração textual. O segundo é a Engenharia de *Prompt*, que atua como a interface de controle e comando, permitindo direcionar e restringir a vasta capacidade dos LLMs para garantir a precisão, a relevância e a reprodutibilidade das saídas. Finalmente, o terceiro pilar é a RAG, uma arquitetura que ancora o processo generativo em fontes de conhecimento externas e confiáveis, servindo como um mecanismo de alinhamento e validação para assegurar que os resultados estejam em conformidade com os referenciais pedagógicos e normativos. A articulação desses três pilares permite transformar a IAGen de uma tecnologia de propósito geral em uma solução especializada e responsável para o domínio crítico da educação.

2.2.1 Grandes Modelos de Linguagem

O trabalho de (VASWANI et al., 2017) introduz a arquitetura *Transformer*, que representa uma ruptura fundamental com os modelos sequenciais anteriores, como as Redes Neurais Recorrentes (RNNs). A inovação central do *Transformer* foi o mecanismo de auto-atenção (do inglês *self-attention*), que permite ao modelo ponderar a importância de diferentes palavras em uma sequência de entrada de forma simultânea, em vez de

processá-las uma a uma. Essa capacidade de processamento paralelo, combinada com a auto-atenção, foi o catalisador que permitiu o escalonamento dos modelos a tamanhos e complexidades sem precedentes.

Formalmente, um LLM é uma rede neural baseada na arquitetura *Transformer*, pré-treinada em um corpus textual massivo e diversificado. Seu objetivo fundamental é aprender as distribuições estatísticas da linguagem a um nível de profundidade que lhe permita não apenas prever a próxima palavra em uma sequência, mas também gerar textos coerentes, contextualmente relevantes e com notável versatilidade estilística.

No âmbito desta dissertação, a seleção dos modelos de linguagem não foi uma escolha de conveniência, mas uma decisão metodológica. Em vez de adotar uma abordagem de “tamanho único”, o método emprega uma orquestração de dois modelos distintos, *deepseek-chat* e *deepseek-reasoner*, alocando cada um à tarefa que melhor corresponde às suas capacidades específicas. Essa estratégia de especialização, que utiliza um modelo para síntese em larga escala e outro para deliberação de alta complexidade, é um diferencial metodológico que visa otimizar tanto a eficiência quanto a qualidade dos resultados em cada etapa do processo.

2.2.1.1 Modelo de Síntese: *deepseek-chat*

Para as tarefas de síntese de conteúdo, como a extração de palavras-chave e a fusão de ementas para criar novas disciplinas, o método utiliza o modelo *deepseek-chat*. Este modelo é baseado na arquitetura *DeepSeek-V3* (DEEPSEEK-AI, 2024), que se destaca por sua estrutura de Mistura de Especialistas (do inglês *Mixture-of-Experts* - *MoE*). A arquitetura MoE permite que o modelo ative seletivamente apenas uma fração de seus parâmetros totais para cada *token* de entrada, resultando em uma eficiência computacional significativamente maior durante o treinamento e a inferência, sem sacrificar o desempenho. A escolha do *deepseek-chat* é justificada por seu desempenho em uma ampla gama de *benchmarks* gerais, o que o torna um motor robusto e eficiente para as tarefas de geração estruturada e de alto volume que caracterizam as fases iniciais da metodologia desta pesquisa.

2.2.1.2 Modelo de Deliberação: *deepseek-reasoner*

Para a tarefa mais crítica do método — a seleção final de disciplinas para a lista da matriz curricular — foi empregado o modelo *deepseek-reasoner*, também conhecido como *DeepSeek-R1* (DEEPSEEK-AI, 2025). Este modelo se diferencia por seu processo de treinamento único, que se baseia em aprendizado por reforço (do inglês *Reinforcement Learning* - *RL*) em larga escala, aplicado diretamente sobre o modelo base, sem uma etapa preliminar de ajuste fino supervisionado (do inglês *Supervised Fine-Tuning* - *SFT*).

Esse método de treinamento incentiva o modelo a desenvolver comportamentos de raciocínio, como a capacidade de auto verificação, reflexão sobre os próprios passos e a geração de longas cadeias de pensamento (do inglês *Chain-of-Thought* - *CoT*) para resolver problemas complexos. Seu desempenho em tarefas de raciocínio, matemática e programação, conforme detalhado em seu relatório técnico, alinha-se diretamente com as demandas da tarefa de seleção curricular. Essa tarefa exige não apenas a compreensão de um documento normativo, mas a capacidade de deliberar sobre múltiplas restrições, avaliar compensações e montar um conjunto coeso e otimizado de componentes, justificando sua escolha como o motor de deliberação especializado desta pesquisa.

2.2.2 Engenharia de *Prompt*

A Engenharia de *Prompt* é a disciplina que se dedica ao projeto de instruções de entrada (*prompts*) precisas e eficazes para guiar e restringir o comportamento dos modelos de LLM. A formalização dessa prática, é um passo crucial para a construção de aplicações de IA confiáveis e escaláveis. O trabalho de (WHITE et al., 2023) foi fundamental nesse processo ao introduzir o conceito de Padrões de *Prompt* (*Prompt Patterns*), que oferecem soluções documentadas e reutilizáveis para problemas comuns de interação com LLMs, de forma análoga aos padrões de projeto (*design patterns*) na engenharia de software.

A metodologia desta dissertação adota essa abordagem sistemática ao se basear no *framework* *RTF* (*Role, Task, Format* - Papel, Tarefa, Formato) para todas as interações com os modelos de IAGen. A adoção do RTF é um pilar de rigor metodológico, pois transforma o uso de LLMs de uma atividade exploratória e de difícil replicação em um componente de um método reproduzível. Ele estabelece um padrão para a comunicação com o modelo, direcionando o modelo a gerar resultados personalizados as entradas do

prompt. O *framework* RTF pode ser decomposto em três elementos fundamentais, cada um correspondendo a um padrão de *prompt* específico:

- **Papel (*Role*):** Este componente, que corresponde ao *Persona Pattern* de (WHITE et al., 2023), instrui o modelo a assumir uma identidade específica, como “especialista em currículo acadêmico”. A instrução do Papel serve para “preparar” o modelo de LLM, fazendo com que ele ative as porções mais relevantes de seu conhecimento e adote a lente analítica apropriada para a tarefa em questão. Ao definir um papel, o pesquisador orienta o modelo a operar dentro de um contexto de expertise específico, aumentando a relevância e a qualidade do resultado.
- **Tarefa (*Taks*):** Este é o núcleo diretivo do *prompt*, onde a atividade a ser executada é especificada. A clareza e a precisão nesta etapa são cruciais para minimizar a ambiguidade e garantir que o LLM execute a operação exata que o pesquisador pretende. Uma tarefa bem definida reduz a probabilidade de o modelo interpretar mal a solicitação ou desviar-se do objetivo.
- **Formato (*Format*):** Este elemento, análogo ao *Template Pattern* (WHITE et al., 2023), é vital para a operacionalização da metodologia. Ele determina a estrutura exata da resposta esperada, como uma lista de termos separados por vírgula ou um objeto JSON. Garantir um formato de saída estruturado e previsível é fundamental para as etapas computacionais subsequentes do fluxo de trabalho, como a análise sintática (*parsing*) das palavras-chave geradas ou a integração automatizada de novas disciplinas ao corpus de dados. Sem essa restrição, a saída textual de forma livre do LLM seria inadequada para o processamento automatizado.

2.2.3 *Retrieval-Augmented Generation (RAG)*

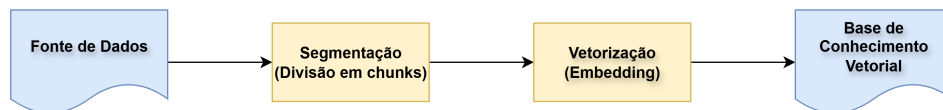
Os Grandes Modelos de Linguagem, apesar de sua impressionante capacidade generativa, possuem limitações inerentes que representam um risco inaceitável em domínios de alta criticidade como a educação. A primeira limitação é que seu conhecimento é estático, confinado aos dados com os quais foram treinados, tornando-os incapazes de acessar informações atuais ou específicas de um domínio que não estava presente em seu treinamento. A segunda, e mais crítica, é a sua propensão a gerar informações factualmente incorretas ou sem fundamento, um fenômeno conhecido como “alucinação”. No contexto

do design curricular, onde as decisões devem ser rigorosamente fundamentadas em padrões autoritativos e diretrizes pedagógicas, depender de um modelo com essas limitações seria metodologicamente e eticamente imprudente.

A arquitetura de Geração Aumentada por Recuperação (RAG) surge como uma solução para essas limitações. Proposta originalmente por (LEWIS et al., 2021), a RAG é uma arquitetura híbrida que combina a vasta memória paramétrica implícita de um LLM com uma memória não paramétrica explícita e externa, como uma base de conhecimento documental, conforme exemplificada na Figura 1. O processo opera em duas fases distintas:

1. Recuperação (*Retrieve*): Dada uma consulta de entrada, um sistema de busca (geralmente um recuperador vetorial) pesquisa em um índice de documentos externos (por exemplo, as diretrizes normatizadas para criação de cursos de graduação) e extrai os trechos de texto semanticamente mais relevantes para a consulta.
2. Aumento e Geração (*Augment and Generate*): Os trechos recuperados são então anexados à consulta original, formando um *prompt* enriquecido e contextualizado. Este *prompt* aumentado é fornecido ao LLM, que gera uma resposta agora “ancorada” nas informações factuais e normativas recuperadas.

Parte 1: Fase de Indexação (Preparação da Base de Conhecimento)



Parte 2: Fase de Recuperação e Geração (Execução da Consulta)

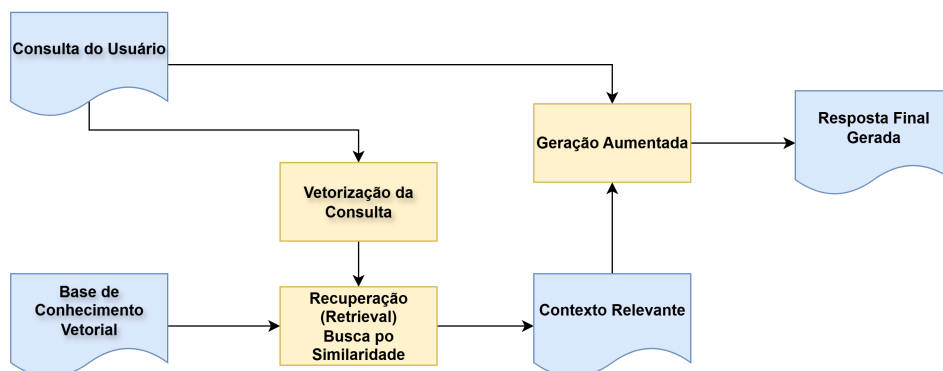


Figura 1 – Exemplo do processo de RAG.

Fonte: (GUPTA et al., 2024)

Na metodologia desta dissertação, a RAG é empregada na etapa de seleção da matriz

curricular final, funcionando como um mecanismo de controle e alinhamento algorítmico. O desafio aqui não é apenas evitar alucinações e vieses, mas garantir que a seleção de disciplinas feita pela IAGen adira estritamente aos princípios pedagógicos, competências e padrões normativos estabelecidos nos documentos de referência. Ao direcionar o modelo de LLM a basear suas decisões de seleção no contexto fornecido pelos trechos recuperados das diretrizes oficiais, o sistema RAG implementa uma forma de alinhamento técnico. Ele restringe o espaço de decisão do modelo, garantindo que seu comportamento seja consistente com a política educacional definida por especialistas humanos.

Essa aplicação da RAG transforma o LLM de um agente puramente criativo e não supervisionado em uma ferramenta que opera dentro dos limites estabelecidos pelo domínio educacional. Conclui-se, portanto, que a RAG é um pilar teórico e técnico indispensável nesta pesquisa, pois assegura que a matriz curricular gerada seja não apenas plausível e coerente, mas também pedagogicamente válida.

2.3 Processamento de Linguagem Natural

O PLN permite a transição dos dados curriculares — inerentemente textuais e não estruturados — para um formato quantitativo e analisável. No contexto desta pesquisa, as técnicas de PLN são fundamentais para interpretar o conteúdo semântico de ementas e nomes de disciplinas, transformando-os em representações numéricas que viabilizam a mensuração de similaridade e a subsequente modelagem do currículo como uma rede de conhecimentos. Esta seção detalha os três processos empregados no método desta pesquisa: o pré-processamento textual, que garante a padronização e a limpeza dos dados, a vetorização, que converte o texto em representações matemáticas densas (*embedding*) e a análise de similaridade entre esses vetores através da técnica de cálculo da similaridade do cosseno.

2.3.1 Pré-processamento Textual

A análise de textos provenientes de fontes do mundo real, como os catálogos de cursos de diversas instituições de ensino superior, exige uma etapa inicial de preparação e normalização. Os dados textuais, em seu estado bruto, são caracterizados por uma

heterogeneidade considerável, contendo variações de capitalização, pontuação, e uma diversidade de palavras comuns (conhecidas como *stopwords*) que, embora essenciais para a estrutura gramatical, carregam pouco significado semântico para fins de análise de conteúdo (JURAFSKY; MARTIN, 2025). A diversidade de fontes e formatos torna o pré-processamento um passo indispensável para mitigar ruídos e garantir a consistência dos dados que alimentarão os modelos (SCHÜTZE et al., 2008).

O objetivo do pré-processamento é, portanto, reduzir a complexidade e a dimensionalidade do vocabulário, padronizando o texto para que as análises semânticas se concentrem nos termos que efetivamente definem o escopo conceitual de cada disciplina. Além de procedimentos como a conversão para letras minúsculas e a remoção de *stopwords*, uma etapa de particular importância é a normalização morfológica, que visa agrupar diferentes formas flexionadas de uma mesma palavra sob um radical comum. Para esta tarefa, duas técnicas principais são amplamente utilizadas: a stemização (*stemming*) e a lematização (*lemmatization*). A escolha entre elas possui implicações diretas na qualidade e na validade dos resultados, especialmente em um domínio semanticamente sensível como a análise curricular (SCHÜTZE et al., 2008).

A stemização é um processo algorítmico que reduz as palavras aos seus radicais, geralmente removendo sufixos. É um método rápido e computacionalmente eficiente, mas pode resultar em radicais que não são palavras reais do idioma (ex: “comput” para “computação”) e pode agrupar palavras com significados distintos de forma incorreta. A lematização, por outro lado, é um processo mais sofisticado que utiliza análise morfológica e conhecimento do léxico de um idioma para reduzir uma palavra à sua forma de dicionário, ou “lema”. Esse processo leva em consideração a classe gramatical da palavra e sempre resulta em uma palavra válida, preservando com maior fidelidade o significado semântico (BALAKRISHNAN; ETHEL, 2014). As diferenças entre as duas técnicas está descri

A análise de ementas curriculares é uma tarefa de alta sensibilidade semântica, onde distinções sutis entre termos como “computação”, “computador” e “computacional” são essenciais, pois representam conceitos distintos e fundamentais dentro de uma estrutura de conhecimento. Para o domínio da análise curricular, onde a precisão semântica é primordial para identificar corretamente as conexões conceituais entre disciplinas, a lematização é a escolha metodológica ideal. A preservação do significado garante que a subsequente medição de similaridade seja mais acurada e pedagogicamente relevante.

Tabela 1 – Comparativo entre stemização e lematização

Características	Lematização	Stemização
Objetivo Principal	Reduzir a palavra à sua forma original.	Reduzir a palavra a seu radical morfológico.
Processo	Baseado em análise morfológica e dicionários.	Baseado em regras heurísticas de remoção de sufixos.
Resultado	Sempre uma palavra válida do idioma.	Pode ser um radical que não é uma palavra.
Custo Computacional	Mais alto, requer recursos linguísticos.	Mais baixo, puramente algorítmico.
Preservação Semântica	Alta. Preserva o significado original.	Moderada. Pode perder nuances semânticas.

2.3.2 Vetorização de Textos e Similaridade de Cosseno

Após o pré-processamento textual, é necessário converter os dados linguísticos em uma forma numérica que permita sua manipulação por algoritmos computacionais. Essa etapa, conhecida como vetorização ou *text embedding*, consiste em transformar palavras, sentenças ou documentos inteiros em vetores de números reais dentro de um espaço matemático de alta dimensionalidade. O princípio fundamental dessa abordagem é que a proximidade entre vetores nesse espaço reflita a semelhança semântica entre os textos: quanto mais próximos os vetores, mais semelhantes os significados. Essa característica torna a vetorização uma ferramenta central para a análise de dados textuais em tarefas como classificação, agrupamento e recomendação. Modelos modernos de vetorização, como o *Bidirectional Encoder Representations from Transformers* (BERT) (DEVLIN et al., 2019), se destacam por serem contextuais, gerando representações distintas para uma mesma palavra dependendo da sentença em que aparece. Isso permite capturar nuances semânticas com elevado grau de fidelidade.

A vetorização de textos evoluiu significativamente, desde modelos baseados em contagem (e.g., TF-IDF), passando por modelos preditivos contextuais (e.g., Word2Vec), até a atual geração de modelos baseados na arquitetura Transformer. Para esta pesquisa, o foco recaiu sobre os modelos Transformer devido à sua capacidade superior de capturar o contexto semântico. A escolha de um modelo específico considerou critérios como a

capacidade de processar textos longos, o desempenho em *benchmarks* e a transparência (código aberto).

Após uma análise comparativa, optou-se pelo modelo *nomic-embed-text* (NUSSBAUM et al., 2024). Sua janela de contexto de 8192 tokens permite a análise integral das ementas sem truncamento. Além disso, o modelo apresenta desempenho competitivo em *benchmarks* de referência como MTEB e LoCo, e sua natureza de código aberto favorece a reprodutibilidade. A Tabela 3 apresenta um comparativo entre o modelo escolhido e outras alternativas notáveis, justificando a seleção.

Tabela 3 – Comparativo entre modelos de vetorização de texto.

Modelo	Abordagem	Contexto Máximo	MTEB	LoCo	Código Aberto
Sentence-BERT (all-mpnet-base-v2)	Bi-Encoder	512 tokens	Bom	Fraco	Sim
OpenAI (text-embedding-3-large)	API Fechada	8192 tokens	Excelente	Ótimo	Não
NOMIC (nomic-embed-text)	Bi-Encoder	8192 tokens	Ótimo	Excelente	Sim

Uma vez que cada disciplina é representada por um vetor de *embedding*, é necessário um método para quantificar a similaridade entre eles. Embora existam diversas métricas para calcular a distância ou a proximidade entre vetores, como a Distância Euclidiana ou a Distância de Manhattan, a similaridade de cosseno é frequentemente preferida para tarefas de PLN.

A razão para esta escolha reside no fato de que, em espaços de alta dimensionalidade como os gerados por *embeddings* de texto, a direção do vetor é um indicador mais confiável de similaridade de conteúdo do que sua magnitude. Métricas como a Distância Euclidiana são sensíveis à magnitude; por exemplo, dois textos sobre o mesmo tópico, mas com comprimentos diferentes (um resumo e um artigo completo), teriam vetores com magnitudes distintas e, portanto, uma grande distância euclidiana, embora sejam semanticamente próximos. A similaridade de cosseno, ao calcular o cosseno do ângulo entre os vetores, ignora a magnitude e foca na orientação. Por essa razão, ela se torna uma medida mais robusta para comparar o conteúdo semântico de textos com tamanhos variados, como as ementas heterogêneas analisadas nesta pesquisa. Formalmente, é definida pela Equação 4.1

$$\text{Similaridade } (D_i, D_j) = \frac{\mathbf{D}_i \cdot \mathbf{D}_j}{\|\mathbf{D}_i\| \|\mathbf{D}_j\|} \quad (2.1)$$

O resultado é um valor que varia de -1 (vetores opostos) a 1 (vetores idênticos), onde valores mais próximos de 1 indicam maior similaridade semântica, conforme demonstrado na Figura 2. A escolha da similaridade de cosseno é particularmente adequada para a análise de textos, pois ela mede a orientação (direção) dos vetores, independentemente de sua magnitude (MANNING, 2009). Isso significa que a métrica é robusta a variações no comprimento dos documentos originais — por exemplo, uma ementa concisa e outra prolixa podem ser consideradas altamente similares se abordarem os mesmos conceitos. A utilização da similaridade de cosseno sobre vetores de *embedding* privilegia a análise do conteúdo semântico, alinhando-se ao objetivo de comparar disciplinas com base no que elas ensinam. É importante notar que a representação vetorial não é exclusivamente semântica, pois modelos *Transformer* modernos também codificam informações sintáticas e contextuais. No entanto, a métrica é projetada para ser robusta a pequenas variações estilísticas ou de estrutura textual. Por exemplo, as representações vetoriais de “o aluno aprenderá a analisar dados” e “a análise de dados será aprendida pelo aluno” tenderão a ser muito próximas no espaço vetorial, pois, apesar da diferença sintática, o núcleo semântico é o mesmo. A similaridade de cosseno, portanto, permite quantificar essa proximidade conceitual de forma eficaz.

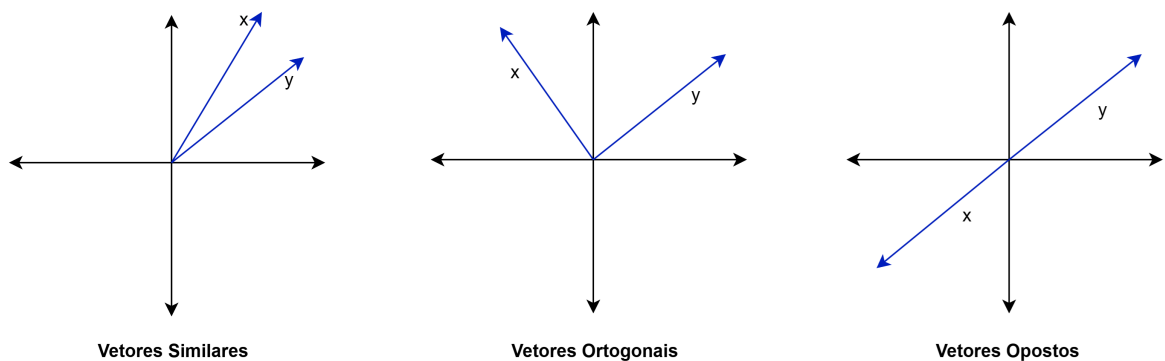


Figura 2 – Exemplo de valores de similaridade.

2.4 Teoria dos Grafos

A Teoria dos Grafos é o campo da matemática que estuda as relações entre objetos de um determinado conjunto. Um grafo G é formalmente definido como um par ordenado $G = (V, E)$, onde V é um conjunto de vértices (ou nós) e E é um conjunto de arestas que

conectam pares de vértices ([CORMEN et al., 2022](#)). No contexto desta pesquisa, os grafos são utilizados como a principal ferramenta para a modelagem da estrutura curricular. Nesta representação, as disciplinas são os vértices da rede, e as arestas representam a similaridade semântica entre elas.

Essa abordagem permite ir além da análise de disciplinas isoladas e passar a investigar o currículo como um sistema complexo de conhecimentos interconectados. A topologia da rede curricular, ou seja, a forma como as disciplinas se organizam e se conectam, revela padrões estruturais que não são aparentes em uma lista linear. A partir da análise dessa topologia, é possível quantificar a coesão de grupos de disciplinas, identificar rotas de aprendizagem e, fundamentalmente, inferir o papel pedagógico de cada componente curricular com base em sua posição na rede. As subseções a seguir detalham os conceitos da Teoria dos Grafos que fundamentam a metodologia deste trabalho.

2.4.1 Análise de Modularidade

A análise de redes complexas frequentemente busca identificar a existência de uma “estrutura de comunidades”, correspondendo a subconjuntos de nós mais densamente conectados entre si do que com o restante da rede. A modularidade é uma métrica que quantifica a qualidade dessa divisão. Proposta por ([NEWMAN; GIRVAN, 2004](#)), a modularidade mede a diferença entre a fração de arestas que existem dentro de uma comunidade e a fração esperada caso as arestas fossem distribuídas aleatoriamente pela rede.

Uma modularidade positiva e alta indica uma estrutura de comunidades forte e bem definida. A fórmula clássica, sem ajuste de resolução, é apresentada na Equação 2.2

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \quad (2.2)$$

[Reichardt e Bornholdt \(2006\)](#) introduziram um parâmetro de resolução γ no modelo de Potts, permitindo ajustar a granularidade das comunidades detectadas. A forma generalizada fica:

$$Q(\gamma) = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \gamma \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \quad (2.3)$$

Onde:

- A_{ij} é um elemento da matriz de adjacência (1 se há aresta entre i e j , 0 caso contrário);
- k_i e k_j são os graus dos nós i e j ;
- m é o número total de arestas do grafo;
- c_i e c_j são as comunidades dos nós i e j ;
- $\delta(c_i, c_j)$ é a delta de Kronecker (1 se $c_i = c_j$, 0 caso contrário);
- γ é o parâmetro de resolução proposto por (REICHARDT; BORNHOLDT, 2006), adotado em variantes como Louvain e Leiden ((TRAAG et al., 2019)).

No âmbito desta dissertação, a análise de modularidade é empregada como um critério técnico para a calibração do limiar de similaridade na etapa de construção do grafo (Seção 4.4). A tarefa de agrupar disciplinas semanticamente similares é, em sua essência, um problema de clusterização ou detecção de comunidade na rede. A maximização da modularidade é uma abordagem consolidada para essa tarefa, pois busca a partição do grafo que melhor representa essa estrutura de *clusters*. O objetivo, portanto, é identificar um limiar que gere uma rede com uma estrutura comunitária intrínseca e estatisticamente significativa, garantindo que os *clusters* de disciplinas para fusão semântica sejam metodologicamente bem fundamentados.

2.4.2 Algoritmos de Busca em Grafos

A exploração sistemática de um grafo é fundamento para diversas análises em redes complexas, notadamente a identificação de componentes conexos e a extração de padrões de conectividade. Nesta dissertação, será apresentado dois algoritmos clássicos descritos em (CORMEN et al., 2022): a Busca em Largura (*Breadth-First Search*, BFS) e a Busca em Profundidade (*Depth-First Search*, DFS).

2.4.2.1 Busca em Largura (BFS)

A BFS parte de um vértice raiz s e “varre” o grafo em camadas, descobrindo primeiro todos os vértices a uma distância $k = 1$, depois aqueles a $k = 2$, e assim por diante. Essa operação é equivalente ao avanço de frentes de onda concêntricas, como

ilustrado na Figura 3. A implementação típica utiliza uma fila FIFO (*First In - First Out*) para gerenciar a fronteira de descoberta e marca cada vértice no momento em que é enfileirado, evitando inserções redundantes e garantindo complexidade $O(|V| + |E|)$.

Algoritmo 1: BFS(G, s)

```

1:  $Q \leftarrow$  fila vazia
2: para todo  $v \in V$ : marcar  $v$  como não visitado
3: marcar  $s$  como visitado;  $Q.enqueue(s)$ 
4: while  $Q$  não está vazia do
5:    $u \leftarrow Q.dequeue()$ 
6:   for cada vizinho  $v$  de  $u$  do
7:     if  $v$  não foi visitado then
8:       marcar  $v$  como visitado
9:        $Q.enqueue(v)$ 
10:    end if
11:  end for
12: end while

```

2.4.2.2 Busca em Profundidade (DFS)

A DFS explora o grafo seguindo um caminho o mais profundo possível antes de retroceder (*backtracking*) ao último nó com vizinhos não visitados. Essa estratégia assemelha-se à exploração de um labirinto, persistindo até encontrar um “beco sem saída” e então retornando. Também tem complexidade $O(|V| + |E|)$ quando implementada adequadamente, marcando cada vértice antes de chamar recursivamente seus vizinhos.

Algoritmo 2: DFS(G, u)

```

1: marcar  $u$  como visitado
2: for cada vizinho  $v$  de  $u$  do
3:   if  $v$  não foi visitado then
4:     DFS( $G, v$ )
5:   end if
6: end for

```

2.4.3 Complexidade e Aplicação Metodológica

Ambos os algoritmos percorrem todas as arestas e vértices alcançáveis em tempo linear, $O(|V| + |E|)$ (CORMEN et al., 2022). A escolha pela BFS na etapa de identificação de componentes conectados do grafo de currículo justifica-se por sua capacidade de:

- Rotular, camada a camada, todos os vértices de uma componente antes de avançar para a próxima;
- Fornecer distâncias mínimas em número de arestas a partir de cada raiz, auxiliando na análise de proximidade temática entre disciplinas;
- Evitar percursos redundantes graças à marcação imediata no enfileiramento.

Assim, na Seção 4.4.2, foi utilizado a BFS iterativamente em cada vértice não visitado para rotular componentes conexos, interpretados como grupos de disciplinas semanticamente equivalentes.

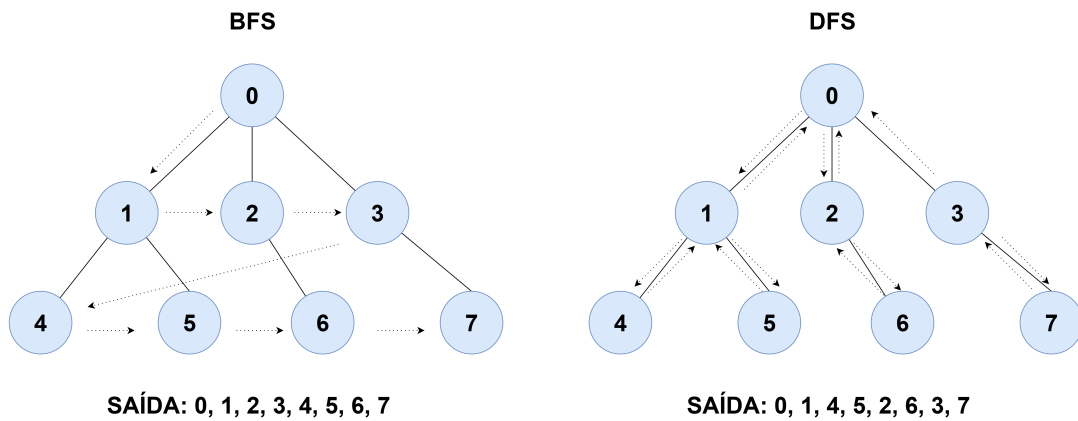


Figura 3 – Comparação esquemática da ordem de exploração da BFS e da DFS em um mesmo grafo.

2.4.4 Métricas de Centralidade como Indicadores Estruturais

Além da detecção de comunidades, a análise de redes complexas recorre a métricas de centralidade para caracterizar o papel estrutural de cada nó no grafo. Na presente dissertação, foram adotadas as definições clássicas de grau, proximidade e intermediação (betweenness), conforme em (OTTE; ROUSSEAU, 2002). Essas medidas permitem não apenas quantificar a importância topológica das disciplinas, mas também inferir sua função pedagógica no currículo.

2.4.5 Definições Formais e Interpretação Pedagógica

A Tabela 4 apresenta um sintese da definição formal das métricas de centralidade e sua aplicação no método proposto neste trabalho, e simplificado na Figura 5.

Tabela 4 – Métricas de Centralidade: definições e interpretação pedagógica

Métrica	Definição Formal	Interpretação Pedagógica
Grau	$C_D(i) = \sum_j m_{ij}$, onde $m_{ij} = 1$ se há aresta entre i e j	Disciplina <i>hub</i> ou fundamental, que se conecta a muitas outras; representa o alicerce conceitual do currículo.
Proximidade	$C_C(i) = \left(\sum_j d(i, j) \right)^{-1}$, onde $d(i, j)$ é o comprimento do menor caminho entre i e j	Disciplina central, com acesso rápido a todo o currículo; item de “fácil alcance” conceitual.
Intermediação	$C_B(i) = \sum_{s \neq i \neq t} \frac{\sigma_{st}(i)}{\sigma_{st}}$, onde σ_{st} é o número de geodésicas entre s, t e $\sigma_{st}(i)$ aquelas que passam por i	Disciplina <i>ponte</i> ou integradora, que conecta módulos distintos do currículo e facilita a interdisciplinaridade.

A classificação dos papéis pedagógicos das disciplinas no método proposto não se baseia em limiares numéricos absolutos, mas sim na posição relativa de cada disciplina no *ranking* de cada uma dessas métricas. A combinação dos valores de centralidade de uma disciplina, em comparação com as demais na mesma rede, permite inferir seu papel estrutural. A tipologia adotada nesta tese correlaciona estes indicadores com as seguintes categorias pedagógicas:

- **Disciplinas de Formação Básica:** Identificadas por possuírem os valores relativamente mais altos de Grau (C_D) e Proximidade (C_C), combinados com valores relativamente baixos de Intermediação (C_B). Sua alta conectividade e fácil acesso as posicionam como o alicerce conceitual do currículo.
- **Disciplinas de Formação Específica (Pontes):** Identificadas primariamente pelos valores relativamente mais altos de Intermediação (C_B), indicando seu papel

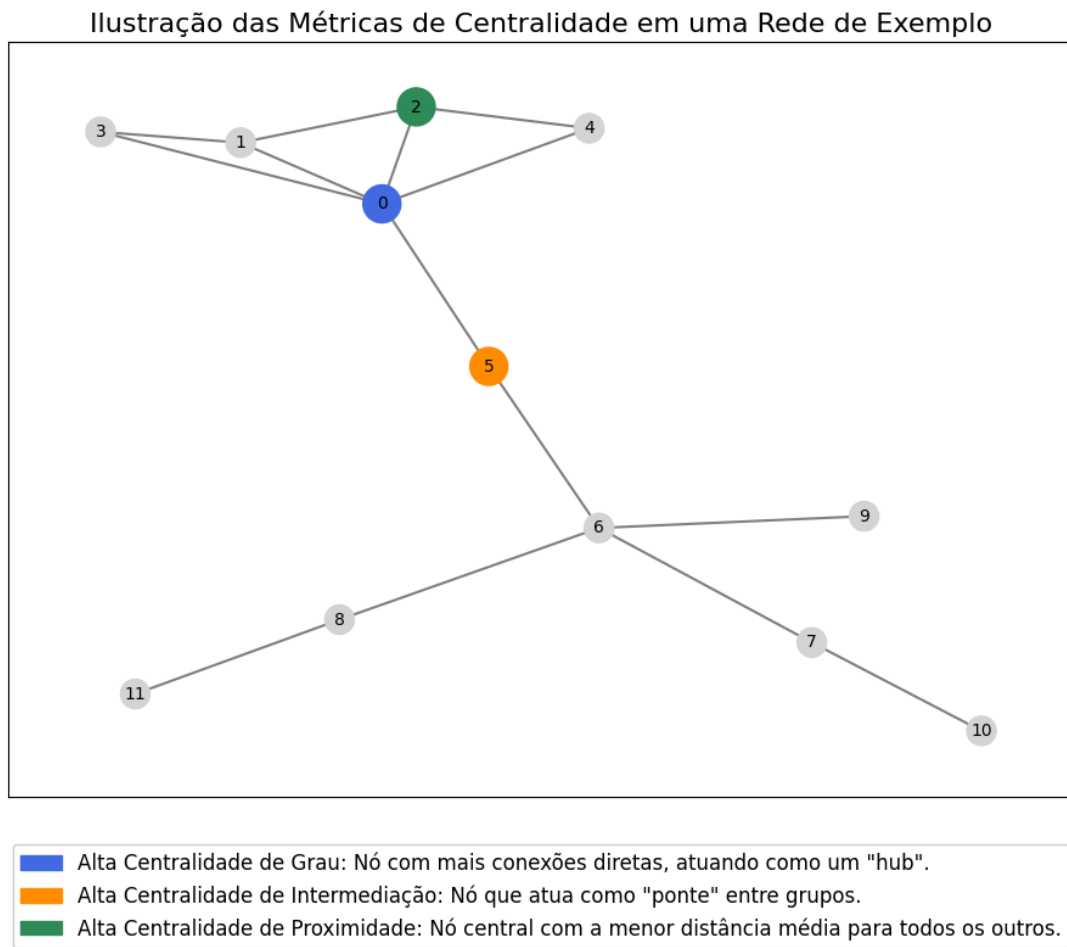


Figura 4 – Representação da função dos nós no grafo.

crucial como conectoras entre diferentes áreas ou módulos do conhecimento.

- **Disciplinas de Formação Complementar (Avançadas):** Caracterizadas por possuírem os valores relativamente mais baixos nas três métricas de centralidade, o que as posiciona na periferia da rede, funcionando como pontos de especialização ou aprofundamento.

Essa tipologia segue recomendações de (WASSERMAN; FAUST, 1994) e (FORTUNATO, 2010), que enfatizam a importância de combinar múltiplas métricas para uma visão abrangente da estrutura de redes curriculares.

2.5 Paradigma Human-in-the-Loop (HITL)

A aplicação de modelos de Inteligência Artificial em domínios de alta complexidade e responsabilidade — como o planejamento educacional — requer mecanismos de validação e controle capazes de garantir qualidade, segurança e alinhamento ético das soluções

geradas. O paradigma Human-in-the-Loop (HITL) surge como um modelo colaborativo que integra a capacidade de processamento de grandes volumes de dados da máquina com o julgamento, a experiência e a intuição do especialista humano (MONARCH, 2021).

Em sua essência, o HITL constitui um ciclo de *feedback* contínuo, onde o sistema de IA processa dados e gera uma proposta preliminar (classificação, recomendação, artefato curricular), então o especialista humano avalia, corrige e refina essa proposta ou a aprova como está e então as decisões e correções do humano retornam ao modelo como novos dados de treinamento, aprimorando seu desempenho futuro (WU, 2021). A Figura

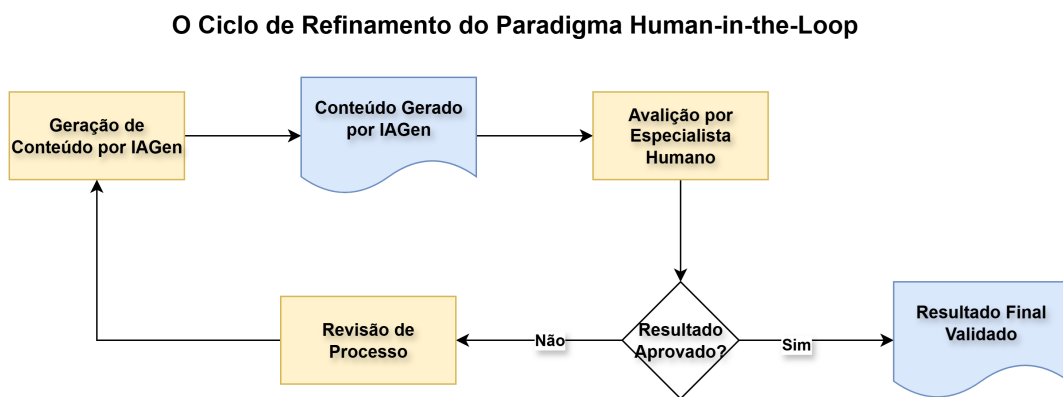


Figura 5 – Representação do paradigma Human-In-The-Loop.

Diferentemente da automação total, o HITL mantém a responsabilidade final com o humano, que atua como validador, árbitro e curador estratégico. Essa atuação sinérgica entre Humano-IA possibilita que a IA identifique padrões em larga escala e propõe soluções em alta velocidade, enquanto que o especialista humano interpreta nuances, lida com ambiguidade e assegura coerência pedagógica e ética (PADOVANO; CARDAMONE, 2024).

A relevância do HITL é ainda mais acentuada no contexto da Inteligência Artificial Generativa. Modelos de linguagem — apesar de sua fluência e capacidade de síntese — podem incorrer em vieses e “alucinações” (informações factualmente incorretas), o que impõe um risco crítico quando aplicados à educação (CHAKRABORTY, 2024; KUMAR et al., 2024a). A supervisão humana, portanto, atua como um mecanismo de controle indispensável para mitigar esses riscos e garantir a robustez e a confiabilidade dos resultados (SANDU et al., 2024a).

No método proposto o paradigma HTIL é operacionalizado por meio de um protocolo de validação explícito e multifásico, aplicado em cada etapa — desde a extração de palavras-chave e a fusão de disciplinas até a seleção final da matriz curricular e sua

representação em grafo. Ao formalizar os pontos de intervenção e as ações corretivas do especialista, o método assegura que a criatividade analítica da IA seja constantemente guiada pela expertise pedagógica, transformando o design curricular em um processo de cocriação responsável e transparente.

3 Revisão da Literatura

A elaboração de matrizes curriculares eficientes, atualizadas e alinhadas às necessidades do mercado e às diretrizes dos órgãos regulamentadores representa um desafio significativo para as instituições de ensino. Em um cenário onde se busca personalizar a construção de disciplinas e currículos, a aplicação de Inteligência Artificial Generativa (IAGen) surge como uma estratégia promissora, capaz de superar obstáculos tradicionais por meio da identificação automatizada de similaridades entre disciplinas, geração de novos conteúdos e organização estrutural do currículo. Diversos estudos têm explorado métodos automatizados com o objetivo de oferecer formações que dialoguem com as demandas contemporâneas, contribuindo para a padronização e modernização dos processos decisórios na gestão educacional.

Neste capítulo, serão discutidos estudos que investigam diversas abordagens para a automação e estruturação de currículos, ressaltando tanto as contribuições quanto as limitações de cada metodologia. Serão examinadas técnicas de extração semântica, modelagem de similaridade e análise de redes, com ênfase em como essas estratégias têm sido aplicadas para desenvolver propostas inovadoras de integração interdisciplinar. Ao final, a Tabela 1 sintetizará os principais trabalhos, métodos e resultados, proporcionando uma visão comparativa dos avanços alcançados na área.

A exploração da IAGen como ferramenta de transformação na educação tem sido objeto de extensas revisões recentes, que buscam consolidar o conhecimento e apontar direções futuras. Nesse contexto, o trabalho de (BAHROUN et al., 2023) apresenta uma revisão abrangente da literatura sobre o uso de IAGen em ambientes educacionais, utilizando uma abordagem de análise bibliométrica e de conteúdo para mapear o cenário de pesquisa. A principal contribuição do estudo é a análise e identificação de *clusters* de pesquisa, destacando a avaliação automatizada e a geração de materiais didáticos como temas proeminentes. A pesquisa revela um crescimento exponencial no interesse pela área, mas também aponta, como limitação, que muitas das aplicações propostas ainda se encontram em estágios conceituais ou de validação preliminar. Este mapeamento conecta-se diretamente ao presente trabalho de duas formas. Primeiramente, ele valida a relevância e a atualidade da investigação sobre o uso da IAGen para tarefas complexas de design educacional. Em segundo lugar, ao posicionar a elaboração de matrizes curriculares

como uma aplicação avançada de geração de conteúdo, este trabalho se apresenta como uma contribuição prática e metodológica que avança da discussão conceitual, identificada como um ponto de atenção por (BAHROUN et al., 2023), para a proposição de um *framework* estruturado, completo e validado.

Corroborando essa visão e aprofundando a investigação no contexto do ensino superior, o trabalho de (CHAKRABORTY, 2024) apresenta uma revisão de escopo que sintetiza 77 artigos para mapear as aplicações e os impactos da IAGen. O estudo vai além do mapeamento de tendências, detalhando categorias de aplicação como a criação de conteúdo personalizado e a automação de avaliações, além de sistematizar os desafios associados à tecnologia. Entre as contribuições relevantes, destacam-se as discussões sobre preocupações éticas, o risco de vieses algorítmicos e a necessidade de desenvolver um letramento em Inteligência Artificial (IA) para educadores e alunos. A conexão com este trabalho se estabelece em dois níveis: primeiro, ao posicionar a geração de matrizes curriculares como uma aplicação de alto impacto na categoria de “criação de conteúdo”; segundo, e mais importante, ao abordar diretamente os desafios éticos. A presente proposta, ao incorporar um *framework Human-in-the-Loop (HITL)* e a utilização de Geração Aumentada por Recuperação (RAG) que oferece um protocolo de validação que diminui os riscos de vieses levantados por (CHAKRABORTY, 2024), assegurando que as etapas do método seja sempre validada através da participação Humano-IA.

Completando este panorama geral sobre o estado da arte, o estudo de (SANDU et al., 2024b) também apresenta uma análise da literatura, focando no papel duplo da IAGen como uma força de otimização e, ao mesmo tempo, como uma fonte de novos desafios para a educação. A análise do trabalho é particularmente útil por estruturar o debate em torno das “oportunidades” — como a automação de tarefas repetitivas e a criação de conteúdo educacional — e dos “desafios”, que incluem a garantia da qualidade e a necessidade de supervisão humana para evitar a propagação de informações incorretas. O trabalho conecta-se à presente tese ao reforçar a premissa de que a automação na criação de conteúdo curricular (uma “oportunidade”) não pode ser irrestrita. A metodologia aqui proposta, com seu HITL em múltiplos estágios, é uma resposta direta à necessidade de “supervisão humana” e “garantia da qualidade” que (SANDU et al., 2024b) identificam como um desafio crítico, posicionando o especialista humano como um validador essencial no processo.

Tendo estabelecido a relevância e os desafios da IAGen na educação através dessas revisões abrangentes, a análise se volta agora para estudos de caso que aplicam estas tecnologias a problemas específicos de design de cursos. Um exemplo prático é encontrado em (KUMAR et al., 2024b), que apresentam um estudo de caso sobre o uso do *ChatGPT* para auxiliar no design de um curso de ciências naturais, utilizando uma análise *SWOT* (*Strengths, Weaknesses, Opportunities, and Threats*) para avaliar os resultados. A principal contribuição do trabalho é a análise pragmática das capacidades e deficiências do modelo para esta tarefa. Entre os pontos fortes, os autores destacam a agilidade na geração de rascunhos de ementas e na criação de atividades de aprendizagem diversificadas. Contudo, como fraqueza fundamental, o estudo aponta o risco significativo de geração de informações factualmente incorretas (alucinações) e a tendência do modelo a produzir conteúdo superficial sem uma curadoria humana rigorosa.

Este estudo de caso conecta-se ao presente trabalho ao demonstrar a viabilidade do uso de IAGen para a tarefa, mas, principalmente, ao expor as limitações de uma abordagem que depende de um único *prompt* a um modelo de propósito geral. A metodologia proposta nesta dissertação busca superar essas deficiências de duas maneiras: primeiro, ao utilizar um processo de RAG que ancora as respostas da Inteligência Artificial em documentos de referência, minimizando o risco de alucinações; e segundo, ao implementar o HITL, que aponta como necessária a curadoria do especialista em cada etapa, abordando diretamente a necessidade de supervisão identificada por (KUMAR et al., 2024b).

Na mesma linha de investigação, (LU; ZOGHI, 2024) apresentam um estudo de caso sobre o uso de IAGen para o design de uma disciplina de final de graduação (*capstone course*), um componente curricular que exige a integração de múltiplos conhecimentos do programa. A análise do estudo corrobora os achados sobre ganhos de eficiência na ideação e estruturação do curso, mas sua principal contribuição é a ênfase em um processo de refinamento iterativo. Os autores descrevem como a qualidade do material gerado foi aprimorada através de múltiplos ciclos de interação com a IA, ajustando os *prompts* para alinhar o resultado com os objetivos de aprendizagem específicos. O trabalho também reforça o papel do especialista humano como o curador final do conteúdo, responsável por garantir a coerência pedagógica.

A conexão com a presente tese é dupla. Primeiro, o processo de refinamento iterativo descrito por (LU; ZOGHI, 2024) é uma forma de validação manual que é semelhante ao

HITL proposto no método, a ser aplicado para validar cada etapa do método e liberar as informações para a próxima etapa, embora de maneira menos formalizada. Segundo, ao demonstrar a eficácia da IAGen para um disciplina *capstone*, o estudo reforça a viabilidade de se aplicar a tecnologia a desafios curriculares de alta complexidade, como a seleção de disciplinas e estruturação de uma matriz curricular inteira, que é o escopo deste trabalho.

Avançando para as aplicações diretas da tecnologia, o trabalho de (CHRIS; SHERIFFDEEN, 2024) apresenta uma análise sobre como a IAGen está moldando o desenvolvimento de currículos e modernizando o conteúdo educacional. A análise dos autores explora o potencial da IAGen para criar experiências de aprendizagem mais dinâmicas e personalizadas. A contribuição do artigo está em discutir as implicações práticas dessa tecnologia, argumentando que ela pode servir como um catalisador para a inovação curricular, ao mesmo tempo em que levanta questões sobre a necessidade de novas competências para os educadores que irão utilizar essas ferramentas.

Este trabalho conecta-se à presente tese ao fornecer um contexto adicional sobre o impacto da IAGen no design curricular. Ele se alinha com os outros estudos de caso práticos, reforçando a relevância do problema de pesquisa. A metodologia desenvolvida nesta dissertação pode ser vista como uma resposta concreta para realizar a “modernização do conteúdo educacional” que os autores discutem, oferecendo um *framework* estruturado para aplicar a IAGen de forma eficaz e governada pelo especialista.

Adicionando uma perspectiva focada em um paradigma pedagógico moderno, o trabalho de (ROS, 2024) apresenta uma metodologia para utilizar a IAGen no design de sílabos (planos de ensino) orientados por competências. A análise do estudo é valiosa por focar não apenas na geração de conteúdo, mas na garantia de que este conteúdo esteja alinhado a um *framework* de competências predefinido, demonstrando como a IAGen pode ser instrumentalizada para auxiliar na padronização e na implementação de abordagens de ensino baseadas em competências em larga escala.

O conceito de alinhar um sílabo a “competências”, como feito por (ROS, 2024), é conceitualmente análogo à etapa de validação por aderência da metodologia apresentada (Seção 4.5), na qual se usa a técnica de RAG para alinhar as disciplinas às “diretrizes” e “competências” descritas nos documentos de referência. Portanto, este trabalho valida a premissa de que a IAGen pode ser efetivamente guiada por documentos normativos, reforçando a abordagem central de validação por alinhamento utilizada nesta dissertação.

Além da geração de novos conteúdos, a literatura também explora o uso da IAGen para analisar e validar estruturas curriculares já existentes. Nesse sentido, (CONKLIN et al., 2024) apresenta o conceito de Mapeamento Curricular Assistido por IA (*AI-Assisted Curriculum Mapping*), que consiste em utilizar a IA para automatizar o processo de mapear os componentes de um currículo (disciplinas, objetivos de aprendizagem) a um conjunto de padrões externos, como diretrizes de acreditação. A principal contribuição desta abordagem, conforme a análise do autor, é a automação de uma tarefa tradicionalmente manual e dispendiosa, permitindo que as instituições realizem análises de lacunas (*gap analysis*) e verifiquem a aderência de seus cursos a padrões de qualidade de forma mais eficiente. A limitação, no entanto, é que o foco da ferramenta é a análise de currículos pré-existent, não a criação de novos.

O trabalho de (CONKLIN et al., 2024) conecta-se à presente pesquisa por validar o uso da IAGen como uma ferramenta eficaz para a análise de alinhamento curricular, que é uma das etapas de validação do método proposto (Seção 4.5). Contudo, a proposta exposta nesse trabalho avança em relação a essa abordagem. Enquanto o Mapeamento Curricular Assistido por IA analisa o que já existe, o *framework* proposta, primeiro gera um currículo inédito e interdisciplinar e, em seguida, aplica uma validação de alinhamento similar, porém nas etapas de validação nas Seções 5.3.3.1 e 5.4.2, demonstra que é possível usar a IAGen para validar matrizes curriculares existentes. Dessa forma, a metodologia apresentada neste trabalho, abrange o ciclo completo, do design e síntese à validação e estruturação, oferecendo uma solução para o problema da elaboração de matrizes curriculares.

Para viabilizar as aplicações de geração e análise curricular discutidas anteriormente, é fundamental a existência de um método robusto para quantificar a similaridade semântica entre os conteúdos educacionais. A literatura em Processamento de Linguagem Natural (PLN) tem se dedicado a este problema. Nesta área, o estudo comparativo de (ABDOUNE et al., 2024) apresenta uma análise da eficácia de diferentes modelos de PLN para medir a similaridade semântica entre recursos educacionais de *MOOCs* (*Massive Open Online Courses*). O trabalho avalia e compara diversos modelos de vetorização de texto (*word embeddings*), como Word2Vec, GloVe e FastText, em conjunto com a métrica de similaridade de cosseno para aferir a proximidade conceitual dos conteúdos.

A principal contribuição da pesquisa é a sua análise empírica, que demonstra a superioridade de abordagens baseadas em *embeddings* para capturar as nuances semânticas

do texto educacional, em detrimento de métodos mais antigos baseados apenas em contagem de palavras. O estudo oferece um panorama sobre as vantagens e desvantagens de cada modelo, fornecendo um guia para a seleção da técnica mais adequada. Este trabalho conecta-se diretamente com a metodologia desta dissertação (Seção 4.3) ao fornecer o embasamento teórico e empírico para a escolha de utilizar a vetorização de textos e a similaridade de cosseno como o motor para a construção da matriz de similaridade curricular. Enquanto (ABDOUNE et al., 2024) comparam modelos como Word2Vec e GloVe, este trabalho aplica o mesmo princípio fundamental, mas utilizando um modelo de *embedding* mais moderno (*nomic-embed-text*) que é guiado pelo contexto da palavra dentro da texto. O estudo de (ABDOUNE et al., 2024), portanto, valida a abordagem de quantificar a relação entre disciplinas através de suas representações vetoriais.

Para operacionalizar as análises estruturais mencionadas, esta dissertação se apoia em conceitos fundamentais da Análise de Redes Sociais (ARS), um campo que fornece o arcabouço teórico para a medição da importância dos nós em uma rede. Neste domínio, o trabalho de (OTTE; ROUSSEAU, 2002) apresenta a Análise de Redes Sociais como uma estratégia para as ciências da informação, detalhando as métricas de centralidade. A principal análise dos autores é a sua definição e interpretação das três principais medidas de centralidade: a Centralidade de Grau como uma medida de conectividade, a Centralidade de Intermediação (*betweenness*) como uma medida do papel de “ponte” no fluxo da rede, e a Centralidade de Proximidade (*closeness*) como uma medida de acesso rápido aos demais componentes da rede.

A metodologia de classificação de disciplinas em papéis pedagógicos (Formação Básica, Específica e Complementar), descrita na Seção 4.6 é uma aplicação direta das definições propostas por (OTTE; ROUSSEAU, 2002). O trabalho deles fornece o embasamento teórico que nos permite traduzir os valores quantitativos de centralidade, calculados para cada disciplina, em uma interpretação qualitativa sobre seu papel estrutural dentro da rede curricular.

A literatura recente tem explorado a aplicação da Análise de Redes Sociais (ARS) e suas métricas de centralidade para investigar diversas estruturas e dinâmicas no contexto educacional e demonstram o uso de métricas como Grau, Intermediação e Proximidade para identificar atores ou componentes curriculares chave (WANG et al., 2025; TSONI et al., 2024; AMORIM; MELO, 2024), analisar padrões de comunicação em fóruns online,

mapear redes de colaboração acadêmica (TSONI et al., 2024; AMORIM; MELO, 2024), e compreender a estrutura geral de redes curriculares, frequentemente revelando padrões como núcleo-periferia. Alguns estudos também buscaram correlacionar essas métricas estruturais com resultados de aprendizagem ou desempenho acadêmico (WANG et al., 2025; TSONI et al., 2024; SADAF et al., 2024).

Investigações mais aprofundadas revelam nuances importantes. O trabalho de Wang et al. (2025), por exemplo, aplicaram Grau, Intermediação e Proximidade em redes curriculares baseadas em suporte a competências e pré-requisitos, identificando cursos básicos de informática como hubs centrais e disciplinas práticas inovadoras como isoladas. Limitações comuns incluem o foco em estudos de caso únicos, redes específicas (citações, fóruns, pré-requisitos), ou o uso das métricas de forma predominantemente descritiva (EVANS; CHEN, 2022; TSONI et al., 2024; WANG et al., 2025).

Embora a literatura valide a relevância das métricas de Grau, Intermediação e Proximidade para caracterizar a importância local, o papel de ponte e a acessibilidade global de nós em redes educacionais (OTTE; ROUSSEAU, 2002; WANG et al., 2025; TSONI et al., 2024; AMORIM; MELO, 2024), a presente dissertação avança ao aplicá-las especificamente a uma rede curricular construída a partir da similaridade semântica entre os componentes. A contribuição distintiva reside na proposição e aplicação de uma heurística formal (detalhada na Seção 4.6.1.2), baseada em quartis, que utiliza a combinação dessas três métricas para classificar sistematicamente todas as disciplinas da matriz em papéis pedagógicos funcionais (Formação Básica, Formação Intermediária, Formação Complementar). Essa abordagem transforma a análise de centralidade de uma ferramenta primariamente descritiva para um *framework* prescritivo e quantitativo, capaz de informar diretamente a estruturação e o design curricular com base na função topológica de cada disciplina dentro da rede de conhecimento.

Uma vez estabelecida a validade da Teoria dos Grafos para a análise curricular, a literatura se diversifica nas abordagens para a construção da própria rede. A metodologia mais tradicional para modelar as conexões entre disciplinas é baseada nas relações formais de pré-requisitos. Um exemplo desta abordagem é o trabalho de (STAVRINIDES; ZUEV, 2023b), que apresentam um método para analisar currículos acadêmicos através da construção de redes de pré-requisitos. Neste modelo, as disciplinas são os nós e uma aresta direcionada é criada de uma disciplina A para uma disciplina B se A for um

pré-requisito formal para B. A principal contribuição do estudo é a sua análise de como esta estrutura de rede pode revelar o fluxo de conhecimento e identificar disciplinas críticas que servem como base para muitas outras ou que atuam como gargalos no percurso do estudante.

A limitação deste estudo reside na dependência da existência de dados de pré-requisitos bem definidos, formais e consistentes, que nem sempre estão disponíveis ou refletem todas as conexões conceituais entre as disciplinas. Este trabalho conecta-se à presente dissertação ao servir como um contraponto metodológico fundamental. Enquanto a abordagem de (STAVRINIDES; ZUEV, 2023b) modela a estrutura *formal* do currículo, a proposta do método apresentado nesse trabalho, é baseada em similaridade semântica entre as disciplinas pertencentes a rede. O método proposto oferece, portanto, uma vantagem em cenários onde os dados de pré-requisitos são inexistentes ou incompletos, e é capaz de descobrir afinidades entre disciplinas que não possuem uma relação formal de dependência, preenchendo uma lacuna importante deixada por abordagens baseadas exclusivamente em pré-requisitos.

Uma outra abordagem para a modelagem de currículos, que busca uma representação das relações entre conceitos, é a utilização de Grafos de Conhecimento (KGs). O survey de (QU et al., 2024) apresenta um mapeamento extensivo das abordagens e aplicações de KGs no domínio da educação. A análise do trabalho categoriza as aplicações de KGs em áreas como sistemas de recomendação e planejamento de percursos de aprendizagem. A contribuição do estudo é a consolidação do estado da arte sobre KGs educacionais, detalhando os métodos para sua construção, que frequentemente envolvem a extração de entidades e relações a partir de fontes textuais para popular uma ontologia predefinida.

Este trabalho é relevante para a presente tese por estabelecer uma distinção clara entre a abordagem de KGs e a metodologia aqui proposta. Enquanto os KGs, conforme descritos por (QU et al., 2024), se baseiam em relações semânticas formais e explícitas (e.g., “conceito A - faz parte de - disciplina B”), o método desta dissertação utiliza um grafo de similaridade não-estruturado. A vantagem da abordagem aqui desenvolvida reside na sua flexibilidade e menor custo de implementação, uma vez que não requer a construção de uma ontologia formal, gerando a estrutura de relações diretamente a partir da semântica nos textos curriculares.

O trabalho de (TUZÓN et al., 2025), apresenta uma abordagem de redes complexas

para analisar a integração entre os currículos de Física e Matemática. O método envolve a construção de uma rede onde os nós representam conceitos extraídos das ementas das disciplinas, permitindo a análise da coerência e da centralidade desses conceitos. A principal contribuição do estudo é a sua análise que utiliza a estrutura da rede para quantificar a integração entre as duas áreas, identificando os conceitos que atuam como pontes interdisciplinares. A abordagem permite uma avaliação orientada por dados da coerência curricular, baseada na forma como os conceitos fundamentais se conectam.

Assim como (TUZÓN et al., 2025) utilizam a análise de redes para identificar conceitos centrais, a metodologia aqui proposta emprega as mesmas métricas de centralidade para identificar os papéis pedagógicos das disciplinas como um todo. Enquanto a análise deles opera em um nível de granularidade mais fino (conceitos), ambos os trabalhos compartilham o princípio fundamental de usar a topologia da rede de conhecimento para inferir a importância estrutural de seus componentes. O trabalho deles, portanto, valida o uso de redes complexas como uma ferramenta para a análise estrutural de currículos científicos.

Aplicando uma abordagem similar de análise de redes a outro domínio de engenharia no contexto brasileiro, o trabalho de (WALKER; REGO, 2025) apresenta um método para a simplificação dos eixos temáticos do curso de Engenharia de Produção usando Teoria dos Grafos. A metodologia consiste na construção de um grafo de disciplinas, onde as arestas representam a correlação de conteúdo, e na aplicação de algoritmos de detecção de comunidade para identificar agrupamentos coesos que possam formar novos eixos curriculares. A principal contribuição do estudo é a sua análise, que demonstra como a detecção de comunidade pode ser usada como uma ferramenta para a reestruturação curricular, oferecendo uma forma quantitativa de agrupar disciplinas por afinidade para propor uma organização mais enxuta e coerente.

A conexão com a presente tese é significativa, pois ambos os trabalhos partem de um princípio semelhante: a construção de um grafo baseado na afinidade de conteúdo entre as disciplinas. Contudo, os objetivos da análise divergem. Enquanto (WALKER; REGO, 2025) utilizam a detecção de comunidade para simplificar e agrupar o currículo em eixos, a metodologia aqui proposta foca na análise de centralidade para classificar o papel pedagógico individual de cada disciplina (Básica, Específica, Complementar). Os trabalhos são, portanto, complementares, demonstrando que o mesmo tipo de grafo curricular pode

ser explorado para diferentes finalidades analíticas, seja a reorganização estrutural ou a classificação funcional de seus componentes.

A análise de grafos em contextos educacionais é também aplicada a um problema adjacente, mas distinto, da elaboração de matrizes curriculares: a geração de trilhas de aprendizagem personalizadas para os estudantes. Nesta frente, o trabalho de [Ramos et al. \(2016\)](#) apresenta uma ferramenta computacional projetada para a identificação e visualização de “Trilhas de Aprendizagem”. A abordagem modela um domínio de conhecimento como um grafo e utiliza algoritmos para extrair sequências de aprendizado que podem ser recomendadas aos alunos. A principal contribuição do estudo é a análise e implementação de uma solução focada no usuário final (o estudante), oferecendo um meio visual para que ele possa navegar por um conjunto de conteúdos ou disciplinas de forma estruturada, apoiando assim sua autonomia e planejamento de estudos.

A conexão com a presente tese se dá pela delimitação de escopo. Enquanto a ferramenta de ([RAMOS et al., 2016](#)) opera sobre um conjunto de disciplinas já existente para criar percursos individuais, a metodologia aqui proposta aborda o problema anterior e mais fundamental: como gerar e estruturar o próprio conjunto de disciplinas que comporá o “mapa” sobre o qual tais trilhas poderão ser traçadas. Desta forma, os dois trabalhos são complementares; a matriz curricular estruturada gerada por este método poderia servir como o *input* ideal para uma ferramenta de identificação de trilhas de aprendizagem.

Complementando a visão focada no estudante, o estudo de ([CHOI et al., 2024](#)) explora o uso de Inteligência Artificial Generativa para o design instrucional automatizado. A abordagem utiliza a IAGen para apoiar a criação de mapas de curso — incluindo objetivos, atividades e avaliações — com base em descrições e resultados de aprendizagem. A principal contribuição da pesquisa é sua análise da aplicação da IAGen ao processo de planejamento pedagógico, demonstrando o potencial da tecnologia como parceira no desenvolvimento de cursos personalizados e escaláveis.

Assim como no caso das trilhas de aprendizagem, este trabalho conecta-se à tese ao delinear claramente os diferentes níveis do problema do planejamento curricular. A solução de ([CHOI et al., 2024](#)) presume a existência de uma matriz curricular fixa e bem definida. A metodologia proposta nesta dissertação, por sua vez, foca na etapa anterior de *criar e estruturar* essa matriz. Juntos, os trabalhos ilustram uma possível cadeia de valor completa: enquanto o método proposto nesta dissertação visa estruturar o conteúdo

curricular (o ‘o quê’), a abordagem de (CHOI et al., 2024) contribui para a organização pedagógica detalhada e alinhada dos componentes instrucionais (o ‘como ensinar’).

Após delimitar o escopo do problema, a revisão se volta para a fronteira da pesquisa que une os dois pilares técnicos desta tese: a Inteligência Artificial Generativa e a Análise de Grafos. O trabalho de (CHEN et al., 2024) apresenta uma exploração abrangente do campo emergente que combina Grandes Modelos de Linguagem (LLMs) com a aprendizagem em grafos (*Learning on Graphs*), mapeando o potencial desta sinergia para resolver problemas complexos. A principal contribuição do estudo é a sua análise e categorização das formas de integração entre as duas tecnologias, mostrando como LLMs podem ser usados para “raciocinar” sobre a estrutura de um grafo e como os grafos podem fornecer um conhecimento estruturado que ancora e aprimora o desempenho dos LLMs.

A metodologia aqui desenvolvida pode ser vista como um estudo de caso prático dentro do campo explorado por (CHEN et al., 2024). Especificamente, o método desta dissertação utiliza LLMs para processar e gerar o conteúdo textual (os nós) e, em seguida, aplica a análise de grafos para entender a estrutura de relações entre eles. Este trabalho, portanto, não apenas utiliza as duas tecnologias de forma isolada, mas as integra em um fluxo coeso, alinhando-se diretamente com a fronteira da pesquisa em IA e demonstrando uma aplicação original dessa sinergia no domínio do design curricular.

Por fim, a integração de tecnologias no design curricular levanta uma questão fundamental sobre o papel do especialista humano no processo. A literatura recente tem convergido para modelos de colaboração, em vez de substituição, como o paradigma mais promissor e responsável. Nesse âmbito, o trabalho de (PADOVANO; CARDAMONE, 2024) apresenta um modelo explícito de colaboração Humano-IA para o desenvolvimento de currículos baseados em competências na área de engenharia industrial. A abordagem utiliza técnicas de PLN para analisar e minerar textos de ementas e tendências educacionais, fornecendo análises estruturados para os planejadores curriculares humanos.

A contribuição central do estudo é a sua análise e formalização do processo de design curricular como uma parceria sinérgica. O trabalho argumenta que a IA se destaca na análise de grandes volumes de dados e na identificação de padrões, enquanto o especialista humano é indispensável para a tomada de decisões estratégicas, a validação pedagógica e a garantia de alinhamento com os valores institucionais. Este artigo oferece a conexão filosófica e metodológica mais forte com a presente tese. O *framework* HITL, que é o

pilar de supervisão e controle de toda a metodologia proposta neste trabalho, é uma implementação prática e rigorosa do modelo de colaboração Humano-IA defendido por (PADOVANO; CARDAMONE, 2024). Enquanto o artigo deles propõe o modelo conceitual, esta dissertação o operacionaliza em um sistema completo e validado, com pontos de intervenção claros e ações de refinamento definidas em cada etapa, conforme detalhado em cada etapa do método. O trabalho deles, portanto, fornece a validação teórica para a escolha do HITL como o paradigma central desta pesquisa.

3.1 Considerações Finais

A análise da literatura recente revela um cenário de intensa exploração e rápido avanço no uso de tecnologias de IA para a otimização de processos educacionais. De um lado, revisões sistemáticas e estudos de caso demonstram um consenso sobre o potencial da Inteligência Artificial Generativa para automatizar e enriquecer a criação de conteúdo, como sílabos e planos de curso. Contudo, esses mesmos estudos alertam para desafios críticos, como o risco de geração de informações factualmente incorretas (alucinações) e a necessidade de uma rigorosa curadoria humana para garantir a qualidade e a profundidade pedagógica do material gerado. Do outro lado, a Teoria dos Grafos se consolida como um paradigma poderoso para modelar e analisar a estrutura de currículos. As abordagens, no entanto, frequentemente se baseiam em dados de pré-requisitos formais, que são rígidos e nem sempre disponíveis, ou em Grafos de Conhecimento, cuja construção pode ser complexa e de alto custo.

A metodologia proposta nesta dissertação se posiciona na interseção dessas frentes, buscando preencher uma lacuna identificada: a ausência de um *framework* integrado que abranja o ciclo completo do design curricular. O diferencial deste trabalho reside na sua abordagem sistêmica para garantir a confiabilidade dos resultados. Em vez de interagir com a IAGen de forma aberta, o método implementa dois mecanismos de controle técnico para direcionar as respostas e mitigar vieses e alucinações. Primeiro, a utilização de uma arquitetura de RAG ancora as decisões da IA em documentos de referência, forçando-a a basear a seleção de disciplinas em fontes de conhecimento confiáveis. Segundo, uma engenharia de *prompt* rigorosa, baseada no *framework* *RTF* (*Role, Task, Format*), é aplicada em todas as interações para restringir o escopo da resposta, aumentando a

precisão e a reprodutibilidade do resultado.

A supervisão de todo este processo técnico é assegurada pelo pilar filosófico do método: um modelo robusto e prático de colaboração Humano-IA. Enquanto a literatura aponta a necessidade de supervisão humana como um desafio, e outros trabalhos defendem conceitualmente a colaboração, esta dissertação operacionaliza essa parceria através de uma proposta de protocolo HITL em múltiplos estágios. O método aqui apresentado contribui, portanto, com um *framework* de ponta a ponta que não apenas gera artefatos curriculares de alta qualidade, mas o faz de maneira controlada, transparente e governada pela expertise pedagógica, oferecendo uma solução concreta e responsável para os desafios do design curricular na era da inteligência artificial.

4 Método e Ferramentas

Neste capítulo, apresentam-se o método desenvolvido e o arcabouço de ferramentas empregado para responder às perguntas de pesquisa e atingir os objetivos propostos. A abordagem metodológica, de natureza quantitativa e qualitativa, foi estruturada em um fluxo de etapas sequenciais que integram técnicas de Inteligência Artificial Generativa (IAGen), Processamento de Linguagem Natural (PLN) e Análise de Grafos, conforme ilustrado na Figura 6.

Neste capítulo, apresentam-se o método desenvolvido e o arcabouço de ferramentas empregado para responder às perguntas de pesquisa e atingir os objetivos propostos. A abordagem metodológica foi estruturada em um fluxo de processo composto por três grandes etapas sequenciais, conforme ilustrado na Figura 6: (1) Preparação e Enriquecimento dos Dados, (2) Geração de Novas Disciplinas, (3) Seleção da Proposta Curricular, e (4) Análise Estrutural e Classificação Pedagógica.

A metodologia desenvolvida propõe um fluxo de trabalho que culmina em um protocolo de aplicação orientado pelo paradigma *Human-in-the-Loop* (HITL), concebido para assegurar a supervisão e validação por um especialista humano (e.g., coordenador de curso) em um cenário de uso real. Conforme detalhado nas seções subsequentes, este protocolo HITL prevê pontos de checagem onde o artefato gerado (e.g., lista de *keywords*, mnovas disciplinas geradas, matriz curricular) seria submetido à avaliação do especialista. Caso o resultado não atenda aos critérios pedagógicos ou de coerência, o protocolo sugere ações de intervenção, como o refinamento de *prompts* ou o ajuste de limiares, em um ciclo iterativo até a aprovação. É importante ressaltar que, durante o desenvolvimento e a validação experimental descritos nesta dissertação, os ciclos de revisão e ajuste foram conduzidos pelo próprio pesquisador, atuando como simulador do especialista de domínio para refinar e verificar a robustez de cada etapa do método e do protocolo HITL proposto.

A aplicação de IAGen é um pilar central desta metodologia, sendo empregada em diferentes fases com técnicas distintas. O método utiliza a IAGen em três momentos cruciais: (i) para o enriquecimento semântico das ementas através da extração de *keywords*, uma tarefa de síntese textual condicionada ao conteúdo de entrada; (ii) para a criação de novas disciplinas por meio da fusão de conteúdos de grupos semanticamente similares; e (iii) para a seleção do conjunto final de disciplinas, a etapa mais complexa, que é realizada

por meio de uma arquitetura de RAG, na qual a IA é contextualizada com informações recuperadas de documentos de referência normativos (e.g., diretrizes da SBC, MEC e CFA). A introdução explícita da arquitetura RAG e do paradigma Human-in-the-Loop (HITL) são os diferenciais que garantem a robustez e a validade pedagógica do processo.

Para a elaboração dos *prompts* destinados aos modelos de IAGen, adotou-se o método de Engenharia de *Prompt* com Papel, Tarefa e Formato de Saída (RTF, do inglês *Role, Task, Format*) (WHITE et al., 2023). Este *framework* foi aplicado de forma específica em cada etapa que demandou o uso da IAGen, conforme detalhado nas seções subsequentes. Todos os *prompts* utilizados estão disponibilizados em repositório público¹. A estrutura do RTF é composta por três elementos fundamentais:

- **Papel (*Role*):** Define a persona ou função que o modelo de IA deve assumir durante a interação (e.g., “especialista em currículo acadêmico”).
- **Tarefa (*Task*):** Especifica a atividade que o modelo deverá executar de forma clara e objetiva.
- **Formato de Saída (*Format*):** Determina a estrutura exata da resposta esperada, como uma lista, um objeto JSON ou uma classificação binária.

As seções a seguir detalham cada uma das etapas do método, bem como o corpus de dados utilizado e o conjunto de ferramentas e tecnologias que viabilizaram a execução dos experimentos.

4.1 Dados Utilizados

A robustez e a capacidade de generalização do método proposto neste trabalho dependem fundamentalmente da qualidade e da diversidade do corpus de dados empregado. Com o intuito de submeter o método a um cenário realista de heterogeneidade terminológica e estrutural, foi constituído um conjunto de dados diversificado, proveniente de matrizes curriculares de diferentes cursos de graduação e instituições de ensino superior. A concepção deste corpus teve como premissa central a inclusão de domínios do conhecimento distintos para validar a eficácia do modelo em identificar similaridades semânticas e promover a integração interdisciplinar. As informações referentes as matrizes curriculares utilizadas podem ser encontradas na Apêndice A.

¹ <https://anonymous.4open.science/r/SBIE_2025-A241/>

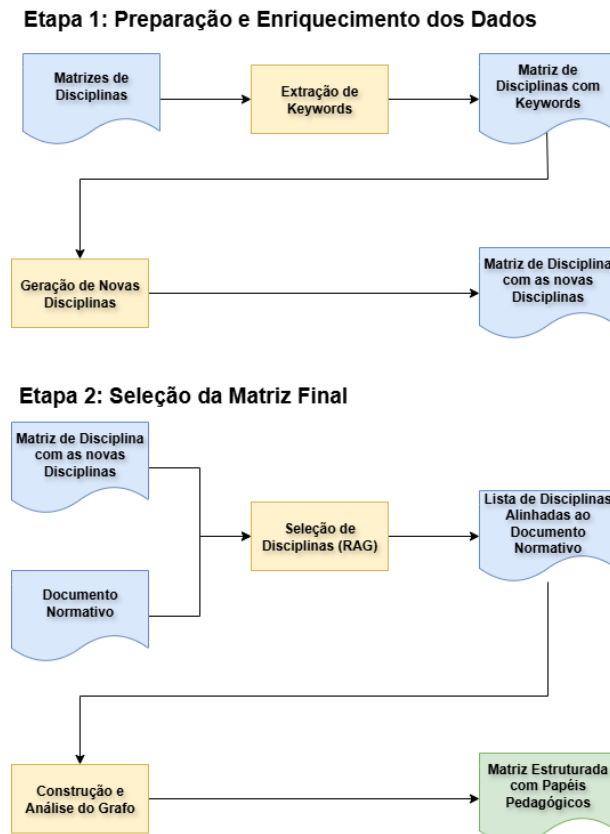


Figura 6 – Fluxograma representando as etapas do método proposto.

Fonte: O autor (2025).

4.1.1 Coleta e Composição do Corpus

O conjunto de dados foi construído a partir da extração manual de informações curriculares disponibilizadas publicamente nos portais oficiais de instituições de ensino superior. Para o estudo de caso central desta pesquisa, a criação de uma matriz para o curso de Ciência de Dados e um curso de Administração, a seleção de cursos buscou um equilíbrio entre especialidade e diversidade. Foram selecionadas as matrizes dos seguintes cursos:

- Da Universidade Federal de Pernambuco (UFPE): Bacharelado em Administração ², Engenharia da Computação ³, Ciências Biológicas ⁴, Sistemas da Informação ⁵ e Ciências da Computação ⁶. A inclusão de cursos de áreas como Administração e Biologia foi estratégica para testar a capacidade do método em lidar com vocabulários

² <<https://www.ufpe.br/administracao-bacharelado-ccsa>>

³ <<https://www.ufpe.br/engenharia-da-computacao-bacharelado-cin>>

⁴ <<https://www.ufpe.br/ciencias-biologicas>>

⁵ <<https://www.ufpe.br/sistemas-de-informacao-bacharelado-cin>>

⁶ <<https://www.ufpe.br/ciencia-da-computacao-bacharelado-cin>>

não diretamente relacionados à computação, no caso da criação do curso de Ciência de Dados.

- Dos cursos de referência em Ciência de Dados: Bacharelado em Ciência de Dados da Universidade Federal da Paraíba (UFPB) ⁷ e da Universidade Federal do Ceará (UFC) ⁸. Estas matrizes servem como uma base de comparação (*baseline*) e validação, por representarem currículos já consolidados na área de interesse.

4.1.2 Pré-processamento e Padronização

As matrizes curriculares coletadas apresentavam uma heterogeneidade estrutural significativa, com variações nos formatos de apresentação e nos campos de informação disponíveis. Diante deste desafio, foi necessário um processo de padronização para criar um *dataset* coeso e funcional para as etapas subsequentes de processamento. Optou-se por extrair e consolidar os três atributos essenciais e comuns à vasta maioria das disciplinas:

1. **Curso de Origem:** Para manter a rastreabilidade da proveniência de cada disciplina.
2. **Nome da Disciplina:** O identificador primário da unidade curricular.
3. **Ementa:** A descrição formal do conteúdo a ser abordado, servindo como a principal fonte de informação semântica.

Após a coleta e padronização inicial, que totalizou 416 disciplinas, foi aplicada uma etapa de filtragem. Foram mantidas no corpus apenas as disciplinas que possuíam ementas oficialmente registradas, removendo disciplinas sem a devida descrição de conteúdo, que poderiam introduzir ruído na análise de similaridade. Ao final deste processo de curadoria, o conjunto de dados final foi consolidado em 406 disciplinas, constituindo a base sobre a qual todas as análises e experimentos descritos neste trabalho foram realizados.

4.2 Extração de Palavras-chave para Enriquecimento Semântico

Uma análise preliminar do corpus revelou que uma parcela significativa das ementas das disciplinas é formulada de maneira sucinta, apresentando descrições que não detalham a totalidade do escopo de conteúdo. Essa característica dos dados de entrada resulta em representações textuais com baixa granularidade semântica, o que representa um risco

⁷ <https://sigaa.ufpb.br/sigaa/public/curso/portal.jsf?id=14289031&lc=pt_BR>

⁸ <https://si3.ufc.br/sigaa/public/curso/curriculo.jsf?lc=pt_BR&id=72460878>

para a acurácia do cálculo de similaridade, uma vez que este depende de representações que capturem com maior fidelidade o escopo da disciplina.

Para mitigar este risco, decorrente de ementas com baixo nível de detalhamento, a primeira etapa do método consiste no enriquecimento e padronização semântica de cada disciplina por meio da extração de um conjunto de palavras-chave (*keywords*). Este processo é conduzido por um modelo de IAGen, instruído a identificar os termos mais relevantes que descrevem o núcleo conceitual da disciplina, considerando tanto seu nome quanto sua ementa. A utilização da IAGen permite inferir e explicitar termos que, embora não literais, são fundamentais ao domínio da disciplina, gerando uma representação mais robusta e menos suscetível a ruídos de termos genéricos.

Foi adotada uma abordagem determinística para a geração, configurando-se o parâmetro de *temperatura* do modelo para 0.0. Essa escolha metodológica é deliberada para garantir a consistência e a reprodutibilidade dos resultados, fazendo com que as respostas sejam mais focadas e alinhadas ao conteúdo de entrada, em detrimento de saídas mais criativas ou aleatórias que poderiam ser obtidas com temperaturas mais altas.

4.2.1 Engenharia de Prompt para Extração

Seguindo o *framework* RTF, foi elaborado um *prompt* específico para esta tarefa. Ao modelo de IAGen foi atribuído o Papel de “especialista em educação”, com a Tarefa de analisar a ementa e extrair as 10 palavras-chave mais representativas, evitando termos genéricos. O Formato de saída foi estritamente definido como uma lista de termos em letras minúsculas, separados por vírgula. O *prompt* exato utilizado foi:

Você é um especialista em educação que ajuda a extrair palavras-chave precisas e relevantes para disciplinas acadêmicas. Analise a ementa da disciplina e extraia as 10 principais palavras-chave que melhor representam o conteúdo da disciplina. As palavras-chave devem ser extraídas diretamente da ementa ou ser termos altamente relevantes para o conteúdo descrito. Evite palavras genéricas ou que não estejam diretamente relacionadas ao texto da ementa. Forneça as palavras-chave em formato de lista, separadas por vírgula, e todas em letras minúsculas.

4.2.2 Protocolo de Validação e Refinamento

A qualidade do módulo de extração de *keywords* foi assegurada por um protocolo que combina um ciclo de refinamento qualitativo, orientado pelo paradigma HITL, com uma validação quantitativa final.

4.2.2.1 Ciclo de Refinamento Qualitativo (HITL).

Nesta etapa, um ciclo iterativo de revisão foi conduzido pelo pesquisador, simulando o papel de especialista de domínio previsto no protocolo HITL proposto, funcionando como uma calibração do método de extração de *keywords*. O pesquisador realizou uma inspeção da coerência pedagógica da lista de *keywords* gerada para cada disciplina. O objetivo desta revisão interna não foi editar a lista diretamente, mas sim avaliar a performance do processo de geração via IAGen. Caso termos inadequados ou a ausência de conceitos essenciais fossem identificados durante esta avaliação, a ação corretiva do pesquisador consistiu em ajustar o *prompt* de extração para guiar o modelo de IAGen a uma melhor performance. Este ciclo de avaliação interna e ajuste foi repetido até que a qualidade da geração fosse considerada satisfatória para todo o corpus, garantindo que o método de extração fosse refinado e validando a viabilidade do passo de revisão por especialista no protocolo HITL.

4.2.2.2 Validação Quantitativa.

Após o ciclo de refinamento, a qualidade do conjunto final de *keywords* é aferida por um indicador quantitativo objetivo. A aderência semântica entre as *keywords* geradas e a ementa original de cada disciplina é medida por meio do cálculo da similaridade de cosseno entre suas respectivas representações vetoriais. O desempenho desta métrica, detalhado no Capítulo 5, serve como a validação formal da eficácia da etapa de extração de *keywords*.

4.3 Construção da Matriz de Similaridade Semântica

Após o enriquecimento semântico, a próxima etapa consiste em quantificar as relações conceituais entre as disciplinas. O objetivo é gerar uma matriz de similaridade

que servirá de base para o agrupamento e a análise de grafos. Para a construção desta matriz, a representação textual de cada disciplina foi consolidada a partir da concatenação de sua Ementa e das *Keywords* previamente extraídas.

A decisão de excluir o atributo “Nome da Disciplina” deste processo foi deliberada. Testes preliminares indicaram que sua inclusão poderia introduzir um viés de similaridade em disciplinas sequenciais (e.g., “Cálculo I” e “Cálculo II”), cujos nomes são textualmente parecidos, mas cujos conteúdos podem apresentar uma distância semântica relevante. A combinação de Ementa e *Keywords* mostrou-se uma representação mais robusta e fiel ao conteúdo pedagógico. A geração desta matriz envolveu três atividades principais: pré-processamento textual, vetorização e o cálculo da similaridade.

Esta etapa do método visa a traduzir o corpus textual das disciplinas em uma estrutura de dados numérica que quantifica as relações de parentesco conceitual entre elas. O resultado é uma matriz de similaridade, que serve como alicerce para as etapas subsequentes de agrupamento e análise de grafos.

O processo inicia-se com a concatenação dos atributos textuais de cada disciplina em um documento único, composto pelo Nome da Disciplina, sua Ementa e as *Keywords* geradas na etapa anterior. Este texto consolidado é então submetido a um *pipeline* de pré-processamento, que inclui a remoção de *stopwords*, normalização de caixa e lematização, a fim de padronizar o conteúdo para a análise.

Em seguida, o texto pré-processado de cada disciplina é transformado em um vetor numérico de alta dimensionalidade (*embedding*) por meio do modelo *nomic-embed-text*. Com as representações vetoriais de todas as disciplinas estabelecidas, a similaridade semântica entre cada par de disciplinas (D_i, D_j) é calculada utilizando a métrica de similaridade de cosseno, formalizada na Equação 4.1.

$$\text{sim}(D_i, D_j) = \frac{\vec{v}_i \cdot \vec{v}_j}{\|\vec{v}_i\| \|\vec{v}_j\|} \quad (4.1)$$

O resultado deste cálculo é um escore no intervalo $[0, 1]$, que é então organizado em uma matriz simétrica $M_{n \times n}$. Nesta matriz, cada elemento M_{ij} armazena o valor de similaridade entre as disciplinas D_i e D_j , fornecendo uma base quantitativa e estruturada para as análises subsequentes. No final é gerada uma matriz semelhante ao exemplo

abaixo:

$$M = \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1n} \\ m_{21} & m_{22} & \cdots & m_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ m_{n1} & m_{n2} & \cdots & m_{nn} \end{bmatrix}$$

4.3.1 Protocolo de Validação e Calibração do Limiar

A construção da matriz de similaridade é um processo computacionalmente determinístico. O paradigma Human-in-the-Loop (HITL), nesta fase, é aplicado de forma estratégica na calibração do limiar de similaridade (T), o hiperparâmetro que governará a formação dos grupos de disciplinas na etapa seguinte.

4.3.2 Ciclo de Refinamento Qualitativo (HITL).

O papel do pesquisador, atuando como avaliador inicial (simulando o especialista de domínio no protocolo HITL proposto), é fundamental na definição do limiar (T) que será usado para considerar duas ou mais disciplinas como semanticamente equivalentes na etapa subsequente (Seção 4.4). Esta é uma decisão crítica, pois impacta diretamente a coerência dos grupos a serem fundidos. Neste estudo, o pesquisador avaliou a qualidade dos agrupamentos resultantes de diferentes limiares (conforme detalhado na análise de sensibilidade da Seção 5.2.0.1), buscando a configuração que apresentasse a maior coerência pedagógica, informada também por métricas quantitativas de estrutura de rede (modularidade).

4.3.3 Suporte à Decisão Quantitativo.

A decisão do especialista não é puramente subjetiva; ela é informada por uma análise quantitativa. Conforme detalhado no Capítulo 5, é realizada uma análise de sensibilidade da modularidade da rede de disciplinas. Esta análise serve como um ferramenta para apoiar a decisão, pois a modularidade quantifica a qualidade da divisão da rede (grafo

das disciplinas) em grupos (subgrafos). O especialista, portanto, utiliza este indicador objetivo para identificar a faixa de limiares que produz a estrutura de comunidades mais coesa e significativa, calibrando sua decisão final com base nesta evidência quantitativa. Este processo assegura que a aplicação da matriz de similaridade esteja alinhada a um critério de coerência pedagógica que é, ao mesmo tempo, validado pela expertise humana e fundamentado em uma análise estrutural da rede.

4.4 Criação de Novas Disciplinas por Fusão Semântica

Com a matriz de similaridade estabelecida, o método avança para uma de suas etapas centrais: a criação de novas disciplinas a partir da fusão de conteúdos semanticamente equivalentes. Este processo visa consolidar conhecimentos dispersos em diferentes matrizes curriculares, gerando componentes curriculares mais coesos. Este subprocesso, dedicado à geração de novas disciplinas, consiste em quatro atividades sequenciais, detalhados a seguir.

4.4.1 Identificação de Pares e Calibração do Limiar de Similaridade

O ponto de partida é a filtragem da matriz de similaridade para identificar pares de disciplinas (D_i, D_j) que exibem uma forte proximidade conceitual. Um limiar (*threshold*) de similaridade, denominado T , é aplicado como critério de corte. Pares cuja similaridade de cosseno iguala ou excede T são considerados candidatos à fusão, conforme a condição apresentada na Equação 4.2.

$$\text{Par Válido}(D_i, D_j) = \begin{cases} 1, & \text{se } \text{sim}(D_i, D_j) \geq T \wedge i \neq j \\ 0, & \text{caso contrário.} \end{cases} \quad (4.2)$$

A escolha de T é uma decisão metodológica crítica que impacta diretamente a granularidade dos agrupamentos. Este limiar não é um valor arbitrário, mas um hiperparâmetro calibrado empiricamente por meio da Análise de Modularidade. O objetivo é identificar o valor de T que otimiza a estrutura comunitária do grafo de disciplinas, maximizando a coesão interna dos grupos (comunidades) e, ao mesmo tempo, evitando a

fragmentação excessiva (limiares elevados) ou a formação de um único componente gigante (limiares baixos). A análise de sensibilidade detalhada e a justificativa para a escolha do limiar T utilizado neste trabalho são apresentadas no Capítulo 5.

4.4.2 Formação de Grupos de Equivalência por Componentes Conectados

Os pares válidos formam as arestas de um grafo não direcionado, no qual as disciplinas são os nós. O próximo passo é particionar esta rede de relações em *clusters*, onde cada *cluster* representa um grupo de disciplinas semanticamente coesas. Para isso, o método identifica os componentes conectados do grafo. Na Teoria dos Grafos, um componente conectado é um subgrafo no qual existe um caminho entre quaisquer dois nós, e que não possui conexão com nenhum nó fora deste subgrafo. Portanto, cada componente conectado encontrado é, por definição, um *cluster* distinto e isolado dos demais.

Para implementar essa extração, foi utilizado o algoritmo de Busca em Largura (*Breadth-First Search* - BFS). O processo, sintetizado no Algoritmo 3, varre sistematicamente o grafo para identificar e extrair cada um desses componentes. Ao final da execução, o resultado não é um único grafo conectado, mas sim um conjunto de múltiplos *clusters* (grupos de equivalência), cada um representando uma área de conhecimento coesa e desconectada das outras.

Algoritmo 3: Formação de Grupos de Equivalência por Componentes Conectados

Input: Matriz de similaridade $M_{n \times n}$, Limiar T

Output: Conjunto de *clusters* $\mathcal{G} = \{G_1, G_2, \dots, G_m\}$

- 1: Construir grafo de adjacência adj a partir de M e T .
 - 2: Inicializar conjunto de *clusters* $\mathcal{G} \leftarrow \emptyset$ e vértices visitados $V \leftarrow \emptyset$.
 - 3: **for all** vértice $i \in \{1, \dots, n\}$ tal que $i \notin V$ **do**
 - 4: $G_k \leftarrow \text{BFS}(adj, i)$ {Executa a busca para encontrar um componente conectado (cluster)}
 - 5: $\mathcal{G} \leftarrow \mathcal{G} \cup \{G_k\}$
 - 6: $V \leftarrow V \cup G_k$ {Marca todos os nós do cluster como visitados}
 - 7: **end for**
 - 8: **return** $\mathcal{G}$
-

4.4.3 Geração de Conteúdo via IAGen

Para cada grupo de equivalência G_k , um modelo de IAGen é instruído a sintetizar uma nova disciplina, atuando como um agregador semântico. A geração é rigorosamente guiada pelo *framework* RTF. Para garantir consistência, o Papel de “especialista em currículo acadêmico” foi fixado para todos os *prompts*. Para cada um dos três componentes da nova disciplina, uma Tarefa e um Formato de saída específicos são definidos para garantir a qualidade e a padronização do resultado:

- **Nome da Disciplina:** O objetivo é criar um título conciso que representasse a fusão conceitual das disciplinas do grupo.

Tarefa: “Analise os nomes das disciplinas a seguir e gere um novo nome de disciplina que represente uma fusão do conteúdo delas.”

Formato: “Apenas retorne o nome da nova disciplina.”

- **Ementa:** O objetivo é consolidar os conteúdos programáticos em um texto coeso e padronizado.

Tarefa: “Analise as ementas das disciplinas a seguir e crie uma nova ementa que englobe os principais tópicos contidos nelas.”

Formato: “Apenas retorne a ementa no formato conciso e padronizado.”

- **Palavras-Chave:** Buscou-se extrair um conjunto de termos unificado e representativo do núcleo de conhecimento do grupo.

Tarefa: “Analise o conjunto de todas as palavras-chave das disciplinas abaixo. A partir deste conjunto, sintetize e retorne uma nova lista com as 10 palavras-chave mais representativas e centrais. A nova lista deve ser concisa, evitar redundâncias desnecessárias e capturar a essência temática do grupo de disciplinas.”

Formato: “Apenas retorne a lista de palavras-chave no mesmo formato.”

Esta abordagem estruturada de engenharia de *prompt* assegura que a IAGen não atue de forma irrestrita, mas sim como uma ferramenta precisa para a síntese de conteúdo curricular, gerando artefatos consistentes e alinhados aos objetivos pedagógicos da fusão.

4.4.4 Protocolo de Validação e Refinamento

A qualidade das disciplinas geradas por fusão semântica é assegurada por um protocolo que combina um ciclo de refinamento qualitativo, orientada pelo paradigma

HITL, com uma validação quantitativa final.

4.4.4.1 Ciclo de Refinamento Qualitativo (HITL).

Para validar a coerência e a pertinência pedagógica das disciplinas sintetizadas, foi realizado um processo iterativo de revisão e refinamento conduzido pelo pesquisador, simulando o ciclo HITL proposto para aplicação real. Neste ciclo de validação interna:

- O pesquisador (atuando como operador do método e avaliador inicial) revisa a disciplina gerada (nome, ementa, *keywords*) sob critérios de qualidade e coerência pedagógica.
- Caso o resultado seja considerado inadequado, ele não intervém tecnicamente, mas fornece um *feedback* qualitativo (e.g., “a ementa omitiu o tópico X”, “o nome não reflete bem o conteúdo Y”).
- De posse desse *feedback*, o Pesquisador realiza a intervenção técnica, que pode ser:
 - Ajustar o limiar de similaridade T , para alterar a composição do grupo de origem.
 - Refinar os *prompts* de geração, para guiar o modelo de IA para gerar um conteúdo mais adequada ao *feedback* recebido.
- O subprocesso de geração é então reexecutado para o grupo em questão, e o novo resultado é submetido a uma nova avaliação pelo pesquisador, repetindo o ciclo até que um resultado satisfatório seja obtido. Este ciclo interno serviu para refinar o método e garantir a qualidade dos artefatos analisados nos resultados, validando a viabilidade do protocolo HITL proposto para uso por especialistas de domínio.

4.4.4.2 Validação Quantitativa.

Após a aprovação no ciclo de refinamento, a qualidade do artefato final é medida de forma objetiva. O objetivo desta validação é medir a fidelidade da disciplina gerada em relação ao seu conteúdo de origem. Para este fim, utiliza-se a métrica de coesão semântica. Essa métrica é operacionalizada como a média da similaridade de cosseno entre a representação vetorial da disciplina recém-criada e a de cada uma das disciplinas originais que compunham seu grupo de equivalência. Um alto valor de coesão, conforme

detalhado no Capítulo 5, indica que a fusão gerada pela IA foi fiel ao escopo semântico do material original.

4.4.5 Integração ao Corpus Expandido

Após a validação, as novas disciplinas são formalmente integradas ao conjunto de dados, que passa a conter tanto as unidades curriculares originais quanto as recém-criadas. Este corpus expandido e enriquecido serve então de base para as etapas subsequentes do método, como a seleção final de disciplinas e a estruturação curricular.

4.5 Seleção de Disciplinas Guiada por Documento de Referência

Após a expansão do corpus com as disciplinas recém-criadas, a etapa seguinte do método consiste em selecionar um subconjunto coeso e alinhado a diretrizes normativas para compor a matriz curricular final. Este processo é realizado por meio de uma abordagem de Geração Aumentada por Recuperação (*Retrieval-Augmented Generation* - RAG), que permite à IAGen tomar decisões informadas com base em um documento de referência, como as diretrizes da Sociedade Brasileira de Computação (SBC) para o curso de Ciência de Dados.

O processo é dividido em duas fases de seleção, precedidas por uma etapa de preparação. A arquitetura deste processo é visualmente detalhada em três diagramas: a Figura 7 ilustra a preparação da base de conhecimento, a Figura 8 detalha a primeira fase de filtragem, e a Figura 9 apresenta a segunda fase de seleção final.

4.5.1 Preparação da Base de Conhecimento

Antes da seleção, é necessário preparar tanto o documento de referência quanto o corpus de disciplinas. O documento normativo é processado, conforme ilustrado na Figura 7, da seguinte forma:

1. **Segmentação Semântica:** O documento é carregado e dividido em fragmentos textuais menores e semanticamente coesos (*chunks*) para permitir uma busca contextual precisa.

2. **Vetorização e Indexação:** Cada *chunk* é convertido em um vetor numérico (*embedding*) por meio de um modelo de linguagem baseado na arquitetura *Transformer*, otimizado para a geração de representações textuais. A escolha de um modelo específico é um parâmetro do método, permitindo a substituição por alternativas mais recentes ou mais adequadas a um domínio particular. O modelo exato utilizado nos experimentos desta tese é detalhado na Seção 4.6.3. Os vetores gerados são então armazenados em um índice vetorial, que permite a realização de buscas semânticas eficientes (um sistema de *retrieval*).

Paralelamente, o corpus de disciplinas (originais e geradas) é preparado de forma análoga, com o texto consolidado de cada disciplina (nome, ementa e *keywords*) sendo convertido em um vetor *embedding*.

Criação da Base de Conhecimento

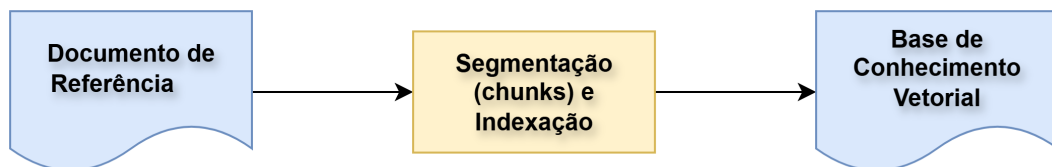


Figura 7 – Diagrama do processo de criação da base de conhecimento a partir do documento de referência.

Fonte: O autor (2025).

4.5.2 Fase 1: Filtragem Individual por Aderência

A primeira fase da seleção, cujo fluxo está detalhado na Figura 8, consiste em uma filtragem individual para avaliar a aderência de cada disciplina ao documento de referência. Para cada disciplina do corpus, o sistema de *retrieval* busca os trechos mais pertinentes na base de conhecimento indexada. Em seguida, a IAGen realiza uma classificação binária, guiada por um *prompt* estruturado no método RTF:

- **Papel:** Classificador binário, focado em uma decisão objetiva.
- **Tarefa:** “Para determinar a aderência seja direto. Se for aderente coloque Sim, porém se a disciplina não for aderente coloque apenas Não.”
- **Formato:** Saída de uma única palavra (“Sim” ou “Não”), sem justificativas, para

processamento automatizado.

Esta etapa preliminar otimiza o processo, descartando disciplinas com baixo alinhamento e reduzindo a complexidade para a fase subsequente.

Fase 1: Filtragem Individual por Aderência

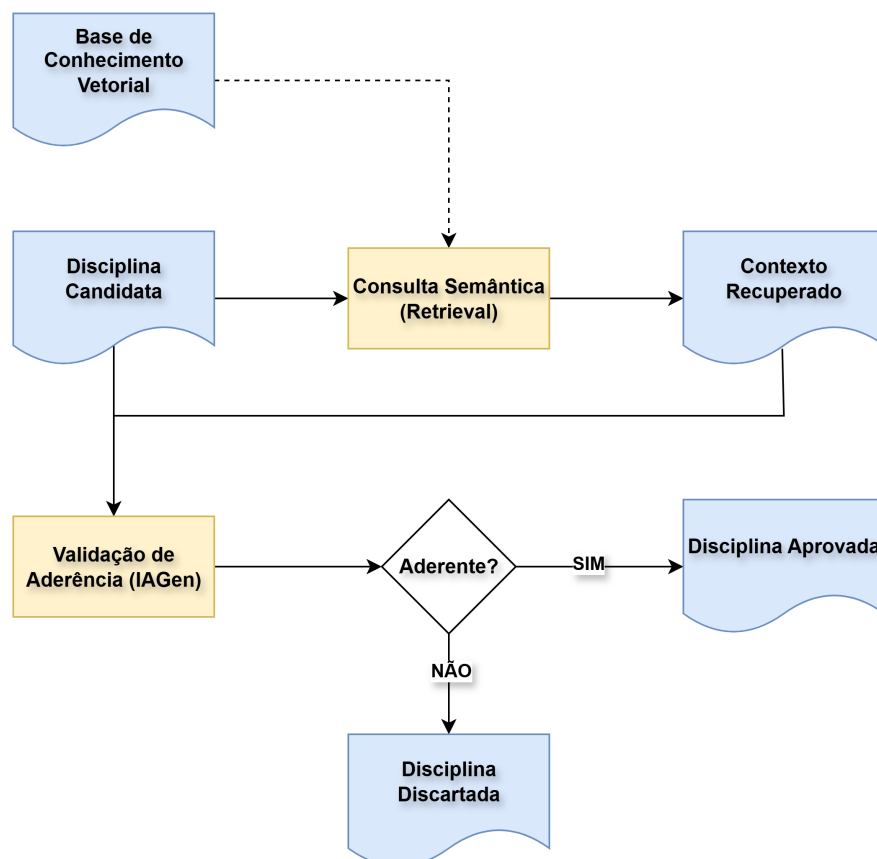


Figura 8 – Fluxo da Fase 1: avaliação individual da aderência de cada disciplina, utilizando a base de conhecimento para contextualizar a decisão da IAGen.

Fonte: O autor (2025).

4.5.3 Fase 2: Seleção Final da Matriz Curricular

As disciplinas aprovadas na primeira fase são submetidas a uma análise consolidada para a seleção final, conforme o processo demonstrado na Figura 9. Nesta etapa, o *retrieval* é direcionado para extrair do documento de referência informações sobre “eixos de formação”, “conteúdos” e “competências” esperadas para o curso. A IAGen, então, assume a persona de um coordenador de curso para realizar a seleção final, seguindo um *prompt* detalhado:

- **Papel:** Coordenador de curso de graduação, com conhecimento em planejamento acadêmico.
- **Tarefa:** “Você [...] precisa analisar uma lista de disciplinas e selecionar as 40 que mais se alinham com o documento de referência. [...] monte uma matriz acadêmica com as disciplinas que mais se alinham com o curso de graduação em Ciência de Dados apresentado no contexto do documento. [...] A matriz curricular deverá ter 40 disciplinas [...] o curso terá 8 semestres e cada semestres terá 5 disciplinas [...] Não selecione nenhuma disciplina de Trabalho de Conclusão de Curso ou Estágio.”
- **Formato:** “Retorne apenas a lista, sem justificativas ou formatações extras, e garanta que cada disciplina esteja em uma linha separada.”

Fase 2: Seleção Final da Matriz Curricular

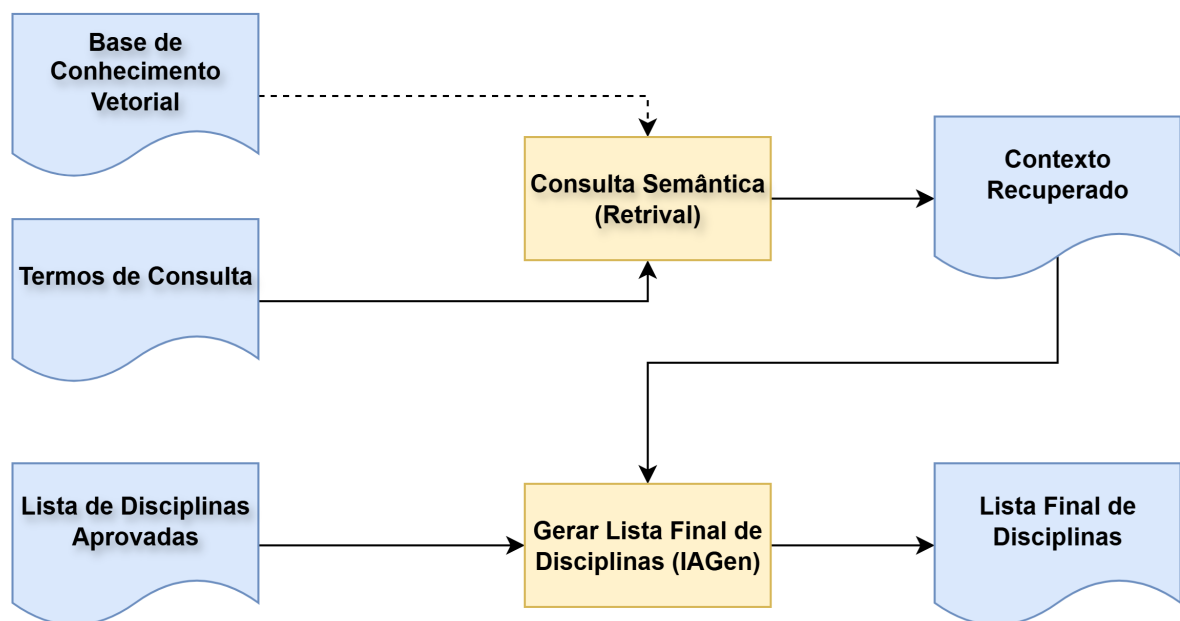


Figura 9 – Fluxo da Fase 2: seleção consolidada da matriz curricular final, utilizando uma consulta específica e a lista de disciplinas pré-aprovadas.

Fonte: O autor (2025).

4.5.4 Protocolo de Validação e Refinamento da Seleção

A seleção do conjunto final de disciplinas é a etapa mais crítica do método. Sua validação nesta pesquisa envolveu duas frentes: um ciclo de refinamento qualitativo interno, realizado pelo pesquisador para simular o ciclo de validação previsto no paradigma HITL,

e uma validação quantitativa final da matriz resultante, que afere a qualidade do artefato gerado.

O ciclo de refinamento interno ocorreu da seguinte forma, simulando os passos propostos para o especialista de domínio:

1. **Revisão da Fase 1:** O pesquisador (simulando o especialista) revisou as classificações binárias de aderência (“Sim” ou “Não”) geradas pela IAGen. Caso uma disciplina fosse classificada de forma percebida como inadequada, o pesquisador reavaliou a necessidade de refinar os parâmetros da consulta de *retrieval* ou o *prompt* de classificação e solicitar uma nova classificação pelo modelo.
2. **Revisão da Fase 2:** A lista final de disciplinas, gerada após a segunda fase de seleção pela IAGen, foi avaliada pelo pesquisador quanto à sua composição e coerência geral (e.g., presença de áreas de conhecimento essenciais, equilíbrio). Se a composição não fosse considerada ideal, o pesquisador ajustou o *prompt* da Fase 2, modificando restrições ou adicionando novas diretrizes, e reexecutou a seleção final.

Este processo iterativo interno visou refinar o método de seleção e assegurar a qualidade pedagógica da matriz curricular analisada nos resultados (Capítulo 5), demonstrando a funcionalidade dos passos de validação propostos no protocolo HITL.

Após a conclusão deste ciclo de refinamento interno, a matriz curricular final foi submetida à validação quantitativa (detalhada no Capítulo 5) para aferir sua coerência e aderência. Este protocolo inclui:

- **Análise de Coerência:** A similaridade de cosseno média entre a matriz gerada e currículos de referência (nacionais e internacionais) é calculada para avaliar a coerência da proposta com práticas consolidadas na área.
- **Análise de Aderência Normativa:** É conduzido um teste de hipótese (e.g., ANOVA ou Kruskal-Wallis, conforme justificado na Seção 4.5.4.1) para verificar se o alinhamento da matriz gerada com documentos de diretrizes oficiais é estatisticamente equivalente ao de currículos de referência, bem como para comparar a coerência com matrizes internacionais.

4.5.4.1 Seleção do Teste de Hipótese Adequado

A seleção do teste de hipótese adequado para cada análise de validação seguiu um protocolo estatístico padrão, baseado primariamente no número de grupos a serem comparados e na verificação dos pressupostos dos dados.

Para a análise de aderência às diretrizes normativas (e.g., SBC, MEC/CFA), que envolve a comparação de três ou mais grupos (Proposta, UFC, UFPB), o método padrão é a Análise de Variância (ANOVA). Contudo, a ANOVA, por ser um teste paramétrico, exige que os dados atendam aos pressupostos de normalidade (avaliado pelo teste de Shapiro-Wilk) e homogeneidade de variâncias (avaliado pelo teste de Levene). Portanto, o protocolo adotado foi: primeiro, verificar os pressupostos; se atendidos, utiliza-se a ANOVA; se violados, emprega-se o teste não paramétrico equivalente, Kruskal-Wallis, que compara as medianas dos grupos sem exigir a normalidade dos dados, garantindo a robustez da análise.

Para a análise de coerência com as matrizes internacionais, que envolve a comparação de apenas dois grupos (“Coerência Externa” vs. “Coerência Interna”), o teste paramétrico padrão seria o Teste t para amostras independentes. No entanto, devido ao tamanho extremamente reduzido das amostras ($N=3$), a verificação dos pressupostos de normalidade é inviável e pouco confiável. Nesses casos, a abordagem estatisticamente mais segura e robusta é a utilização do teste não paramétrico equivalente, o Teste de Mann-Whitney U, que compara as distribuições dos dois grupos sem fazer suposições sobre a normalidade dos dados.

4.6 Estruturação Curricular e Análise de Papéis Pedagógicos via Grafos

Com a lista final de disciplinas definida, a etapa conclusiva do método consiste em organizar esses componentes em uma estrutura curricular coerente. O objetivo é organizar, a partir de uma lista linear, e modelar o currículo como uma rede de conhecimentos interconectados, na qual cada disciplina possui um papel dentro da rede. Para isso, emprega-se a Análise de Grafos, representando as disciplinas como nós e suas relações semânticas como arestas.

4.6.1 Construção do Grafo Curricular e Calibração do Limiar

O primeiro passo é a construção de um grafo não direcionado a partir das disciplinas selecionadas. A matriz de similaridade, previamente calculada, é filtrada para conter apenas as disciplinas da matriz final. A existência de uma aresta entre dois nós (disciplinas) é determinada por um limiar de similaridade.

A construção do grafo a partir da matriz de similaridade requer a definição de um limiar (T), que é um hiperparâmetro crítico que define a topologia da rede. A escolha de um valor para T envolve a seguinte análise: um limiar muito baixo resulta em uma rede excessivamente densa, na qual as distinções estruturais entre as disciplinas são ofuscadas pelo excesso de conexões (“ruído”). Por outro lado, um limiar muito alto leva à fragmentação da rede em múltiplos componentes desconectados, o que impede a análise do currículo como um todo e invalida o cálculo de métricas de centralidade do grafo.

Portanto, o objetivo da definição de T é encontrar uma topologia de grafo que seja “pedagogicamente significativa”, ou seja, que seja ao mesmo tempo conectada, permitindo a análise do currículo como um sistema integrado, e estruturalmente heterogênea, permitindo a diferenciação dos papéis das disciplinas por meio das métricas de centralidade. Para encontrar este ponto de equilíbrio, foi conduzida uma análise de sensibilidade, cujos resultados são detalhados no Capítulo 5, avaliando como a estrutura da rede responde à variação do limiar para a escolha do valor mais adequado.

4.6.1.1 Análise de Centralidade para Definição de Papéis Pedagógicos

Uma vez construído o grafo, o papel estrutural de cada disciplina é quantificado por meio de métricas de centralidade, que medem a sua importância dentro da rede. As métricas utilizadas são:

- **Centralidade de Grau (*Degree Centrality*):** Mede o número de conexões diretas de uma disciplina. Um grau elevado sugere que a disciplina é um *hub* conceitual, conectada a muitas outras áreas do currículo.
- **Centralidade de Intermediação (*Betweenness Centrality*):** Quantifica a frequência com que uma disciplina atua como ponte nos caminhos mais curtos entre outras duas disciplinas. Uma alta intermediação indica que a disciplina é crucial para conectar diferentes eixos de conhecimento.

- **Centralidade de Proximidade (*Closeness Centrality*):** Avalia a “distância” média de uma disciplina a todas as outras na rede. Uma alta proximidade sugere que o conteúdo da disciplina é rapidamente acessível a partir de qualquer outro ponto do currículo.

4.6.1.2 Classificação das Disciplinas em Eixos Formativos

A combinação dos valores de centralidade, calculados para cada disciplina na rede curricular, permite a sua classificação em três papéis pedagógicos distintos. Para garantir uma classificação objetiva e adaptável a diferentes estruturas curriculares, esta tipologia não se baseia em limiares numéricos absolutos, mas sim na posição relativa de cada disciplina nos *rankings* de cada métrica.

Para isso, foi definida uma heurística de classificação que utiliza os quartis da distribuição de cada métrica para definir os limiares de “valor baixo” e “valor alto”:

- **Valor Baixo:** Um valor de centralidade é considerado baixo se for menor ou igual ao primeiro quartil ($Q1$) da sua respectiva distribuição ($Valor \leq Q1$).
- **Valor Alto:** Um valor de centralidade é considerado alto se for maior ou igual ao terceiro quartil ($Q3$) da sua respectiva distribuição ($Valor \geq Q3$).

Seja $P(d)$ a função que atribui um Papel Pedagógico à disciplina d , pertencente ao conjunto de todas as disciplinas D , onde $C_D(d)$, $C_B(d)$ e $C_C(d)$ são seus valores de Grau, Intermediação e Proximidade, e $Q1$ e $Q3$ são os quartis das respectivas distribuições. A classificação segue uma ordem de precedência:

1. Disciplinas de Formação Complementar (Avançadas):

São as primeiras a serem identificadas. Uma disciplina é classificada nesta categoria se apresentar valores baixos nas três métricas de centralidade simultaneamente.

- Grau (C_D): $\leq Q1$
- Intermediação (C_B): $\leq Q1$
- Proximidade (C_C): $\leq Q1$

A baixa conectividade, a ausência de papel de ponte e o difícil acesso conceitual posicionam essas disciplinas na periferia da rede, validando seu papel como tópicos de especialização ou aprofundamento.

2. Disciplinas de Formação Intermediária (Núcleo Conector):

Após a identificação das disciplinas avançadas, as intermediárias são classificadas. Uma disciplina se enquadra nesta categoria se atender a pelo menos uma das seguintes condições:

- **Perfil “Ponte”:** Apresenta um valor alto de Intermediação, independentemente dos outros valores.
 - Intermediação (C_B): $\geq Q3$
- **Perfil “Pilar”:** Apresenta valores altos nas três métricas de centralidade.
 - Grau (C_D): $\geq Q3$
 - Intermediação (C_B): $\geq Q3$
 - Proximidade (C_C): $\geq Q3$

O alto valor de intermediação indica um papel crucial como conector entre diferentes módulos de conhecimento. Disciplinas que são simultaneamente hubs, pontes e de fácil acesso formam o núcleo integrador do curso.

3. Disciplinas de Formação Básica:

Todas as disciplinas que não foram classificadas nas duas categorias anteriores são, por exclusão, definidas como de Formação Básica. Contudo, seu perfil característico, conforme a descrição, é:

- Grau (C_D): $\geq Q3$
- Intermediação (C_B): $< Q3$

A alta conectividade (Grau), combinados com um baixo papel de intermediação, posicionam estas disciplinas como o alicerce fundamental do curso, sendo conceitos centrais que não necessariamente conectam áreas distintas.

Esta heurística é formalmente definida pela seguinte função matemática, onde $P(d) \in \{\text{Complementar, Intermediária, Básica}\}$:

$$P(d) = \begin{cases} \text{Complementar} & \text{se } (C_D(d) \leq Q1(C_D)) \wedge (C_B(d) \leq Q1(C_B)) \wedge (C_C(d) \leq Q1(C_C)) \\ \text{Intermediária Ponte} & \text{se } (C_B(d) \geq Q3(C_B)) \\ \text{Intermediária Pilar} & \text{se } (C_D(d) \geq Q1(C_D)) \wedge (C_B(d) \geq Q1(C_B)) \wedge (C_C(d) \geq Q1(C_C)) \\ \text{Básica} & \text{se } (C_D(d) \geq Q1(C_D)) \wedge (C_B(d) \leq Q1(C_B)) \end{cases}$$

Esta tipologia, que considera a performance relativa dos nós em múltiplas métricas, permite uma análise estrutural do currículo, orientando a sua organização sequencial ao longo dos semestres.

4.6.2 Protocolo de Validação e Refinamento da Estrutura Curricular

A estrutura curricular gerada pela análise de grafos é o artefato final do método e, como tal, é submetida a um protocolo final de validação e refinamento que combina a expertise pedagógica do especialista com a análise objetiva das propriedades da rede.

4.6.2.1 Ciclo de Refinamento Qualitativo (HITL).

Este ciclo interno, simulando o protocolo HITL proposto, visa garantir que a organização final do currículo, baseada na classificação dos papéis pedagógicos, seja pedagogicamente sólida e estruturalmente coerente. O pesquisador (atuando como especialista) realizou uma avaliação em duas frentes sobre a estrutura gerada:

- **Análise da Classificação:** Verificou se o papel pedagógico atribuído a cada disciplina (Básica, Intermediária/Específica, Complementar/Avançada), resultante

da heurística da Seção 4.6.1.2, era consistente com uma interpretação pedagógica plausível da natureza do conteúdo e da sua posição na rede.

- **Análise das Conexões e Estrutura:** Inspeccionou as conexões mais fortes do grafo e a estrutura geral revelada pelas visualizações (e.g., núcleo-periferia) para garantir que representassem relações conceituais válidas e uma organização coerente.

Caso fossem identificadas inconsistências (e.g., uma disciplina considerada fundamental classificada como “Avançada”, conexões semânticas inesperadas muito fortes), o pesquisador avaliou a necessidade de intervenção técnica. A principal intervenção consistiu em ajustar o limiar de similaridade (T) usado para construir o grafo, e então reexecutar todo o processo de análise estrutural (cálculo de métricas e classificação). Este ciclo iterativo de avaliação e ajuste pelo pesquisador foi repetido até que a classificação e a estrutura do grafo fossem consideradas coerentes, validando a abordagem proposta para uso futuro com especialistas de domínio.

4.6.2.2 Validação Quantitativa e Estrutural.

Após a aprovação do especialista no ciclo HITL, a validação final do artefato é realizada por meio da análise quantitativa e visual de suas propriedades estruturais, conforme detalhado no Capítulo 5. Esta validação inclui a análise das métricas de centralidade de cada disciplina, que serve como a evidência numérica para a classificação, e a análise da topologia da rede através de diferentes *layouts* de visualização (e.g., direcionado por força e hierárquico), que permitem confirmar qualitativamente a coerência da estrutura, como a formação de um núcleo-periferia.

4.6.3 Ambiente Experimental e Ferramentas

A implementação do método descrito neste capítulo foi realizada utilizando um conjunto de ferramentas e tecnologias de código aberto, selecionadas para garantir a reprodutibilidade, transparência e eficácia da pesquisa. Esta seção detalha o arcabouço técnico empregado na execução dos experimentos.

A orquestração de todo o processo foi gerenciada pelo *framework* *LangChain*⁹,

⁹ <<https://github.com/langchain-ai/langchain>>

utilizado para construir o fluxo que integra a leitura de documentos de referência, a criação de índices vetoriais e as chamadas aos modelos de linguagem. A linguagem de programação base foi *Python*.

As principais bibliotecas e suas funções específicas foram:

- **Processamento de Linguagem Natural:** Para as tarefas de pré-processamento textual em português, foi utilizada a biblioteca *spaCy* (v3.8.4).
- **Machine Learning:** A biblioteca *scikit-learn* (v1.5.0) foi empregada para os cálculos de similaridade de cosseno.
- **Busca Vetorial:** A criação de índices vetoriais para a recuperação eficiente de informações foi realizada com *FAISS* (*Facebook AI Similarity Search*), integrado ao fluxo pelo *LangChain*.
- **Análise de Redes e Dados:** As bibliotecas *NetworkX*¹⁰, *pandas*¹¹, *numpy*¹² e *scipy*¹³ foram utilizadas para a manipulação de dados e a análise estrutural dos grafos curriculares.

Os modelos de Inteligência Artificial (IA) foram um componente central do método. A seleção dos modelos baseou-se em critérios de alta performance em *benchmarks* e na relação custo-benefício. Os modelos utilizados foram:

- **Modelo de Vetorização (*Embedding*):** A conversão de textos em vetores foi realizada pelo modelo *open-source nomic-embed-text* (NUSSBAUM et al., 2024), escolhido por sua robustez, resultados superiores e contexto de entrada ampliado. Este modelo foi executado localmente através da plataforma *Ollama*¹⁴.
- **Modelos de Geração de Conteúdo:** Para tarefas como extração de *keywords* e criação de novas disciplinas, foi empregado o modelo *deepseek-chat (v3)* (DEEPSEEK-AI, 2024). Para a tarefa mais complexa de seleção final da matriz curricular, foi utilizado o modelo *deepseek-reasoner (v1)* (DEEPSEEK-AI, 2025), que possui maior capacidade de raciocínio.

Todos os artefatos computacionais, incluindo *scripts* e *prompts*, estão disponibilizados em repositório público para garantir a total transparência e a possibilidade de replicação do estudo¹⁵.

¹⁰ <<https://networkx.org/documentation/stable/tutorial.html>>

¹¹ <<https://pandas.pydata.org/docs/>>

¹² <<https://numpy.org/doc/stable/>>

¹³ <<https://docs.scipy.org/doc/scipy/>>

¹⁴ <<https://github.com/ollama/ollama>>

¹⁵ <https://anonymous.4open.science/r/SBIE_2025-A241/>

4.7 Limitações da Metodologia

Apesar do método proposto apresentado e dos resultados que serão apresentados no Capítulo 5, é importante reconhecer as limitações do método.

- **Dependência do Modelo de IA:** Os resultados específicos são contingentes aos modelos de linguagem e de vetorização utilizados nos experimentos (e.g., `deepseek`, `nomic-embed-text`). A rápida evolução da área de IA significa que a mudança na versão de um modelo ou a substituição por um LLM diferente podem resultar em saídas distintas. Embora o método proposto não dependa de um modelo específico, a reprodução exata dos artefatos gerados está atrelada às versões específicas das ferramentas no momento da execução.
- **Qualidade e Acessibilidade dos Dados de Entrada:** A eficácia do método é intrinsecamente dependente da qualidade e da representatividade do corpus de dados. Vieses, imprecisões ou falta de diversidade nas ementas e matrizes curriculares utilizadas como fonte podem ser herdados pelo processo. Adicionalmente, a própria obtenção de dados curriculares abertos e completos é um desafio significativo, pois muitas instituições não disponibilizam suas matrizes e ementas de forma estruturada e de fácil acesso, o que pode limitar a abrangência de futuras aplicações do método.
- **Sensibilidade aos Hiperparâmetros:** O método depende da calibração de hiperparâmetros críticos, como o limiar de similaridade (T). Como demonstrado na análise de sensibilidade, a escolha deste valor impacta a topologia do grafo e a análise estrutural. A definição do limiar ótimo requer uma análise cuidadosa e pode variar dependendo do corpus e do objetivo específico da análise.
- **Ausência de *Ground Truth*:** Uma limitação metodológica deste estudo reside na ausência de uma comparação quantitativa direta da classificação automática de papéis pedagógicos com um *ground truth* estabelecido por especialistas humanos. A validação da correspondência entre o papel estrutural identificado (baseado nas métricas de centralidade) e o papel pedagógico esperado foi realizada pela análise da coerência da classificação com a estrutura topológica da rede. A criação de um *ground truth* manual robusto para comparação direta apresenta desafios significativos, incluindo a subjetividade inerente à classificação pedagógica, a necessidade de consenso entre múltiplos especialistas e a dificuldade em definir critérios padronizados, além de demandar um esforço considerável que contraria o objetivo de escalabilidade da

abordagem computacional. Adicionalmente, o método visa oferecer uma classificação *data-driven* baseada na estrutura de conectividade semântica, que pode revelar papéis estruturais não imediatamente óbvios em uma análise manual tradicional, servindo como uma ferramenta de suporte que complementa, em vez de apenas replicar, a visão do especialista. A realização de um estudo comparativo com um *ground truth* manual validado permanece como uma oportunidade para trabalhos futuros.

4.8 Considerações sobre o Método Proposto

Ao longo deste capítulo, foi detalhado um método multifacetado para dar suporte à elaboração e estruturação de matrizes curriculares. A abordagem se distingue por articular, de forma sinérgica, três pilares tecnológicos — Processamento de Linguagem Natural para análise semântica, IAGen para a síntese de conteúdo e Análise de Grafos para a modelagem estrutural — sob um princípio norteador de validação contínua por um especialista.

O percurso metodológico demonstra como representações textuais não estruturadas, como ementas e nomes de disciplinas, podem ser sistematicamente transformadas em artefatos quantitativos e relacionais. A construção de uma matriz de similaridade semântica não é um fim em si mesma, mas a fundação que possibilita, em primeiro lugar, a identificação de sobreposições e a fusão coerente de disciplinas. Subsequentemente, a mesma base relacional permite a construção de um grafo curricular, traduzindo a proximidade conceitual em topologia de rede e revelando os papéis pedagógicos de cada componente curricular através de métricas de centralidade.

O diferencial mais significativo desta abordagem, contudo, reside na integração sinérgica de múltiplas tecnologias em um framework de ponta a ponta, que abrange desde a análise semântica de fontes heterogêneas até a geração e estruturação de uma matriz curricular validada. A robustez deste framework é garantida por um princípio fundamental que permeia todo o processo: o paradigma *Human-in-the-Loop (HITL)*. O método, portanto, não se propõe a substituir o especialista em educação, mas a aumentar sua capacidade analítica e criativa, com o HITL atuando como a camada de validação que assegura a qualidade e a validade pedagógica de cada etapa. A Tabela 5 detalha como este princípio é operacionalizado, revelando que cada fase do método é ancorada não apenas na avaliação subjetiva do especialista, mas também em métricas quantitativas objetivas,

formando um protocolo de validação duplo.

Tabela 5 – Protocolo de Validação Quantitativa no Fluxo Metodológico.

Etapa do Método	Técnica Utilizada	Métrica de Validação Quantitativa Aplicada
Extração de Keywords	IAGen	Análise de Similaridade de Cosseno (Keywords vs. Ementa).
Criação de Disciplinas	Análise de Grafos e IAGen	Análise de Coesão Semântica (Similaridade da disciplina gerada com seu grupo de origem).
Seleção de Disciplinas	IAGen com RAG	Análise Comparativa com matrizes de referência e Teste de Hipótese (ANOVA/Kruskal-Wallis).
Estruturação Curricular	Análise de Grafos	Análise de Sensibilidade da topologia do grafo (Conectividade e Estrutura).

O *framework* metodológico aqui apresentado oferece um caminho reproduzível e transparente para o design curricular. Ele equipa os planejadores educacionais com uma ferramenta capaz de processar, sintetizar e estruturar vastas quantidades de informação, permitindo que o foco humano seja direcionado para as dimensões mais estratégicas do planejamento pedagógico. A integração dessas técnicas, governada pelo julgamento de especialistas, aponta para um futuro promissor na colaboração entre a IA e a inteligência humana na concepção de percursos formativos mais coesos, atualizados e eficazes.

5 Resultados e Análise Experimental

Este capítulo apresenta os resultados empíricos obtidos a partir da aplicação do método detalhado no capítulo anterior. O objetivo aqui é validar, por meio de métricas objetivas e análises quantitativas, o desempenho de cada etapa do *framework* proposto, demonstrando sua aplicabilidade prática na concepção de matrizes curriculares.

A apresentação está organizada de forma a espelhar o fluxo metodológico. Inicialmente, são detalhados os resultados das etapas fundamentais de processamento — a validação da extração de *keywords* e da criação de novas disciplinas — cujos processos são a base para qualquer aplicação do método. Em seguida, o capítulo se aprofunda na análise de dois estudos de caso distintos, que servem como prova de conceito e teste de generalização da abordagem:

1. **Estudo de Caso I:** A criação de uma matriz curricular para um curso de Bacharelado em Ciência de Dados, um domínio de alta especialização e diretamente alinhado às tecnologias do método.
2. **Estudo de Caso II:** A criação de uma matriz curricular para um curso de Bacharelado em Administração, um domínio de conhecimento distinto, utilizado para avaliar a robustez e a capacidade de adaptação do método.

Para cada estudo de caso, são avaliados os artefatos finais: a lista de disciplinas selecionada em aderência a documentos de referência e a estrutura curricular gerada pela análise de grafos. Os resultados aqui apresentados fornecem as evidências quantitativas que validam o desempenho de cada componente do método, além de evidenciarem sua flexibilidade em diferentes domínios do conhecimento.

5.1 Estudo de Caso I: Validação da Extração de Keywords

O primeiro estudo experimental teve como objetivo validar a eficácia do módulo de enriquecimento semântico, cuja técnica central é a extração de palavras-chave (*keywords*) via IAGen, conforme descrito na Seção 4.2. A análise foi conduzida sobre o corpus completo, composto por 236 disciplinas. A validação quantitativa consistiu em aferir a qualidade e a relevância semântica das *keywords* geradas, verificando se os termos extraídos representavam fielmente o conteúdo das ementas originais. Para isso, foi calculada a

similaridade de cosseno entre o vetor das *keywords* e o vetor da ementa correspondente para cada disciplina.

A análise demonstrou uma forte aderência semântica, com uma similaridade média de 0,8862 e uma baixa variabilidade, indicada por um desvio padrão de apenas 0,0414. Esses valores indicam que, na maioria dos casos, o conjunto de *keywords* gerado manteve um alto grau de alinhamento com o conteúdo original, e o método apresentou um desempenho consistente em todo o conjunto de dados. O histograma apresentado na Figura 10 ilustra visualmente essa consistência.

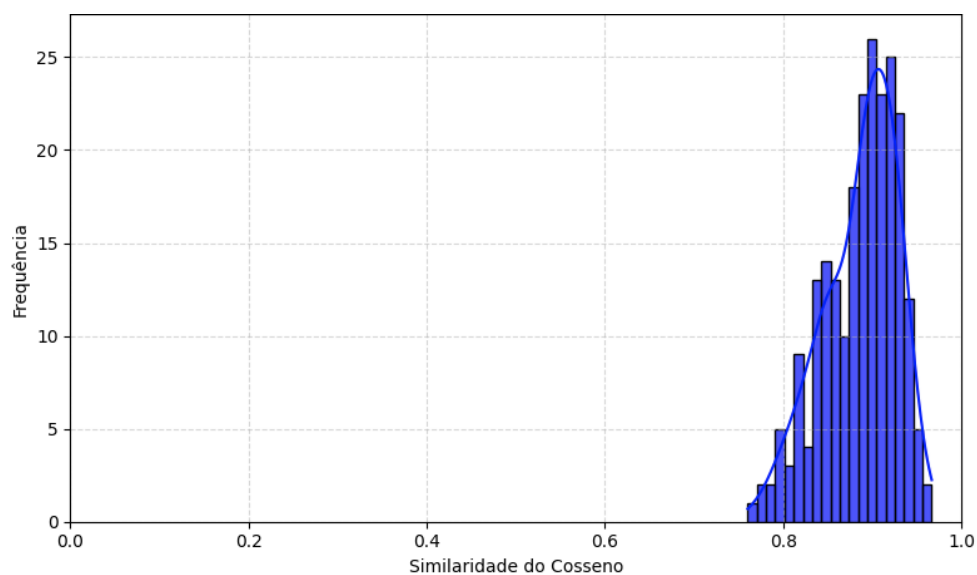


Figura 10 – Distribuição das similaridades de cosseno entre as *keywords* geradas e as ementas originais.

Fonte: O autor (2025).

Observa-se na figura que a grande maioria dos valores de similaridade se concentra na faixa de alta performance, com um pico de frequência próximo a 0,90. A distribuição acentuada à direita, com uma cauda esquerda quase inexistente, reforça o achado de que o método produz resultados consistentemente elevados.

Além da validação numérica, o método demonstrou a capacidade de realizar a expansão conceitual do conteúdo. Por exemplo, para a disciplina de *Gerenciamento de Dados e Informações*, cuja ementa original era excessivamente sucinta (“Algoritmos e estrutura de dados”), a IAGen expandiu o vocabulário para um conjunto de termos mais abrangente e específico: *dados*, *informação*, *gerenciamento*, *algoritmos*, *estruturas*, *banco de dados*, *modelagem*, *análise*, *processamento*, *armazenamento*.

Essa capacidade de inferir e explicitar termos relevantes, mesmo quando não estão

literais na ementa, valida a eficácia desta etapa. Os resultados confirmam que o processo de extração de *keywords* gerou uma base de dados com maior granularidade semântica. Com isso, a representação textual de cada disciplina se torna mais explícita em seu significado, o que é fundamental para o sucesso das etapas posteriores de análise de similaridade e estruturação curricular.

5.2 Estudo de Caso II: Validação da Criação de Novas Disciplinas por Fusão

O segundo estudo experimental teve como objetivo validar o módulo de fusão de conteúdo, processo que agrupa disciplinas semanticamente similares para, a partir desses grupos, gerar novos componentes curriculares consolidados. Esta seção apresenta os resultados quantitativos e qualitativos que validam tanto o processo de formação dos grupos quanto a qualidade do conteúdo gerado por fusão.

5.2.0.1 Calibração do Limiar de Agrupamento via Análise de Sensibilidade

A escolha do limiar de similaridade (T) é uma decisão crítica que define a granularidade dos agrupamentos de disciplinas. Para definir um valor ótimo para este hiperparâmetro, foi realizada uma análise de sensibilidade. Esta análise trata a rede de disciplinas como um grafo e avalia como sua estrutura muda em diferentes limiares. O critério para a avaliação foi a **modularidade**, uma métrica que quantifica a qualidade da divisão de uma rede em *clusters* ou “comunidades”. Em termos práticos, uma modularidade alta indica que o grafo foi bem dividido em grupos de disciplinas que são mais similares entre si do que com o restante da rede, validando a coesão dos *clusters* formados.

Foram testados limiares de 0.50 a 0.95, com incremento de 0.05. Os resultados, apresentados na Tabela 6, mostram que em limiares baixos (≤ 0.70), a rede permanece como um único componente com modularidade nula, indicando que não há uma estrutura de *clusters* discernível. A partir de $T = 0.75$, a rede começa a se fragmentar em comunidades distintas, e a modularidade se torna positiva.

Tabela 6 – Análise de Sensibilidade da Estrutura do Grafo com Diferentes Limiares de Similaridade.

Limiar (T)	Nós no Grafo	Arestas	Comunidades	Modularidade
0.50	405	81.217	1	~ 0.0
0.60	405	53.860	1	~ 0.0
0.70	405	7.636	1	~ 0.0
0.75	394	1.883	8	0.015
0.80	287	440	48	0.836
0.85	145	111	57	0.963
0.90	80	56	34	0.960
0.95	65	43	29	0.954

A análise da Tabela 6 revela um importante *trade-off* entre a qualidade da modularidade e a abrangência da análise. Embora o pico de modularidade ocorra em $T = 0.85$ (0.963), este limiar resulta em uma fragmentação severa da rede, mantendo apenas 145 das 405 disciplinas iniciais na análise. Com base nisso, o limiar de $T = 0.80$ foi escolhido como o ponto de equilíbrio ideal para este estudo. Ele oferece um valor de modularidade elevado (0.836), indicando uma excelente estrutura de comunidades, ao mesmo tempo em que mantém um número substancial de disciplinas (287) na análise, permitindo um processo de fusão mais abrangente. Com este limiar, foram formados 48 grupos de equivalência.

É importante ressaltar que, na metodologia proposta, o limiar T é um parâmetro configurável. Um especialista que replique o método pode testar diferentes valores, como os que geram maior número de comunidades (e.g., $T = 0.85$), para explorar diferentes níveis de granularidade no agrupamento, dependendo dos objetivos específicos do design curricular.

5.2.1 Validação da Coesão Semântica das Disciplinas Geradas

Após a formação dos 48 grupos, a IA Generativa foi utilizada para sintetizar uma nova disciplina para cada grupo. Para validar a qualidade dessa síntese, foi calculada a

coesão semântica: a média da similaridade de cosseno entre a disciplina recém-criada e cada um dos membros originais de seu grupo.

Os resultados demonstram uma alta fidelidade da síntese ao conteúdo original. A distribuição das médias de similaridade para cada grupo gerado, ilustrada na Figura 11, concentra-se massivamente na faixa de alta similaridade, com a maioria dos valores situando-se entre 0.85 e 0.95.

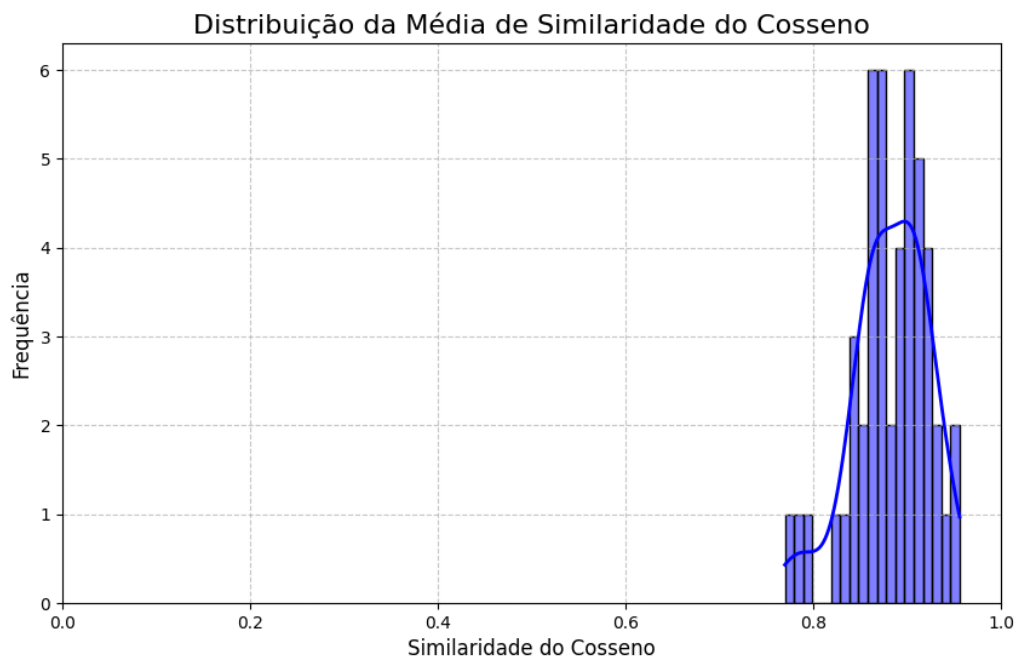


Figura 11 – Distribuição da média de similaridade de cosseno entre as disciplinas geradas e seus respectivos grupos de origem.

Fonte: O autor (2025).

A análise dos dados revela uma média de coesão geral de 0.8845, com um desvio padrão de 0.0412. Este resultado quantitativo valida que o conteúdo gerado pela IA manteve uma forte conexão semântica com os materiais de origem, confirmando a eficácia do método de fusão para criar componentes curriculares coesos e representativos.

5.2.2 Análise Qualitativa: Exemplos de Fusão Semântica na Prática

Esta seção apresenta uma análise qualitativa, através da apresentação de exemplos de disciplinas geradas a partir da fusão das disciplinas dentro dos grupos. O objetivo é inspecionar os artefatos gerados e oferecer uma visão da capacidade do método de criar

disciplinas a partir de outras disciplinas similares entre si. Foram selecionados três grupos de disciplinas, extraídos dos 48 grupos formados, cada um escolhido para ilustrar uma capacidade distinta e estratégica do módulo de fusão de conteúdo. Juntos, estes casos práticos oferecem evidências de como o processo de fusão opera em diferentes cenários e validam a coerência dos componentes curriculares gerados.

5.2.2.1 Exemplo de Integração Vertical: Comunidade G35

O primeiro exemplo ilustra a capacidade do método de realizar a integração vertical, ou seja, de sintetizar uma disciplina de base consolidada a partir de componentes introdutórios de diferentes áreas do conhecimento. A Lista 7 apresenta os dados originais das três disciplinas que formaram a Comunidade G35.

Tabela 7 – Dados Originais das Disciplinas da Comunidade G35.

Disciplina (Curso de Origem)	Ementa Original	Keywords Originais
Tecnologia da Informação (<i>Administração</i>)	Introdução; Arquitetura básica de um computador; Conceitos básicos de computação; Informatização nas empresas; Algoritmos estruturados; Prática de programação.	Arquitetura, Computador, Computação, Informática, Empresas, Algoritmos, Programação, Hardware, Software, Sistemas
Infraestrutura de Hardware (<i>Eng. da Computação</i>)	Introdução à programação; Sistemas digitais.	Hardware, Infraestrutura, Programação, Sistemas Digitais, Arquitetura de Computadores, Processadores, Memória, Dispositivos de Entrada e Saída, Barramento, Clock
Introdução ao Computador (<i>Ciência de Dados - UFPB</i>)	Evolução histórica dos computadores; Arquitetura básica de um computador; Sistemas de numeração e representação de dados; Introdução aos sistemas operacionais; Introdução às redes de computadores e à internet.	Histórica, Computadores, Arquitetura, Sistemas, Numeração, Dados, Operacionais, Redes, Internet, Representação

A partir dos dados da Lista 7, o modelo de IAGen foi instruído a realizar a fusão semântica, gerando o seguinte componente curricular:

Nome Gerado: “Tecnologia da Informação e Infraestrutura de Hardware Computacional”

Ementa Gerada: *“Introdução à Computação: conceitos básicos, evolução histórica e arquitetura dos computadores, sistemas de numeração e representação de dados. Fundamentos de programação: algoritmos estruturados, prática de codificação e introdução à lógica de programação. Visão geral de sistemas digitais, operacionais e redes de computadores, incluindo a internet, aplicados ao contexto de informatização nas empresas.”*

Keywords Geradas: *Arquitetura, Computação, Hardware, Software, Sistemas, Programação, Algoritmos, Processadores, Memória, Redes*

A ementa gerada não é uma simples concatenação, mas uma reestruturação lógica do conteúdo. Ela herda conceitos de cada fonte, como “evolução histórica” e “sistemas de numeração” da disciplina da UFPB, “algoritmos estruturados” da disciplina de Administração, e “sistemas digitais” da disciplina de Engenharia da Computação. Além disso, ela organiza esses tópicos em uma sequência pedagógica clara (“Introdução à Computação”, “Fundamentos de Programação”, “Visão Geral”), resultando em uma ementa completa e bem estruturada que consegue unir os conceitos de cada uma das disciplinas pertencentes ao grupo.

5.2.2.2 Exemplo de Consolidação Horizontal: Comunidade G41

Este segundo exemplo ilustra a capacidade do método de realizar a consolidação horizontal, ou seja, de identificar disciplinas com alta sobreposição de conteúdo em diferentes currículos e fundi-las em um único componente curricular, otimizando a matriz e eliminando redundâncias. A Lista 8 apresenta os dados originais das duas disciplinas que formaram a Comunidade G41.

Tabela 8 – Dados Originais das Disciplinas da Comunidade G41.

Disciplina (Curso de Origem)	Ementa Original	Keywords Originais
Administração Estratégica (<i>Administração</i>)	Administração Estratégica: modelos conceituais; Análise do ambiente externo e interno: recursos, capacidades e competências; Estratégias funcionais, de negócios e corporativas; Vantagem competitiva; Implementação de estratégias; Estratégia e empreendedorismo; Estratégia internacional.	Administração estratégica, Modelos conceituais, Análise ambiental, Recursos, Capacidades, Competências, Estratégias corporativas, Vantagem competitiva, Implementação, Estratégia internacional
Estratégia e Planejamento (<i>Sistemas de Informação</i>)	Modelos conceituais de estratégia empresarial; Análise do ambiente externo; Estrutura industrial; Análise de competidores; Sistema de inteligência sobre concorrentes; Análise de portfólio; Avaliação dos recursos competitivos; Definição de objetivos e metas; Implementação e avaliação.	Estratégia, Planejamento, Análise, Ambiente externo, Estrutura industrial, Competidores, Inteligência, Portfólio, Recursos competitivos, Implementação

A partir destes dados, o modelo de IAGen gerou o seguinte componente curricular consolidado:

Nome Gerado: “Estratégia, Planejamento e Administração Estratégica”

Ementa Gerada: *“A disciplina integra os fundamentos da administração estratégica, abordando modelos conceituais e a análise dos ambientes externo e interno, incluindo estrutura industrial, competidores e recursos competitivos. Explora a formulação de estratégias corporativas, empresariais e funcionais, com foco na vantagem competitiva e na implementação de ações. Aborda métodos de planejamento estratégico, definição de objetivos e metas, além da avaliação de desempenho. Também discute estratégias internacionais e a relação entre estratégia e empreendedorismo, proporcionando uma visão abrangente para a tomada de decisões em contextos dinâmicos.”*

Keywords Geradas: *Estratégia, Planejamento, Análise ambiental, Recursos, Competências, Vantagem competitiva, Implementação, Ambiente externo, Estrutura industrial, Inteligência competitiva*

A análise comparativa evidencia a capacidade do método como uma ferramenta de otimização curricular. A ementa gerada sintetiza com sucesso os tópicos comuns a ambas as disciplinas de origem, como “modelos conceituais” e “análise do ambiente externo”. Ela enriquece o conteúdo ao integrar conceitos específicos de cada uma das fontes ao incorporar a “estrutura industrial” e “competidores” da disciplina de Sistemas de Informação, ao mesmo tempo em que preserva os tópicos de “vantagem competitiva” e “estratégia internacional” da disciplina de Administração. O resultado é um componente curricular único que, em uma reforma curricular real, poderia liberar espaço na grade para outros conteúdos, demonstrando o potencial do método para a consolidação e eliminação de redundâncias.

5.2.2.3 Exemplo de Síntese de Tópicos Avançados: Comunidade G46

Este último exemplo demonstra a capacidade do método de lidar com a complexidade e a sobreposição de disciplinas técnicas avançadas. O processo identificou quatro componentes curriculares de dois cursos de Ciência de Dados, todos dentro do domínio de Inteligência Artificial, e os sintetizou em uma única disciplina avançada. A [Lista 9](#) apresenta os dados originais.

Tabela 9 – Dados Originais das Disciplinas da Comunidade G46.

Disciplina (Curso de Origem)	Ementa Original	Keywords Originais
Redes Neurais Artificiais (<i>Ciência de Dados - UFC</i>)	Introdução às Redes Neurais: Aspectos históricos, Estrutura de um neurônio artificial, Aprendizado supervisionado e não-supervisionado; Algoritmos de aprendizado: perceptron, retropropagação de erros (backpropagation), mapas auto-organizáveis (SOM).	Redes neurais, Neurônio artificial, Aprendizado supervisionado, Não-supervisionado, Perceptron, Retropropagação, Mapas auto-organizáveis, MLP, Redes convolucionais
Aprendizado Profundo (<i>Ciência de Dados - UFC</i>)	Redes Multi-Layer Perceptron; Convolutional Neural Networks; Redes auto-associativas (Auto-encoders); Redes Neurais Recorrentes; Aprendizado por reforço; Implementação em Keras e Tensorflow.	Redes neurais convolucionais, Redes neurais recorrentes, Aprendizado por reforço, MLP, Auto-encoders, Transferência de aprendizado, Keras, Tensorflow
Aprendizado Profundo (<i>Ciência de Dados - UFPB</i>)	Fundamentos matemáticos; Perceptrons e multi-Layer perceptrons; Redes feedforward; Backpropagation; Regularização; Convolutional neural networks (CNNs); Recurrent neural networks (RNNs); LSTMs; Generative adversarial networks (GANs); Transfer learning.	Redes neurais, Perceptrons, Backpropagation, CNNs, RNNs, LSTMs, GANs, Transfer learning, Deep autoencoders
Inteligência Computacional (<i>Ciência de Dados - UFPB</i>)	Redes Neurais; Computação Evolucionária: Algoritmo Genético (AG); Otimização Combinatorial; Mineração de Dados; Lógica Fuzzy; Conjuntos Fuzzy; Sistemas Fuzzy.	Redes neurais, Neurônio artificial, Aprendizado supervisionado, Algoritmos genéticos, Otimização combinatorial, Mineração de dados, Lógica fuzzy, Conjuntos fuzzy

A partir da fusão destes quatro componentes, o modelo de IAGen gerou a seguinte disciplina:

Nome Gerado: “Inteligência Computacional e Aprendizado Profundo com Redes Neurais”

Ementa Gerada: “A disciplina aborda os fundamentos teóricos e práticos das redes neurais, explorando aspectos históricos, estrutura de neurônios artificiais e princípios de aprendizado. São estudados os tipos de aprendizado supervisionado e não-supervisionado, além de algoritmos

como perceptron, mínimos quadrados, retropropagação (backpropagation) e aprendizado competitivo, incluindo mapas auto-organizáveis (SOM). Os modelos de redes neurais abrangem Adaline, MLP, RBF, Hopfield, redes convolucionais (CNNs), redes probabilísticas e autoencoders. Também são abordadas redes recorrentes (RNNs, LSTMs), redes de aprendizado por reforço e modelos generativos (GANs). A disciplina inclui técnicas de regularização, avaliação de desempenho e transferência de aprendizado, com implementação prática em plataformas como TensorFlow e Keras, juntamente com noções de computação evolucionária (algoritmos genéticos) e lógica fuzzy.”

Keywords Geradas: *Redes neurais, Neurônio artificial, Aprendizado supervisionado, Aprendizado não-supervisionado, Perceptron, Retropropagação, Redes convolucionais, Redes recorrentes, Aprendizado por reforço, Transferência de aprendizado*

Este exemplo demonstra a capacidade do método de lidar com a alta complexidade e sobreposição de tópicos técnicos. A ementa gerada não é uma simples concatenação, mas uma síntese hierárquica do conhecimento. Ela começa com os fundamentos de Redes Neurais (presentes em três das quatro fontes), progride para as arquiteturas específicas de Aprendizado Profundo, como CNNs, RNNs, LSTMs e GANs (extraídas das disciplinas de “Aprendizado Profundo”), e finaliza integrando os tópicos de computação evolucionária e lógica fuzzy (provenientes da disciplina de “Inteligência Computacional”).

5.3 Estudo de Caso III: Criação da Matriz Curricular para Ciência de Dados

O presente estudo de caso da pesquisa detalha a geração e a validação de uma matriz curricular completa para um curso de Bacharelado em Ciência de Dados. A aplicação do método resultou em uma lista final de 40 disciplinas, denominada “Proposta”. Uma análise de sua composição revela seu caráter híbrido e interdisciplinar: 12 disciplinas são componentes inéditos criados pelo processo de fusão de disciplinas, 6 foram selecionadas de cursos correlatos de computação da UFPE (Engenharia da Computação, Ciência da Computação e Sistemas de Informação), e 22 foram selecionadas dos currículos de referência em Ciência de Dados da UFC e da UFPB.

A validação desta matriz curricular Proposta foi conduzida por meio de um protocolo em três frentes distintas, visando aferir sua qualidade em diferentes aspectos:

1. **Análise de Coerência Interna:** Uma comparação de similaridade semântica com os currículos da UFC e da UFPB. O objetivo aqui não é um teste contra dados inéditos, mas sim validar a capacidade do método de harmonizar e integrar componentes de múltiplas origens em um todo que permanece fiel aos seus principais referenciais.
2. **Validação Externa com Padrões Globais:** Uma análise de similaridade comparativa com os currículos de três universidades internacionais (*ETH Zurich*, *National University of Singapore* e *University of Toronto*). Este teste avalia o alinhamento da “Proposta” com outros cursos de Ciência de Dados aplicados em outros países, utilizando dados inéditos para o método. Adicionalmente, esta análise é aprofundada por um teste de hipótese não paramétrico (Mann-Whitney U) para verificar se a coerência da “Proposta” com o cenário internacional é estatisticamente equivalente à coerência entre os próprios cursos internacionais.
3. **Análise de Aderência Normativa Nacional:** Um teste de hipótese formal para avaliar se o alinhamento da “Proposta” com as diretrizes da Sociedade Brasileira de Computação (SBC) é estatisticamente equivalente ao dos cursos de referência nacionais.

As seções a seguir detalham os resultados obtidos em cada uma dessas três frentes de validação.

5.3.1 Composição da Matriz Curricular Proposta

A matriz final não se restringiu a extrair disciplinas apenas dos cursos de referência em Ciência de Dados, mas integrou componentes curriculares de áreas correlatas da computação, como Engenharia da Computação, Sistemas de Informação e Ciência da Computação. Além disso, 12 das 40 disciplinas (30% do total) são componentes inéditos, criados pelo processo de fusão de disciplinas descrito anteriormente. A Tabela 10 detalha a origem das disciplinas que compõem a matriz curricular proposta.

A composição detalhada na Tabela 10 é uma evidência da capacidade do método em atuar como um integrador curricular. Ele identifica e seleciona disciplinas relevantes de

Tabela 10 – Distribuição da Origem das Disciplinas na Matriz Curricular Proposta.

Fonte de Origem	Nº de Disciplinas
Ciência de Dados (UFPB)	13
Disciplinas Inéditas (Geradas por IA)	12
Ciência de Dados (UFC)	9
Engenharia da Computação (UFPE)	4
Ciência da Computação (UFPE)	1
Sistemas de Informação (UFPE)	1
Total	40

múltiplos contextos e as harmoniza com novos componentes gerados por fusão, resultando em uma proposta curricular multifacetada. A seguir, esta matriz será submetida às análises de coerência, validação com matrizes internacionais e aderência normativa nacional.

5.3.2 Análise de Coerência com Matrizes de Referência

Para avaliar se a matriz “Proposta” é semanticamente coerente com currículos já consolidados, foi calculada a Similaridade do Cosseno entre ela e matrizes de referência nacionais e internacionais, conforme a metodologia descrita na Seção 4.5.4. A análise completa, que permite múltiplas comparações, está sintetizada na Tabela 11. A interpretação dos resultados foi dividida em três frentes: uma análise de coerência com as fontes nacionais, uma validação externa com padrões globais e uma verificação de especificidade com domínios distintos.

Tabela 11 – Análise Consolidada de Similaridade de Conteúdo com Matrizes Nacionais e Internacionais.

Comparação	Ementa	Keywords	Ementa + Key.	Disciplina + Ementa + Key.
Análise da Proposta x Referências Nacionais (Coerência Interna)				
Proposta x UFPB	0,7236	0,5853	0,4972	0,5436
Proposta x UFC	0,6294	0,7123	0,4333	0,6417
Análise da Proposta x Referências Internacionais (Validação Externa)				
Proposta x ETH Zurich	0,6352	0,7897	0,6785	0,7116
Proposta x Singapura (NUS)	0,6878	0,8181	0,7082	0,6832
Proposta x Toronto	0,7111	0,6980	0,6706	0,6530
Baselines de Referência				
UFPB x UFC	0,7567	0,6959	0,6274	0,7701
ETH Zurich x Singapura	0,6220	0,7897	0,6943	0,6478
ETH Zurich x Toronto	0,6916	0,7597	0,7114	0,6798
Singapura x Toronto	0,7970	0,8134	0,7305	0,7213
UFPB x Singapura	0,6625	0,7306	0,6504	0,6236
UFPB x Toronto	0,7380	0,7335	0,6965	0,6998
UFPB x ETH Zurich	0,6341	0,6953	0,6857	0,6320
UFC x Singapura	0,6892	0,7607	0,6769	0,6226
UFC x Toronto	0,7200	0,6756	0,6823	0,6522
UFC x ETH Zurich	0,6457	0,6934	0,6783	0,6515

Primeiramente, a análise de coerência com as referências nacionais (UFC e UFPB) valida a capacidade do método de integrar componentes de forma harmoniosa. É importante notar que a “Proposta” é uma matriz formada por disciplinas de fontes distintas, contendo 22 disciplinas selecionadas destes dois cursos. Como tal, uma alta similaridade é esperada e confirma que, mesmo após a introdução de 45% de conteúdo inédito ou de outros cursos,

o currículo final manteve-se fiel aos seus principais referenciais.

A validação externa com padrões globais constitui o teste de alinhamento mais rigoroso, pois compara a “Proposta” com currículos internacionais (ETH Zurich, Singapura e Toronto) que não foram utilizados no processo de geração. Os resultados revelam que a matriz “Proposta” não apenas apresenta um grau de similaridade comparável ao dos *baselines* internacionais, mas frequentemente demonstra um alinhamento com o cenário global superior ao dos próprios cursos de referência brasileiros. Por exemplo, considerando a combinação completa de atributos (“Disc.+Ementa+Key.”), a similaridade da “Proposta” com a ETH Zurich é de 0,7116, um valor superior ao da UFC com a ETH Zurich (0,6515) e ao da UFPB com a ETH Zurich (0,6320). Este padrão se repete em outras comparações, como com a Universidade de Singapura (NUS).

Estes dados sugerem que o método, ao se basear em um corpus diversificado e ao sintetizar componentes inéditos, foi capaz de gerar um currículo que não apenas reflete a prática nacional, mas que também se aproxima de forma mais acentuada do consenso e do estado da arte internacional em Ciência de Dados do que os currículos de referência utilizados.

Como forma de controle e para validar a especificidade da métrica, a matriz Proposta também foi comparada com cursos de domínios distintos: Administração (ADM) e Ciências Biológicas (CBIO). Os resultados, com valores de similaridade consistentemente mais baixos (e.g., 0,5266 para ADM e 0,5667 para CBIO na combinação completa), confirmam que a análise está de fato capturando a proximidade conceitual específica do domínio de Ciência de Dados.

Estes dados fornecem evidências de que o currículo gerado pelo método é coerente com a prática nacional, mas também possui um alinhamento semântico com programas semelhantes em instituições de ensino superior internacionais.

5.3.3 Análise de Aderência e Coerência por Testes de Hipótese

Para complementar a análise descritiva de similaridade, foram conduzidos testes de hipótese para validar formalmente o alinhamento da matriz “Proposta”. A validação foi realizada em duas frentes: uma análise de aderência às diretrizes normativas nacionais e uma análise de coerência em relação a um conjunto de currículos internacionais de

referência.

5.3.3.1 Aderência às Diretrizes Nacionais (SBC)

A primeira validação formal verificou se o alinhamento da matriz “Proposta” com as diretrizes da Sociedade Brasileira de Computação (SBC) é estatisticamente equivalente ao dos cursos de referência (UFC e UFPB). O método consiste em calcular a similaridade de cada disciplina das três matrizes com o documento da SBC e, em seguida, comparar as médias das três distribuições de *scores*. A hipótese nula (H_0) para cada teste é que não há diferença estatisticamente significativa entre as médias de alinhamento dos três grupos. Conforme o protocolo metodológico, o teste ANOVA ou seu equivalente não paramétrico, Kruskal-Wallis, foi utilizado dependendo da verificação dos pressupostos dos dados. A Tabela 13 apresenta os resultados.

Tabela 13 – Resultados dos Testes de Hipótese para Alinhamento com as Diretrizes da SBC.

Atributo(s) Analisados	Teste Utilizado	Estatística (F ou H)	p-valor	Decisão Estatística ($\alpha = 0.05$)
Apenas Nome da Disciplina	ANOVA	5,3923	0,0061	Rejeita-se H_0
Apenas Ementa	Kruskal-Wallis	3,5050	0,1733	Não se rejeita H_0
Apenas Keywords	Kruskal-Wallis	3,1925	0,2027	Não se rejeita H_0
Ementa + Keywords	Kruskal-Wallis	3,5016	0,1736	Não se rejeita H_0
Disciplina + Ementa	ANOVA	1,8576	0,1619	Não se rejeita H_0
Disciplina + Keywords	ANOVA	1,3101	0,2748	Não se rejeita H_0
Disciplina + Ementa + Keywords	ANOVA	1,9706	0,1452	Não se rejeita H_0

Com exceção da análise que considera apenas o nome da disciplina, em todas as outras seis combinações, o p-valor foi superior ao nível de significância ($\alpha = 0.05$), levando à não rejeição da hipótese nula. Este resultado valida a capacidade do método de gerar uma matriz curricular com um grau de aderência às diretrizes normativas nacionais estatisticamente indistinguível daquelas criadas por especialistas.

5.3.3.2 Validação da Coerência em Contexto Global

A segunda validação formal buscou responder à questão: *a matriz “Proposta” é tão coerente com o cenário internacional quanto os próprios cursos internacionais são*

coerentes entre si? Para isso, foi realizado um teste de hipótese comparando dois grupos de *scores* de similaridade: o de “Coerência Externa” (Proposta vs. Internacionais) e o de “Coerência Interna” (Internacionais vs. Internacionais). A hipótese nula (H_0) foi a de que não há diferença estatística entre as medianas dos dois grupos. Dado o tamanho reduzido das amostras ($N=3$), utilizou-se o teste não paramétrico de Mann-Whitney U. A Tabela 14 resume os dados e o resultado do teste.

Tabela 14 – Teste de Hipótese (Mann-Whitney U) para Coerência com Matrizes Internacionais.

Componente	Valor
Grupo de Coerência Externa (Proposta vs. Internacionais)	[0,8034; 0,8100; 0,8122] Média: 0,8085
Grupo de Coerência Interna (Internacionais vs. Internacionais)	[0,7838; 0,8002; 0,8205] Média: 0,8015
Resultado do Teste	
Estatística U de Mann-Whitney	6,0000
p-valor	0,7000
Decisão Estatística ($\alpha = 0.05$)	Não se rejeita H_0

Com um p-valor de 0.70, substancialmente maior que o nível de significância de 0.05, a hipótese nula não foi rejeitada. Este resultado fornece a evidência estatística de que o método proposta é capaz de gerar uma lista de disciplinas para a matriz “Proposta” que se integra ao ecossistema de referência internacional de forma estatisticamente semelhante da integração que os próprios membros desse ecossistema têm entre si, validando a relevância da lista de disciplinas selecionadas pelo método.

5.3.4 Análise Estrutural e Validação do Grafo Curricular

A etapa final deste estudo de caso consistiu em analisar a estrutura da matriz curricular de 40 disciplinas por meio de um grafo. O objetivo era validar a capacidade do método em organizar as disciplinas e atribuir-lhes papéis pedagógicos (Formação Básica,

Específica e Complementar) com base em sua conectividade semântica.

5.3.4.1 Construção e Classificação do Grafo.

A construção do grafo, que utiliza as disciplinas como nós e a similaridade semântica como arestas, requer a definição de um limiar de conectividade (T). Conforme discutido na análise metodológica, para que as métricas de centralidade globais (Proximidade e Intermediação) sejam válidas, é fundamental que o grafo seja totalmente conectado. Portanto, foi adotado o limiar de $T = 0,6957$ (a média de similaridade da matriz), que se mostrou o valor ideal para garantir a conectividade de todos os 40 nós, resultando em um grafo com 394 arestas.

Com o grafo construído, foram calculadas as métricas de centralidade para cada disciplina, permitindo sua classificação nos eixos formativos. As métricas de centralidade para cada disciplina da lista pode ser visualizadas no Apêndice B.1. A Tabela 15 apresenta o resultado final desta classificação, organizando a matriz curricular proposta.

Tabela 15 – Classificação Final das Disciplinas da Matriz Proposta por Papel Pedagógico.

Papel Pedagógico	Disciplinas
Formação Básica	• Informática Teórica • Lógica para Computação • Paradigmas de Linguagens Computacionais • Introdução à Ciência de Dados • Probabilidade e Estatística • Programação Orientada a Objetos • Banco de Dados I • Análise de Regressão I • Cálculo Diferencial e Integral I • Cálculo Diferencial e Integral II • Programação Estruturada • Pesquisa Operacional • Aprendizagem de Máquina • Aprendizado Profundo • Análise Multivariada e Aprendizagem Não Supervisionado • Álgebra Linear Computacional e Geometria Analítica Aplicada • Estatística Computacional Aplicada: Probabilidade e Inferência • Fundamentos de Programação e Engenharia de Software • Matemática Discreta e Lógica Computacional • Métodos Numéricos e Equações Diferenciais Computacionais • Fundamentos e Laboratório de Ciência de Dados com Big Data e Visualização • Inteligência Computacional • Aprendizagem de Máquina e Reconhecimento de Padrões

Tabela 15 – (Continuação) Classificação Final das Disciplinas da Matriz Proposta por Papel Pedagógico.

Papel Pedagógico	Disciplinas
Formação Específica (Pontes)	• Estatística e Probabilidade para Computação • Matemática Discreta • Fundamentos de Programação • Otimização Contínua • Laboratório de Ciência de Dados • Estrutura de Dados • Análise e Projeto de Algoritmos • Otimização Não-Linear • Visualização de Dados • Big Data • Algoritmos, Estruturas de Dados e Teoria da Computação com Otimização • Fundamentos e Aplicações da Inteligência Artificial
Formação Complementar (Avançadas)	• Cálculo Diferencial e Integral III • Gerenciamento de Dados e Informação • Mineração Estatística de Dados • Gerenciamento e Administração de Bancos de Dados e Informação • Sistemas Inteligentes Avançados

5.3.4.2 Análise Visual da Estrutura Curricular.

Para validar e interpretar a classificação apresentada na Tabela 15, o grafo foi visualizado utilizando dois *layouts* distintos. O primeiro, um *layout* hierárquico (Figura 12), organiza os nós de acordo com o fluxo pedagógico proposto. O segundo, um *layout* direcionado por força (Figura 13), revela a estrutura de conexões intrínseca da rede.

A visualização hierárquica demonstra um fluxo claro e lógico em três estágios (Básica $\rightarrow$ Específica $\rightarrow$ Complementar). Observa-se uma alta densidade de arestas conectando a camada Básica à Específica, validando o papel da primeira como alicerce

do curso. A camada Específica, por sua vez, funciona como um “*hub* de distribuição”, recebendo conhecimento da base e direcionando-o para as especializações, o que reforça seu papel de pilar e ponte.

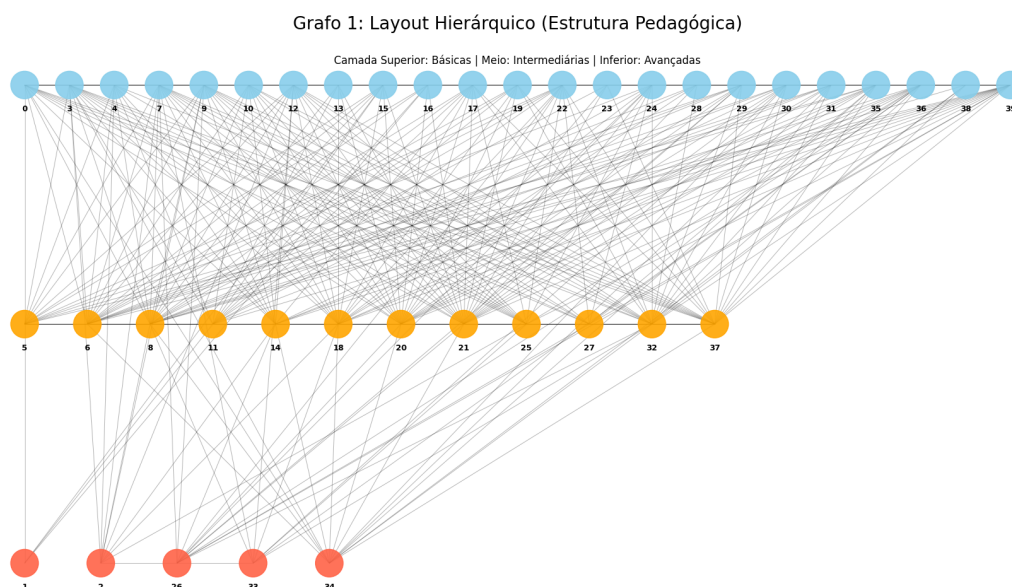


Figura 12 – Grafo do currículo de Ciência de Dados em *layout* hierárquico, ilustrando o fluxo pedagógico proposto.

Fonte: O autor (2025).

A análise do *layout* direcionado por força (Figura 13) revela a verdade estrutural da rede: uma clara organização de núcleo-periferia. O núcleo, densamente conectado, é formado por uma mistura de disciplinas de Formação Básica (azuis) e Específica (laranjas). Este é uma análise crucial, mostrando que as disciplinas de base permanecem estruturalmente centrais ao longo do curso. Por outro lado, as disciplinas de Formação Complementar (vermelhas) situam-se, quase sem exceção, nas extremidades do grafo. Sua posição na periferia valida de forma poderosa a sua classificação como disciplinas “terminais” ou de especialização, que funcionam como pontos finais do fluxo de conhecimento.

A análise combinada das duas visualizações, portanto, valida o método de classificação de forma robusta. O Grafo Direcionado por Força (Figura 13) confirma que a classificação, baseada puramente em métricas, corresponde a papéis estruturais reais na rede, com uma segregação nítida entre o núcleo (Básicas e Específicas) e a periferia (Complementares). Por sua vez, o Grafo Hierárquico (Figura 12) demonstra que esta mesma classificação, estruturalmente válida, pode ser organizada em um fluxo pedagógico coerente. A forte correspondência entre a análise estrutural e o modelo pedagógico confere grande

solidez à metodologia e aos resultados apresentados na Tabela 15. Os valores numéricos detalhados de cada métrica para todas as 40 disciplinas encontram-se no Apêndice B.1.

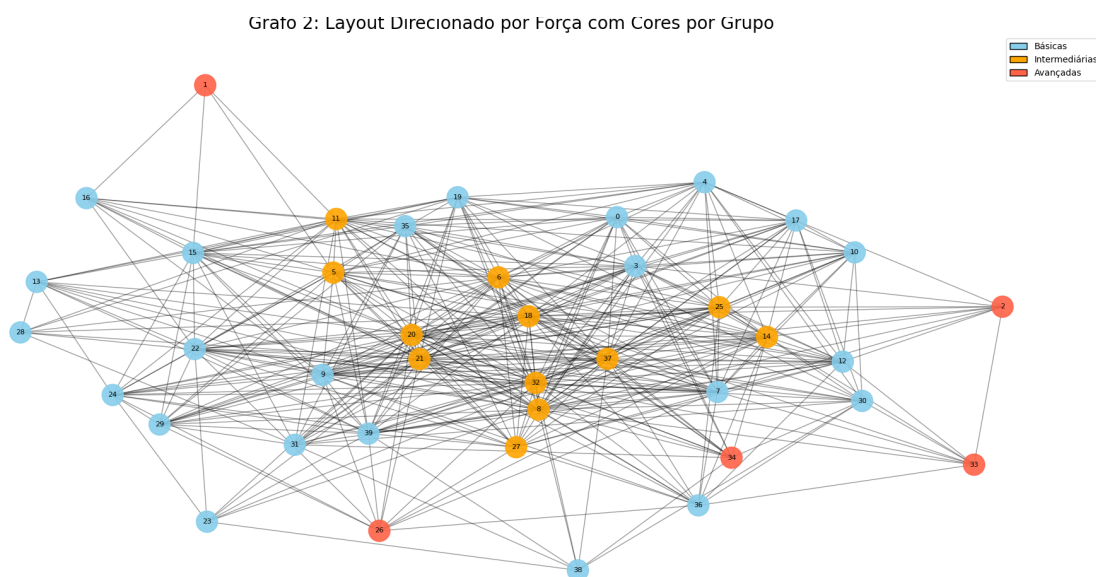


Figura 13 – Grafo do currículo de Ciência de Dados em *layout* direcionado por força, revelando a estrutura de núcleo-periferia.

Fonte: O autor (2025).

5.4 Estudo de Caso IV: Criação da Matriz Curricular para Administração

Para testar a generalização e a robustez do método em um domínio não tecnológico, foi conduzido um segundo estudo de caso focado na criação de uma matriz curricular para um curso de Bacharelado em Administração. Para esta tarefa, o processo de Geração Aumentada por Recuperação (RAG) foi alimentado com dois documentos de referência: as Diretrizes Curriculares Nacionais (DCN) do MEC para cursos de Administração e o guia de diretrizes do Conselho Federal de Administração (CFA).

A validação da matriz de 40 disciplinas gerada (“Proposta”) seguiu o mesmo protocolo do estudo de caso anterior, com uma análise de coerência contra matrizes de referência e uma análise de aderência às diretrizes normativas.

5.4.1 Análise de Coerência com Matrizes de Referência

A coerência da matriz “Proposta” foi avaliada por meio de sua similaridade semântica com os currículos de Administração de duas universidades federais: a UFPE e a UFPB¹. A comparação entre os dois currículos de referência (UFPE x UFPB) serviu como *baseline*. Os resultados estão sintetizados na Tabela 17.

Tabela 16 – Análise Comparativa de Similaridade para o Estudo de Caso de Administração.

Atributo(s)	Proposta x	Proposta x	UFPE x UFPB
	UFPB	UFPE	(Baseline)
Ementa	0,5239	0,6087	0,6045
<i>Keywords</i>	0,7930	0,7038	0,7471
Ementa + <i>Keywords</i>	0,5808	0,6669	0,6054
Disciplina + Ementa + <i>Keywords</i>	0,6368	0,5997	0,6775

Os dados demonstram que a “Proposta” gerada pelo método alcançou um nível de coerência em relação aos currículos existentes. Destaca-se que a similaridade baseada em *Keywords* entre a “Proposta” e a UFPB (0,7930) foi consideravelmente superior ao *baseline* (0,7471), indicando que o método foi capaz de capturar e sintetizar os conceitos-chave da área de forma eficaz. Nas outras combinações, os valores se mantêm na mesma ordem de grandeza do *baseline*, validando a relevância da matriz gerada.

5.4.2 Análise de Aderência às Diretrizes Normativas (MEC/CFA)

A validação formal foi realizada comparando o alinhamento das três matrizes (Proposta, UFPE e UFPB) com os documentos de referência do MEC e do CFA. Primeiramente, foi realizada uma análise descritiva das médias de similaridade de cada matriz com os documentos.

A análise descritiva da Tabela 17 já aponta para uma forte equivalência: as médias de alinhamento da “Proposta” são virtualmente idênticas às das matrizes de referência em

¹ <<https://sigaa.ufpb.br/sigaa/link/public/curso/curriculo/2558867>>

Tabela 17 – Análise Comparativa de Similaridade para o Estudo de Caso de Administração.

Atributo(s) Analisados	Proposta x UFPB	Proposta x UFPE	UFPE x UFPB (Baseline)
Ementa	0,5239	0,6087	0,6045
Keywords	0,7930	0,7038	0,7471
Ementa + Keywords	0,5808	0,6669	0,6054
Disciplina + Ementa + Keywords	0,6368	0,5997	0,6775

todas as combinações de atributos. Para confirmar essa observação, foi conduzido o teste de hipótese (ANOVA ou Kruskal-Wallis), cujos resultados estão na Tabela 18.

Tabela 18 – Resultados dos Testes de Hipótese para Alinhamento com as Diretrizes de Administração.

Atributo(s) Analisados	Teste Utilizado	Estatística (F ou H)	p-valor	Decisão Estatística ($\alpha = 0.05$)
Apenas Ementa	Kruskal-Wallis	6,0850	0,0477	Rejeita-se H_0
Apenas Keywords	ANOVA	0,7211	0,4886	Não se rejeita H_0
Apenas Nome da Disciplina	ANOVA	0,4515	0,6379	Não se rejeita H_0
Ementa + Keywords	ANOVA	0,3040	0,7385	Não se rejeita H_0
Disciplina + Ementa	ANOVA	0,7801	0,4609	Não se rejeita H_0
Disciplina + Keywords	ANOVA	0,5807	0,5613	Não se rejeita H_0
Disciplina + Ementa + Keywords	Kruskal-Wallis	0,0031	0,9985	Não se rejeita H_0

Com exceção da análise baseada apenas na “Ementa”, que apontou uma diferença estatística sutil, todos os outros seis testes não mostraram diferença significativa entre o alinhamento das três matrizes ($p > 0.05$). Este resultado comprova que a matriz gerada pelo método possui um grau de aderência às diretrizes normativas da área de Administração

que é estatisticamente indistinguível dos currículos de universidades de referência.

5.4.3 Análise Estrutural e Validação do Currículo de Administração

Para finalizar o segundo estudo de caso, a matriz de 40 disciplinas de Administração foi submetida à análise estrutural via grafos. O objetivo foi validar se o método, mesmo em um domínio não tecnológico, é capaz de gerar um currículo com uma estrutura interna coesa e com papéis pedagógicos discerníveis.

5.4.3.1 Construção e Classificação do Grafo.

Seguindo o mesmo critério do estudo de caso anterior, a construção do grafo priorizou a conectividade da rede para garantir a validade da análise de centralidade. Foi utilizado o limiar de $T = 0,6798$ (a média de similaridade da matriz), que resultou em um grafo totalmente conectado com 40 nós e 388 arestas. As métricas de centralidade para cada disciplina da lista pode ser visualizadas no Apêndice [B.1](#). Com base neste grafo, as disciplinas foram classificadas em seus respectivos papéis pedagógicos, conforme detalhado na Tabela [20](#).

Tabela 20 – Classificação Final das Disciplinas da Matriz de Administração por Papel Pedagógico.

Papel Pedagógico	Disciplinas
Formação Básica	• Introdução à Administração • Ética e Filosofia para o Administrador • Comunicação Empresarial • Teoria Geral da Administração • Administração Pública • Contabilidade Geral A • Administração no Contexto Brasileiro • Gestão e Inovação Tecnológica • Finanças II • Gestão de Pessoas I • Metodologia da Pesquisa em Administração • Administração de Marketing II • Governança e RSE • Sistemas de Informações • Administração Estratégica • Logística Empresarial • Administração de Cargos e Salários • Análise e Estratégia de Investimento • Avaliação de Desempenho • Direito Administrativo • Novos Negócios
Formação Específica (Pontes)	• Obrigações Empresariais • Administração Tributária • Economia das Empresas • Contabilidade Gerencial • Comportamento Organizacional • Processo Decisório • Administração da Qualidade • Gestão de Pessoas II • Administração de Vendas • Orçamento Empresarial

Tabela 20 – (Continuação) Classificação Final das Disciplinas da Matriz de Administração.

Papel Pedagógico	Disciplinas
Formação Complementar (Avançadas)	• Fundamentos de Sociologia • Psicologia I • Estatística Aplicada à Administração • Administração de Políticas Públicas • Introdução à Finanças • Finanças I • Administração de Marketing I • Administração de Operações • Metodologia Científica e Filosofia da Ciência

5.4.3.2 Análise Visual da Estrutura Curricular.

A validação da classificação apresentada na tabela foi realizada por meio da inspeção de duas visualizações do grafo. A Figura 14 apresenta o *layout* direcionado por força, que revela a estrutura de conexões intrínseca da rede, enquanto a Figura 15 organiza os nós em um *layout* hierárquico para ilustrar o fluxo pedagógico.

O *layout* de força (Figura 14) fornece a evidência estrutural mais forte. É imediatamente visível uma nítida segregação das cores, formando uma estrutura de *layout* núcleo-periferia *layout*. O centro da rede é composto por uma mistura densa de disciplinas Básicas (azuis) e Intermediárias (laranjas), confirmando que estes grupos formam o coração do curso. As disciplinas Avançadas (vermelhas), por outro lado, estão inequivocamente posicionadas nas bordas externas, o que valida visualmente seu papel periférico ou terminal, conforme previsto pelas baixas métricas de centralidade.

O *layout* hierárquico (Figura 15), por sua vez, “desembaraça” a complexidade do núcleo para contar a história do fluxo pedagógico. Ele comunica instantaneamente a progressão em três estágios: da base de conhecimento ampla (Formação Básica), passando por um núcleo conector (Formação Específica) e culminando nas especializações (Formação Complementar). A “cascata” de arestas mostra que a camada Intermediária é o

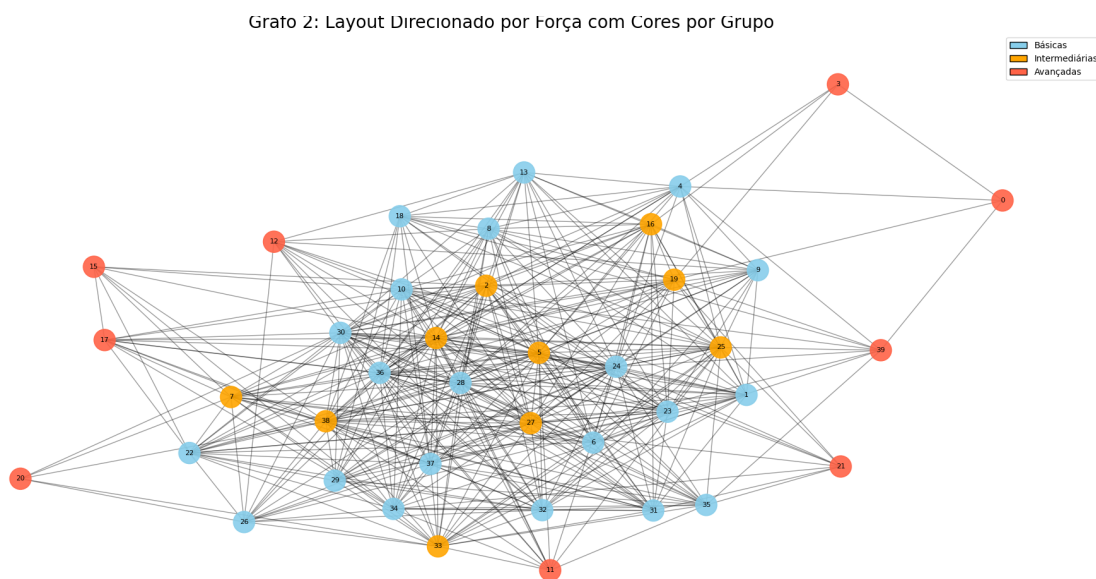


Figura 14 – Grafo do currículo de Administração em *layout* direcionado por força, revelando a estrutura de núcleo-periferia e a segregação dos papéis pedagógicos.

Fonte: O autor (2025).

principal “alvo” das disciplinas de base e a principal “fonte” para as avançadas, confirmando visualmente seu papel de ponte.

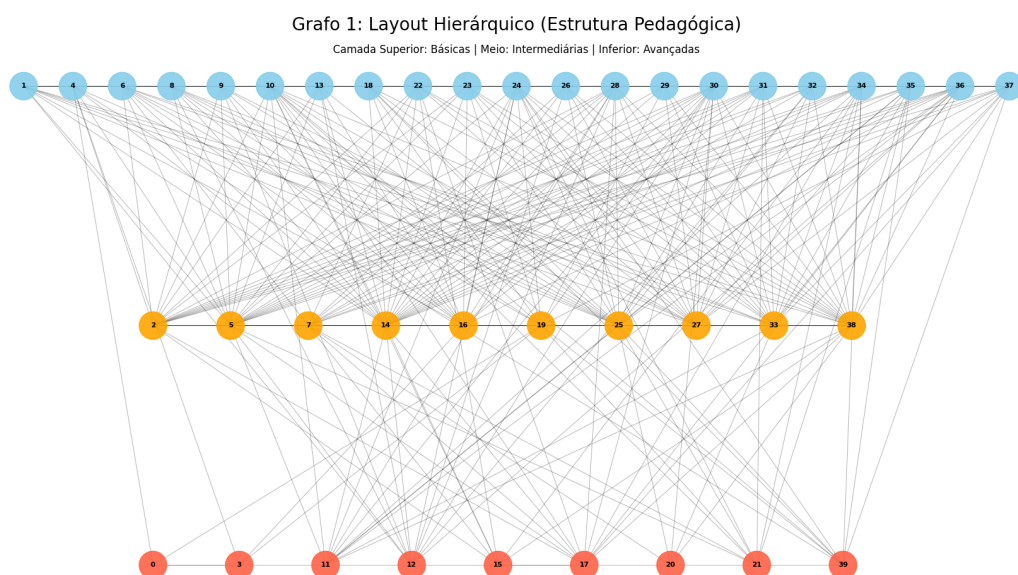


Figura 15 – Grafo do currículo de Administração em *layout* hierárquico, ilustrando o fluxo pedagógico proposto.

Fonte: O autor (2025).

A aplicação bem-sucedida do método de classificação e a coerência das visualizações resultantes em um domínio distinto como a Administração validam a robustez e a

capacidade de generalização da abordagem proposta.

5.5 Discussão Consolidada dos Resultados

Após a apresentação detalhada das validações em cada etapa do método e nos estudos de caso, esta seção final tem como objetivo consolidar os principais achados e conectá-los diretamente às perguntas de pesquisa que guiaram esta dissertação. A Tabela 21 apresenta um resumo sinóptico dos principais resultados quantitativos obtidos, servindo como base para a discussão a seguir.

Tabela 21 – Visão Geral dos Principais Resultados de Validação do Método.

Etapa de Validação	Métrica Principal	Resultado Chave	Conclusão da Validação
Extração de Keywords	Similaridade de Cosseno	Média = 0,8862 (DP = 0,0414)	Geração consistente e semanticamente alinhada.
Criação de Disciplinas	Coesão Semântica	Média = 0,8845 (DP = 0,0412)	Fusão de conteúdo gera disciplinas coesas e fiéis à origem.
Seleção (Ciên. de Dados)	Teste de Hipótese (ANOVA/Kruskal)	$p > 0.05$ (em 6 de 7 testes)	Aderência normativa estatisticamente equivalente a currículos de referência.
Estrutura (Ciên. de Dados)	Análise de Centralidade	Estrutura de núcleo-periferia	Método classifica papéis pedagógicos de forma coerente.
Seleção (Administração)	Teste de Hipótese (ANOVA/Kruskal)	$p > 0.05$ (em 6 de 7 testes)	Validação em domínio distinto, comprovando a generalização do método.
Estrutura (Administração)	Análise de Centralidade	Estrutura de núcleo-periferia	Classificação coerente, validando a robustez do método.

A análise aprofundada dos resultados permite responder de forma robusta às questões de pesquisa.

5.5.1 Resposta à Pergunta de Pesquisa 1: Geração e Validação de Componentes Curriculares Coesos

A primeira pergunta de pesquisa questionava: *“De que forma a combinação de IAGen e similaridade semântica permite gerar e validar novos componentes curriculares coesos a partir da fusão de conteúdos de fontes heterogêneas?”*. Os estudos 1 e 2 (Seções 5.1 e 5.2) forneceram as evidências para responder a esta questão. O método demonstrou ser capaz de gerar novos componentes ao primeiro enriquecer semanticamente o corpus com *keywords* (com uma alta similaridade média de 0,8862 com as ementas originais), e em seguida, utilizar a análise de modularidade para identificar *clusters* de disciplinas com alta afinidade conceitual (T=0.80, modularidade=0.836). A IAGen, então, atua sobre

estes *clusters* para sintetizar novas disciplinas.

A validação destes componentes foi assegurada por uma métrica de coesão semântica, que revelou uma alta fidelidade média de 0,8845 entre as disciplinas geradas e suas fontes. A análise qualitativa dos exemplos de fusão (Seção 5.2.2) tornou este resultado tangível, mostrando que o método é capaz de realizar tanto a integração vertical (criando disciplinas de base a partir de componentes introdutórios) quanto a consolidação horizontal (unificando disciplinas com alta sobreposição). Portanto, a combinação de IAGen e similaridade semântica permite a geração e validação de componentes coesos através de um processo estruturado de enriquecimento, agrupamento por afinidade e síntese validada quantitativamente.

5.5.2 Resposta à Pergunta de Pesquisa 2: Desempenho na Elaboração de Matrizes Curriculares Alinhadas e Interdisciplinares

A segunda pergunta de pesquisa avaliava: *“Qual o desempenho da utilização de IAGen e Análise de Grafos para elaborar matrizes curriculares interdisciplinares alinhadas a referenciais formativos?”*. O desempenho do método foi validado em múltiplas frentes nos dois estudos de caso. A capacidade de gerar matrizes **interdisciplinares** foi evidenciada na composição da matriz de Ciência de Dados, que integrou componentes de quatro cursos distintos, além de 12 disciplinas inéditas (30% do currículo).

O desempenho em gerar matrizes alinhadas a referenciais formativos foi o ponto central da validação. A análise de coerência (Tabela 11) revelou um achado crucial: a matriz “Proposta” não apenas se mostrou coerente com os referenciais nacionais, mas demonstrou um alinhamento com os padrões internacionais superior ao dos próprios cursos brasileiros de referência, sugerindo que o método é capaz de gerar currículos de relevância global. Este alinhamento foi formalmente validado pelos testes de hipótese, que comprovaram que a aderência da “Propost” às diretrizes da SBC é estatisticamente indistinguível ($p > 0.05$) daquela de cursos consolidados.

Por fim, o desempenho da Análise de Grafos na “elaboração” da matriz foi confirmado pela análise estrutural. A metodologia de classificação por centralidade foi capaz de organizar a lista de disciplinas em papéis pedagógicos que se mostraram perfeitamente alinhados à estrutura de núcleo-periferia da rede, conforme revelado pelas visualizações dos

grafos (Figuras 13 e 14). A replicação bem-sucedida de todas estas validações no estudo de caso de Administração atesta o alto desempenho e a capacidade de generalização do método proposto.

6 Considerações Finais

Este capítulo final apresenta as considerações finais desta dissertação, que se propôs a apresentar uma solução possível ao desafio da elaboração e estruturação de matrizes curriculares no ensino superior através da proposição e validação de um método inédito que integra Inteligência Artificial Generativa e Análise de Grafos para apoiar os especialistas educacionais nesse processo. Após a apresentação da metodologia e a análise dos resultados, o objetivo aqui é consolidar os achados, discutir as contribuições da pesquisa para a ciência e para a prática institucional, e refletir sobre o alcance e o impacto do trabalho. Para tanto, esta seção irá sintetizar os resultados em resposta aos objetivos propostos, detalhar as contribuições do estudo, e, por fim, apontar direções para futuras investigações na área.

A presente dissertação, portanto, desenvolveu e validou um método computacional que parte de fontes curriculares heterogêneas para chegar a uma matriz curricular estruturada e validada. O método inicia com o enriquecimento semântico dos componentes curriculares através de IAGen executando a tarefa de geração de *keywords* a partir das ementas das disciplinas, seguido pela construção de uma matriz de similaridade que quantifica as relações conceituais entre as disciplinas. A partir desta matriz, o método utiliza técnicas de grafos para agrupar disciplinas similares e assim formar subgrafos, que representam grupos de disciplinas com similaridade acima de um limiar obtido por análise de modularidade da matriz das disciplinas, após isso, gera novos componentes curriculares por meio de fusão semântica e, em seguida, após incorporar as novas disciplinas ao corpus de disciplinas, emprega uma arquitetura de Geração Aumentada por Recuperação (RAG) para extrair informações relevantes de Documentos de Referência na formação de cursos, combinados com modelos de IAGEN, para selecionar um conjunto final de disciplinas. Por fim, a matriz resultante é novamente modelada como um grafo para uma análise estrutural que revela os papéis pedagógicos de cada disciplina, através do papel de cada componente curricular dentro do grafo.

A execução deste método permitiu alcançar todos os objetivos traçados para esta pesquisa. O objetivo de enriquecer os componentes curriculares, através das *keywords* foi atingido e validado pela alta similaridade (média de 0,8862) entre as *keywords* geradas e as ementas originais. O processo de geração de novas disciplinas por fusão também foi validado, alcançando uma coesão semântica média de 0,8845 em relação aos seus grupos de origem,

compostos por disciplinas similares. O objetivo de selecionar um conjunto de disciplinas alinhado a normativas foi cumprido e validado por testes de hipótese, que demonstraram que as matrizes geradas possuem uma aderência estatisticamente indistinguível ($p > 0.05$ na maioria dos testes) daquela de cursos de referência, já estabelecidos em instituições de ensino superior, além de apresentar valores estatisticamente indistinguível de ementas de cursos internacionais. A análise estrutural via grafos foi bem-sucedida ao revelar uma organização coerente de núcleo-periferia e ao classificar as disciplinas em papéis pedagógicos que correspondem à sua função no currículo, a partir de suas métricas de centralidade. Finalmente, a generalização e robustez do método foram avaliadas com sucesso através da aplicação completa do método em dois domínios distintos, Ciência de Dados e Administração, cumprindo o último objetivo específico e, por consequência, o objetivo geral desta dissertação.

Ao alcançar os objetivos propostos, esta pesquisa oferece contribuições originais em múltiplas frentes. A principal contribuição metodológica é a proposição e validação de um método integrado de ponta a ponta para o suporte ao design curricular e criação de disciplinas a partir de disciplinas semelhantes. Diferentemente de abordagens na literatura que aplicam técnicas de forma isolada, este trabalho demonstrou um fluxo que articula a análise semântica via PLN, a síntese de conteúdo via IAGen — incluindo a arquitetura RAG para alinhamento com documentos normativos — e a análise estrutural via Análise dos Grafos. Do ponto de vista prático e tecnológico, a pesquisa contribui com um método de alta aplicabilidade e generalização. A validação bem-sucedida em dois domínios distintos como Ciência de Dados e Administração comprova a eficácia do método. Por fim, no plano conceitual, esta dissertação contribui com a formalização de um protocolo de colaboração Humano-IA para uma tarefa educacional complexa. O modelo HITL detalhado, com seus pontos de validação e ações de refinamento explícitas, serve como um exemplo prático para o desenvolvimento de sistemas de IA mais responsáveis, transparentes e orientados pela expertise humana em domínios de alto impacto social. Embora a validação experimental desta dissertação tenha simulado esse ciclo com o pesquisador, o protocolo HITL oferece um modelo prático para a aplicação responsável da ferramenta em instituições de ensino.

Ao alcançar os objetivos propostos, esta pesquisa oferece contribuições originais em múltiplas frentes. A principal contribuição metodológica é a apresentação e validação de um método integrado de ponta a ponta para o suporte ao design curricular. Diferentemente de

abordagens na literatura que aplicam técnicas de forma isolada, este trabalho demonstrou um fluxo que articula a análise semântica via PLN, a síntese de conteúdo via IAGen e, a Análise de Grafos. A contribuição da Análise de Grafos é fundamental para o método: primeiramente, ela é utilizada para informar a geração de novas disciplinas, ao modelar a rede de similaridades e permitir a identificação de grupos coesos que servem como base para a criação das novas disciplinas a partir das disciplinas semelhantes dentro de cada grupo (subgrafo) encontrado. Em segundo lugar, ela é empregada para validar e estruturar a matriz final, onde a análise de centralidade revela a estrutura de núcleo-periferia do currículo e permite a classificação objetiva das disciplinas em seus papéis pedagógicos (Básica, Específica e Complementar). A análise dessas métricas fornece ao especialista uma visão da estrutura da matriz, que passa de uma lista de disciplinas para uma estrutura de disciplinas classificadas de acordo com seu papel dentro da matriz.

Do ponto de vista prático e tecnológico, a pesquisa contribui com um método aplicável e generalizado, cuja eficácia foi comprovada em dois domínios distintos (Ciência de Dados e Administração). Por fim, no plano conceitual, a dissertação contribui com a formalização de um protocolo de colaboração Humano-IA para uma tarefa educacional complexa. O modelo HITL detalhado, com seus pontos de validação e ações de refinamento explícitas, serve como um exemplo prático para o desenvolvimento de sistemas de IA mais responsáveis e transparentes, orientados pela expertise humana, buscando prevenir vieses e alucinações dos modelos de IAGen.

Apesar das contribuições apresentadas, é fundamental reconhecer as limitações inerentes ao presente estudo, que abrem caminho para futuras investigações. Primeiramente, a qualidade e a diversidade dos currículos gerados são intrinsecamente dependentes do corpus de dados utilizado como fonte. O método pode, potencialmente, herdar e amplificar vieses ou lacunas de conhecimento presentes nas matrizes curriculares que o alimentam. Em segundo lugar, os resultados são contingentes aos modelos de Inteligência Artificial específicos empregados (*deepseek-chat*, *deepseek-reasoner*, *nomic-embed-text*). O desempenho e as características das respostas podem variar com a utilização de outros modelos de linguagem ou de vetorização. Adicionalmente, a validação do método, embora robusta em seus critérios comparativos e normativos, não se estendeu a uma implementação em um cenário real; não foram medidos os impactos de um currículo gerado pelo método em indicadores de aprendizagem ou na trajetória profissional de estudantes. Por fim,

embora o paradigma HITL seja uma orientação metodológica, o processo de validação pelo especialista humano representa um potencial gargalo em termos de escalabilidade, caso o método fosse aplicado para gerar ou reformular centenas de currículos simultaneamente.

As limitações reconhecidas neste estudo não invalidam os resultados, mas, ao contrário, delineiam um roteiro claro e promissor para futuras investigações. Como perspectivas, no que tange aos modelos de linguagem, propõe-se ir além da exploração de modelos mais recentes e investigar abordagens de IAGen mais especializadas. Isso inclui a possibilidade de realizar o *fine-tuning* de um LLM com um corpus focado em documentos pedagógicos e curriculares, ou a aplicação de técnicas de otimização de *prompts*, como o *prompt tuning* (soft prompts), em vez da chamada direta à API. Tais abordagens poderiam conferir maior especialização e performance ao modelo para as tarefas de síntese e análise. Adicionalmente, a análise estrutural pode ser aprimorada ao se ponderar as arestas no grafo, incorporando variáveis como nível de dificuldade, carga horária e pré-requisitos formais, tornando o mapeamento mais preciso e multifacetado. Pretende-se também estender o método à geração de componentes curriculares complementares — como planos de ensino detalhados, trilhas de aprendizagem e sugestões de atividades avaliativas —, expandindo o escopo do método para além da matriz curricular. Uma possível direção para trabalhos futuros é a aplicação do método na prática institucional. A validação definitiva do valor desta abordagem viria da condução de um estudo de caso em um cenário real, permitindo medir não apenas a qualidade do artefato gerado, mas também o impacto da ferramenta na eficiência do processo e na satisfação dos especialistas envolvidos. Tal estudo permitiria validar o método não apenas como uma prova de conceito, mas como uma solução de alto impacto para os desafios do design curricular contemporâneo.

Referências

- ABDOUNE, R.; LAZIB, L.; DAHMANI-BOUARAB, F.; MIMOUNI, N. Semantic alignment in disciplinary tutoring system: leveraging sentence transformer technology. **Inteligencia Artificial. Ibero-American Journal of Artificial Intelligence**, v. 28, n. 75, p. 46–62, 2024.
- AMORIM, C.; MELO, T. de. Analisando o impacto de paulo freire no ensino brasileiro de ciência, tecnologia e sociedade por meio de citações. **Indagatio Didactica**, v. 16, n. 3, p. 113–132, Nov. 2024. Disponível em: <<https://proa.ua.pt/index.php/id/article/view/38436>>.
- AYTEKIN, M. C.; SAYGIN, Y. et al. Ace: Ai-assisted construction of educational knowledge graphs with prerequisite relations. **Journal of Educational Data Mining**, v. 16, n. 2, p. 85–114, 2024.
- BAHROUN, Z.; ANANE, C.; AHMED, V.; ZACCA, A. Transforming education: A comprehensive review of generative artificial intelligence in educational settings through bibliometric and content analysis. **Sustainability**, Multidisciplinary Digital Publishing Institute, v. 15, n. 17, p. 12983, 2023.
- BALAKRISHNAN, V.; ETHEL, L.-Y. Stemming and lemmatization: A comparison of retrieval performances. **Lecture Notes on Software Engineering**, v. 2, p. 262–267, 01 2014.
- BIANCHESSI, C. (Ed.). **Educação, Currículo, Cultura Digital: reflexões para a escola na atualidade**. Bagé, RS: Editora Bagai, 2021. Disponível em: <<https://educapes.capes.gov.br/bitstream/capes/584678/2/Editora%20BAGAI-%20Educa%C3%A7%C3%A3o%20Curr%C3%ADculo%20Cultura%20Digital.pdf>>.
- CHAKRABORTY, S. **Generative AI in Modern Education Society**. 2024. Disponível em: <<https://arxiv.org/abs/2412.08666>>.
- CHEN, Z.; MAO, H.; LI, H.; JIN, W.; WEN, H.; WEI, X.; WANG, S.; YIN, D.; FAN, W.; LIU, H. et al. Exploring the potential of large language models (llms) in learning on graphs. **ACM SIGKDD Explorations Newsletter**, ACM New York, NY, USA, v. 25, n. 2, p. 42–61, 2024.
- CHOI, G. W.; KIM, S. H.; LEE, D.; MOON, J. Utilizing generative ai for instructional design: Exploring strengths, weaknesses, opportunities, and threats. **TechTrends**, Springer, v. 68, n. 4, p. 832–844, 2024.
- CHRIS, E.; SHERIFFDEEN, K. Shaping ai-driven curriculum development: How generative ai is modern educational content. 11 2024.
- CONKLIN, S.; DORGAN, T.; BARRETO, D. Is ai the new course creator. **Discover Education**, Springer, v. 3, n. 1, p. 285, 2024.
- CORMEN, T. H.; LEISERSON, C. E.; RIVEST, R. L.; STEIN, C. **Introduction to algorithms**. [S.l.]: MIT press, 2022.

DEEPSEEK-AI. **DeepSeek-V3 Technical Report**. 2024. Disponível em: <<https://arxiv.org/abs/2412.19437>>.

_____. **DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning**. 2025. Disponível em: <<https://arxiv.org/abs/2501.12948>>.

DEVLIN, J.; CHANG, M.-W.; LEE, K.; TOUTANOVA, K. **BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding**. 2019. Disponível em: <<https://arxiv.org/abs/1810.04805>>.

DOWNES, S. An introduction to connective knowledge. Innsbruck university press, 2008.

EVANS, T. S.; CHEN, B. Linking the network centrality measures closeness and degree. **Communications Physics**, Nature Publishing Group UK London, v. 5, n. 1, p. 172, 2022.

FORTUNATO, S. Community detection in graphs. **Physics Reports**, Elsevier, v. 486, n. 3–5, p. 75–174, 2010.

FREITAS, E. L. S. X.; PINTO, I. I. B. S.; ISOTANI, S.; MARQUES, L.; DERMEVAL, D.; SILVA, A. P. d.; MELLO, R. F. Offline artificial intelligence for education: a path to a more inclusive field. **Revista Brasileira de Informática na Educação - RBIE**, 2023.

GUPTA, S.; RANJAN, R.; SINGH, S. N. **A Comprehensive Survey of Retrieval-Augmented Generation (RAG): Evolution, Current Landscape and Future Directions**. 2024. Disponível em: <<https://arxiv.org/abs/2410.12837>>.

JURAFSKY, D.; MARTIN, J. H. **Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models**. 3rd. ed. [s.n.], 2025. Online manuscript released January 12, 2025. Disponível em: <<https://web.stanford.edu/~jurafsky/slp3/>>.

KUMAR, J. A.; ZHUANG, M.; THOMAS, S. Chatgpt for natural sciences course design: Insights from a strengths, weaknesses, opportunities, and threats analysis. **Natural Sciences Education**, v. 53, n. 2, p. e70003, 2024.

_____. Chatgpt for natural sciences course design: Insights from a strengths, weaknesses, opportunities, and threats analysis. **Natural Sciences Education**, Wiley Online Library, v. 53, n. 2, p. e70003, 2024.

LEWIS, P.; PEREZ, E.; PIKTUS, A.; PETRONI, F.; KARPUKHIN, V.; GOYAL, N.; KÜTTLER, H.; LEWIS, M.; YIH, W. tau; ROCKTÄSCHEL, T.; RIEDEL, S.; KIELA, D. Retrieval-augmented generation for knowledge-intensive nlp tasks. **Advances in Neural Information Processing Systems (NeurIPS)** 33, 2020. Disponível em: <<https://arxiv.org/abs/2005.11401>>.

_____. **Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks**. 2021. Disponível em: <<https://arxiv.org/abs/2005.11401>>.

LOPES, A. C. Teorias pós-críticas, política e currículo. **Educação, sociedade & culturas**, n. 39, p. 7–23, 2013.

LU, W.; ZOGHI, B. B. Using generative ai for a graduate level capstone course design—a case study. In: **2024 ASEE Annual Conference & Exposition**. [S.l.: s.n.], 2024.

- MANNING, C. D. **An introduction to information retrieval**. [S.l.: s.n.], 2009.
- MIAO, F.; HOLMES, W.; UNESCO. **Guidance for Generative AI in Education and Research**. [S.l.], 2023. Publicly available guidance on GenAI in education (UNESCO, 2023). Disponível em: <<https://unesdoc.unesco.org/ark:/48223/pf0000386693>>.
- MONARCH, R. **Human-in-the-Loop Machine Learning**. [S.l.]: Manning Publications, 2021.
- MOSQUEIRA-REY, E.; HERNÁNDEZ-PEREIRA, E.; ALONSO-RÍOS, D.; BOBES-BASCARÁN, J.; FERNÁNDEZ-LEAL Ángel. Human-in-the-loop machine learning: a state of the art. **Artificial Intelligence Review**, v. 56, p. 3005–3054, 2022. Disponível em: <<https://link.springer.com/article/10.1007/s10462-022-10246-w>>.
- NEWMAN, M. E. J.; GIRVAN, M. Finding and evaluating community structure in networks. **Phys. Rev. E**, American Physical Society, v. 69, p. 026113, Feb 2004. Disponível em: <<https://link.aps.org/doi/10.1103/PhysRevE.69.026113>>.
- NUSSBAUM, Z.; MORRIS, J. X.; DUDERSTADT, B.; MULYAR, A. Nomic embed: Training a reproducible long context text embedder. **arXiv preprint arXiv:2402.01613**, 2024.
- OTTE, E.; ROUSSEAU, R. Social network analysis: a powerful strategy, also for the information sciences. **Journal of information Science**, Sage Publications Sage CA: Thousand Oaks, CA, v. 28, n. 6, p. 441–453, 2002.
- PADOVANO, A.; CARDAMONE, M. Towards human-ai collaboration in the competency-based curriculum development process: The case of industrial engineering and management education. **Computers and Education: Artificial Intelligence**, Elsevier, v. 7, p. 100256, 2024.
- PINAR, W. F. **What is curriculum theory?** [S.l.]: Routledge, 2019.
- QU, K.; LI, K. C.; WONG, B. T.; WU, M. M.; LIU, M. A survey of knowledge graph approaches and applications in education. **Electronics**, MDPI, v. 13, n. 13, p. 2537, 2024.
- RAMOS, D. B. et al. Uma ferramenta baseada em grafo para identificação e visualização de trilhas de aprendizagem. Universidade Federal do Amazonas, 2016.
- REICHARDT, J.; BORNHOLDT, S. Statistical mechanics of community detection. **Physical Review E**, American Physical Society (APS), v. 74, n. 1, jul. 2006. ISSN 1550-2376. Disponível em: <<http://dx.doi.org/10.1103/PhysRevE.74.016110>>.
- ROS, M. Z. Ia generativa y chatgpt en educación: Un reto para la evaluación y¿ una nueva pedagogía? **Revista Paraguaya de Educación a Distancia (REPED)**, v. 5, n. 1, p. 12–44, 2024.
- SADAF, A.; OLESOVA, L.; CHOI, H. Empowering learning networks: Insights from social network analysis in inquiry-based discussions. **Online Learning Journal**, Online Learning Consortium (OLC), p. 200, 2024.
- SANDU, R.; GIDE, E.; ELKHODR, M. The role and impact of chatgpt in educational practices: Insights from an australian higher education case study. **Discover Education**, v. 3, n. 1, p. 71, 2024.

_____. The role and impact of chatgpt in educational practices: insights from an australian higher education case study. **Discover Education**, Springer, v. 3, n. 1, p. 71, 2024.

SCHENEIDER, T. F.; CECHINEL, C.; MELLO, R. F. L. d.; FREITAS, E. L. S. X.; FALCÃO, T. P. d. R.; RAMOS, V. F. C. Análise do nível de maturidade na adoção de learning analytics em instituições de ensino superior do sul e sudeste brasileiro. **Revista Brasileira de Informática na Educação**, v. 31, p. 1005–1030, dez. 2023. Disponível em: <<https://journals-sol.sbc.org.br/index.php/rbie/article/view/3058>>.

SCHÜTZE, H.; MANNING, C. D.; RAGHAVAN, P. **Introduction to information retrieval**. [S.l.]: Cambridge University Press Cambridge, 2008. v. 39.

SIEMENS, G. Connectivism: a learning theory for the digital age. **Int J Instr Technol Dis Learn**, 2025.

STAVRINIDES, P.; ZUEV, K. M. Course-prerequisite networks for analyzing and understanding academic curricula. **Applied Network Science**, v. 8, n. 1, p. 19, 2023.

_____. Course-prerequisite networks for analyzing and understanding academic curricula. **Applied Network Science**, Springer, v. 8, n. 1, p. 19, 2023.

TRAAG, V. A.; WALTMAN, L.; ECK, N. J. van. From louvain to leiden: guaranteeing well-connected communities. **Scientific Reports**, Springer Science and Business Media LLC, v. 9, n. 1, mar. 2019. ISSN 2045-2322. Disponível em: <<http://dx.doi.org/10.1038/s41598-019-41695-z>>.

TSONI, R.; PAXINOU, E.; GKOUALALAS-DIVANIS, A.; KARAPIPERIS, D.; KALLES, D.; VERYKIOS, V. S. Exploiting properties of student networks to enhance learning in distance education. **Information**, MDPI, v. 15, n. 4, p. 234, 2024.

TUZÓN, P.; SALVÀ, A. S.; FERNÁNDEZ-GRACIA, J. Complex networks approach to curriculum analysis and subject integration: A case study on physics and mathematics. **Physical Review Physics Education Research**, American Physical Society (APS), v. 21, n. 1, jun. 2025. ISSN 2469-9896. Disponível em: <<http://dx.doi.org/10.1103/PhysRevPhysEducRes.21.010161>>.

VASILEIOU, S. L.; YEOH, W. Trace-cs: Trustworthy reasoning for contrastive explanations in course scheduling problems. **CoRR**, abs/2409.03671, 2024. Disponível em: <<https://arxiv.org/abs/2409.03671>>.

VASWANI, A.; SHAZEER, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GOMEZ, A. N.; KAISER, L.; POLOSUKHIN, I. Attention is all you need. **CoRR**, abs/1706.03762, 2017. Disponível em: <<http://arxiv.org/abs/1706.03762>>.

WALKER, R. guiar; REGO, R. B. Simplificação dos eixos da engenharia de produção usando teoria dos grafos. 2025.

WANG, Y.; ZHAN, Z.; WANG, H. Network analysis of outcome-based education curriculum system: A case study of environmental design programs in medium-sized cities. **Sustainability**, MDPI, v. 17, n. 15, p. 7091, 2025.

WASSERMAN, S.; FAUST, K. **Social Network Analysis: Methods and Applications**. [S.l.]: Cambridge University Press, 1994.

WHITE, J.; FU, Q.; HAYS, S.; SANDBORN, M.; OLEA, C.; GILBERT, H.; ELNASHAR, A.; SPENCER-SMITH, J.; SCHMIDT, D. C. **A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT**. 2023.

WU, X. e. a. A survey of human-in-the-loop for machine learning. **arXiv preprint arXiv:2108.00941**, 2021. <<https://arxiv.org/abs/2108.00941>>.

YANG, B.; GHAREBHAYGLOO, M.; RONDI, H. R.; HORTIS, E.; LOSTALO, E. Z.; HUANG, X.; JOHNSON, M. D.; GAO, F.; DELIS, G.; MILENKOVIC, O.; ERCAL, G. Comparative analysis of course prerequisite networks for five midwestern public institutions. **Applied Network Science**, v. 9, p. 25, 2024. Disponível em: <<https://appliednetsci.springeropen.com/articles/10.1007/s41109-024-00637-z>>.

APÊNDICES

APÊNDICE A – Matrizes Curriculares Utilizadas na Pesquisa

A.1 Matriz Curricular do Curso de Bacharelado em Administração - Universidade Federal de Pernambuco (UFPE)

Fonte: <<https://www.ufpe.br/administracao-bacharelado-ccsa>>

Tabela 22 – Exemplo de disciplinas utilizadas do curso de Administração da UFPE.

Nome da Disciplina	Ementa Original
Introdução à Administração	Antecedentes históricos da Administração; Processo administrativo; Planejamento e controle estratégicos; Direção; Áreas funcionais e ambientes organizacionais; Novas tendências em Administração; Ética na Administração.
Obrigações Empresariais	Conceito de pessoas física e jurídica; Capacidade civil; Distinção entre direito público e privado; Responsabilidade civil frente ao direito de empresa; Formação dos contratos no novo código civil: sujeitos, objeto e finalidade; Obrigações: conceito e tipos; Títulos de crédito: conceito e tipos; Nova lei de falência: conceitos básicos e limites legais; Estrutura do poder judiciário e legislativo brasileiro: atribuições e competências restritivas aos poderes.
Fundamentos de Sociologia	A Sociologia e outras ciências; Modelos de análise sociológica; O social e a sociedade; Conceitos sociológicos básicos.
Ética e Filosofia para o Administrador	Conceitos fundamentais de ética e filosofia; A pessoa humana; O problema moral; Ética e responsabilidade social; Deontologia do administrador; A ética nas funções gerenciais.

Tabela 22 – (Continuação) Disciplinas obrigatórias utilizadas do curso de Administração da UFPE.

Nome da Disciplina	Ementa Original
Administração Tributária	A função social dos tributos; O sistema tributário nacional; Espécies de tributos; Obrigações tributárias; Administração tributária; Impostos federais, estaduais e municipais.
Comunicação Empresarial	O processo de comunicação; Comunicação interpessoal e organizacional; Barreiras à comunicação; Comunicação escrita e oral; Comunicação e tecnologia da informação.
Teoria Geral da Administração	O campo da Administração; As organizações e o ambiente; A evolução do pensamento em Administração; As abordagens clássica, humanística, burocrática, estruturalista, comportamental, sistêmica, contingencial e as novas abordagens.
Administração Pública	Evolução e características da administração pública brasileira; A administração pública burocrática e a crise do Estado; A administração pública gerencial; O processo de reforma do Estado; A gestão pública contemporânea.
Contabilidade Geral A	Princípios fundamentais e normas de contabilidade; O patrimônio; Escrituração contábil; O balanço patrimonial e a demonstração do resultado do exercício.
Administração de Políticas Públicas	O Estado e as políticas públicas; O ciclo de políticas públicas; Análise e avaliação de políticas públicas; Gestão de políticas públicas.
Administração no Contexto Brasileiro	A formação histórica, social e econômica do Brasil; O Estado o desenvolvimento; As organizações e a administração no Brasil; A cultura brasileira e o gerenciamento.

A.2 Matriz Curricular do Curso de Bacharelado em Ciências Biológicas - Universidade Federal de Pernambuco (UFPE)

Fonte: <<https://www.ufpe.br/ciencias-biologicas>>

Tabela 23 – Exemplo de disciplinas utilizadas do curso de Ciências Biológicas da UFPE.

Nome da Disciplina	Ementa Original
Biofísica A	Em destaque: conceitos, propriedades e métodos de análise biofísica; processos de termodinâmica; propriedades, características e aplicações de biomoléculas.
Biologia Celular A	Estudo da célula eucariótica sob o ponto de vista morfológico; Introdução às principais técnicas aplicadas à morfologia; Utilização de modelos de células eucarióticas.
Bioquímica I	Estudo das propriedades físico-químicas da água e das soluções; Estrutura e função de aminoácidos, proteínas, carboidratos, lipídios e ácidos nucleicos; Processos de biossíntese e degradação; Regulação enzimática.
Bioética	Abordagem histórica da ética na ciência e na pesquisa; Bioética e meio ambiente; Questões atuais de bioética; Noções básicas de legislação e de profissional frente a dilemas bioéticos relacionados ao ser humano.
Direitos Humanos e Relações Étnico-Raciais	Conceitos de direitos humanos e suas respectivas fundamentações históricas e éticas; Processos de promoção, proteção e defesa; Aplicação na vida cotidiana e nas relações étnico-raciais.
Geologia	Compreensão sobre os processos geológicos atuantes no interior e na superfície da Terra, assim como a interação entre ambos e sua relação com o tempo geológico.
Introdução ao Laboratório	Conceitos básicos sobre técnicas analíticas gerais utilizadas nos laboratórios das Ciências Biológicas.

Tabela 23 – (Continuação) Disciplinas obrigatórias utilizadas do curso de Ciências Biológicas da UFPE.

Nome da Disciplina	Ementa Original
Metodologia Científica e Filosofia da Ciência	Apresentação e caracterização da metodologia científica; Descrição das fases de um projeto científico, desde a elaboração do projeto até a redação dos resultados; Fundamentação do conhecimento científico; Apresentação das principais características e limitações do método científico.
Micro-organismos	Conceitos básicos de vírus, bactérias, arqueobactérias e protozoários; Abordagem histórica e importância dos micro-organismos; Diversidade de grupos bacterianos; Vírus: estrutura, replicação, genética e patogenicidade; Arquitetura da célula procariótica; Fisiologia bacteriana; Crescimento bacteriano; Variabilidade genética; Controle de micro-organismos.
Sistemática e Biogeografia	A disciplina introduz os conceitos básicos de taxonomia, sistemática filogenética e biogeografia; regras de nomenclatura biológica; conceitos e terminologias básicas de homologia, caracteres e grupos filogenéticos; tipos de árvores filogenéticas; dará noções básicas de construção de árvores filogenéticas; Introdução à Biogeografia; História da Biogeografia; Espécies, extinção, dispersão, endemismo, províncias e tipos de biomas.

A.3 Matriz Curricular do Curso de Bacharelado em Ciências da Computação - Universidade Federal de Pernambuco (UFPE)

Fonte: <<https://www.ufpe.br/ciencia-da-computacao-bacharelado-ci>>

Tabela 24 – Exemplo de disciplinas utilizadas do curso de Ciências da Computação da UFPE.

Nome da Disciplina	Ementa Original
Álgebra Vetorial e Linear para Computação	Espaços vetoriais; transformações lineares e aplicações.
Cálculo Diferencial e Integral I	Limites e continuidade de funções; Derivadas; Aplicações da derivada; Teorema de Rolle; Teorema do valor médio e teorema do valor médio generalizado; Integrais definidas; Integrais indefinidas; Técnicas de integração.
Introdução à Computação	O que é a Internet?
Introdução à Programação	Conceitos básicos de linguagens de programação; Conceitos básicos de programação orientada a objetos e qualidade de software; Estruturas de controle e estrutura de dados compostas; Ambiente de programação; Conceitos básicos de programação imperativa; Padrões e arquiteturas de software convencionais; Estilo de programação; Projeto de implementação.
Matemática Discreta para Computação	Introdução; Indução e recursão; Teoria dos conjuntos; Álgebra discreta; Teoria dos números; Combinatória; Funções geradoras.
Algoritmos e Estruturas de Dados	Introdução à análise de algoritmos; Elementos de estruturas de dados; Algoritmos de busca; Algoritmos geométricos; Noções de buscas por exaustão e problema NP-completos e aplicações.
Estatística e Probabilidade para Computação	Estatística descritiva; Probabilidade; Estatística inferencial.
Física para Computação	Revisão de mecânica; Eletricidade e eletromagnetismo; Oscilações; Ondas; Conceitos de termodinâmica.

Tabela 24 – (Continuação) Disciplinas obrigatórias utilizadas do curso de Ciências da Computação da UFPE.

Nome da Disciplina	Ementa Original
Lógica para Computação	Introdução; Lógica proposicional; Lógica dos predicados; Aplicações práticas para computação; Tópicos avançados: lógicas -clássicas.
Sistemas Digitais	Álgebra Booleana; VHDL; Circuitos combinacionais; Circuitos sequenciais.

A.4 Matriz Curricular do Curso de Bacharelado em Engenharia da Computação - Universidade Federal de Pernambuco (UFPE)

Fonte: <<https://www.ufpe.br/engenharia-da-computacao-bacharelado-cin>>

Tabela 25 – Exemplo de disciplinas utilizadas do curso de Engenharia da Computação da UFPE.

Nome da Disciplina	Ementa Original
Álgebra Vetorial e Linear para Computação	Espaços vetoriais; Transformações lineares e aplicações.
Cálculo Diferencial e Integral I	Limites e continuidade de funções; Derivadas; Aplicações da derivada; Teorema de Rolle; Teorema do valor médio e teorema do valor médio generalizado; Integrais definidas; Integrais indefinidas; Técnicas de integração.
Cálculo Diferencial e Integral II	Técnicas de integração; Diferenciabilidade em duas variáveis; Integração em duas variáveis.
Cálculo Diferencial e Integral III	Séries; Fórmula de Taylor; Máximos e mínimos; Integrais triplas; Integrais de linha e de superfície; Teoremas de Green, Gauss e Stokes.

Tabela 25 – (Continuação) Disciplinas obrigatórias utilizadas do curso de Engenharia da Computação da UFPE.

Nome da Disciplina	Ementa Original
Física Experimental I	Cinemática; Energia mecânica; Colisões; Dinâmica dos corpos rígidos; Oscilações e ondas; Hidrostática termodinâmica.
Física Geral I	Movimento em uma dimensão; Vetores; Movimento em um plano; Dinâmica da partícula; Trabalho e energia; Conservação da energia; Conservação do momentum linear; Choques; Cinemática da rotação; Dinâmica da rotação.
Física Geral II	Gravitação; Flúidos; Movimento oscilatório; Ondas; Superposição e interferência de ondas harmônicas; Termologia; Leis da termodinâmica; Teoria cinética dos gases; Expansão térmica.
Física Geral III	Campo elétrico; Potencial elétrico; Capacitores e dielétricos; Circuitos elétricos; Campo magnético; Lei de Ampère; Indução eletromagnética.
Introdução à Computação	O que é a Internet?
Algoritmos e Estruturas de Dados	Introdução à análise de algoritmos; Elementos de estruturas de dados; Algoritmos de busca; Algoritmos geométricos; Noções de buscas por exaustão e problema NP-completos e aplicações.

A.5 Matriz Curricular do Curso de Bacharelado em Sistema da Informação - Universidade Federal de Pernambuco (UFPE)

Fonte: <<https://www.ufpe.br/sistemas-de-informacao-bacharelado-cin>>

Tabela 26 – Exemplo de disciplinas utilizadas do curso de Sistemas de Informação da UFPE.

Nome da Disciplina	Ementa Original
Administração Contemporânea	Fundamentos teóricos; Design organizacional - Estruturas organizacionais e processos organizacionais; Racionalidade gerencial; Novas tendências em administração; Redes organizacionais; Inovação organizacional.
Cálculo Diferencial e Integral I	Limites e continuidade de funções; Derivadas; Aplicações da derivada; Teorema de Rolle; Teorema do valor médio e teorema do valor médio generalizado; Integrais definidas; Integrais indefinidas; Técnicas de integração.
Introdução aos Sistemas de Informação	Conceitos básicos; Evolução histórica; Introdução aos sistemas de informação; Introdução às tecnologias de informação e comunicação; Empresas digitais; Questões éticas e sociais; O mercado.
Matemática Discreta para Computação	Introdução; Indução e recursão; Teoria dos conjuntos; Álgebra discreta; Teoria dos números; Combinatória; Funções geradoras.
Programação I	Conceitos básicos de algoritmos; Conceitos básicos de linguagens de programação; Conceitos básicos de programação orientada a objetos e qualidade de software; Ambiente de programação; Estruturas de controle e estrutura de dados compostas; Arquivos; Conceitos básicos de estruturas de dados; Projeto de implementação.
Algoritmos e Estruturas de Dados	Conceitos básicos de algoritmos; Algoritmos e estruturas de dados dinâmicas básicas; Técnicas de construção de algoritmos; Conceitos de complexidade de algoritmo.

Tabela 26 – (Continuação) Disciplinas obrigatórias utilizadas do curso de Sistemas de Informação da UFPE.

Nome da Disciplina	Ementa Original
Análise das Organizações	Entendimento da composição organizacional; A conceituação da organização como entidade social; Configurações organizacionais.
Estatística e Probabilidade para Computação	Estatística descritiva; Probabilidade; Estatística inferencial; Eventos; Experimentos aleatórios; Teoria da amostragem.
Inglês para Computação	Vocabulário e expressões idiomáticas inglesas em Ciências da Computação e Tecnologia da Informação; Interpretação de textos técnico-científicos em inglês; Compreensão de apresentações orais técnico-científicas em inglês; Estilística da redação de documentos técnico-científicos em inglês; Estilística de seminários orais técnico-científicos em inglês.
Lógica para Computação	Introdução; Lógica proposicional; Lógica dos predicados; Aplicações práticas para computação; Tópicos avançados: lógicas não clássicas.

A.6 Matriz Curricular do Curso de Bacharelado em Ciência de Dados - Universidade Federal do Ceará (UFC)

Fonte: <https://si3.ufc.br/sigaa/public/curso/curriculo.jsf?lc=pt_BR&id=72460878>

Tabela 27 – Exemplo de disciplinas utilizadas do curso de Ciência de Dados da UFC.

Nome da Disciplina	Ementa Original
Matemática Discreta	Raciocínio lógico; Lógica matemática; Técnicas de demonstração: exaustiva, direta, contraposição, absurdo, indução (fraca e forte); Indução Finita; Conjuntos; Relações e funções; Contagem; Relações: fechos, ordem parcial e total, relações e classes de equivalência; Noções de teoria dos grafos: terminologia, alguns grafos especiais, isomorfismo, conectividade, árvores (definição e propriedades).
Função em Uma Variável	Funções reais de uma variável; Limite e continuidade; Derivação; Aplicações da derivada; Máximo e mínimo de funções; Integração; Técnicas de integração.
Introdução à Ciência de Dados	Perfil do cientista de dados; Relatórios reprodutíveis em ciência de dados; Coleta e pré-processamento de dados; Análise exploratória de dados; Aprendizado supervisionado/não-supervisionado; Ética e privacidade no contexto de big data e ciência dos dados.
Fundamentos de Programação	Algoritmos e Lógica de Programação; Construção de algoritmos utilizando uma linguagem de programação; Linguagens Interpretadas e Compiladas; Tipos de dados primitivos; Instruções de entrada e saída de dados; Estruturas de condição; Estruturas de repetição; Funções, Passagem de Parâmetros, Recursividade; Representação de Dados (Listas, Vetores e Matrizes, Strings e Tuplas, Manipulação de Arquivos, Conjuntos, Dicionários).

Tabela 27 – (Continuação) Disciplinas obrigatórias utilizadas do curso de Ciência de Dados da UFC.

Nome da Disciplina	Ementa Original
Metodologia da Pesquisa Científica	Base conceitual e metodológica da pesquisa científica: ciência e método científico, tipos de conhecimento; conceito de ciência; classificação e divisão da ciência; Tipos de pesquisa na área da computação; Trabalhos acadêmicos: tipos, características e diretrizes para elaboração; Revisão da literatura; Ferramentas de Busca Bibliográfica; Metodologia quantitativa e qualitativa; Medidas de Qualidade de Periódicos e Veículos de Impacto; Publicação Científica; Plágio.
Álgebra Linear e Geometria Analítica	Sistemas lineares; Coordenadas no plano e no espaço; Vetores, produto escalar, determinantes, produto vetorial, produto misto; Equações de retas e planos; Espaços vetoriais; Transformações lineares, matrizes, diagonalização; Produtos internos.
Função em Várias Variáveis	Cálculo diferencial para funções de várias variáveis; Cálculo vetorial: Curvas, superfícies e parametrizações; Máximos e mínimos de funções de várias variáveis; Integração: Dupla e tripla.
Probabilidade e Estatística	Análise Exploratória de Dados; Análise Bidimensional; Noções de amostragem; Probabilidade e Variáveis Aleatórias; Inferência estatística.
Estruturas de Dados I	Tipos abstratos de dados, conceitos, operações, representações e manipulação; Vetores, matrizes, listas, pilhas e filas; Listas ligadas, circulares e duplamente ligadas; Estruturas para representação de árvores; Árvores binárias e suas aplicações; Estruturas de representação de grafos (matriz de adjacência e de incidência); Estruturas de dados para conjuntos disjuntos (Union/Find); Pesquisa e ordenação de dados.

Tabela 27 – (Continuação) Disciplinas obrigatórias utilizadas do curso de Ciência de Dados da UFC.

Nome da Disciplina	Ementa Original
Programação Orientada a Objetos	Introdução aos conceitos fundamentais da programação orientada a objetos; Diagramas de Classes da UML; Conceitos fundamentais de orientação a objetos: classes, objetos, atributos/estados, métodos/comportamentos, mensagens; Tipos e classes; Polimorfismo; Abstração, Generalização, Subclasses, Superclasses, Instanciação; Herança; Encapsulamento; Agregação; Composição; Construtores e Destrutores; Ligação estática e dinâmica; Herança múltipla e interfaces. Uso de bibliotecas em linguagens orientadas a objetos. Aplicações e estudos de caso.

A.7 Matriz Curricular do Curso de Bacharelado em Ciência de Dados - Universidade Federal da Paraíba (UFPB)

Fonte: <https://sigaa.ufpb.br/sigaa/public/curso/portal.jsf?id=14289031&lc=pt_BR>

Tabela 28 – Exemplo de disciplinas utilizadas do curso de Ciência de Dados da UFPB.

Nome da Disciplina	Ementa Original
Cálculo Vetorial e Geometria Analítica	Vetores no espaço; Retas e Planos; Cônicas; Superfícies Quádricas.
Cálculo Diferencial e Integral I	Funções reais de uma variável real; Limite e Continuidade; Derivadas: conceito, regras e aplicações.

Tabela 28 – (Continuação) Disciplinas obrigatórias utilizadas do curso de Ciência de Dados da UFPB.

Nome da Disciplina	Ementa Original
Introdução à Programação	Histórico das linguagens de programação; Descrição e construção de algoritmos; Metodologia de programação; Introdução à uma linguagem de programação moderna: tipos elementares e compostos de dados; operadores; expressões e funções; mecanismos de passagem de parâmetros; variáveis e comandos; procedimentos; recursividade; tipos definidos pelo programador e tipos abstratos de dados; noções de estruturas dinâmicas de dados; Aplicações práticas.
Introdução ao Computador	Evolução histórica dos computadores; Arquitetura básica de um computador; Sistemas de numeração e representação de dados; Introdução aos sistemas operacionais; Introdução às redes de computadores e à internet.
Metodologia do Trabalho Científico	Conhecimento e ciência; A ciência moderna e o contexto sociocultural; Ciência e método científico; Técnica de estudo: técnica de leitura, de resumir e elaborar fichamentos; Produção científica e apresentação estética de trabalhos acadêmicos: position paper, resenhas, relatórios, ensaios, artigo e monografias.
Cálculo Diferencial e Integral II	Integral de funções uma variável real; Funções reais de várias variáveis: limite e continuidade; Derivadas Parciais e Diferenciabilidade; Regra da Cadeia e derivação implícita; Máximos e Mínimos; Multiplicadores de Lagrange.
Introdução à Álgebra Linear	Espaços Vetoriais; Aplicações Lineares e Matrizes; Diagonalização de Operadores; Produto Interno.

Tabela 28 – (Continuação) Disciplinas obrigatórias utilizadas do curso de Ciência de Dados da UFPB.

Nome da Disciplina	Ementa Original
Matemática Discreta	Introdução à lógica proposicional; Contagem e combinatória: princípios de adição, princípios do produto, listas, fatorial, arranjo, permutações, combinações, com e sem repetição de elementos, subconjuntos e triângulo de Pascal; Relações: conceito, funções como relações, propriedades, equivalências; Indução Matemática: princípios, indução forte, visão recursiva, indução estrutural, recorrências; Sequências e somatórios; Inferência e prova: regras de inferência, prova direta, prova por indução, prova por contradição, prova por construção e prova por absurdo.
Programação Estruturada	Introdução à linguagem C: tipos elementares; operadores e expressões; Estruturas de controle: decisão repetição; Ponteiros e funções; Passagem por valor e referência; Recursividade; Arranjos e matrizes; Strings; Programação multiarquivo; Processamento de arquivos; Tipos definidos pelo programador e tipos abstratos de dados; Alocação dinâmica de memória; Sockets; Aplicações práticas.
Introdução à Ciência de Dados	Conceitos introdutórios de Ciência de Dados: Big Data e Data Driven Economy; a ciência dos dados e as organizações; inteligência estratégica baseada em dados; mineração de dados; aprendizagem de máquina; aprendizagem estatística; redes neurais; Análise exploratória de dados: tipos de variáveis; limpeza de dados; dados faltantes; outliers; distribuições de frequências; medidas de posição e dispersão; Ferramentas computacionais para tratamento, análise e visualização de dados: histogramas, Boxplot.

APÊNDICE B – Métricas de Centralidade Detalhadas

Este apêndice apresenta os valores detalhados das métricas de centralidade calculadas para cada uma das disciplinas das matrizes curriculares geradas nos estudos de caso, servindo como a base de dados para a análise estrutural e a classificação de papéis pedagógicos apresentada no Capítulo 5.

B.1 Estudo de Caso: Ciência de Dados

Tabela 29 – Métricas de Centralidade Calculadas para as Disciplinas da Matriz de Ciência de Dados (T=0,6957).

Disciplina	Grau	Intermediação	Proximidade
Otimização Não-Linear	0.8718	0.0121	0.8864
Análise e Projeto de Algoritmos	0.8462	0.0472	0.8667
Fundamentos de Programação	0.8205	0.0405	0.8478
Algoritmos, Estruturas de Dados e Teoria da Computação com Otimização	0.8205	0.0418	0.8478
Matemática Discreta	0.7692	0.0256	0.8125
Estrutura de Dados	0.7692	0.0432	0.8125
Fundamentos e Aplicações da Inteligência Artificial	0.7692	0.0148	0.8125
Lógica para Computação	0.6667	0.0108	0.7500
Introdução à Ciência de Dados	0.6410	0.0013	0.7358
Probabilidade e Estatística	0.6410	0.0121	0.7358
Laboratório de Ciência de Dados	0.6410	0.0364	0.7358
Visualização de Dados	0.6154	0.0324	0.7222
Aprendizagem de Máquina e Reconhecimento de Padrões	0.6154	0.0108	0.7222

Tabela 29 – (Continuação) Métricas de Centralidade Calculadas para Ciência de Dados.

Disciplina	Grau	Intermediação	Proximidade
Estatística e Probabilidade para Computação	0.5897	0.0297	0.7091
Otimização Contínua	0.5897	0.0270	0.7091
Aprendizagem de Máquina	0.5641	0.0121	0.6964
Banco de Dados I	0.5385	0.0067	0.6842
Cálculo Diferencial e Integral I	0.5385	0.0135	0.6842
Informática Teórica	0.5128	0.0081	0.6724
Big Data	0.5128	0.0256	0.6724
Paradigmas de Linguagens Computacionais	0.4872	0.0135	0.6610
Pesquisa Operacional	0.4872	0.0067	0.6610
Programação Orientada a Objetos	0.4615	0.0000	0.6500
Programação Estruturada	0.4615	0.0013	0.6500
Métodos Numéricos e Equações Diferenciais Computacionais	0.4615	0.0094	0.6500
Análise Multivariada e Aprendizagem Não Supervisionado	0.4359	0.0081	0.6393
Estatística Computacional Aplicada: Probabilidade e Inferência	0.4359	0.0054	0.6393
Fundamentos de Programação e Engenharia de Software	0.4103	0.0121	0.6190
Matemática Discreta e Lógica Computacional	0.4103	0.0094	0.6290
Fundamentos e Laboratório de Ciência de Dados com Big Data e Visualização	0.3846	0.0054	0.6094
Sistemas Inteligentes Avançados	0.3590	0.0000	0.6000
Análise de Regressão I	0.3077	0.0040	0.5909

Tabela 29 – (Continuação) Métricas de Centralidade Calculadas para Ciência de Dados.

Disciplina	Grau	Intermediação	Proximidade
Mineração Estatística de Dados	0.3077	0.0000	0.5821
Gerenciamento de Dados e Informação	0.2564	0.0000	0.5652
Cálculo Diferencial e Integral II	0.2564	0.0013	0.5571
Gerenciamento e Administração de Bancos de Dados e Informação	0.2308	0.0000	0.5493
Aprendizado Profundo	0.2051	0.0013	0.5493
Álgebra Linear Computacional e Geometria Analítica Aplicada	0.2051	0.0027	0.5571
Inteligência Computacional e Aprendizado Profundo com Redes Neurais	0.2051	0.0013	0.5417
Cálculo Diferencial e Integral III	0.1026	0.0000	0.4756

B.2 Estudo de Caso: Administração

Tabela 30 – Métricas de Centralidade Calculadas para as Disciplinas da Matriz de Administração (T=0,6798).

Disciplina	Grau	Intermediação	Proximidade
Contabilidade Gerencial	0.8718	0.0216	0.8864
Direito Administrativo	0.8205	0.0162	0.8478
Governança e RSAE	0.7949	0.0094	0.8298
Administração Estratégica	0.7949	0.0162	0.8298
Administração Tributária	0.7692	0.0337	0.8125
Metodologia da Pesquisa em Administração	0.7436	0.0175	0.7959

Tabela 30 – (Continuação) Métricas de Centralidade Calculadas para Administração.

Disciplina	Grau	Intermediação	Proximidade
Orçamento Empresarial	0.7436	0.0351	0.7959
Administração da Qualidade	0.6923	0.0243	0.7647
Gestão de Pessoas II	0.6667	0.0202	0.7500
Obrigações Empresariais	0.6154	0.0270	0.7222
Gestão de Pessoas I	0.6154	0.0108	0.7222
Administração de Vendas	0.6154	0.0243	0.6964
Administração de Cargos e Salários	0.5641	0.0135	0.6724
Análise e Estratégia de Investimento	0.5641	0.0108	0.6842
Novos Negócios	0.5641	0.0148	0.6964
Introdução à Administração	0.5385	0.0054	0.6842
Economia das Empresas	0.5128	0.0391	0.6724
Contabilidade Geral A	0.5128	0.0175	0.6724
Comportamento Organizacional	0.5128	0.0405	0.6724
Finanças II	0.5128	0.0040	0.6500
Logística Empresarial	0.5128	0.0081	0.6610
Teoria Geral da Administração	0.4872	0.0081	0.6610
Processo Decisório	0.4872	0.0472	0.6610
Comunicação Empresarial	0.4615	0.0108	0.6500
Avaliação de Desempenho	0.4615	0.0108	0.6393
Administração Pública	0.4359	0.0000	0.6393
Administração no Contexto Brasileiro	0.4359	0.0013	0.6393
Ética e Filosofia para o Administrador	0.4103	0.0175	0.6290

Tabela 30 – (Continuação) Métricas de Centralidade Calculadas para Administração.

Disciplina	Grau	Intermediação	Proximidade
Administração de Marketing II	0.4103	0.0121	0.6094
Sistemas de Informações	0.4103	0.0067	0.6290
Gestão e Inovação Tecnológica	0.3846	0.0081	0.6190
Finanças I	0.3333	0.0000	0.5821
Estatística Aplicada à Administração	0.3077	0.0000	0.5735
Administração de Políticas Públicas	0.3077	0.0067	0.5735
Metodologia Científica e Filosofia da Ciência	0.2564	0.0067	0.5652
Administração de Operações	0.2308	0.0094	0.5652
Introdução à Finanças	0.2051	0.0000	0.5417
Administração de Marketing I	0.1282	0.0000	0.5132
Fundamentos de Sociologia	0.1026	0.0013	0.4643
Psicologia I	0.1026	0.0013	0.4588