Universidade Federal de São Carlos– UFSCar
Centro de Ciências Exatas e de Tecnologia– CCET
Departamento de Computação– DC
Programa de Pós-Graduação em Ciência da Computação– PPGCC

Renan Sanches Saraiva dos Santos

Visualização de Trajetórias Acadêmicas: uma abordagem em dados complexos

São Carlos
2025

Renan Sanches Saraiva dos Santos

Visualização de Trajetórias Acadêmicas: uma abordagem em dados complexos

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Centro de Ciências Exatas e de Tecnologia da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação.

Área de concentração: Banco de Dados

Orientador: Prof. Dr. Renato Bueno

São Carlos
2025

Este trabalho é dedicado à minha esposa Aline e a minha filha Luísa, meus maiores incentivos.

Agradecimentos

Expresso minha gratidão a todos que, direta ou indiretamente, contribuíram para a realização desta dissertação e para minha jornada no Mestrado.

Primeiramente, agradeço a Deus, por me conceder a vida e a saúde, e por iluminar meu caminho, permitindo-me vivenciar e concluir esta significativa e desafiadora etapa acadêmica.

Agradeço ao meu orientador, Prof. Dr. Renato Bueno. Sua vasta experiência, o conhecimento compartilhado em cada reunião e a paciência demonstrada foram cruciais para o desenvolvimento deste trabalho.

À minha amada esposa, Aline, o meu mais profundo reconhecimento. Seu apoio incondicional, sua capacidade de me dar forças nos momentos mais difíceis e sua persistência em nunca me deixar desistir foram o pilar que sustentou esta jornada. Seu amor, compreensão e sacrifício são inestimáveis para esta conquista.

À minha filha, Luísa, agradeço por me ensinar, a cada dia, a ser uma pessoa melhor. Sua alegria, inocência e o simples fato de ser seu pai são minha maior inspiração e motivação.

Aos meus pais, agradeço por sempre me incentivar a estudar e terem tratado a minha educação como prioridade.

Ao meu irmão, Rodolfo, um agradecimento especial pelo incentivo final.

Por fim, agradeço aos meus amigos da SVCOM da EESC-USP, pelo companheirismo, apoio e por me fazerem acreditar.

“Se você pode sonhar,
pode realizar.

Resumo

As instituições de ensino superior acumulam, ao longo do tempo, extensos volumes de dados acadêmicos, que constituem valiosas fontes de informação para compreender o comportamento de estudantes, turmas e cursos. Entretanto, a crescente complexidade e volume desses dados demandam abordagens inovadoras para extrair conhecimento relevante e apoiar a tomada de decisões estratégicas, como a otimização da oferta de disciplinas, a redução da evasão e a melhoria de práticas pedagógicas.

Esta dissertação propõe uma abordagem para modelar o percurso formativo dos estudantes como dados complexos em evolução temporal, representando-o por meio de trajetórias em um espaço métrico. A partir dessa modelagem, são exploradas técnicas de visualização interativas que permitem analisar a evolução acadêmica dos estudantes ao longo dos semestres, identificando padrões recorrentes, tendências e possíveis situações de risco, como evasão ou atraso na graduação.

Adicionalmente, o trabalho apresenta mecanismos para delimitar e refinar os dados a serem visualizados, utilizando filtros e estratégias de consulta, bem como técnicas de sumarização de trajetórias, que possibilitam a análise do comportamento médio de grupos personalizados de estudantes. A proposta foi validada por meio de um estudo de caso com dados reais de um curso de graduação, demonstrando sua aplicabilidade e potencial para subsidiar análises comparativas e apoiar processos de gestão educacional.

Palavras-chave: Visualização de dados educacionais. Trajetórias acadêmicas. Dados complexos. Espaço métrico .

Abstract

Higher education institutions continuously accumulate extensive volumes of academic data, which constitute valuable resources for understanding the behavior of students, cohorts, and academic programs. However, the growing complexity and scale of these datasets demand innovative approaches to extract meaningful insights and support strategic decision-making, such as optimizing course offerings, reducing dropout rates, and improving pedagogical practices.

This dissertation proposes an approach to model students' academic journeys as complex data undergoing temporal evolution, representing them as trajectories in a metric space. Building upon this model, interactive visualization techniques are explored to analyze the academic progression of students across semesters, enabling the identification of recurring patterns, trends, and potential risk situations, such as dropout or delayed graduation.

Additionally, the work introduces mechanisms to delimit and refine the data to be visualized through filtering and querying strategies, as well as trajectory summarization techniques that allow the analysis of average behaviors within customizable student groups. The proposed approach was validated through a case study involving real-world data from an undergraduate program, demonstrating its applicability and potential to support comparative analyses and inform educational management processes.

Keywords: Educational data visualization. Academic trajectories. Complex data. Trajectory summarization.

Lista de ilustrações

Figura 1 – Variáveis de uma visualização 32
Figura 2 – Exemplos de representações gráficas: gráfico de barras, histograma, boxplot, gráfico de pontos, nuvem de palavras e gráfico de pizza. 34
Figura 3 – Exemplos de representações gráficas: gráfico de dispersão e gráfico de bolhas. 34
Figura 4 – Exemplos de representações gráficas: gráfico linha e coordenadas paralelas. 35
Figura 5 – Exemplo de uma visualização com redução de dimensionalidade através do algoritmo MDS clássico. 40
Figura 6 – Fluxo geral do método proposto. 54
Figura 7 – Representação genérica de trajetória acadêmica em espaço 3D 64
Figura 8 – Interface geral do protótipo STVis 78
Figura 9 – Visualização de instâncias temporais individuais de cada estudante após concluir o primeiro semestre. 81
Figura 10 – Visualização ampliada e rotacionada das instâncias temporais individuais de cada estudante do banco de dados após concluir o primeiro semestre. Ao passar o *mouse* sobre a esfera, é possível identificar o estudante que está sendo representado 82
Figura 11 – Trajetória simples de um estudante ao longo de quatro semestres. . . . 83
Figura 12 – Trajetória simples de um estudante ao longo de quatro semestres exibindo somente as instancias temporais inicial e final. 83
Figura 13 – Trajetória média da classe Formado4. 84
Figura 14 – Visão geral da turma de 1996 nos três primeiros semestres do curso. . . 86
Figura 15 – Trajetória do estudante 100074 até o 3º semestre e as 5-NN tratórias . 87
Figura 16 – Trajetória do estudante 100074 até o 3º semestre e as 5-NN tratórias com *zoom* reduzido 88

Figura 17 – Trajetórias sumarizadas das classes Formado, Formado4, CAP<6 e Evadido 89
Figura 18 – Trajetórias sumarizadas das classes Formado, Formado4, Cap<6 e Evadido 90
Figura 19 – Análise comparativa da trajetória de 2 estudantes-alvo no início do quinto semestre com a tendência futura dos vizinhos e classes sumarizadas. 91
Figura 20 – Análise do cenário por outro ângulo. 92
Figura 21 – Trajetória completa dos estudantes 100175 e 100176 e das classes sumarizadas. 93
Figura 22 – Comparação entre resultados da consulta por similaridade métrica (à esquerda) e métrico-temporal (à direita) para o estudante 100509. . . . 95
Figura 23 – Comparação entre resultados da consulta por similaridade métrica (à esquerda) e métrico-temporal (à direita) para o estudante 100358. . . . 96
Figura 24 – Consulta híbrida: comparação entre o estudante 100058 (em amarelo) e estudantes reprovados em Construção de Algoritmos e Programação (CAP) com trajetórias similares. 97
Figura 25 – Consulta por similaridade com uma trajetória sumarizada como alvo. . 99
Figura 26 – Consulta por similaridade pontual: análise da trajetória do estudante 100510 no segundo semestre. 100

Lista de tabelas

Tabela 1 – Matriz de distâncias entre instâncias temporais de dois estudantes. . . 62
Tabela 2 – Coordenadas 3D obtidas pelo MDS para as instâncias. 62
Tabela 3 – Disciplinas obrigatórias e respectivas posições no vetor de características 75
Tabela 4 – Representação da instância temporal de um estudante em vetores de características 76
Tabela 5 – Representação da classe sumarizada `Formado4` no semestre 1 80

Lista de siglas

BCC Bacharelado em Ciência da Computação

CAP Construção de Algoritmos e Programação

L-MDS Landmark Multidimensional Scaling

MDS Multidimensional Scalling

PCA Principal Component Analysis

SNE *Stochastic Neighbor Embedding*

t-SNE t-Distributed stochastic neighbor embedding

UFSCar Universidade Federal de São Carlos

UMAP Uniform Manifold Approximation and Projection

Sumário

1 INTRODUÇÃO 23
1.1 Contexto 23
1.2 Motivação 25
1.3 Objetivos 26
1.4 Hipótese 27
1.5 Organização do Trabalho 27

2 FUNDAMENTAÇÃO TEÓRICA 29
2.1 Considerações Iniciais 29
2.2 Visualização de dados 29
2.2.1 Classificação das visualizações 30
2.2.2 Características de uma boa visualização 31
2.2.3 Visualizações convencionais 2D e 3D 32
2.2.4 Técnicas de interação 36
2.3 Técnicas de Projeção Multidimensional 38
2.3.1 *Multidimensional Scalling* 39
2.3.2 *Principal Component Analysis* 40
2.3.3 *t-Distributed stochastic neighbor embedding* 40
2.3.4 Considerações finais 41
2.4 Dados complexos 42
2.4.1 Espaço métrico 43
2.4.2 Dados complexos em evolução temporal 44
2.4.3 Espaço métrico-temporal 44
2.5 Visualização de Dados Complexos 45
2.5.1 Visualização de Dados Complexos Estáticos 46
2.5.2 Visualização de Dados Complexos em Evolução Temporal 46
2.5.3 Síntese e Aplicação na Presente Pesquisa 47

2.6 Trabalhos Relacionados 48
2.6.1 No domínio acadêmico 48
2.6.2 Em outros domínios 50
2.7 Considerações Finais 52

3 MÉTODO PARA VISUALIZAÇÃO DE TRAJETÓRIAS ACADÊMICAS 53
3.1 Considerações iniciais 53
3.2 Modelagem de Estudantes como Dados Complexos 54
3.3 Trajetórias Acadêmicas 55
3.4 Representação de estudantes no espaço métrico 56
3.4.1 Representação Métrico-Temporal de Instâncias de Estudantes 58
3.4.2 Projeção para um Espaço Vetorial 60
3.5 Grupos de Estudantes e Sumarização de Trajetórias 60
3.6 Aplicação de Técnica de Projeção Multidimensional para visualização 61
3.7 Visualização de Trajetórias Acadêmicas 63
3.7.1 Representação Visual das Trajetórias 63
3.7.2 Técnicas de Interação na Visualização 64
3.7.3 Flexibilidade da Abordagem Visual 65
3.8 Filtros e Estratégias de Consulta de Trajetórias 65
3.8.1 Consultas Baseadas em Atributos Relacionais 66
3.8.2 Consulta por Similaridade de Trajetória 66
3.8.3 Consulta por Similaridade Pontual de Trajetória 67
3.8.4 Consultas Híbridas 68
3.9 Considerações Finais 69

4 IMPLEMENTAÇÃO DO ESTUDO DE CASO 71
4.1 Conjunto de Dados Utilizado 71
4.2 Representação Utilizada neste Estudo 73
4.2.1 Exemplo de Instâncias Temporais 76
4.3 Protótipo STVis 76
4.4 Criação e Sumarização de Grupos de Estudantes 79
4.5 Visualizações Iniciais Geradas 81
4.5.1 Instância Temporal dos estudantes 81
4.5.2 Trajetória de Estudante 82
4.5.3 Trajetória Média de Classe Sumarizada 83
4.6 Considerações Finais 84

5 CENÁRIOS EXPLORATÓRIOS E RESULTADOS 85
5.1 Considerações iniciais 85
5.2 Cenário 1 — Análise Geral de uma Turma e Identificação de Estudante Atípico 86
5.3 Cenário 2 — Criação de Classes Personalizadas para Validação de Hipóteses 88
5.4 Cenário 3 — Consulta por Similaridade de Trajetória com Comparação a Classes Sumarizadas 90
5.5 Cenário 4 — Comparação entre Representações Métrica e Métrico-Temporal 93
5.6 Cenário 5 — Aplicação da Consulta Híbrida: Reprovação em CAP e Trajetórias Similares 96
5.7 Cenário 6: Identificação de Estudantes com um perfil específico 98
5.8 Cenário 7: Detecção de Risco Acadêmico em Semestre Específico100
5.9 Considerações finais 101

6 CONCLUSÕES 103
6.1 Contribuições do Trabalho 104
6.2 Limitações e Trabalhos Futuros 105

REFERÊNCIAS 107

Capítulo 1

Introdução

1.1 Contexto

Cada vez mais as organizações investem recursos para extrair conhecimento a partir de dados, com o objetivo de apoiar decisões estratégicas. As técnicas de análise de dados evoluíram, permitindo análises preditivas e prescritivas com maior precisão. O planejamento e as decisões táticas passam a ser baseados em informações disponibilizadas por sistemas de informação, por meio de relatórios, gráficos e *dashboards* (SILAHTAROĞLU; ALAYOGLU, 2016). Esses *dashboards* concentram diversas técnicas de visualização em uma mesma interface, tornando-se essenciais nesta era da informação ao proporcionarem acesso rápido e visual a dados provenientes de múltiplas fontes (ZHUANG; CONCANNON; MANLEY, 2022).

As técnicas de visualização de dados surgem, nesse contexto, como um recurso fundamental para tornar a interpretação de grandes volumes de dados mais simples e clara, proporcionando, em conjunto com a inteligência humana, percepções importantes, como a identificação de anomalias, o surgimento de padrões e correlações (CHEN et al., 2020). A representação visual das informações reduz o esforço cognitivo necessário para a execução de tarefas analíticas (KEIM et al., 2006).

No campo educacional, as instituições produzem uma quantidade massiva de dados a partir de registros acadêmicos, sistemas de aprendizagem online, avaliações e atividades administrativas. Entretanto, o potencial desses dados ainda é subutilizado, sobretudo para subsidiar decisões estratégicas e a formulação de políticas educacionais (RAUT; HAJARE, 2025; JAYARATNA et al., 2020). As visualizações podem contribuir com diversas intervenções, ao oferecerem orientação suficiente, destacarem informações críticas ou modificarem a forma como os usuários interagem com os dados, promovendo decisões

mais informadas e racionais (DIMARA; STASKO, 2022). Por exemplo, podem auxiliar educadores e administradores de cursos na análise das atividades e informações de uso dos estudantes, proporcionando uma visão geral da aprendizagem (ROMERO; VENTURA, 2010), bem como na identificação de preditores de desempenho, detecção precoce de estudantes em risco e no planejamento de intervenções personalizadas (RAUT; HAJARE, 2025).

Tradicionalmente, as representações visuais no contexto educacional concentram-se em gráficos de barras, linhas, tabelas, redes e *scatterplots*, usualmente organizados em *dashboards* (BODILY; VERBERT, 2017). Além dessas, outras técnicas como gráficos de coordenadas paralelas, árvores de decisão, *treemaps* e ferramentas de planejamento e reflexão também são utilizadas (BODILY; VERBERT, 2017). Entretanto, o contínuo crescimento da quantidade de dados exige abordagens novas ou inovadoras para compreender os padrões de valor embutidos nesses dados (SIEMENS; LONG, 2011).

Adicionalmente, há uma grande diversidade de sistemas e ambientes educacionais, como salas de aula tradicionais, ambientes de *e-learning*, sistemas de gestão da aprendizagem, sistemas de hipermídia adaptativa, tutores inteligentes, fóruns, redes sociais, ambientes de jogos educacionais, entre outros. Cada um desses ambientes gera dados distintos, com diferentes estruturas e significados, o que aumenta a complexidade e os desafios associados à sua análise integrada (ROMERO; VENTURA, 2010; ZHANG et al., 2022). Em cenários de alta complexidade, torna-se necessário que os analistas consigam interpretar vastos volumes de dados que se manifestam de múltiplas fontes, exibem diversas dimensões e evoluem continuamente no tempo. Essa capacidade de compreensão é essencial para embasar a tomada de decisões estratégicas, especialmente em momentos cruciais (KEIM et al., 2006).

A conversão de dados heterogêneos em visualizações que geram insights significativos é um desafio complexo que vai além da simples otimização da capacidade de processamento computacional. Existem diversas maneiras de representar os dados e a identificação da técnica mais adequada para um contexto específico nem sempre é evidente (KEIM et al., 2006).

No caso específico dos dados educacionais, a jornada de um estudante em uma universidade pode ser concebida como uma trajetória, que vai acumulando diferentes dados acadêmicos e de desempenho, associados a registros temporais (JAYARATNA et al., 2020). Dados orientados ao tempo são frequentemente multivariados, ou seja, compostos por diversas variáveis dependentes do tempo (AIGNER et al., 2011). Assim, é possível modelar o estado do estudante, ao longo de seu percurso acadêmico, como uma entidade complexa, composta por múltiplas variáveis que evoluem no tempo. Apresentar e interpretar visualmente um grande número de variáveis de forma simultânea pode ser uma tarefa complexa (AIGNER et al., 2011).

As visualizações temporais de dados de estudantes são úteis para revelar os compor-

tamentos de aprendizagem ao longo do tempo, maximizando a explicação dos hábitos e tendências desses estudantes (ZHANG et al., 2022). Contudo, a análise dessas trajetórias requer considerar a complexidade dos dados envolvidos. Dados complexos são geralmente representados como elementos em espaços de alta dimensão ou métricos, sendo que uma medida da complexidade dos dados é o número de atributos associados a cada instância: quanto mais atributos, maior a complexidade (HASUGIAN et al., 2023; BUENO et al., 2010; PAULOVICH et al., 2008).

O espaço métrico é um conceito geral que pode ser aplicado tanto a vetores (representações dimensionais), quanto a objetos como *strings* ou figuras, que não podem ser facilmente representados como vetores (BRIN, 1995). Nesses casos, a proximidade entre os objetos é definida unicamente por uma função de distância que satisfaça as propriedades de positividade, identidade, simetria e desigualdade triangular (CHEN et al., 2022). No entanto, a interpretação visual de espaços métricos é desafiadora e a abordagem mais comum para visualizar dados métricos consiste em mapeá-los em um espaço vetorial de duas ou três dimensões, preservando, tanto quanto possível, as distâncias entre os elementos (BUENO et al., 2010).

Nesse sentido, ao tratar os estudantes como dados complexos em evolução temporal, a análise de seus percursos formativos pode ser modelada por meio de trajetórias representadas em espaços métricos, possibilitando assim a interpretação de dados com alta complexidade. Essa representação possibilita visualizar, comparar e recuperar trajetórias semelhantes, facilitando a identificação de padrões e tendências ao longo do tempo. Assim, o método proposto neste trabalho busca viabilizar a geração de *insights* sobre o comportamento dos estudantes ao longo de sua trajetória, favorecendo a identificação de padrões de desempenho, tendências de progressão e potenciais situações de risco, como a evasão ou o atraso na conclusão do curso.

1.2 Motivação

A visualização de dados surgiu há muito tempo, antes mesmo de existirem equipamentos computacionais. Os primeiros gráficos estatísticos foram apresentados por William Playfair durante o século XVIII, para visualizar indicadores econômicos (TUFTE, 2001).

A área de visualização de dados tornou-se extremamente importante na medida em que há cada vez mais dados a serem analisados. A dificuldade e a limitação cognitiva do ser humano em analisar e comparar dados tabulares é parte importante da necessidade de técnicas e ferramentas de visualização .

No contexto acadêmico, há muitos dados, de diferentes sistemas, que podem ser utilizados para compreender a evolução dos estudantes ao longo do tempo a fim de melhorar a gestão educacional (RAUT; HAJARE, 2025; JAYARATNA et al., 2020). A análise do percurso formativo de estudantes pode revelar padrões de sucesso e insucesso, tendências

de evasão, trajetórias atípicas e características comuns entre grupos com desfechos semelhantes. Essas informações podem ser utilizadas por coordenadores de curso, gestores institucionais e pesquisadores da área de educação para subsidiar decisões pedagógicas, intervenções antecipadas e políticas de apoio ao estudante (ROMERO; VENTURA, 2010).

As abordagens tradicionais de análise educacional frequentemente tratam os estudantes como entidades estáticas, representadas por atributos isolados ou resumos estatísticos (ROMERO; VENTURA, 2010). Esse tipo de modelagem ignora a natureza temporal do percurso acadêmico dos estudantes e os vários dados que podem representá-lo ao longo de sua jornada estudantil. Desta maneira, para entender a dinamicidade do seu comportamento, torna-se necessário representá-lo como dados complexos em evolução temporal, de modo a capturar sua trajetória acadêmica ao longo dos semestres.

Assim, este trabalho é motivado pela necessidade de aplicar diferentes formas de representação de dados e explorar visualmente os dados dos estudantes, fornecendo técnicas que permitam analisar a evolução temporal a fim de apoiar análises comparativas, diagnósticos e tomadas de decisão baseadas em evidências.

1.3 Objetivos

Esta pesquisa teve como objetivo principal propor uma nova abordagem que favoreça a exploração visual de trajetórias acadêmicas, tratando os estudantes como dados complexos em evolução temporal. Dessa forma, busca-se possibilitar a geração de *insights* que apoiem a identificação de padrões, a tomada de decisões estratégicas e a antecipação de riscos relacionados ao comportamento dos estudantes ao longo de seu percurso acadêmico.

Os objetivos específicos desta pesquisa foram:

a) propor uma modelagem genérica do percurso acadêmico dos estudantes, onde possa ser utilizado dados heterogêneos para representá-lo como trajetórias temporais, permitindo a comparação e análise da evolução acadêmica;

b) propor e implementar técnicas de filtragem e consulta de trajetórias, de modo a possibilitar diferentes formas de exploração e análise dos dados acadêmicos;

c) utilizar técnicas de sumarização de trajetórias para gerar representações médias de grupos de estudantes com características comuns, favorecendo análises coletivas;

d) desenvolver um protótipo interativo que viabilize a exploração visual das trajetórias acadêmicas, com suporte a diferentes estratégias de análise e consulta;

e) realizar um estudo de caso com dados reais de estudantes de um curso de graduação, demonstrando a aplicabilidade e a efetividade da abordagem proposta na análise de dados educacionais;

1.4 Hipótese

Este trabalho parte da hipótese de que a visualização de trajetórias de estudantes modelados a partir de dados complexos em evolução temporal pode fornecer *insights* sobre o percurso acadêmico dos estudantes, favorecendo a identificação de padrões de desempenho, tendências de progressão e potenciais situações de risco, como evasão ou atraso na conclusão do curso. Acredita-se que esta abordagem, aliada a mecanismos de filtragem e sumarização de dados, seja capaz de compor diferentes cenários de análise do percurso formativo dos estudantes, ampliando o potencial de interpretação e tomada de decisão por parte de gestores, coordenadores e pesquisadores da área educacional.

1.5 Organização do Trabalho

O restante deste trabalho está organizado de modo a apresentar progressivamente a fundamentação teórica, a proposta metodológica, o estudo de caso realizado e os resultados obtidos.

O Capítulo 2 aborda a fundamentação teórica necessária para a compreensão desta pesquisa, discutindo conceitos relacionados à visualização de dados, dados complexos e em evolução temporal, representação métrica e métrico-temporal, técnicas de projeção multidimensional e trabalhos relacionados.

O Capítulo 3 descreve o método proposto para a modelagem, análise e visualização das trajetórias acadêmicas. Apresenta o processo de modelagem dos estudantes como dados complexos em evolução temporal, o cálculo das distâncias, a projeção para espaços multidimensionais e a definição das técnicas de filtragem de trajetórias e sumarização de grupos.

O Capítulo 4 detalha o estudo de caso efetuado sobre dados reais, aplicando o método proposto. É apresentado o conjunto de dados, a modelagem dos estudantes e o protótipo que viabiliza a aplicação das técnicas.

O Capítulo 5 apresenta uma série de cenários exploratórios elaborados para demonstrar a aplicabilidade e a eficácia da abordagem proposta. Cada cenário evidencia potenciais usos da técnica, suas vantagens e limitações na análise de trajetórias acadêmicas.

Por fim, o Capítulo 6 sintetiza as principais conclusões do trabalho, responde à hipótese formulada, discute as contribuições alcançadas, as limitações identificadas e aponta perspectivas para trabalhos futuros.

Capítulo 2

Fundamentação teórica

2.1 Considerações Iniciais

Este capítulo apresenta a fundamentação teórica que subsidia a proposta desenvolvida nesta dissertação. São abordados os principais conceitos e técnicas relacionadas à visualização de dados, dados complexos, representação métrica, além de aspectos sobre a modelagem e visualização de dados complexos. A construção deste referencial visa contextualizar o estudo e estabelecer os fundamentos conceituais e metodológicos que orientam as escolhas realizadas ao longo da pesquisa.

2.2 Visualização de dados

Este é o principal conceito que envolve este trabalho. A visualização de dados é abrangente — além da construção de gráficos de barras, pizza e linha — e está relacionada diretamente à transmissão clara e eficiente de informações a partir de um grande volume de dados, sem que as complexidades sejam notadas pelo leitor.

Esse campo de estudo se encontra na interseção entre as áreas de dados e de design. Ambas possuem muita relevância dentro deste conceito e precisam estar devidamente equilibradas para gerar bons resultados. A área de *design* diz respeito à estética, psicologia das cores, alinhamento de elementos visuais, dentre outros aspectos relacionados à parte gráfica e a área de dados, envolve pré-processamento, banco de dados e estatística. A exploração visual de dados é interdisciplinar e engloba áreas de sistemas de banco de dados, análise estatística, psicologia perceptiva e visualização científica (BÖHLEN et al., 2008). De acordo com Wilke (2019), “a visualização de dados é parte arte e parte ciência. O desafio é acertar na arte sem errar na ciência, e vice-versa”.

É importante elencar a utilidade e relevância de apresentar visualmente as informações. Iliinsky e Steele (2011) citam que a visualização se faz útil para examinar, entender e transmitir informações e descrevem as seguintes funções, objetivos e características desta área:

a) é possível transmitir rapidamente ao cérebro humano uma grande quantidade de informação através de sistemas visuais;

b) as visualizações aproveitam-se da capacidade instintiva do cérebro em identificar padrões e relacionamentos de dados visuais;

c) as visualizações instigam novos questionamentos, a necessidade de exploração de dados e a descoberta de subproblemas;

d) possuem grande capacidade de identificar exceções e tendências em um campo abrangente, com grande escala de dados.

O desenvolvimento de uma pesquisa na área de visualização pode ser orientado ao problema ou à técnica em si. Quando é orientado ao problema, o principal objetivo é trabalhar com usuários reais, para resolver problemas do mundo real, utilizando técnicas já existentes. Quando é orientado à técnica, o objetivo é desenvolver algoritmos ou técnicas novas e melhores sem estabelecer conexão com as necessidades dos usuários (SEDLMAIR; MEYER; MUNZNER, 2012).

Esta dissertação se situa na interseção dessas duas abordagens. Nosso objetivo é propor um novo método para a representação e análise visual de trajetórias de estudantes, que, embora utilize e adapte técnicas conhecidas da área, é intrinsecamente guiado pela necessidade de resolver problemas práticos e levantar soluções estratégicas para instituições de ensino.

2.2.1 Classificação das visualizações

As visualizações possuem diferentes classes e tipos. Iliinsky e Steele (2011) classifica as visualizações em dois termos que são utilizados em diferentes contextos: a visualização do tipo “Infográfico” e a visualização do tipo “Visualização de dados (DataViz)”.

Infográfico trata-se de uma ilustração mais elaborada, rica esteticamente, com desenhos feitos de forma manual e específica para os dados que estão no contexto da elaboração do infográfico. A criação é feita em softwares de ilustração e a alteração e atualização dos dados são muito trabalhosa.

Já no termo Visualização de Dados (DataViz) ou Visualização de Informações (InfoViz), são utilizados algoritmos e métodos computacionais para gerar as visualizações, os dados são facilmente inseridos e alterados, são flexíveis para serem utilizadas com novos conjuntos de dados e têm estética mais séria e acadêmica. Pelos objetivos e propósitos traçados para esta pesquisa, o foco será a visualização de dados (DataViz).

Kim, Zhu e Chen (2016) descreve uma divisão do significado dos termos visualização de informações de visualização de dados. Visualização de informação são representações visuais interativas de dados abstratos, geradas por computadores com o objetivo de facilitar a cognição humana. Já visualização de dados pode ser considerada um subdomínio do termo anterior, cujo objetivo é fazer com que as pessoas entendam o significado dos "dados" apresentando-os visualmente, através de gráficos estatísticos e outras representações, permitindo realizar comparações e identificar causalidade.

Ainda é possível, dentro destas classes definidas acima, classificar as visualizações com relação ao seu propósito. Iliinsky e Steele (2011) citam que existem três tipos com finalidades diferentes: a exploratória, explanatória e híbrida. É importante saber a diferença de abordagem entre elas para utilizá-las adequadamente.

A visualização exploratória se faz importante quando há uma grande quantidade de dados e não é possível identificar com clareza suas características e o que eles representam. Assim, a transformação dos dados em um modo visual possibilita identificar padrões, tendências, anomalias, subproblemas e entender seu comportamento. Esse tipo de visualização é utilizado no início de uma análise de dados, quando ainda há pouca informação conhecida.

A visualização explanatória é voltada para explicar informações que já são conhecidas, ou seja, o processo de análise já foi feito e já sabe-se o que realmente é relevante a ser apresentado. Este tipo de visualização geralmente é utilizado no final de um processo de análise de dados, quando a intenção é divulgar conhecimento adquirido acerca dos dados. É importante focar na ideia que se quer transmitir e eliminar as informações distrativas na elaboração destas visualizações.

A visualização híbrida trata-se do meio termo entre a exploratória e a explanatória. Neste tipo, já foi realizada a etapa de análise, com as informações já conhecidas e a visualização está pronta para o usuário final. O que difere a visualização híbrida da explanatória, é que ela permite a interação do usuário, normalmente por meio de uma interface gráfica. Há uma certa liberdade ao leitor para modificar os parâmetros, refazer consultas e realizar descobertas próprias sobre os dados.

Aliada às classificações, há diversas características que formam uma boa visualização. A seguir, são apresentadas as principais características conforme estabelecido por Tufte (2001).

2.2.2 Características de uma boa visualização

Os principais objetivos a serem alcançados por uma boa visualização gráfica são (TUFTE, 2001):

a) levar o leitor a pensar sobre a representação visual propriamente dita e não sobre detalhes técnicos de *design* ou tecnologia;

b) evitar distorções do que os dados estão dizendo;

c) tornar grandes conjuntos de dados coerentes e exibi-los em um espaço pequeno;

d) instigar a comparação entre diferentes grupos de dados;

e) revelar os dados em vários níveis de granularidade, desde uma visão abrangente até um detalhe minucioso;

f) estar com seu propósito bem definido;

g) ter integração entre as descrições estatísticas e verbais do conjunto de dados;

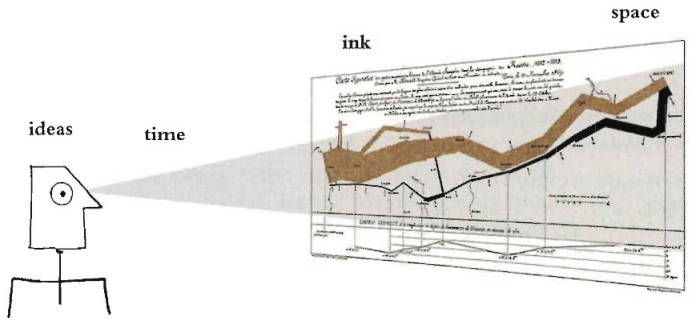
h) comunicar ideias complexas com clareza, eficiência e precisão;

i) fornecer rapidamente ao leitor o maior número de ideias possíveis no menor espaço, utilizando a menor quantidade de tinta;

j) dizer a verdade sobre os dados.

A Figura 1 mostra um leitor se deparando com uma representação visual e as variáveis presentes no processo de visualização: ideia, tempo, tinta e espaço.

Figura 1 – Variáveis de uma visualização



Fonte: (TUFTE, 2001)

2.2.3 Visualizações convencionais 2D e 3D

É possível visualizar dados de 1, 2, 3 ou mais dimensões e utilizar a técnica juntamente com a representação visual correta para auxiliar na interpretação do que os dados têm a dizer e gerar descobertas relevantes no domínio explorado. Há várias formas de classificar as técnicas de visualização de acordo com o seu tipo. Keim (2002) classifica desta forma:

a) Projeções 2D/3D convencionais - esta classe corresponde às técnicas triviais de visualização, como gráfico de barras, linhas, relações x-y (x-y-z), etc;

b) Projeções geométricas - corresponde as técnicas que transformam conjunto de dados multidimensionais para serem visualizados em 2 ou 3 dimensões. Classificam-se nesta categoria as técnicas de coordenadas paralelas, matriz de dispersão, etc;

c) Baseadas em ícones - são as técnicas que mapeiam dados multidimensionais em características de figuras ou ícones, como um rosto humano. A técnica mais conhecida desta classe é a *Chernoff faces* que utiliza aspectos do rosto humano para representar dados;

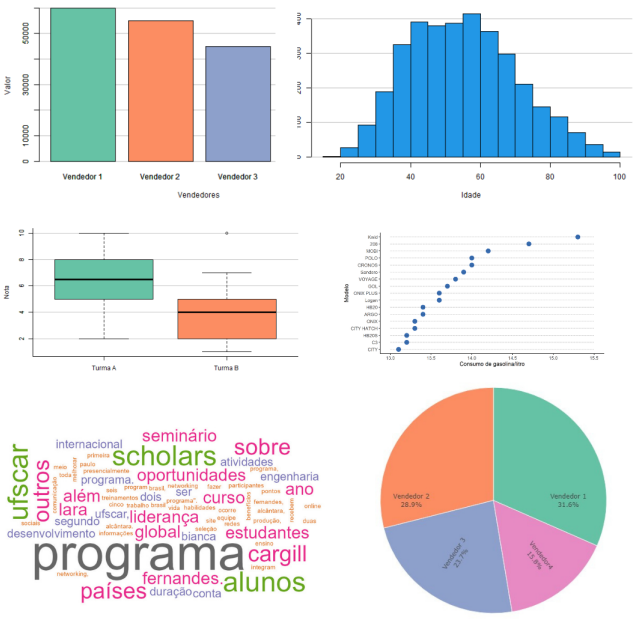
d) Orientadas a pixels - esta é uma técnica para visualizar dados multidimensionais que particiona a janela de visualização em várias sub-janelas, onde cada qual representa uma dimensão dos dados e cada valor é representado por um pixel colorido (KEIM, 2000). Permite projetar um grande conjunto de dados ao mesmo tempo, sendo que a limitação de valores a serem visualizados é de acordo com o tamanho da resolução do equipamento onde está sendo projetado, pois utiliza somente um pixel para representar cada valor e há também a limitação da percepção humana (KEIM; KRIEGEL; ANKERST, 1995). Classificam-se nessa categoria as técnicas de espiral, padrão recursivo e segmentos circulares (KEIM, 2000)

e) Técnicas hierárquicas: correspondem as técnicas que particionam os dados de forma hierárquica, onde um sistema de coordenadas é colocado dentro de outro sistema de coordenadas, criando níveis. Um exemplo é o mapa de árvores, que cria retângulos recursivamente representando a hierarquia dos dados.

Existem muitas formas de visualizações gráficas, como é possível verificar em Harris (1999). A seguir, serão apresentadas algumas técnicas de visualizações. Dentre as técnicas de projeção 2D/3D convencionais destacadas por Keim (2002), encontram-se os gráficos mais utilizados em análises exploratórias e comunicações visuais de dados. Estas técnicas são amplamente utilizadas para a representação e análise de dados em diferentes domínios, incluindo o educacional (ROMERO; VENTURA, 2010). Muitas são voltadas para a exploração inicial dos dados, oferecendo formas simples e intuitivas de sumarizar informações, identificar padrões gerais e comunicar resultados (ILIINSKY; STEELE, 2011; WILKE, 2019; HARRIS, 1999).

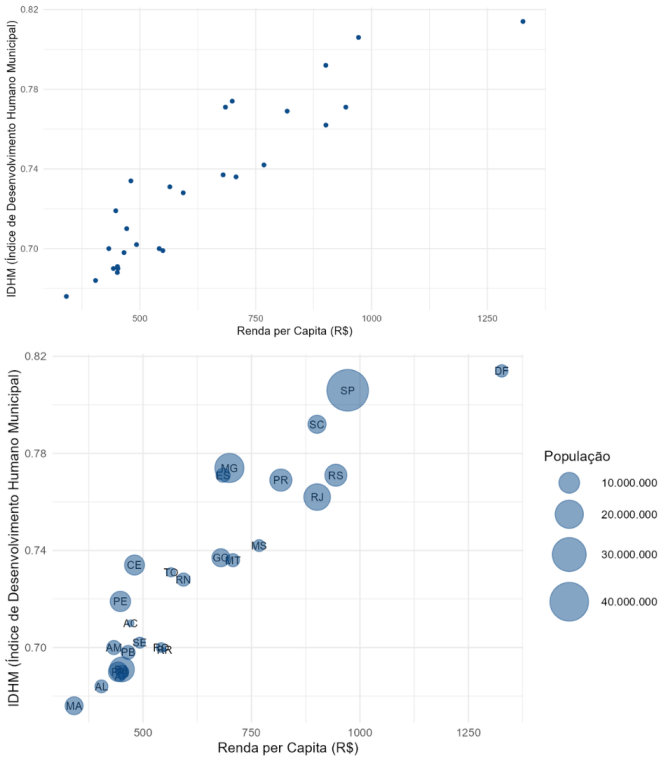
As Figuras 2, 3 e 4 agrupam exemplos de algumas dessas representações gráficas e as mais utilizadas em dashboards educacionais são: gráfico de barras, linhas, dispersão, pizza, nuvem de palavras e linha do tempo (BODILY; VERBERT, 2017).

Figura 2 – Exemplos de representações gráficas: gráfico de barras, histograma, boxplot, gráfico de pontos, nuvem de palavras e gráfico de pizza.



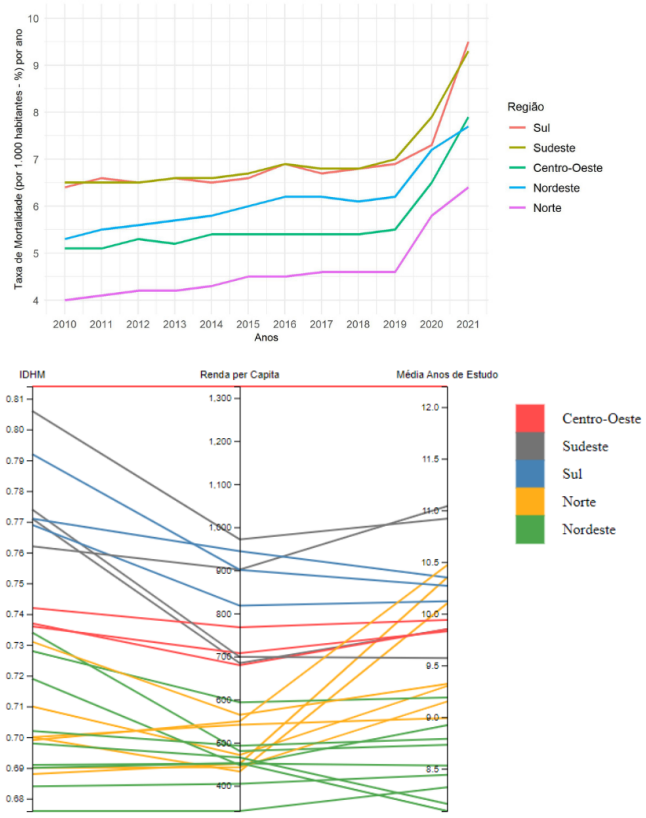
Fonte: Elaborada pelo autor

Figura 3 – Exemplos de representações gráficas: gráfico de dispersão e gráfico de bolhas.



Fonte: Elaborada pelo autor

Figura 4 – Exemplos de representações gráficas: gráfico linha e coordenadas paralelas.



Fonte: Elaborada pelo autor

Essas técnicas, apesar de tradicionais, podem ser utilizadas como recursos complementares na análise exploratória, fornecendo uma visão geral e sumarizada de aspectos pontuais dos dados acadêmicos.

Algumas dessas técnicas podem ser utilizadas para representar dados com mais de duas dimensões ou com dimensão temporal, incluindo o gráfico de dispersão, o gráfico de linhas e o gráfico de coordenadas paralelas.

O **gráfico de dispersão** permite explorar relações e correlações entre variáveis quantitativas, sendo fundamental para identificar tendências ou agrupamentos em dados multidimensionais. Quando estendido ao formato de *gráfico de bolhas*, possibilita a visualização simultânea de três variáveis, adicionando uma dimensão extra pelo tamanho dos círculos (HARRIS, 1999; WILKE, 2019).

O **gráfico de linhas** é especialmente adequado para a visualização de séries temporais, possibilitando a identificação de tendências, ciclos ou mudanças no comportamento dos dados ao longo do tempo. Esta técnica é frequentemente utilizada na análise de evolução do desempenho acadêmico, sendo apropriada para representar o rendimento dos estudantes ao longo de seus semestres (ILIINSKY; STEELE, 2011; WILKE, 2019).

O **gráfico de coordenadas paralelas** é uma técnica avançada para visualização de dados multidimensionais, permitindo a inspeção simultânea de diversas variáveis e

suas inter-relações. Sua aplicação é relevante para a comparação de perfis acadêmicos complexos, compostos por múltiplos atributos, como notas, frequência e outras métricas de desempenho (KEIM, 2002; LI, 2015).

2.2.4 Técnicas de interação

As visualizações por si só não garantem uma análise eficiente dos dados ao usuário, sendo necessárias técnicas de interação e distorção para que seja possível a manipulação e ajuste das visualizações geradas de acordo com as necessidades, ampliando a capacidade de análise dos dados. As técnicas de interação permitem criar diferentes cenários para um melhor entendimento dos dados, onde as imagens são alteradas e o usuário consegue observar o que acontece ao modificar a visualização (NASCIMENTO; FERREIRA, 2011). Portanto, o usuário consegue interagir diretamente e participar do processo de exploração visual. E, na exploração de grandes conjuntos de dados, é essencial aplicar técnicas que permitam a segmentação do conjunto total de elementos, de modo a concentrar a análise em subconjuntos de interesse. Essa segmentação pode ser realizada por meio da navegação direta sobre os dados ou pela especificação de uma consulta sobre atributos do subconjunto desejado (KEIM, 2002).

Há várias formas de classificar e apresentar as formas de interação com as visualizações. A seguir, são descritas algumas abordagens propostas por diferentes autores.

Shneiderman (1996) aborda as interações, estabelecendo sete tarefas que as visualizações devem apresentar:

a) Visão geral: uma visão geral de todo o conjunto de dados deve ser apresentada;

b) Zoom: deve permitir ampliar o detalhamento dos dados numa região de interesse, apresentando funções para ampliar e reduzir o zoom;

c) Filtro: permitir aplicar consultas dinâmicas sobre o conjunto de dados para descartar dados desnecessários;

d) Detalhes sob demanda: ao selecionar um item ou grupo, os detalhes devem aparecer quando necessários;

e) Relacionamento: formas de identificar relações entre os dados devem ser apresentadas, como ao selecionar uma característica, registros com a mesma característica devem ser destacados;

f) Histórico: no processo de exploração de dados, são realizadas várias tarefas e um histórico para permitir desfazer ou refazer ações deve ser apresentado;

g) Extração: deve ser permitido extrair porções de dados através de consultas e salvar o resultado em arquivo;

Já Keim (2002) classifica estas técnicas da seguinte forma:

a) Projeção Interativa: o objetivo desta técnica é modificar dinamicamente a projeção de dados multidimensionais para uma visualização bidimensional, por exemplo, a geração de uma série de gráficos de dispersão para demonstrar o relacionamento de várias dimensões.

b) Filtragem Interativa: é a técnica que permite ao usuário filtrar os dados e focar em subconjunto de dados de interesse. Pode ser feita através de uma consulta definindo valores de propriedades ou navegando pelo conjunto de dados diretamente.

c) Zoom Interativo: trata-se da técnica de apresentar os dados numa visão geral, ampla e em alto nível mas com a possibilidade de ir ampliando a visualização para que os dados sejam apresentados com mais detalhes ao aumentar o nível do zoom. Isso possibilita a exploração detalhada para uma região de interesse.

d) Distorção Interativa: é uma técnica que suporta a exploração de dados mantendo uma visão geral do conjunto enquanto se faz um detalhamento de partes dos dados. O objetivo é exibir porções de dados em alto nível de detalhes e manter uma visão original ao mesmo tempo.

e) Seleção e Ligação: o objetivo desta técnica de interação é combinar diversos métodos de visualização para aprimorar a exploração dos dados. Várias visualizações conectadas fornecem mais informações para detectar dependências e correlações do que uma visualização independente. Ao selecionar um ponto de interesse em uma visualização, as outras são atualizadas automaticamente com o dado selecionado

E Böhlen et al. (2008) cita que as principais técnicas de interação são:

a) Animação: consiste em ter uma visão geral e explorar de forma ampla os dados com diferentes níveis de granularidade.

b) Análise Condicional: é a possibilidade de dividir os dados em diferentes partes e analisá-los de forma independente, sob determinada condição. Os subconjuntos de dados são geralmente originados pela filtragem de valores de variáveis categóricas.

c) Equalização: trata-se do processo de ajuste na distribuição dos dados para uma melhor representatividade das classes, evitando que as visualizações sejam dominadas por padrões não interessantes.

d) Janelamento: consiste na ampliação de uma parte do conjunto de dados para investigação com mais detalhes, permitindo a visualização num nível menor. É um "zoom" inteligente nos dados.

Há outras técnicas de interação abordadas na literatura, como: baseadas em toque e fala (SAKTHEESWARAN; SRINIVASAN; STASKO, 2020), gestos no ar (*mid-air gestures*) para manipulação de visualização 3D (KOSTIC et al., 2024), *Stack Zoom* e *ChronoLenses* para explorar séries temporais (WALKER; BORGO; JONES, 2016), entre

outras. Portanto, existem diversas técnicas de interação que podem ser aplicadas às visualizações, sendo importante analisar o conjunto de dados e identificar as técnicas que realmente agregam valor e aumentam a capacidade de exploração dos dados pelo usuário no domínio da aplicação.

2.3 Técnicas de Projeção Multidimensional

Embora algumas representações gráficas como *heatmap*, *chernoffaces*, coordenadas paralelas e gráfico de bolhas, permitam visualizar dados com três ou até mais dimensões diretamente, conforme o número de dimensões aumenta, são necessárias técnicas especializadas em dados multidimensionais (com muitas características e atributos), onde busca-se encontrar similaridade e vizinhança dos dados.

O número de atributos de um objeto indica sua dimensão e os que são considerados multidimensionais possuem geralmente acima de três. Esta técnica, de acordo com Neves (2016), "mapeia as instâncias de dados em elementos gráficos num espaço geralmente bidimensional, preservando alguma informação sobre as relações de distância ou similaridade entre elas de forma a revelar as estruturas existentes".

O resultado da aplicação desta técnica geralmente é um conjunto de pontos apresentados num espaço plano, onde os pontos representam as instâncias de dados e a posição ou distância entre os pontos indicam a similaridade de cada instância (NEVES et al., 2015).

A denominação de projeção de dados multidimensionais se dá para a técnica de reduzir a dimensionalidade dos objetos, para 1, 2 ou 3 dimensões, mantendo as medidas de similaridade e dissimilaridade entre as instâncias (JOIA et al., 2011). Ou seja, um objeto de múltiplos atributos é transformado em um objeto mais simples, sem perder suas características. A redução da dimensionalidade permite que seja criada uma visualização simplificada (bi ou tridimensional) de objetos complexos, possibilitando a utilização da capacidade humana de interpretação visual, para reconhecimento de padrões e formação de grupos (PAULOVICH et al., 2008).

Existem dois tipos principais de redução da dimensionalidade, a seleção de atributos e a extração de atributos (PUDIL; HOVOVICOVA, 1998). Na seleção de atributos, são descartados os atributos que não descrevem os dados ou não são importantes para a análise que está sendo feita, diminuindo assim as dimensões do conjunto de dados. Já na extração de atributos, é realizado um processo de mapeamento, onde os dados são transformados por uma função que mantém as características principais dos dados originais, mas com redução na dimensionalidade para que possam ser projetados em um plano cartesiano. As técnicas de projeção multidimensional tratam a abordagem da extração de atributos (ORTIGOSSA; DIAS; NASCIMENTO, 2022).

As técnicas de projeção podem ser classificadas quanto à sua natureza, podendo ser local ou global (ORTIGOSSA; DIAS; NASCIMENTO, 2022). As técnicas locais buscam

preservar relacionamentos de vizinhança, ou seja, a proximidade entre as instâncias vizinhas é melhor retratada. É indicada quando o objetivo for separar grupos e identificá-los visualmente. Já as técnicas globais buscam preservar a média das distâncias entre todas as instâncias no espaço transformado (FADEL et al., 2015).

Existem várias técnicas para projetar dados multidimensionais, sendo um assunto muito estudado na literatura. A técnica a ser utilizada depende do tipo e tamanho do conjunto de dados e do resultado que se está buscando. As mais conhecidas e tradicionais são (SMELSER; MILLER; KOBOUROV, 2024; FLEXA et al., 2021; JOIA et al., 2011) : Principal Component Analysis (PCA), Multidimensional Scalling (MDS) e t-Distributed stochastic neighbor embedding (t-SNE).

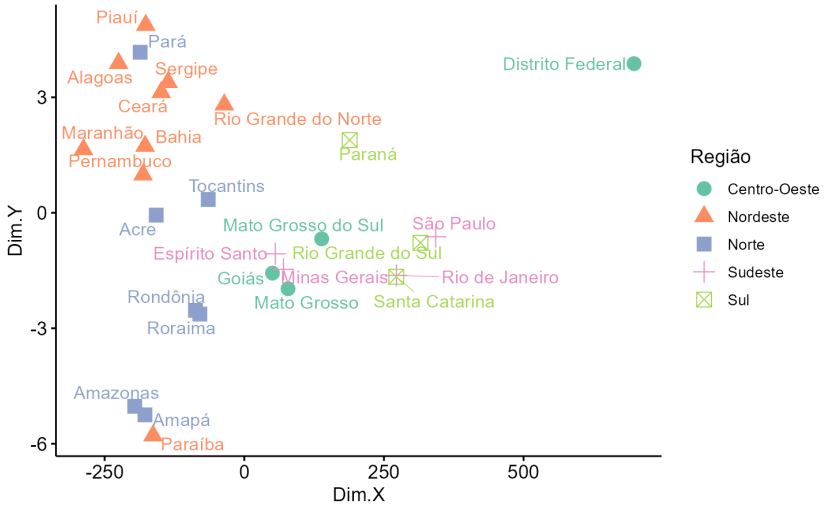
2.3.1 *Multidimensional Scalling*

Trata-se de uma técnica bem conhecida que teve sua origem em meados do século XX e é considerada uma família de técnicas de projeção multidimensional (JOIA et al., 2011). Paulovich et al. (2008) define como uma "classe de técnicas destinadas a mapear instâncias pertencentes a um espaço m-dimensional em instâncias em um espaço d-dimensional ($d \leq m$), esforçando-se para manter algumas relações de distância".

O método mais conhecido desta família de técnicas é o *Classic Scalling*, onde é feita a decomposição espectral da matriz gerada das distâncias dos objetos multidimensionais. Possui boa precisão na preservação da distância global, mas é muito custosa computacionalmente, de ordem cúbica O(n^3) não sendo indicada para grande volume de dados (ORTIGOSSA; DIAS; NASCIMENTO, 2022).

Muitos métodos foram desenvolvidos para aprimorar o desempenho da MDS Clássica, tentando preservar a precisão da distância, como o Landmark Multidimensional Scaling (L-MDS), Pivot-MDS, Pekalska (NEVES et al., 2015). Estes métodos empregam estratégias similares, iniciando com um conjunto menor de instâncias e depois interpolando as instâncias restantes na projeção final. A Figura 5 demonstra um gráfico de dispersão de duas dimensões originado de dados multidimensionais. Para o exemplo, foram utilizados os dados dos estados brasileiros relacionados à renda per capita, IDHM, média de anos de estudo, taxa de analfabetismo, IDEB e aplicado o algoritmo de Multidimensional Scalling clássico para reduzir a dimensionalidade, permitindo a projeção num espaço bidimensional. É possível realizar uma análise comparativa na proximidade entre os estados de acordo com os indicadores analisados e percebe-se que a maior parte está próxima a estados da mesma região.

Figura 5 – Exemplo de uma visualização com redução de dimensionalidade através do algoritmo MDS clássico.



Fonte: Dados do AtlasBR (2021) e elaborada pelo autor

2.3.2 *Principal Component Analysis*

A *Principal Component Analysis* é uma técnica de projeção do tipo linear e foi uma das primeiras utilizadas para redução de dimensionalidade, surgida no início do século XX. O seu funcionamento se dá através da busca por componentes principais, que são as combinações de retas ortogonais que melhor representam a variabilidade dos dados (NEVES, 2016). A busca é um processo matemático que inicializa-se com o cálculo da matriz de covariância entre os atributos, seguida de uma decomposição espectral para encontrar autovetores. A quantidade de componentes principais gerados pode ser igual ao número de dimensões, mas os primeiros componentes carregam a parte mais significativa da variação dos dados (ORTIGOSSA; DIAS; NASCIMENTO, 2022). A PCA preserva mais características globais do que locais e as vantagens desta técnica são: capacidade de encontrar tendências e padrões, eliminar ruídos e produzir a variabilidade dos dados em poucas dimensões. Já como desvantagens, tem-se o fato de não representar relações não lineares entre as dimensões e, como em visualizações gráficas são utilizados somente os primeiros componentes principais gerados, um conjunto de dados com vários grupos que possuem variâncias distintas não é bem representado (ORTIGOSSA; DIAS; NASCIMENTO, 2022).

2.3.3 *t-Distributed stochastic neighbor embedding*

Esta técnica foi proposta por Maaten e Hinton (2008) para visualizar dados de alta dimensão em espaços bi e tridimensionais e trata-se de um aprimoramento da técnica base *Stochastic Neighbor Embedding* (SNE) através da forma inteligente que lida com o

problema do congestionamento utilizando a distribuição t (HUANG et al., 2023). As técnicas de incorporação estocásticas de vizinhos calculam uma matriz de similaridade N x N em ambos os espaços, original e de baixa dimensão, onde a semelhança entre os pares de instâncias forma uma distribuição de probabilidade. Pares de instâncias similares possuem uma alta probabilidade dentro da distribuição e pares com pouca similaridade possuem uma baixa probabilidade (MAATEN, 2014). A t-SNE é capaz de representar de maneira satisfatória a estrutura local dos dados, enquanto ainda demonstra a presença de clusters em várias escalas através da preservação das estruturas globais. Contudo, possui complexidade computacional e de memória quadrática O(n^2) , tornando inviável aplicá-la a um grande conjunto de dados (MAATEN; HINTON, 2008). O bom resultado deste algoritmo é suscetível ao ajuste fino de pelo menos três hiperparâmetros pertencentes à função de custo (perplexidade) e de otimização (exagero e taxa de aprendizagem) (FLEXA et al., 2021).

2.3.4 Considerações finais

Existem inúmeras técnicas tradicionais e de projeção multidimensional para representar dados visualmente. As técnicas tradicionais limitam-se a utilizar uma pequena quantidade de atributos para representar as instâncias. Quando os dados possuem dezenas ou centenas de atributos — ou ainda, quando se trata de dados complexos —, as técnicas tradicionais passam a não ser suficientes para comparar as instâncias, necessitando da utilização de técnicas sofisticadas que utilizam algoritmos para transformar os dados e representá-los visualmente em projeções bi e tridimensionais.

No domínio acadêmico, essas técnicas têm sido utilizadas para identificar agrupamentos e prever o desempenho de estudantes (ALHAZMI; SHENEAMER, 2023), bem como para análises de evasão (HEGDE; PRAGEETH, 2018), entre outras aplicações. A análise do comportamento acadêmico requer o processamento de múltiplos atributos heterogêneos, como dados demográficos, educacionais, comportamentais e sociais — incluindo idade, gênero, status socioeconômico, desempenho acadêmico anterior, notas em avaliações, frequência às aulas, interações em plataformas de aprendizagem e redes de relacionamento entre colegas. Combinar e processar essas variáveis diversas é um desafio significativo que demanda o uso de técnicas avançadas de engenharia de atributos e redução de dimensionalidade (VERMA; SRIVASTAVA; BHARTI, 2024).

Nesse contexto, a compilação de todas essas informações para fins de análise visual exige a transformação e projeção desses dados para espaços com dimensões adequadas à interpretação humana. Considerando que a proposta desta dissertação busca estabelecer uma modelagem genérica para representar estudantes, permitindo sua caracterização a partir de diversos atributos ao longo de sua trajetória acadêmica, o emprego de técnicas de projeção multidimensional é necessário. Essas técnicas viabilizam a compilação e representação de dados complexos em espaços bi ou tridimensionais, adequados à visualização

e à exploração analítica das trajetórias acadêmicas.

A seção 2.4 detalha o conceito de dados complexos (conceito central na modelagem genérica dos estudantes proposta por este trabalho) abordando suas principais características, a utilização de espaço métrico e métrico-temporal e as técnicas adequadas para sua análise e representação.

2.4 Dados complexos

Dados complexos (imagens, áudios, vídeos, séries temporais e outros) possuem características diferentes de dados convencionais. Dados tradicionais ou convencionais são representados por números, textos curtos e datas, que podem ser ordenados. Já os dados complexos ou não convencionais não possuem relação de ordem total entre os elementos, não sendo possível utilizar operadores de comparação relacionais ($<, >, \leq, \geq$) para recuperá-los numa consulta (CHÁVEZ et al., 2001). Assim, são necessárias formas de consultas que estabelecem o conceito de proximidade ou similaridade para recuperar estes dados. A similaridade é baseada numa função de distância que é calculada entre os elementos, e estes são comumente representados em um espaço métrico. No espaço métrico, estão disponíveis apenas informações de distância entre os elementos (CHÁVEZ et al., 2001).

A complexidade de um conjunto de dados pode ser compreendida pela sua dimensionalidade, definida pelo número de atributos que caracterizam cada unidade de informação. Ilustrando com um exemplo de censo demográfico, cada registro individual — contendo variáveis como idade, sexo, escolaridade, ocupação e renda — pode ser concebido como um ponto em um espaço de (m) dimensões, onde (m) corresponde ao total de atributos observados (PAULOVICH et al., 2008).

A consulta de dados complexos pode ser viabilizada através da extração de um conjunto de características que os representa. O conjunto de características extraídas, denominado vetor de características, é empregado como representação dos dados originais para realizar consultas por similaridade (BöHM; BERCHTOLD; KEIM, 2001).

Uma consulta por similaridade busca recuperar objetos de um banco de dados que sejam considerados "semelhantes" a um objeto de referência. A (des)similaridade entre esses objetos é computada utilizando-se uma função de distância métrica (d) (CIACCIA; PATELLA; ZEZULA, 1997). As principais abordagens de consultas por similaridade são a consulta por abrangência (*range query*) e a consulta por vizinhos mais próximos (*k-Nearest Neighbor query*) (BöHM; BERCHTOLD; KEIM, 2001).

Portanto, dados complexos são representados em espaços métricos para permitir buscas por aproximação baseadas em uma função de distância (NAVARRO, 1999). A seguir, serão apresentados os conceitos relacionados ao espaço métrico e às métricas de similaridade, bem como as abordagens para tratar dados em evolução temporal, o espaço

métrico-temporal e por fim, as técnicas de visualização de dados complexos.

2.4.1 Espaço métrico

Um espaço métrico representa um conceito matemático abrangente, onde a noção de proximidade entre quaisquer dois objetos é estabelecida exclusivamente por uma função de distância. Essa função, para ser considerada uma métrica válida, deve aderir a quatro postulados fundamentais (CHEN et al., 2022):

- **Não-negatividade:** $d(x, y) \geq 0$
- **Identidade :** $d(x, y) = 0 \iff x = y$
- **Simetria:** $d(x, y) = d(y, x)$
- **Desigualdade triangular:** $d(x, z) \leq d(x, y) + d(y, z)$

Espaço métrico é um conceito genérico e permite sua aplicação não apenas a entidades facilmente representáveis como vetores (onde a distância Euclidiana é um exemplo comum), mas também a estruturas mais complexas, como sequências de caracteres (strings) ou grafos, as quais não se traduzem diretamente em representações vetoriais. Dentro de tal espaço, a tarefa de identificar "vizinhos próximos" consiste em selecionar, de um conjunto de dados finito, aqueles elementos que se encontram a uma distância especificada de um ponto de referência (BRIN, 1995). Quanto menor a distância entre dois objetos, mais similares eles são (NAVARRO, 1999).

Formalmente, um espaço métrico é definido na literatura como um par (M, d), onde M é um conjunto não vazio e $d : D \times D \rightarrow R$ é uma métrica em M (LIMA, 1977) que respeita as propriedades da simetria, identidade, não-negatividade e desigualdade triangular (CHEN et al., 2022).

Uma métrica apropriada influencia diretamente no desempenho e na qualidade das buscas durante a recuperação de dados baseada em similaridade. É ela quem determina o quão próximos dois objetos estão, e essa escolha depende do tipo de dado e do contexto em que estão inseridos. As principais métricas utilizadas em espaços vetoriais são a Euclidiana, *Manhattan, Minkowski e Chebyshev* (ARANTES, 2005).

A distância Euclidiana calcula a raiz quadrada da soma dos quadrados das diferenças entre os atributos de dois objetos e é utilizada geralmente quando os dados são homogêneos e contínuos. Por sua vez, a distância *Manhattan* realiza a soma dos valores absolutos das diferenças e é adequada quando os deslocamentos ocorrem em trajetos ortogonais. A distância de *Minkowski* é uma generalização das anteriores e é definida como a raiz de ordem p da soma das diferenças elevadas a p; quando $p = 1$ equivale à Manhattan, e quando $p = 2$, à Euclidiana. Já a distância de *Chebyshev* considera a maior diferença absoluta entre os atributos comparados, sendo útil em situações onde a variação máxima

entre as dimensões é o fator mais relevante para a análise (ARANTES, 2005; DIAS et al., 2017).

Apesar de eficaz na comparação entre objetos, o espaço métrico não contempla a dimensão temporal, essencial em dados que evoluem ao longo do tempo. A seguir, será abordado o conceito de dados complexos em evolução temporal, que aborda formas de representar e analisar variações no tempo em dados métricos.

2.4.2 Dados complexos em evolução temporal

Em muitos domínios que utilizam dados complexos, como medicina, finanças, meteorologia e ciências ambientais, a informação de tempo é essencial para detectar anomalias, prever comportamentos futuros e tomar decisões baseadas na evolução temporal. No entanto, representar dados complexos com a dimensão do tempo apresenta desafios, especialmente quando são dados métricos, onde estão disponíveis apenas as distâncias entre os elementos (BUENO et al., 2009). Segundo Ward, Grinstein e Keim (2015), esses desafios estão relacionados diretamente com a representação eficiente das mudanças, o tratamento da alta dimensionalidade e a identificação de padrões temporais, como tendências, ciclos e anomalias.

Além disso, quando queremos comparar diferentes conjuntos de dados que evoluem no tempo, uma forma é utilizar métricas e estruturas de dados que levem em conta essa dimensão temporal. No caso de dados multidimensionais, é possível acrescentar uma dimensão específica para representar a informação temporal (BUENO et al., 2010).

Uma outra abordagem é quando o conjunto de dados é inerentemente métrico e é necessário representar a dimensão temporal. Para isso, foi estabelecido o conceito de espaço métrico-temporal, como uma extensão do espaço métrico, que incorpora a dimensão do tempo na análise de similaridade, fornecendo modelos capazes de representar variações temporais e espaciais ao mesmo tempo (BUENO et al., 2009).

2.4.3 Espaço métrico-temporal

O espaço métrico-temporal, conforme proposto por Bueno et al. (2009), é uma extensão conceitual do espaço métrico, cuja finalidade é integrar simultaneamente aspectos de similaridade métrica e proximidade temporal entre objetos. No espaço métrico, a proximidade entre duas instâncias é definida exclusivamente com base em uma função de distância d_m, que opera sobre o conjunto de características ou atributos dos objetos. Embora eficiente para muitas aplicações, tal abordagem desconsidera a informação temporal que frequentemente está associada aos dados.

Para suprir essa limitação, o espaço métrico-temporal adiciona uma componente temporal explícita ao processo de comparação entre objetos. Assim, além da função de distância métrica tradicional, incorpora-se uma função de distância temporal d_t, responsável

por mensurar a separação entre os momentos ou ciclos temporais nos quais as instâncias ocorrem. A combinação desses dois componentes resulta em uma função composta que permite avaliar a dissimilaridade global entre dois objetos, considerando simultaneamente a diferença métrica e a diferença temporal.

Formalmente, a função de distância métrico-temporal é definida a seguir:

$$\Delta(u_i, u_j) = (w_s \cdot \delta_s(s_i, s_j)) + (w_t \cdot \delta_t(t_i, t_j)) \tag{1}$$

O termo $\delta_s(s_i, s_j)$ representa a distância métrica entre as estruturas ou atributos das instâncias, enquanto $\delta_t(t_i, t_j)$ corresponde à diferença temporal entre os momentos associados às mesmas. Os pesos w_s e w_t definem a importância relativa desses dois componentes na distância final (BUENO et al., 2009).

O cálculo dos pesos dos componentes métrico (w_s) e temporal (w_t) na definição do espaço métrico-temporal visa ajustar a contribuição relativa de cada um na função de similaridade composta, evitando distorções provocadas por redundâncias ou correlações internas. Para isso, normaliza-se os componentes métrico e temporal de acordo com suas respectivas dimensões intrínsecas (BUENO et al., 2010).

Essa definição confere ao espaço métrico-temporal flexibilidade na relação de similaridade entre objetos representados no espaço métrico, permitindo sua aplicação em diversos domínios em que tanto a estrutura interna dos objetos quanto a sua evolução temporal devem ser consideradas para análises mais precisas e contextualizadas.

2.5 Visualização de Dados Complexos

A visualização de dados complexos é uma abordagem essencial para apoiar a análise exploratória e a interpretação de dados caracterizados por múltiplos atributos, estruturas não triviais ou relações não lineares. Dados complexos são frequentemente representados como elementos em espaços de alta dimensão ou métricos, onde não é trivial obter uma representação visual direta (BUENO et al., 2010).

Em ambientes onde os dados podem ser representados como vetores numéricos de múltiplos atributos, técnicas de redução de dimensionalidade são amplamente utilizadas para projetar esses dados em espaços de duas ou três dimensões, possibilitando a interpretação humana. Métodos como o *Principal Component Analysis*,*t-Distributed stochastic neighbor embedding*, *Multidimensional Scalling* e *Uniform Manifold Approximation and Projection* são comumente empregados nesse contexto (SMELSER; MILLER; KOBOUROV, 2024). Esses algoritmos utilizam um conjunto de dados de alta dimensão como entrada ou uma matriz de distância entre suas instâncias (MILLER et al., 2024). Após essa etapa, técnicas gráficas baseadas em eixos (gráficos de dispersão e gráficos de coordenadas paralelas), glifos, *pixels* e representações hierárquicas, juntamente com animação e percepção, podem ser utilizadas para construir a visualização (LIU et al., 2017).

Contudo, quando os dados são representados em espaços métricos — ou seja, quando a única informação disponível entre os elementos é a sua proximidade relativa, definida por uma função de distância que obedece aos axiomas de positividade, simetria, identidade e desigualdade triangular (CHEN et al., 2022) —, a visualização torna-se ainda mais desafiadora. Nesses casos, não há necessariamente uma representação espacial direta dos objetos, pois eles podem não ter uma forma vetorial ou geométrica conhecida (BUENO et al., 2010).

A maneira usual de visualizar dados puramente métricos é mapear o espaço métrico em um espaço vetorial, geralmente de duas ou três dimensões, de modo que as distâncias entre os elementos sejam preservadas o máximo possível. Esse mapeamento é realizado através de técnicas específicas, como o MDS clássico, que busca minimizar o erro entre as distâncias reais no espaço métrico e as distâncias resultantes no espaço projetado (BUENO et al., 2010).

2.5.1 Visualização de Dados Complexos Estáticos

A maior parte das técnicas de visualização de dados complexos disponíveis na literatura trata de representações estáticas (LIU et al., 2017), isto é, em que não há variação explícita das instâncias ao longo do tempo. Nesses casos, o objetivo é explorar a estrutura dos dados, identificar agrupamentos, *outliers* ou padrões por meio da análise espacial resultante após a projeção, como feito em Barioni et al. (2002), Dias et al. (2017) e Xia et al. (2023).

A busca por padrões interessantes na relação entre dados complexos é comumente exploratória, visto que, em geral, características ou padrões adequados para fornecer informações úteis são inicialmente desconhecidos. Durante esse processo, um problema crucial é como identificar se uma relação descoberta é útil e como interpretar os resultados. Técnicas de visualização de dados têm sido empregadas com sucesso para auxiliar nessa análise (BUENO et al., 2010).

Exemplos incluem a análise e visualização de dados de documentos (PAULOVICH et al., 2008), de cidades inteligentes (LOPES; NETO; MARTINS, 2020), imagens (PAIVA; BUENO, 2019) entre outros.

2.5.2 Visualização de Dados Complexos em Evolução Temporal

Independentemente do domínio de aplicação, muitas vezes é necessário analisar dados complexos que evoluem ao longo do tempo, como séries temporais multivariadas ou trajetórias de entidades cujas características mudam em função do tempo.

De modo geral, a maneira mais comum de realizar a visualização de dados complexos em evolução temporal é analisar a dimensão temporal separadamente como uma dimensão

adicional (LIU et al., 2017) onde a análise, então, é guiada pela observação de como o objeto se comporta ao longo da linha do tempo (AIGNER et al., 2011).

Quando são dados métricos, é possível mapear o componente métrico dos dados complexos para um espaço bi ou tridimensional e observar as variações nas posições espaciais que os elementos sofrem ao longo do tempo. Uma outra abordagem, quando é necessário aproximar elementos que tenham informações temporais semelhantes, é incorporar os componentes métricos e temporais em um espaço homogêneo e, assim, mapear esse espaço de informações métricas e temporais para um espaço dimensional com a finalidade de visualização. É possível conectar esses elementos mapeados com segmentos de linhas para indicar a movimentação temporal (BUENO et al., 2010).

A interpretação dessas mudanças temporais é fundamental para diversos domínios. Por exemplo, em estudos médicos, o estado de um paciente pode ser composto por uma sequência de exames realizados ao longo de um período, variando de acordo com diagnósticos e prognósticos. Assim, o conjunto de informações a cada momento define um estado, e a trajetória médica do paciente pode ser representada como uma sequência de elementos em um espaço 3D (BUENO et al., 2010).

2.5.3 Síntese e Aplicação na Presente Pesquisa

Tradicionalmente, as representações visuais no ambiente educacional concentram-se em gráficos de barras, linhas, tabelas e scatterplots, muitas vezes organizados em dashboards (BODILY; VERBERT, 2017). Contudo, esses formatos geralmente tratam os dados de forma estática e univariada, não contemplando a evolução temporal do estudante. A jornada de um estudante em uma universidade é similar a uma trajetória que agrega diferentes informações acadêmicas e comportamentais ao longo do tempo (JAYARATNA et al., 2020).

O presente trabalho adota essa perspectiva para modelar e visualizar as trajetórias acadêmicas de estudantes como dados complexos em evolução temporal. Ao longo do percurso acadêmico, o estudante acumula variações que modificam seu estado, como uma entidade dinâmica. Muitos fatores podem ser considerados para representar o estudante ao longo do tempo: notas, frequência, desempenho em atividades extracurriculares, interação em um sistema de aprendizagem *online*, entre outros. Além disso, informações estáticas, como gênero, tipo de escola no ensino médio e forma de ingresso na universidade, podem ser incorporadas para compor uma representação mais rica.

Dadas essas possibilidades, o estudante pode ser modelado como uma instância complexa, cujas características evoluem temporalmente ao longo do percurso formativo. Assim, sua trajetória pode ser representada através de técnicas já conhecidas para visualização de dados complexos com evolução temporal. A comparação de trajetórias é realizada com base no conceito de similaridade, utilizando uma função de distância apropriada.

A proposta deste trabalho, portanto, se fundamenta na incorporação dessas técnicas ao domínio da análise de trajetórias acadêmicas, buscando possibilitar a geração de *insights* sobre o comportamento dos estudantes ao longo do tempo, favorecendo a identificação de padrões de desempenho, tendências de progressão e potenciais situações de risco, como evasão ou atraso na conclusão do curso.

2.6 Trabalhos Relacionados

A visualização e análise de dados complexos têm sido amplamente exploradas em diversos domínios, com foco na identificação de padrões e tendências em grandes volumes de dados.

No domínio acadêmico, embora existam iniciativas que utilizam visualizações para apoiar a gestão acadêmica e o acompanhamento do desempenho estudantil, grande parte das abordagens se baseia em análises estáticas e agregadas, utilizando filtros e indicadores pontuais. Alguns trabalhos analisam o estudante em evolução temporal, levando em consideração apenas as médias das notas ao longo do tempo, sem considerar outros dados ou relações de similaridade entre estudantes, o que limita a capacidade de capturar padrões de progressão relacionados à trajetória acadêmica. Embora, em outros campos, a representação e a comparação de trajetórias sejam consolidadas, no domínio educacional, ainda é inicial a aplicação de técnicas que tratem os estudantes como entidades complexas em evolução temporal.

Esta seção apresenta e discute os principais trabalhos que abordam a visualização de dados educacionais e de dados complexos. Além disso, são destacadas propostas de outros domínios que servem de referência metodológica para esta dissertação, especialmente no que diz respeito ao uso de representações métricas e técnicas de consulta por similaridade.

2.6.1 No domínio acadêmico

Diversos estudos têm buscado aplicar técnicas de visualização e análise de dados educacionais para apoiar a gestão acadêmica e o acompanhamento do desempenho estudantil. Contudo, grande parte dessas abordagens permanece limitada a análises agregadas e estáticas, sem considerar a natureza dinâmica das trajetórias acadêmicas.

No trabalho de Kunzayila (2022) foram apresentadas visualizações para análise de dados educacionais no contexto acadêmico, financeiro e produção científica, para todos os cursos da universidade de Kimpa Vita (UNIKIV) de Angola. Foram geradas diversas visualizações através da ferramenta PowerBi, como: desempenho geral dos alunos, taxa de alunos formados, taxa de alunos reprovados por turma, ano letivo e disciplina, taxa de alunos aprovados por turma, entre outras. As visualizações propostas são generalizadas, com dados resumidos em “dashboards” para cada contexto citado, onde são apresentados

vários tipos de gráficos, como os de linha, pizza, barras, dispersão, segmentação, entre outros. A contribuição deste estudo foi desenvolver uma ferramenta de monitoramento e gerenciamento dos resultados dos alunos através de visualizações de dados de forma agregada, para auxiliar os utilizadores na tomada de decisão. Apresenta algumas limitações, como: não há visualizações para análises preditivas, não há possibilidade de avançar a análise dos dados ao ponto de chegar até a um aluno específico e relacioná-lo a outros dados, e também por depender de uma ferramenta comercial.

Em Guerra Laura e Arciniegas (2019) um estudo é realizado numa universidade que substituirá um curso de graduação por outro, com mudanças na grade curricular e no nome do curso. Para apoiar o processo de mudança, foi apresentado um sistema de visualização de dados para gestão acadêmica a fim de obter informações a respeito dos alunos para identificar problemas no desempenho acadêmico e aumentar o número de egressos. É realizado um processo de extração e transformação dos dados para serem utilizados na ferramenta Tableau, responsável pela análise e geração das visualizações. A ferramenta proposta utiliza-se de gráficos de linhas, barras, mosaico e bolhas, sendo apresentadas visualizações como: evolução acadêmica do aluno, rendimento médio das disciplinas por nível ou semestre, rendimento médio do curso ao longo do tempo e rendimento médio de todas as disciplinas do curso. Estas visualizações ajudam a identificar a evolução individual do aluno e as disciplinas, anos e semestres em que os estudantes têm maiores dificuldades de aprovação. Alguns padrões são encontrados como: os primeiros semestres são mais difíceis para o estudante, melhorando após o quarto semestre; disciplinas como cálculo diferencial e programação possuem alta reprovação e disciplinas dos primeiros semestres relacionadas à matemática e programação possuem baixo rendimento médio. O estudo permite análises sobre desempenho acadêmico e apoio à gestão de cursos, porém não explora a individualidade das trajetórias estudantis ao longo do tempo.

Santos, Ponti e Rodrigues (2022) propuseram o uso de relatórios digitais para apoiar a identificação de dados relacionados à evasão universitária. A proposta utiliza relatórios digitais para exibir indicadores como taxas de evasão por curso, perfil dos estudantes e desempenho por disciplina, permitindo que diferentes perfis de usuários acompanhem informações históricas e tomem decisões com base em evidências. Os relatórios foram avaliados por diferentes perfis de usuários e mostraram potencial para apoiar a gestão educacional. No entanto, o trabalho realiza uma análise estática dos dados, baseada em agregações e filtros aplicados aos registros. Ao desconsiderar a dimensão temporal, limita-se a compreensão da trajetória acadêmica dos alunos, impossibilitando a detecção de padrões que se manifestam gradualmente.

Já Guerra et al. (2019) apresentam uma ferramenta de visualização de dados acadêmicos chamada TrAC que captura a dimensão temporal, desenvolvida para apoiar os coordenadores de curso na análise da trajetória acadêmica dos estudantes. A solução permite sobrepor o histórico escolar do aluno à estrutura curricular fixa dos cursos, exibindo

visualmente as aprovações e reprovações, dados anteriores de outros cursos realizados, notas históricas do curso e relações de pré-requisitos das disciplinas. Contudo, seu modelo visual se limita à exibição estática do percurso individual sobre uma grade curricular fixa, sem considerar a comparação direta entre trajetórias, a análise de similaridade entre estudantes ou a identificação de padrões coletivos.

Ortega, Bravo e Peña (2019) propõem uma forma de visualização de trajetórias acadêmicas utilizando coordenadas paralelas, com o objetivo de apoiar a gestão acadêmica e permitir uma compreensão visual do progresso dos estudantes ao longo do tempo. O estudo foi realizado com dados de estudantes de uma universidade equatoriana, acompanhados ao longo de 11 períodos acadêmicos. A técnica utilizada possibilitou identificar tendências de desempenho, padrões de evasão e relações entre variáveis sociodemográficas e acadêmicas. A visualização permite interações como filtragem por curso, faculdade e intervalo de notas. A proposta evidencia a importância da visualização de dados multidimensionais e temporais na gestão educacional, utilizando as coordenadas paralelas como ferramenta central. Entretanto, a solução apresentada possui algumas limitações, como: realiza uma visualização 2D diretamente no espaço das variáveis originais, utilizando a média de notas por período em cada eixo da representação gráfica, não oferecendo a possibilidade de acrescentar outros dados que impactam o percurso acadêmico dos estudantes; consultas são pré-definidas sobre filtros sociodemográficos e notas, não atuam considerando a similaridade entre os dados, impossibilitando a criação de cenários personalizados para exploração; a visualização da trajetória não considera grupos personalizados de estudantes para análise comparativa.

O trabalho de Trimm, Rheingans e desJardins (2012) propõe uma abordagem para representar e analisar as trajetórias de estudantes no formato de séries temporais com foco na identificação de padrões coletivos. A proposta representa cada estudante como uma trajetória bidimensional, onde o eixo horizontal corresponde ao semestre e o eixo vertical à média das notas das disciplinas cursadas. Também permite traçar a trajetória média baseando-se na média geral das notas dos estudantes de um grupo. A coloração das trajetórias é utilizada para reforçar informações relevantes, como desempenho acadêmico, facilitando a identificação de comportamentos comuns ou desvios. Porém, abordagem se limita a análises agregadas e representações fixas, além da modelagem simplificada da trajetória dos estudantes, baseando-se no semestre e na média das disciplinas cursadas.

2.6.2 Em outros domínios

Em domínios diversos, já há mais estudos sobre a análise e visualização de trajetórias complexas, com propostas metodológicas que inspiram a presente dissertação.

Em Junior e Bueno (2021) foi proposta uma técnica de consultas de similaridade utilizando a representação gráfica de elementos complexos em diferentes posições no tempo (trajetória). Na consulta por similaridade tradicional, é definido um único objeto como

centro de busca, já na consulta por similaridade baseada em trajetória proposta, é definido um conjunto de instâncias temporais de um objeto como alvo de busca e são retornadas as trajetórias próximas à trajetória do elemento alvo. Para representar a trajetória das imagens, são extraídos vetores de características da situação das imagens em cada posição do tempo e convertidos para o espaço métrico. Em seguida, é aplicado o MDS para projetar as trajetórias em um espaço 3D, permitindo a visualização da consulta. Este trabalho também apresenta uma técnica para sumarizar as trajetórias relacionadas dos dados classificados e permitir a comparação das trajetórias médias de classes das imagens.

Em Bueno et al. (2010) são apresentadas três técnicas para visualização de dados métricos que evoluem ao longo do tempo: visualização baseada em linha do tempo, visualização baseada na variação do espaço e visualização métrico-temporal. A primeira técnica apresentada mapeia os dados métricos para 2 dimensões e usa o terceiro eixo para representar o tempo, mostrando a evolução dos dados em uma linha do tempo. A segunda visualização permite visualizar as variações nas posições espaciais dos elementos ao longo do tempo. Os dados métricos são mapeados pra um espaço 3D usando o algoritmo FastMap e a informação temporal é usada para ligar, através de segmentos de linhas, todas as instâncias do mesmo elemento durante o tempo. A terceira técnica é baseada no conceito de espaço métrico-temporal, onde é calculado o peso do componente métrico e do componente temporal para formar o elemento no espaço métrico. O espaço métrico é então mapeado para o espaço 3D e segmentos de linhas são usados para conectar as instâncias do mesmo elemento nos diferentes tempos disponíveis. Os autores demonstraram que estas técnicas podem ajudar a analisar e minerar visualmente tendências em conjuntos de dados complexos e em evolução temporal.

Em Paiva, Malaquias e Bueno (2020) foram aplicadas técnicas para visualização de dados complexos em evolução temporal com estimativa de trajetória. Os autores propõem técnicas para estimar a representação visual de um objeto em um instante específico, utilizando um conjunto de imagens do mesmo objeto capturadas ao longo do tempo. Através da dimensão de temporalidade, verifica-se a situação dos dados em diferentes pontos no tempo e é possível estimar uma condição futura de dados complexos. A técnica realiza uma interpolação entre as imagens disponíveis na consulta, gerando uma estimativa da posição da imagem em um espaço tridimensional. A partir dessa estimativa, é aplicada uma consulta k-NN baseada na trajetória, permitindo a recuperação de imagens similares e sua visualização em um ambiente 3D.

A análise dos trabalhos relacionados evidencia que, embora diversas abordagens existam para visualização de trajetórias em diferentes domínios, há uma lacuna na literatura quanto à representação de trajetórias acadêmicas modeladas por dados complexos e representadas em espaço métrico com foco em similaridade para exploração interativa e visual. Nenhuma das propostas analisadas combina de forma integral a modelagem de dados de estudantes em dados complexos em evolução temporal e a representação e aná-

lise visual de trajetórias acadêmicas baseando-se em similaridade, elementos centrais do método apresentado nesta dissertação.

2.7 Considerações Finais

Este capítulo apresentou a base teórica necessária para a compreensão e desenvolvimento da abordagem proposta nesta dissertação. A revisão dos principais conceitos relacionados à visualização de dados complexos e à representação em espaços métricos permitiu identificar as potencialidades e limitações das técnicas atualmente disponíveis no domínio acadêmico.

Foram abordadas todas as técnicas necessárias para tratar os estudantes como entidades complexas em evolução temporal, cuja trajetória pode ser modelada e analisada por meio de representações métricas e visualizações projetadas em espaços de menor dimensionalidade.

Assim, os fundamentos apresentados orientam a estruturação do método proposto, que integra conceitos e técnicas discutidas neste capítulo de maneira inovadora, visando potencializar a exploração visual e analítica das trajetórias acadêmicas. No próximo capítulo, será detalhado o método desenvolvido, bem como a sua aplicação prática no estudo de caso.

Capítulo 3

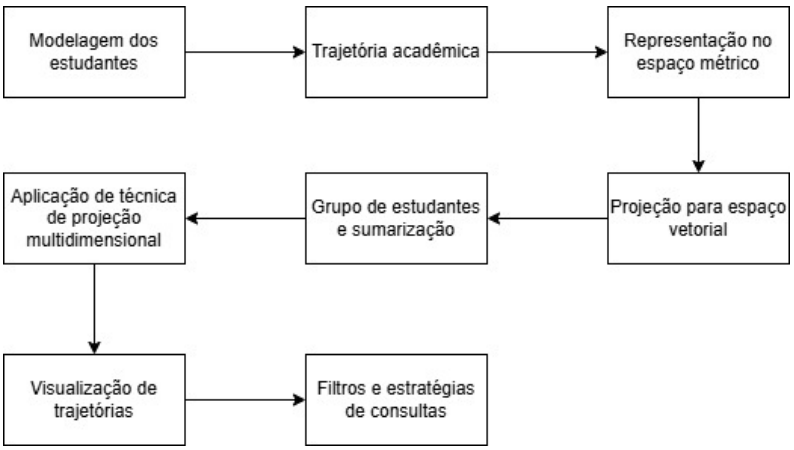
Método para visualização de trajetórias acadêmicas

Este capítulo descreve detalhadamente o método desenvolvido para a modelagem, representação e visualização das trajetórias acadêmicas dos estudantes.

3.1 Considerações iniciais

A abordagem adotada considera modelar os estudantes de maneira genérica, como dados complexos em evolução temporal e propõe sua representação em espaço métrico, para que, em seguida, seja possível utilizar um algoritmo para projetá-lo em um espaço multidimensional, de modo a viabilizar a aplicação de técnicas de visualização e análise exploratória, considerando a similaridade entre os dados. Para isso, foram definidos processos específicos para: (i) modelagem de estudantes como dado complexo, (ii) trajetória acadêmica, (iii) representação de estudantes no espaço métrico, (iv) projeção para espaço vetorial, (v) grupos de estudantes e sumarização de trajetórias, (vi) aplicação de técnicas de projeção multidimensional, (vii) visualização de trajetórias acadêmicas, e (viii) filtros e estratégias de consulta. A Figura 3.1 ilustra o fluxo geral do método proposto.

Figura 6 – Fluxo geral do método proposto.



Fonte: Elaborada pelo autor

As seções a seguir detalham cada uma das etapas descritas.

3.2 Modelagem de Estudantes como Dados Complexos

A fim de representar o percurso formativo dos estudantes ao longo do tempo, este trabalho propõe a modelagem de cada estudante como um dado complexo em evolução temporal. Essa modelagem ocorre por meio da construção de *instâncias temporais*, que corresponde à fotografia do estado de um estudante em um determinado momento do curso (semestre letivo). Além disso, a técnica permite que a instância temporal represente também um conjunto de instâncias de estudantes sumarizadas conforme será detalhado na seção 3.5. Formalmente, uma instância temporal (i_t) é definida como um par ordenado:

$$i_t = (I_t, t) \tag{2}$$

onde (I_t) representa um conjunto qualquer de informações que descreve o estado do estudante ou grupo no período (t), e (t) é o carimbo de tempo, indicando o período acadêmico.

A técnica proposta é genérica e permite que as instâncias temporais sejam representadas por qualquer tipo de dado ou combinação de dados – como números, textos, datas, imagens – desde que seja possível definir uma função de distância entre essas instâncias, conforme aborda o conceito de espaço métrico, na seção 2.4.1. Essa premissa torna a técnica aplicável a diversos domínios e contextos analíticos, ampliando seu potencial de uso.

Por exemplo, um estudante poderia ser representado por:

- **Representação vetorial com notas e faltas**, compondo um vetor numérico com as notas finais e o número de presenças em todas as disciplinas cursadas até o

semestre de referência. Essa forma de representação permite a utilização de métricas conhecidas como a *distância euclidiana*, *distância de Manhattan* ou *distância de Minkowski* para medir a similaridade entre duas instâncias.

- **Representações textuais**, como análise semântica de redações ou justificativas acadêmicas. A instância pode ser representada por um relatório textual contendo informações qualitativas do desempenho do estudante em um ciclo formativo (por exemplo, avaliação de docentes, atividades extracurriculares, descrição de dificuldades ou avanços). Neste caso, a similaridade entre instâncias pode ser obtida por algoritmos de comparação textual, como o *Cosine Similarity* entre vetores TF-IDF (MANNING; RAGHAVAN; SCHüTZE, 2008), ou modelos baseados em *word embeddings* como *BERTScore* (ZHANG et al., 2020);

- **Representações híbridas**, combinando desempenho acadêmico com dados de perfil socioeconômico, tipo de escola no ensino médio, ou forma de ingresso na universidade, desde que uma função de distância adequada seja definida para tratar o espaço heterogêneo;

A única exigência para que uma representação seja compatível com a abordagem proposta é que se possa calcular uma função de similaridade ou dissimilaridade entre os elementos (vide seção 3.4). Desta maneira, a modelagem proposta para as instâncias temporais — de estudantes ou grupos — é flexível, permitindo adaptar o método para diferentes contextos institucionais, níveis de ensino e objetivos analíticos. A escolha da representação mais adequada depende da disponibilidade dos dados e da semântica que se deseja capturar nas análises.

3.3 Trajetórias Acadêmicas

Ao longo de sua formação, um estudante vivencia mudanças graduais e cumulativas em seu desempenho, envolvimento e progresso acadêmico (JAYARATNA et al., 2020). Essa característica de evolução temporal permite modelagem como uma entidade dinâmica, cujas transformações podem ser analisadas em diferentes pontos do tempo (BUENO et al., 2010). Nesse contexto, propõe-se a modelagem do percurso formativo do estudante como uma *trajetória acadêmica*.

Formalmente, a trajetória acadêmica (T_e) de um estudante (e) é definida como um conjunto de instâncias temporais:

$$T_e = \{i_{e,t_1}, i_{e,t_2}, ..., i_{e,t_{n_e}}\} \tag{3}$$

onde (i_{e,t_k}) é a instância temporal do estudante (e) no período acadêmico (t_k), conforme definido na Seção 3.2, e (n_e) é o número total de instâncias temporais na trajetória do

estudante (e). É importante notar que a ordem dos elementos neste conjunto não implica uma ordenação da trajetória; a ordenação é determinada pelos carimbos de tempo ((t)) dentro de cada instância temporal.

Cada instância temporal (i_{e,t_k}) representa o estado do estudante (e) no período acadêmico (t_k), conforme discutido na Seção 3.2. Dessa forma, a trajetória reflete, ao longo do tempo, a evolução dos dados que representam o estudante, permitindo a análise não apenas de seu desempenho isolado, mas também das tendências e padrões que emergem durante sua formação. Como uma instância temporal pode representar grupos de estudantes, uma trajetória acadêmica também pode ser de grupos sintetizados de estudantes.

A modelagem do percurso acadêmico dos estudantes através de trajetória possibilita:

- **Análise temporal estruturada:** Permite que a análise do percurso acadêmico leve em consideração a sequência e a progressão dos eventos, possibilitando inferências sobre tendências de melhora, estagnação ou declínio de desempenho.
- **Compreensão progressiva:** Ao analisar a trajetória de um estudante como um todo, e não apenas suas instâncias isoladas, torna-se possível compreender melhor o contexto e o ritmo de sua evolução.
- **Comparação entre estudantes:** A modelagem por trajetórias possibilita comparar estudantes com perfis de evolução similares, mesmo que suas características específicas sejam distintas, favorecendo a identificação de padrões compartilhados e potenciais riscos.
- **Generalização analítica:** A abordagem de trajetórias pode ser aplicada a qualquer domínio educacional em que dados temporais estejam disponíveis, permitindo sua adaptação a diferentes contextos institucionais, níveis de ensino e objetivos analíticos.
- **Comparação entre grupos de estudantes:** A abordagem também permite comparar a trajetória individual de estudantes a trajetória média de grupos de estudantes criados a partir de critérios customizados, favorecendo análises comparativas contextualizadas.

Com base nessa modelagem de trajetória acadêmica, torna-se possível aplicar técnicas de análise e visualização de dados complexos em evolução temporal (JUNIOR; BUENO, 2021), conforme será detalhado nas seções seguintes.

3.4 Representação de estudantes no espaço métrico

As instituições educacionais atuais produzem uma grande quantidade de dados em registros de estudantes, avaliações, sistemas de aprendizagem online e outras atividades administrativas (RAUT; HAJARE, 2025). O crescimento contínuo da quantidade de

dados cria um ambiente no qual abordagens novas ou inovadoras são necessárias para entender os padrões de valor existentes nos dados (SIEMENS; LONG, 2011). Muitos desses dados podem ser utilizados para representar os estudantes ao longo de sua jornada acadêmica. Como a proposta deste trabalho é propor uma modelagem genérica, a abordagem que se encaixa é a representação dos estudantes no espaço métrico.

A representação das instâncias temporais em um espaço métrico é uma etapa central do método proposto, pois possibilita tratar os estudantes como dados complexos — dimensionais ou adimensionais — a partir de uma estrutura que preserva relações de similaridade, conforme abordado na seção 2.4.1. Ao adotar um espaço métrico, não se exige uma representação em um espaço vetorial tradicional, bastando que seja definida uma função de distância entre pares de instâncias (CIACCIA; PATELLA; ZEZULA, 1997). Essa abordagem permite flexibilidade na modelagem, acomodando dados numéricos, textuais, categóricos ou multimodais. O cálculo dessas distâncias fornece a base para operações posteriores, como a projeção multidimensional e a visualização das trajetórias acadêmicas em um espaço de baixa dimensão.

A representação das instâncias temporais (de um estudante ou grupo) em um espaço métrico permite aplicar técnicas de comparação e análise baseadas em similaridade (NAVARRO, 1999). Por exemplo, é possível modelar as instâncias temporais dos estudantes por vetores numéricos, como vetores de notas em disciplinas, e assim, calcular a distância entre duas instâncias utilizando-se métricas clássicas, como a *distância euclidiana*. Através dessa abordagem, é possível representar a dissimilaridade entre vetores no espaço multidimensional. Além da distância euclidiana, outras funções de distância podem ser utilizadas dependendo do tipo de dados, como a distância de Manhattan para dados numéricos, a distância de Hamming para dados categóricos (ARGIENTO; FILIPPI-MAZZOLA; PACI, 2024), ou medidas de similaridade textual como o *Cosine Similarity* (MANNING; RAGHAVAN; SCHüTZE, 2008) e o BERTScore (ZHANG et al., 2020) para dados textuais. A escolha da função de distância influencia diretamente os resultados das análises de similaridade (BRIN, 1995) e das técnicas de projeção multidimensional (LIU et al., 2017).

Considere o exemplo hipotético de duas instâncias temporais A e B refletindo o estado de dois estudantes ao final do segundo semestre letivo. Estas instâncias foram representadas através de vetores de características de mesmo tamanho compostos por notas em todas as disciplinas do semestre:

$$\vec{A} = [7.0, 8.5, 6.0, 9.0, 7.5, 8.0, 6.5, 7.0]$$

$$\vec{B} = [6.5, 9.0, 5.5, 8.5, 7.0, 8.5, 7.0, 6.0]$$

A distância euclidiana entre $\vec{A}$ e $\vec{B}$ é dada por:

$$d(\vec{A}, \vec{B}) = \sqrt{\sum_{i=1}^{n}(a_i - b_i)^2} \tag{4}$$

Aplicando os valores:

$$\begin{aligned} d(\vec{A}, \vec{B}) &= \sqrt{(7.0-6.5)^2 + (8.5-9.0)^2 + (6.0-5.5)^2 + (9.0-8.5)^2} \\ &\quad + \sqrt{(7.5-7.0)^2 + (8.0-8.5)^2 + (6.5-7.0)^2 + (7.0-6.0)^2} \\ &= \sqrt{0.25+0.25+0.25+0.25+0.25+0.25+0.25+1.00} \\ &= \sqrt{2.75} \approx 1.66 \end{aligned}$$

Esse valor representa a dissimilaridade entre os dois estudantes com base em seu desempenho no segundo semestre. Valores menores indicam maior similaridade, enquanto valores maiores representam percursos mais distintos.

A utilização do espaço métrico permite aplicar algoritmos de busca por similaridade, como do tipo *k-Nearest Neighbors* ou consultas por abrangência (CHÁVEZ et al., 2001), além de representar dados de entrada para técnicas de projeção multidimensional que viabilizam a visualização gráfica das trajetórias. A escolha adequada da função de distância (como a euclidiana neste exemplo) é crucial para garantir a eficácia desses algoritmos e técnicas.

Nas seções seguintes, será apresentada a extensão desse modelo para incorporar a dimensão temporal explicitamente no cálculo de distância, resultando na *representação métrico-temporal*, bem como a criação e sumarização de grupos de estudantes e a aplicação de técnicas de visualização para exploração das trajetórias acadêmicas.

3.4.1 Representação Métrico-Temporal de Instâncias de Estudantes

Uma abordagem adicional desta dissertação consiste em incorporar a distância métrico-temporal na representação das trajetórias de estudantes no espaço métrico. Esta escolha metodológica foi a única encontrada na literatura para lidar com dados complexos que evoluem ao longo do tempo (vide seção 2.4.2).

A representação métrico-temporal estende o conceito de representação métrica ao incorporar a proximidade temporal entre as instâncias temporais no cálculo da distância, conforme definido na seção 2.4.3. Essa abordagem permite uma análise mais contextualizada de como as características acadêmicas dos estudantes evoluem ao longo do tempo, considerando tanto "o que" muda quanto "quando" muda.

No contexto de trajetórias acadêmicas, a representação métrico-temporal permite reconhecer a similaridade entre os estados acadêmicos de duas instâncias temporais, não

apenas pela representação métrica (do conjunto de dados que as representam), mas também pelo momento temporal (semestre letivo, ano letivo ou outra identificação de ciclo temporal) da informação métrica. Por exemplo, dois estudantes com notas similares no primeiro semestre podem ser considerados mais similares do que dois estudantes com as mesmas notas, mas um as obteve no 1º semestre e o outro no 3º semestre.

Essa abordagem permite capturar variações na progressão acadêmica que uma abordagem puramente métrica pode perder. É possível diferenciar estudantes que seguem trajetórias similares no mesmo período de tempo e aqueles que exibem padrões semelhantes, mas em diferentes estágios de sua jornada acadêmica.

Aproveitando o exemplo da seção 3.4, de dois estudantes A e B no 2º semestre, considere que os estudantes estejam em semestres diferentes, onde deseja-se comparar o estudante A no 2º semestre com o estudante B no 4º semestre.

Estudante A (2º Semestre):
$\vec{A_2} = [7.0, 8.5, 6.0, 9.0, 7.5, 8.0, 6.5, 7.0], t_A = 2$
Estudante B (4º Semestre):
$\vec{B_4} = [6.5, 9.0, 5.5, 8.5, 7.0, 8.5, 7.0, 6.0], t_B = 4$

1. **Distância Métrica:**

   Foi calculada a distância Euclidiana no exemplo anterior (sem influência temporal): $d_m(\vec{A_2}, \vec{B_4}) \approx 1.66$

2. **Distância Temporal:**

   A distância temporal é a diferença absoluta em semestres:

   $d_t(A_2, B_4) = |t_A - t_B| = |2 - 4| = 2$

3. **Distância Ponderada:**

   Assume-se que, para este exemplo, a distância métrica é ligeiramente mais importante do que a distância temporal, então define-se $\alpha = 0.6$ e $\beta = 0.4$. Os valores usados na ponderação dos pesos são obtidos através do cálculo da dimensão intrínseca dos componentes métricos e temporais dividido pelas distâncias máximas dentro de cada componente, conforme proposto por Bueno et al. (2009).

   Assim, a distância métrico-temporal é:

   $d_{mt}(A_2, B_4) = 0.6 \cdot 1.66 + 0.4 \cdot 2 = 0.996 + 0.8 = 1.796$

Caso os mesmos estudantes A e B estivessem ambos no 2º semestre, a distância métrico-temporal seria calculada da seguinte forma:

1. A distância métrica (d_m) permaneceria a mesma, pois é calculada com base nos vetores de notas, que não se alteram: $d_m(\vec{A_2}, \vec{B_2}) \approx 1.66$

2. A distância temporal (d_t) seria zero, pois ambos os estudantes estão no mesmo semestre: $d_t(A_2, B_2) = |2 - 2| = 0$

3. A distância métrico-temporal (d_{mt}) seria então: $d_{mt}(A_2, B_2) = 0.6 \cdot 1.66 + 0.4 \cdot 0 = 0.996$

Neste caso, a distância métrico-temporal (0.996) entre estudantes do mesmo semestre é menor do que a distância métrico-temporal (1.796) de estudantes em semestres diferentes, devido à influência do tempo no cálculo.

3.4.2 Projeção para um Espaço Vetorial

A abordagem proposta nesta dissertação assume a representação dos estudantes em um espaço métrico, onde apenas a informação de distância entre os elementos está disponível. Para possibilitar a sumarização de grupos e operações posteriores — como a inclusão de representações sumarizadas em buscas por similaridade — é necessário mapear esse espaço métrico para um espaço vetorial intermediário, no qual as instâncias sejam expressas como vetores numéricos de tamanho fixo.

Esse mapeamento é realizado utilizando técnicas de projeção multidimensional, como o MDS, que transformam as relações de distância do espaço métrico original em representações com coordenadas reais em um espaço de d dimensões. A escolha de d corresponde à *dimensão intrínseca* do conjunto de dados, ou seja, ao número mínimo de dimensões capaz de preservar adequadamente a distribuição das distâncias.

Para a determinação da dimensão intrínseca, podem ser empregados métodos baseados na *teoria dos fractais*, como o *Box-Counting*, que estima a *dimensão fractal de Hausdorff* (D_0), a *dimensão de correlação* (D_2) (BELUSSI; FALOUTSOS, 1998), ou o método do *Distance Exponent* (TRAINA; TRAINA; FALOUTSOS, 2000).

Neste espaço vetorial mapeado, são realizadas as operações de sumarização e filtragem (vide seção 3.8). Como exceção, quando a representação dos estudantes já é vetorial, com vetores numéricos de tamanho fixo, não há necessidade de realizar o mapeamento. Nesse caso, a operação de cálculo da média pode ser realizada diretamente no espaço métrico original, pois a representação já satisfaz os requisitos necessários de dimensionalidade fixa.

3.5 Grupos de Estudantes e Sumarização de Trajetórias

Este trabalho utiliza a proposta de sumarização de classes apresentada por Junior e Bueno (2021), adaptando-a ao domínio acadêmico para possibilitar a modelagem e visualização de trajetórias representando grupos de estudantes. Um grupo pode ser definido a partir de critérios específicos, como desfecho acadêmico (e.g., Formados, Evadidos),

desempenho em disciplinas-chave (e.g., reprovados em Cálculo), tempo de conclusão do curso, atributos relacionais (e.g., ingressantes por ações afirmativas, por região do país, por renda familiar, etc.), entre outros fatores. A análise da trajetória média de grupos permite identificar padrões coletivos e tendências comuns, auxiliando na formulação de estratégias institucionais.

A técnica proposta permite tanto a criação de grupos de estudantes em tempo de execução, quanto a definição de grupos pré-calculados, cujas trajetórias médias são armazenadas no banco de dados. Essas trajetórias sumarizadas podem ser recuperadas de maneira idêntica às trajetórias individuais de estudantes, pois são tratadas no mesmo espaço mapeado que as instâncias dos estudantes. Assim, a técnica considera a trajetória média de um grupo como uma entidade com o mesmo status de uma trajetória individual, permitindo que grupos sumarizados sejam comparados, buscados e visualizados.

A construção de trajetórias médias de grupos tem múltiplos objetivos:

- ❏ Comparar padrões médios de diferentes grupos (por exemplo, entre formados e evadidos, reprovados em cálculo, etc);
- ❏ Identificar desvios individuais em relação ao comportamento coletivo;
- ❏ Visualizar o progresso médio de um grupo em relação a outro;
- ❏ Subsidiar ações institucionais com base em dados empíricos consolidados.

As trajetórias sumarizadas são compatíveis com todas as etapas subsequentes do método, podendo ser projetadas no espaço tridimensional, visualizadas e comparadas com trajetórias individuais, conforme abordado nas Seções 3.6 e 3.7.

3.6 Aplicação de Técnica de Projeção Multidimensional para visualização

A representação das trajetórias acadêmicas de estudantes ou grupos em espaço métrico estabelece apenas as relações de distância entre as instâncias temporais, não definindo coordenadas espaciais explícitas que possibilitem sua visualização direta. Para permitir a construção de visualizações que aproveitem a capacidade de análise e interpretação humana, é necessário aplicar uma técnica de projeção multidimensional que transforme a matriz de distâncias métricas em um conjunto de coordenadas em um espaço dimensional, como 2D ou 3D (LIU et al., 2017).

Neste trabalho, utiliza-se a técnica de *Multidimensional Scalling* clássico para projetar as trajetórias acadêmicas em um espaço tridimensional. Embora seja custoso computacionalmente (FALOUTSOS; LIN, 1995), o algoritmo tenta preservar ao máximo a distância

em pares entre os pontos de dados (SMELSER; MILLER; KOBOUROV, 2024) e o conjunto de dados usado nessa pesquisa é relativamente pequeno.

Para exemplificar a aplicação do MDS, considere duas trajetórias acadêmicas, cada uma composta por duas instâncias temporais consecutivas, representando dois estudantes em diferentes momentos do curso. As instâncias temporais foram representadas por dados complexos quaisquer (como vetores, textos ou combinações de atributos). A Tabela 1 apresenta uma matriz de distâncias fictícia entre as quatro instâncias, calculada com base em uma função de distância genérica, que pode ser, por exemplo, a distância Euclidiana no caso da representação através de vetores numéricos.

Tabela 1 – Matriz de distâncias entre instâncias temporais de dois estudantes.

	A_1	A_2	B_1	B_2
A_1	0.00	1.20	2.00	2.60
A_2	1.20	0.00	1.90	2.10
B_1	2.00	1.90	0.00	1.10
B_2	2.60	2.10	1.10	0.00

Fonte: Elaborada pelo autor

Após a aplicação do MDS, as quatro instâncias são mapeadas para coordenadas em um espaço tridimensional, conforme ilustrado na Tabela 2.

Tabela 2 – Coordenadas 3D obtidas pelo MDS para as instâncias.

Instância	x	y	z
A_1	-1.2	0.8	0.1
A_2	-0.3	0.5	0.0
B_1	0.7	-0.6	-0.2
B_2	1.1	-0.8	-0.3

Fonte: Elaborada pelo autor

Com base nessas coordenadas, é possível representar graficamente as trajetórias dos estudantes em um ambiente tridimensional. Por exemplo, conectando as instâncias $A_1 \rightarrow A_2$ e $B_1 \rightarrow B_2$ com segmentos de linha, formam-se duas trajetórias espaciais distintas. Esse tipo de visualização é compatível com a abordagem baseada em linhas conectadas (BUENO et al., 2010) e permite interpretar visualmente a evolução dos estudantes ao longo do tempo.

Portanto, essa etapa de projeção é responsável por converter trajetórias acadêmicas representadas no espaço métrico para trajetórias acadêmicas representadas em um espaço de baixa dimensão, habilitando a criação de representações visuais das trajetórias.

3.7 Visualização de Trajetórias Acadêmicas

O mapeamento visual desempenha um papel essencial na conversão do resultado da transformação de dados em estruturas visuais para renderização (LIU et al., 2017). É neste processo que são geradas as representações gráficas.

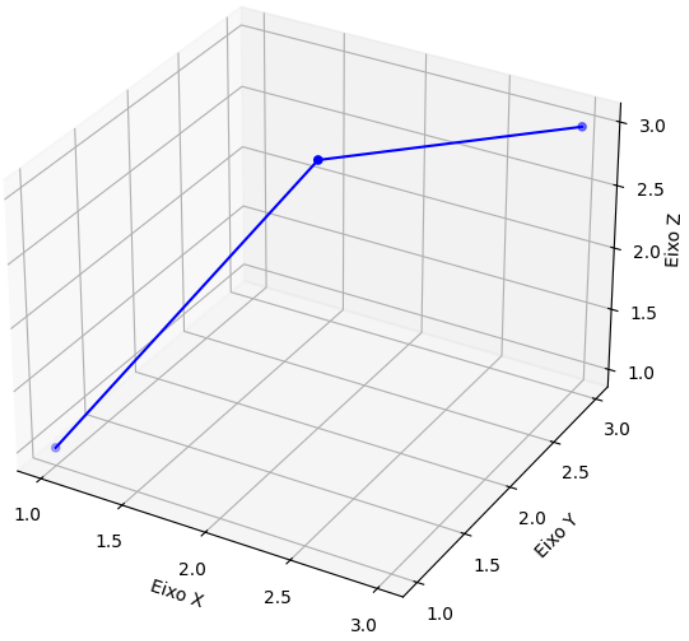
A proposta deste trabalho é apresentar graficamente as trajetórias acadêmicas que foram previamente representadas no espaço métrico e projetá-las em um espaço tridimensional. Essa representação visual permite explorar, de forma intuitiva, a evolução de estudantes ou grupos de estudantes ao longo do tempo, facilitando a identificação de padrões, desvios, tendências de progresso e potenciais riscos acadêmicos.

3.7.1 Representação Visual das Trajetórias

Cada trajetória visual corresponde a uma sequência de *instâncias temporais*, que pode representar um estudante individual ou a média de um grupo de estudantes, conforme detalhado na seção 3.5. Na representação gráfica proposta por este trabalho, as instâncias temporais são representadas por pontos posicionados em um ambiente tridimensional, conforme as coordenadas geradas pela técnica de projeção multidimensional, similar à técnica empregada em Junior e Bueno (2021). Cada ponto corresponde a um momento específico (por exemplo, um semestre letivo) da trajetória. Os pontos são conectados por segmentos de linha que representam a continuidade temporal (BUENO et al., 2010), formando uma **trajetória visual**.

A Figura 7 apresenta uma visualização genérica de uma trajetória acadêmica representada no espaço tridimensional, onde os pontos representam instâncias temporais, e as linhas conectando os pontos indicam a evolução ao longo dos semestres.

Figura 7 – Representação genérica de trajetória acadêmica em espaço 3D



Fonte: Elaborada pelo autor

Ao representar os dados de forma espacial, a visualização tridimensional permite a análise comparativa de diferentes trajetórias acadêmicas e possibilita a identificação de padrões de comportamento coletivo ou individual. Trajetórias próximas no espaço 3D indicam similaridade entre os estados acadêmicos utilizados para representar as instâncias temporais.

Esse tipo de visualização pode revelar:

- ❏ Agrupamentos naturais de estudantes com evolução acadêmica semelhante;
- ❏ Padrões de progresso acelerado ou atrasado;
- ❏ Casos atípicos que podem indicar situações de risco (evasão, trancamento, desempenho inferior à média);
- ❏ Divergência repentina entre trajetórias previamente similares;
- ❏ Proximidade entre a trajetória de um estudante e a trajetória média de uma classe (ex.: estudantes formados, evadidos ou com perfil específico).

3.7.2 Técnicas de Interação na Visualização

Este trabalho propõe o uso de técnicas de interação no ambiente de visualização 3D, para ampliar a capacidade analítica das informações visuais (KEIM, 2002). Como apre-

sentado na seção 2.2.4, recursos como *zoom*, *pan* e *rotação* permitem manipular a perspectiva da visualização, possibilitando a inspeção detalhada de agrupamentos, a comparação entre diferentes regiões do espaço projetado e o foco em trajetórias específicas.

O recurso de rotação permite observar a configuração espacial das trajetórias sob diferentes ângulos e possibilita revelar padrões que podem não ser perceptíveis em uma única perspectiva. Já o zoom viabiliza uma análise em diferentes níveis de granularidade, desde um panorama geral até detalhes individuais de uma trajetória.

3.7.3 Flexibilidade da Abordagem Visual

A técnica de visualização adotada neste trabalho é independente do tipo de dado originalmente utilizado para modelar as instâncias temporais, uma vez que atua sobre o espaço projetado a partir das distâncias. Assim, independentemente de a instância ter sido representada por notas, textos, imagens ou vetores multimodais, a visualização reflete apenas as relações de similaridade capturadas na etapa de modelagem métrica.

Além disso, a abordagem visual é flexível quanto à natureza da trajetória, podendo representar tanto o percurso individual de um estudante quanto trajetórias médias de grupos definidos por critérios customizados. Essa flexibilidade permite a construção de cenários exploratórios variados, adaptados a diferentes demandas institucionais e analíticas.

3.8 Filtros e Estratégias de Consulta de Trajetórias

A visualização simultânea de todas as trajetórias acadêmicas disponíveis pode dificultar a análise, devido à grande quantidade de informações e sobreposição visual. Para tornar a exploração mais eficiente, este trabalho propõe mecanismos de filtragem que permitem restringir a visualização a subconjuntos de interesse (KEIM, 2002). As consultas podem ser categorizadas em três grupos: baseadas em atributos relacionais, por similaridade de trajetória e híbridas.

O principal objetivo das estratégias de filtragem é permitir ao analista explorar cenários específicos a partir de critérios objetivos. A filtragem permite, por exemplo, isolar grupos de estudantes com características semelhantes, comparar o comportamento de estudantes ativos com estudantes inativos que já encerraram o ciclo acadêmico, ou avaliar a evolução de estudantes com base em eventos que acontecem durante o percurso acadêmico, por exemplo, uma reprovação em uma disciplina-chave do curso.

A técnica considera a representação sumarizada de grupos de estudantes como idêntica à representação dos estudantes individuais, possibilitando que trajetórias de grupos sumarizados sejam recuperadas nas consultas como se fossem trajetórias de estudantes. Assim, além de permitir consultas direcionadas a estudantes individuais, também permite operações de filtragem a grupos sumarizados, enriquecendo as possibilidades analíticas.

3.8.1 Consultas Baseadas em Atributos Relacionais

Consultas baseadas em atributos relacionais utilizam diretamente os atributos armazenados no banco de dados dos estudantes, como curso, forma de ingresso, status acadêmico ou desempenho em disciplinas específicas. Essas consultas podem ser expressas por filtros em linguagem SQL.

Por exemplo, pode-se filtrar as trajetórias apenas "Estudantes do sexo feminino que ingressaram em 2023 no curso de Bacharelado em Ciência da Computação" para verificar se há algum comportamento atípico que necessite de intervenção. A visualização destas trajetórias permite observar a evolução geral das estudantes selecionadas pela consulta.

3.8.2 Consulta por Similaridade de Trajetória

A consulta por similaridade de trajetória, adaptada da proposta de Junior e Bueno (2021) para o domínio de trajetórias acadêmicas, permite recuperar estudantes cujos percursos acadêmicos apresentam semelhança com uma ou mais trajetórias-alvo, que podem ser de estudantes ou grupos. Essa similaridade é avaliada no espaço multidimensional intermediário, resultante do mapeamento do espaço métrico original, conforme descrito na Seção 3.5. Esse espaço possibilita que tanto trajetórias individuais quanto trajetórias sumarizadas de grupos sejam tratadas de forma homogênea nas consultas por similaridade, utilizando a representação métrica ou a representação métrico-temporal (conforme definido na Seção 3.4.1), que considera também a proximidade temporal entre as instâncias.

Considere um conjunto de trajetórias $S = \{T_1, T_2, \ldots, T_n\}$, onde cada trajetória T_i representa a sequência de instâncias temporais de um estudante i:

$$T_i = \{i_{i1}, i_{i2}, \ldots, i_{im_i}\} \tag{5}$$

Seja $T_a = \{i_{a1}, i_{a2}, \ldots, i_{am_a}\}$ a trajetória-alvo, composta por m_a instâncias temporais. A consulta consiste em buscar, para cada instância $i_{a_t} \in T_a$, os k elementos mais próximos do conjunto S, considerando todas as instâncias temporais disponíveis no banco de dados, em qualquer tempo.

Para cada uma das m_a instâncias i_{a_t} da trajetória-alvo, obtêm-se os k vizinhos mais próximos kNN(i_{a_t}). Em seguida, para cada vizinho identificado, sua trajetória completa é incluída no conjunto de resultados.

Assim, o conjunto total de trajetórias retornadas pode conter até $k \cdot m_a$ elementos distintos, considerando que diferentes instâncias da trajetória-alvo podem recuperar estudantes diferentes como vizinhos.

Essa abordagem pode ser utilizada para identificar perfis estudantis com evolução acadêmica similar, prever possíveis desfechos e embasar ações de suporte com base em padrões históricos de trajetória.

Adicionalmente, esta consulta pode ser aplicada para comparar estudantes ainda ativos, cujas trajetórias estão em andamento, com estudantes já concluídos, cujos desfechos acadêmicos são conhecidos. Isso possibilita a inferência de possíveis tendências de conclusão, como formação ou evasão, mesmo em semestres futuros que o estudante ainda não vivenciou. Ao observar que as trajetórias mais próximas à trajetória do estudante ativo possuem, em sua maioria, um determinado tipo de desfecho, o analista pode antecipar possíveis riscos e direcionar ações de intervenção.

3.8.3 Consulta por Similaridade Pontual de Trajetória

A consulta por similaridade pontual é uma extensão da estratégia anterior, permitindo ao analista selecionar tempos específicos da trajetória-alvo para a busca por similaridade. Isso oferece maior controle sobre a análise, possibilitando a investigação de contextos acadêmicos em momentos específicos do percurso.

Formalmente, considere uma trajetória-alvo $T_a = \{i_{a1}, i_{a2}, \ldots, i_{am_a}\}$ e um subconjunto de tempos $P \subseteq \{1, 2, \ldots, m_a\}$, que define as instâncias $\{i_{a_t} \mid t \in P\}$ como pontos de consulta.

Para cada instância i_{a_t} selecionada como referência para a busca, a consulta busca os k vizinhos mais próximos utilizando a distância métrica ou métrico-temporal, considerando todas as instâncias temporais disponíveis no banco de dados, independentemente do tempo em que foram geradas.

O resultado consiste na recuperação das trajetórias completas dos estudantes que aparecem como vizinhos em algum dos tempos selecionados. Assim, o conjunto de respostas pode conter até $k \cdot |P|$ trajetórias.

Essa consulta pode ser utilizada para recuperar trajetórias similares em semestres-chave para o percurso acadêmico. Por exemplo, hipoteticamente, o segundo semestre de um curso de graduação é considerado o mais difícil pelos estudantes. Ao investigar a trajetória de um estudante que está no terceiro semestre, buscando por trajetórias similares especificamente no segundo semestre, pode-se verificar o desfecho acadêmico destes estudantes e validar com quais perfis ele mais se aproximava naquele momento. Dessa forma, inferir possíveis tendências futuras.

Ambas as estratégias descritas são flexíveis quanto à representação dos estudantes e compatíveis com diferentes formas de modelagem, pois consideram a similaridade no espaço métrico. A distinção central reside no nível de controle oferecido ao analista: enquanto a consulta por similaridade original (JUNIOR; BUENO, 2021) considera toda a trajetória-alvo, a versão pontual permite focar a análise em pontos críticos do percurso formativo.

3.8.4 Consultas Híbridas

As consultas híbridas combinam filtros relacionais, que operam sobre atributos explícitos dos estudantes registrados no banco de dados, com a consulta por similaridade de trajetória, baseada nas representações métricas (CHÁVEZ et al., 2001) ou métrico-temporais (BUENO et al., 2009) das instâncias temporais.

O processo ocorre em duas etapas:

1. **Filtragem relacional**: inicialmente, aplica-se um ou mais filtros relacionais sobre os atributos descritivos dos estudantes, como forma de ingresso, período de entrada, desfecho acadêmico, gênero ou qualquer outro campo disponível no banco de dados. Essa etapa delimita o subconjunto de estudantes que será considerado na consulta subsequente.

2. **Consulta por similaridade**: em seguida, executa-se a consulta por similaridade de trajetória sobre esse subconjunto previamente filtrado. A busca é realizada com base em uma ou mais trajetórias-alvo, que pode ser de estudantes individuais ou de grupos sumarizado. A métrica utilizada pode ser exclusivamente métrica ou métrico-temporal, conforme configurado pelo analista.

Essa abordagem permite que a similaridade seja avaliada dentro de grupos específicos, evitando comparações entre trajetórias de estudantes que possuem características muito heterogêneas.

Considere o seguinte exemplo: um analista deseja investigar a situação de um estudante específico X, ingressante a partir de cota racial, que apresentou baixo desempenho acadêmico no primeiro semestre.

A consulta híbrida poderia ser formulada da seguinte maneira:

- **Filtro relacional**: selecionar apenas estudantes que ingressaram através de cotas raciais.

- **Consulta por similaridade**: dentro desse subconjunto, identificar as 5 trajetórias mais similares à trajetória do estudante X, que apresentou baixo desempenho acadêmico no primeiro semestre.

Essa consulta híbrida tem como objetivo principal identificar outros estudantes com um perfil de ingresso similar (cotas raciais) que possam estar seguindo um percurso acadêmico análogo ao do estudante X, que já demonstrou dificuldades iniciais. A análise visual dessas 5 trajetórias similares pode revelar padrões de desafio comuns a esse grupo específico, permitindo à instituição desenvolver e aplicar programas de apoio ou intervenções pedagógicas direcionadas e preventivas.

As consultas híbridas ampliam as possibilidades de exploração dos dados, permitindo focar em subgrupos específicos de estudantes, comparar trajetórias entre diferentes perfis

demográficos ou acadêmicos e reduzir o espaço de busca, tornando as consultas mais eficientes e direcionadas.

3.9 Considerações Finais

O capítulo apresentou, de forma estruturada, o método proposto para a modelagem, representação, projeção e visualização de trajetórias acadêmicas de estudantes, considerando-os como dados complexos em evolução temporal. A abordagem é flexível e extensível, permitindo a aplicação a diferentes tipos de dados e contextos educacionais.

Os principais aspectos abordados foram: a construção de instâncias temporais, a representação métrica e métrico-temporal, a projeção tridimensional das trajetórias, os mecanismos de visualização, os filtros de consulta e a possibilidade de sumarização de trajetórias de grupos de estudantes.

As estratégias propostas visam não apenas representar o percurso dos estudantes, mas também possibilitar análises visuais capazes de apoiar decisões pedagógicas, administrativas e políticas institucionais.

No próximo capítulo, será apresentado um estudo de caso com dados reais de um curso de graduação em Ciência da Computação, demonstrando a aplicação prática do método e os potenciais analíticos das técnicas desenvolvidas.

Capítulo 4

Implementação do Estudo de Caso

Este capítulo apresenta os detalhes técnicos da implementação do método proposto utilizando dados reais fornecidos pela Universidade Federal de São Carlos (UFSCar). São descritos o conjunto de dados utilizado, o protótipo desenvolvido para exploração visual das trajetórias, as estratégias adotadas para a modelagem das instâncias temporais e o processo de criação e sumarização de grupos de estudantes.

4.1 Conjunto de Dados Utilizado

Os dados utilizados neste estudo de caso foram extraídos de um sistema acadêmico da UFSCar e contêm registros completos de 660 estudantes do curso de Bacharelado em Ciência da Computação (BCC), matriculados em dez turmas consecutivas. O BCC passou por uma mudança em sua grade curricular em 2019 (UNIVERSIDADE FEDERAL DE SÃO CARLOS. DEPARTAMENTO DE COMPUTAÇÃO, 2018), sendo que, para este estudo, foi utilizada a grade anterior a esta mudança por conter maior quantidade de dados históricos disponíveis para análise, e não foram incluídos no estudo os estudantes que cursavam durante o período de transição de grades. Todos os estudantes considerados já encerraram seu ciclo acadêmico e têm desfecho acadêmico conhecido.

Todos os dados pessoais e identificadores foram anonimizados pela coordenação do curso, assegurando o cumprimento das normas de proteção de dados vigentes.

Três tabelas foram disponibilizadas, já estruturadas e normalizadas conforme boas práticas de modelagem relacional. As tabelas e seus respectivos atributos são descritos a seguir:

- **Tabela *estudante***: armazena dados identificadores e informações acadêmicas dos

estudantes.

  - `rafake` (inteiro): identificação anonimizada do estudante.
  - `nomefake` (texto): nome anonimizado do estudante.
  - `anoingresso` (inteiro): ano de ingresso anonimizado na universidade.
  - `semestreingresso` (inteiro): semestre de ingresso no ano.
  - `status` (texto): desfecho acadêmico, podendo assumir os valores: "Formado", "Evadido", "Transferência Interna", "Transferência Externa", "Jubilado", "Perda de Vaga Rematrícula", "Perda de Vaga Desempenho Mínimo", "Cancelado".
  - `anoegresso` (inteiro): ano de saída do estudante (anonimizado).
  - `semestreegresso` (inteiro): semestre de saída dentro do ano.

- **Tabela *disciplina***: armazena os dados das disciplinas oferecidas no curso.
  - `codigo` (inteiro): código identificador da disciplina.
  - `nome` (texto): nome da disciplina.
  - `créditos` (inteiro): número de créditos integralizadores da disciplina.
  - `tipo` (texto): tipo da disciplina, podendo ser "Obrigatória", "Optativa" ou *null*.

- **Tabela *matricula***: registra o histórico de matrícula dos estudantes nas disciplinas ao longo dos períodos.
  - `rafake` (inteiro): identificação anonimizada do estudante.
  - `codigo_disciplina` (inteiro): código identificador da disciplina.
  - `anofake` (inteiro): ano em que a disciplina foi cursada (anonimizado).
  - `semestre` (inteiro): semestre em que a disciplina foi cursada dentro do ano.
  - `nota` (numérico): nota final obtida pelo estudante.
  - `frequencia` (inteiro): frequência na disciplina.
  - `resultado` (texto): resultado da matrícula, com possíveis valores como "Aprovado", "Reprovado por Nota", "Reprovado por Frequência", "Reprovado por Nota e Frequência", "Desistente", "Cancelado", "Reconhecido", "Suspenso".

Adaptações para Suporte à Implementação

Os dados disponibilizados foram importados para um banco de dados do sistema gerenciador de banco de dados PostgreSQL versão 15.1[1]. Embora os dados tenham sido

[1] https://www.postgresql.org/

disponibilizados em formato normalizado, algumas adaptações foram realizadas para servir de suporte à implementação das técnicas propostas nesta dissertação. As modificações incluem:

- **Criação das tabelas auxiliares *classe* e *classe_estudante***: utilizadas para associar cada estudante a uma ou mais classes de agrupamento, conforme será detalhado na seção 4.4.
  - `classe`: `codigo` (inteiro), `nome` (texto).
  - `classe_aluno`: `rafake` (inteiro), `codigo_classe` (inteiro).
- **Criação das tabelas auxiliares *turma* e *turma_estudante***: utilizadas para associar os estudantes às turmas de ingresso no curso.
  - `turma`: `codigo` (inteiro), `nome` (texto).
  - `turma_estudante`: `rafake` (inteiro), `codigo_turma` (inteiro).
- **Criação da tabela auxiliar *instancia_temporal***: utilizada para dar suporte ao modelo de representação de instâncias temporais de estudantes utilizadas neste trabalho, conforme será apresentado na seção 4.2.
  - `instancia_temporal`: `codigo_instancia` (inteiro), `semestre` (inteiro), `vetor` vetor, `tipo` texto. O atributo tipo identifica o tipo da instância temporal, podendo ser estudante, classe ou turma. O atributo vetor utiliza uma tipagem específica, fornecida pela extensão PGVector [2], que possibilita o uso de dados vetoriais em um banco de dados relacional.
- **Adição do atributo `posicao_no_vetor`** na tabela *disciplina*: utilizado para dar suporte à maneira como serão modelados os estudantes neste estudo, conforme será apresentado na seção 4.2.
- **Adição do atributo `periodo`** na tabela *matricula*: representa o período acadêmico no qual o estudante cursou a disciplina, consolidando o ano e o semestre em um único identificador sequencial, para facilitar o ordenamento temporal das instâncias.

4.2 Representação Utilizada neste Estudo

A representação das instâncias temporais dos estudantes pode ser feita de diversas maneiras, conforme a técnica genérica proposta e detalhada na seção 3.2. Os dados disponibilizados para este estudo permitiriam representar o estudante, por exemplo, pelas

[2] https://github.com/pgvector/pgvector

notas em todas as disciplinas, pelas notas e frequência, pelas notas em disciplinas optativas, ou pelas notas em disciplinas obrigatórias.

Para fins de validação da abordagem e construção dos experimentos neste trabalho, optou-se por uma representação vetorial específica: as notas dos estudantes em disciplinas obrigatórias do curso de Bacharelado em Ciência da Computação da Universidade Federal de São Carlos (UFSCar). Cada vetor de características representa uma instância temporal do estudante ao final de um semestre, contendo a média final de todas as disciplinas obrigatórias cursadas até aquele momento. O vetor de características tem dimensão fixa de 44 posições, correspondendo à quantidade total de disciplinas obrigatórias do curso. A Tabela 3 apresenta a ordem e os nomes das disciplinas considerados na formação dos vetores.

Tabela 3 – Disciplinas obrigatórias e respectivas posições no vetor de características

Posição	**Disciplina**
0	CÁLCULO 1
1	CONSTRUÇÃO DE ALGORITMOS E PROGRAMAÇÃO
2	FUNDAMENTOS DE FÍSICA PARA A COMPUTAÇÃO
3	GEOMETRIA ANALÍTICA
4	INTRODUÇÃO À LÓGICA
5	ORIENTAÇÃO PROFISSIONAL EM COMPUTAÇÃO
6	ÁLGEBRA LINEAR 1
7	CÁLCULO DIFERENCIAL E SÉRIES
8	CIRCUITOS DIGITAIS
9	ESTRUTURAS DISCRETAS
10	INTRODUÇÃO A PROBABILIDADE 1
11	LABORATÓRIO DE CIRCUITOS DIGITAIS
12	PROGRAMAÇÃO DE COMPUTADORES
13	ARQUITETURA E ORGANIZAÇÃO DE COMPUTADORES 1
14	CÁLCULO NUMÉRICO
15	ESTRUTURAS DE DADOS
16	INTRODUÇÃO AOS SISTEMAS DE INFORMAÇÃO
17	LABORATÓRIO DE ARQUITETURA E ORGANIZAÇÃO DE COMPUTADORES 1
18	ARQUITETURA E ORGANIZAÇÃO DE COMPUTADORES 2
19	BANCO DE DADOS
20	ENGENHARIA DE SOFTWARE 1
21	LABORATÓRIO DE ARQUITETURA E ORGANIZAÇÃO DE COMPUTADORES 2

Continua na próxima página...

Tabela 3 – Disciplinas obrigatórias e respectivas posições no vetor de características

Posição	**Disciplina**
22	LINGUAGENS FORMAIS E AUTÔMATOS
23	ORGANIZAÇÃO E RECUPERAÇÃO DA INFORMAÇÃO
24	PROJETO E ANÁLISE DE ALGORITMOS
25	TEORIA DOS GRAFOS
26	ADMINISTRAÇÃO DE EMPRESAS 1
27	CONSTRUÇÃO DE COMPILADORES 1
28	ENGENHARIA DE SOFTWARE 2
29	LABORATÓRIO DE BANCO DE DADOS
30	PARADIGMAS DE LINGUAGENS DE PROGRAMAÇÃO
31	PROJETO ACADÊMICO EM COMPUTAÇÃO
32	SISTEMAS OPERACIONAIS 1
33	COMPUTAÇÃO GRÁFICA E MULTIMÍDIA
34	CONSTRUÇÃO DE COMPILADORES 2
35	INTELIGÊNCIA ARTIFICIAL
36	METODOLOGIAS DE DESENVOLVIMENTO DE SISTEMAS 1
37	REDES DE COMPUTADORES
38	SISTEMAS OPERACIONAIS 2
39	DESENVOLVIMENTO DE SOFTWARE PARA A WEB
40	SISTEMAS DISTRIBUÍDOS
41	ESTÁGIO EM COMPUTAÇÃO
42	SEMINÁRIOS EM COMPUTAÇÃO
43	TRABALHO DE GRADUAÇÃO

Conclusão da Tabela 3

Fonte: Elaborada pelo autor

A maneira como os estudantes são representados neste estudo, com vetores de dimensão fixa, possibilita a representação diretamente em um espaço vetorial. Por esta razão, não é necessário realizar qualquer mapeamento adicional para a execução das operações de sumarização de grupos ou consulta por similaridade. As operações podem ser realizadas diretamente nesse espaço vetorial, com as distâncias entre as instâncias sendo calculadas a partir das métricas apropriadas sobre os vetores originais.

Diferentemente de cenários em que os dados são puramente métricos — exigindo mapeamento para um espaço multidimensional — nesta modelagem, as instâncias já estão aptas a serem processadas diretamente nas etapas de consulta, sumarização e projeção para visualização.

4.2.1 Exemplo de Instâncias Temporais

A Tabela 4 ilustra três exemplos de instâncias temporais de um estudante representadas por vetores de características. Os valores correspondem às médias finais nas disciplinas obrigatórias cursadas em cada posição do tempo.

Tabela 4 – Representação da instância temporal de um estudante em vetores de características

Estudante	Semestre	Vetor de Características
100009	1	[8, 8.7, 9.1, 9.2, 8.7, 8.4, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
100009	2	[8, 8.7, 9.1, 9.2, 8.7, 8.4, 9.2, 7.6, 8.9, 7.3, 8.3, 8.5, 8.4, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
100009	3	[8, 8.7, 9.1, 9.2, 8.7, 8.4, 9.2, 7.6, 8.9, 7.3, 8.3, 8.5, 8.4, 9.6, 8, 9, 7.8, 9, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]

Fonte: Elaborada pelo autor

Ao longo do tempo, novos valores vão sendo adicionados às posições correspondentes à medida que o estudante avança no curso e novas disciplinas são cursadas. Com isso, forma-se uma trajetória acadêmica composta por um conjunto de vetores de características, permitindo a análise de sua evolução temporal.

Tratamento de Disciplinas Não Cursadas e Casos Especiais

- **Disciplinas não cursadas:** atribuída nota 0, mantendo a posição da disciplina no vetor.
- **Disciplinas cursadas antes da matrícula (aluno especial):** consideradas como pertencentes ao primeiro semestre do curso regular.
- **Disciplinas repetidas:** quando o estudante cursa mais de uma vez a disciplina, se a nota obtida for superior, a nota é atualizada a partir da instância temporal correspondente ao semestre em que cursou novamente. Se a nota for inferior, permanece a nota que estava.

Essa estratégia representa o acúmulo de desempenho do estudante até cada ponto de sua trajetória acadêmica.

4.3 Protótipo STVis

Para ilustrar a aplicação das técnicas propostas, foi desenvolvido um protótipo, STVis (*Student Trajectory Visualization*), que implementa funcionalidades para gerar contextos

exploratórios para a visualização de trajetórias dos estudantes. Sua implementação foi realizada no framework Apache NetBeans IDE 18 [3], utilizando a linguagem Java.

O sistema é composto por três módulos principais:

- **Importação e preparação dos dados**: esta etapa é executada apenas na primeira utilização do sistema. É responsável pela ingestão e processamento dos dados fornecidos pela instituição, estruturando as instâncias temporais dos estudantes. Também são geradas e armazenadas as trajetórias médias de classes e turmas de estudantes, que poderão ser utilizadas nas consultas e visualizações.

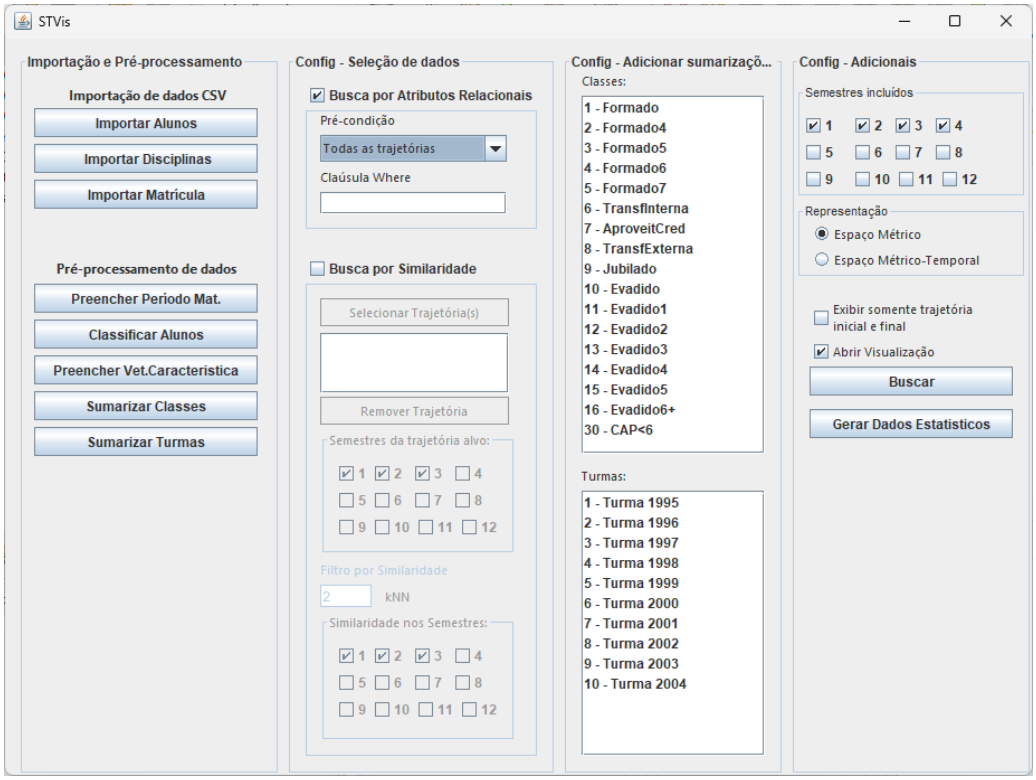
- **Configuração das consultas**: nesta etapa o usuário define os critérios da análise a ser realizada. Estão disponíveis as três estratégias de filtragens abordadas na seção 3.8. A seleção de trajetórias-alvo pode ser feita com base em estudantes individuais ou em grupos. É possível incluir no resultado as trajetórias sumarizadas de classes e turmas previamente definidas, o que auxilia na interpretação contextualizada do comportamento dos estudantes. Por fim, o usuário configura os parâmetros de visualização: seleção dos semestres a serem exibidos, escolha entre representação métrica ou métrico-temporal, e opção de ocultar instâncias intermediárias — exibindo apenas os pontos inicial e final de cada trajetória.

- **Visualização 3D interativa**: ambiente gráfico tridimensional em que as trajetórias são renderizadas e podem ser exploradas por meio de técnicas de interação. São suportadas funcionalidades como rotação, *zoom*, seleção e destaque de elementos. A interação direta com o ambiente gráfico permite ao analista realizar explorações visuais dinâmicas, facilitando a compreensão de padrões, semelhanças e anomalias nas trajetórias dos estudantes. Este ambiente de visualização foi implementado através de uma ferramenta desenvolvida no *framework* Unity 3D [4] em sua versão gratuita 2022.3.9.f1, como parte do protótipo. Este é um poderoso *framework* de fácil uso para criação de jogos em 3D e aplicações em realidade virtual para várias plataformas (LI, 2020).

A Figura 8 apresenta a interface do protótipo STVis.

[3] https://netbeans.apache.org/
[4] https://unity.com/pt

Figura 8 – Interface geral do protótipo STVis



Fonte: Elaborada pelo autor

O fluxo completo executado pelo STVis para a exibição das trajetórias no ambiente 3D segue as seguintes etapas:

1. As trajetórias selecionadas na consulta são recuperadas do banco de dados.
2. Cada instância temporal da trajetória, da maneira como é representada neste estudo, já está no espaço vetorial, então não é necessário o mapeamento para o espaço multidimensional. O protótipo realiza o cálculo das distâncias entre as instâncias no espaço métrico original dos dados.
3. A partir dessas distâncias, constrói-se a matriz de distância que servirá como entrada para a técnica de projeção multidimensional.
4. O algoritmo *Multidimensional Scaling* (MDS) clássico é então aplicado, realizando a projeção do espaço métrico para um espaço tridimensional.
5. Finalmente, os dados projetados são enviados ao ambiente de visualização interativa, permitindo a representação gráfica das trajetórias e sua exploração visual.

Embora o MDS clássico seja reconhecido por sua complexidade computacional (ORTIGOSSA; DIAS; NASCIMENTO, 2022), com ordem de processamento $O(N^3)$, no contexto deste estudo de caso, a quantidade de dados empregada, compreendendo as trajetórias de

660 estudantes e todas as suas instâncias temporais, demonstrou ser computacionalmente viável.

A execução da projeção foi realizada em um ambiente de desenvolvimento e teste padrão, apresentando tempos de processamento considerados aceitáveis para fins exploratórios. Além disso, a escolha do MDS se justifica pela sua capacidade de preservar as relações globais de dissimilaridade entre as instâncias temporais, garantindo que a visualização tridimensional reflita adequadamente as similaridades e dissimilaridades estruturais presentes no espaço métrico original.

4.4 Criação e Sumarização de Grupos de Estudantes

Com o objetivo de ampliar as possibilidades analíticas deste estudo de caso, foram criados grupos de estudantes com base em critérios acadêmicos, temporais e de desempenho. Esses grupos foram utilizados tanto para visualização quanto como suporte às estratégias de comparação e consulta.

Definição de Classes e Turmas de Estudantes

As classificações foram organizadas em dois tipos principais:

- **Classes de estudantes**, baseadas em critérios como tempo de conclusão, reprovações e tipo de desligamento. As classes criadas incluem:
  - **Formado**, **Formado4**, **Formado5**, **Formado6**, **Formado7** – estudantes que se formaram, agrupados também pelo número de anos até a conclusão;
  - **Evadido**, **Evadido1**, **Evadido2**, **Evadido3**, **Evadido4**, **Evadido5**, **Evadido6+** – estudantes evadidos, categorizados conforme o semestre de saída;
  - **Cap<6** – estudantes que obtiveram nota inferior a 6 na disciplina "Construção de Algoritmos e Programação";
  - **TransfInterna**, **TransfExterna**, **Jubilado**, **AproveitCred**.

  Para as classes de evasão, foram considerados estudantes com status "Perda de Vaga Rematrícula", "Perda de Vaga Desempenho Mínimo" ou "Cancelado" no banco de dados.

- **Turmas de ingresso**, definidas com base no atributo `anoingresso` dos estudantes. Foram criadas dez turmas consecutivas: *Turma 1995* até *Turma 2004*, com os anos sendo anonimizados para preservar a identidade institucional.

As informações de associação entre estudantes, classes e turmas foram estruturadas em tabelas auxiliares no banco de dados:

- `classe(codigo, nome)`
- `classe_aluno(rafake, codigo_classe)`
- `turma(codigo, nome)`
- `turma_aluno(rafake, codigo_turma)`

Um mesmo estudante pode pertencer a múltiplas classes simultaneamente. Por exemplo, um aluno pode ser classificado como `Formado`, `Formado5` e `Cap<6`, indicando que se formou em 5 anos e teve baixo desempenho na disciplina de CAP. As classes representam diferentes características de interesse sobre o percurso formativo dos estudantes.

Além das classes pré-definidas, o método permite a criação de novas classes personalizadas em tempo de execução, com base em qualquer critério definido pelo analista, ampliando a flexibilidade e o potencial de exploração dos dados.

Processo de Sumarização de Grupos

Com os grupos estabelecidos, foi realizado o processo de **sumarização de trajetórias** conforme descrito na Seção 3.5. A trajetória média de um grupo foi construída a partir da agregação das instâncias temporais dos estudantes que o compõem. Para cada semestre letivo, calculou-se a *instância temporal média* utilizando a média aritmética dos vetores de características numéricas (notas em disciplinas obrigatórias). Destaca-se que, nesta aplicação, a representação dos estudantes já corresponde a um espaço vetorial de dimensão fixa, o que possibilita que a operação de sumarização seja realizada diretamente sobre as representações originais, sem necessidade de mapeamento prévio para um espaço multidimensional intermediário.

A sumarização foi aplicada tanto para as classes de estudantes quanto para as turmas. A partir disso, foi possível gerar visualizações que comparam trajetórias individuais com trajetórias médias de classes ou turmas. Essa comparação visual permite identificar, por exemplo, estudantes que estão distantes da média de sua turma, o que pode indicar situações de risco e motivar ações de acompanhamento.

A sumarização foi pré-calculada e armazenada no banco de dados, de forma a estar disponível para inclusão em qualquer consulta no protótipo STVis. Como exemplo, a Tabela 5 mostra a instância temporal média da classe `Formado4` no primeiro semestre.

Tabela 5 – Representação da classe sumarizada `Formado4` no semestre 1

Classe	Semestre	Vetor de Características
Formado4	1	[6.16, 7.39, 6.69, 6.72, 7.23, 8.4, 0, 0, 0, 0, 0, 0, 0.06, 0.11, 0, 0.06, 0.21, 0.06, 0.05, 0, 0, 0.07, 0, 0, 0, 0, 0, 0, 0, 0, 0.06, 0.06, 0.02, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]

Fonte: Elaborada pelo autor

Este vetor representa a média das notas dos estudantes da classe `Formado4` nas disciplinas obrigatórias consideradas até o primeiro semestre. Disciplinas não cursadas naquele momento são representadas com zero.

Com esse recurso, é possível explorar visualmente a convergência ou divergência de estudantes em relação a perfis de referência, enriquecendo a análise dos dados acadêmicos e apoiando a tomada de decisões educacionais.

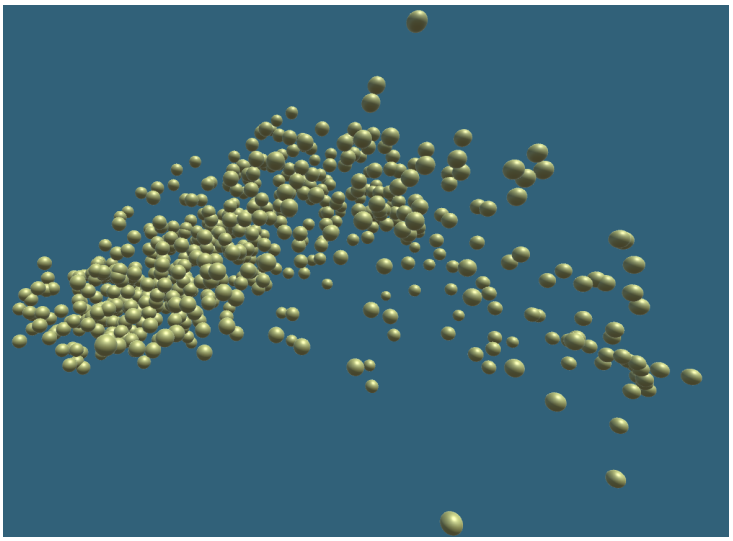
4.5 Visualizações Iniciais Geradas

Com o objetivo de validar a implementação do método proposto e ilustrar as possibilidades visuais oferecidas, foram geradas algumas visualizações iniciais utilizando dados reais dos estudantes. Essas visualizações demonstram como as instâncias temporais, trajetórias individuais e trajetórias sumarizadas são representadas graficamente no ambiente tridimensional.

4.5.1 Instância Temporal dos estudantes

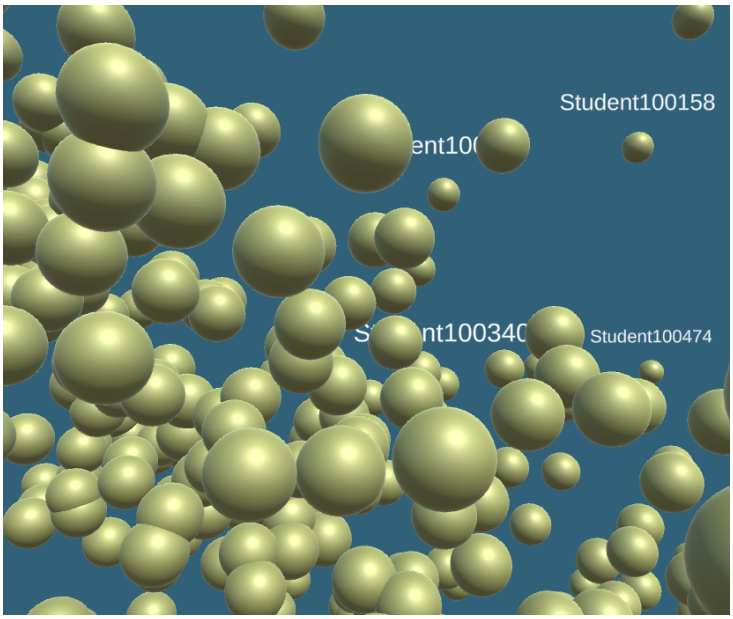
A Figura 9 apresenta a representação de instâncias temporais de todos os estudantes do banco de dados, correspondente ao primeiro semestre no curso. Essas instâncias são exibidas como esferas coloridas posicionadas no espaço tridimensional. Como parte do método proposto para a exploração visual, o ambiente de visualização incorpora técnicas de interação que permitem ao usuário manipular a perspectiva e o foco da análise. A Figura 10 ilustra o cenário da Figura 9 rotacionado e expandido.

Figura 9 – Visualização de instâncias temporais individuais de cada estudante após concluir o primeiro semestre.



Fonte: Elaborada pelo autor

Figura 10 – Visualização ampliada e rotacionada das instâncias temporais individuais de cada estudante do banco de dados após concluir o primeiro semestre. Ao passar o *mouse* sobre a esfera, é possível identificar o estudante que está sendo representado

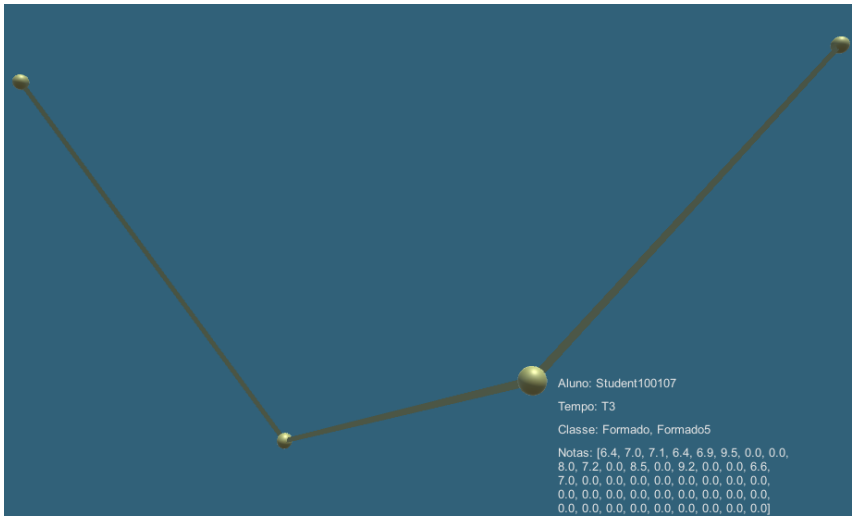


Fonte: Elaborada pelo autor

4.5.2 Trajetória de Estudante

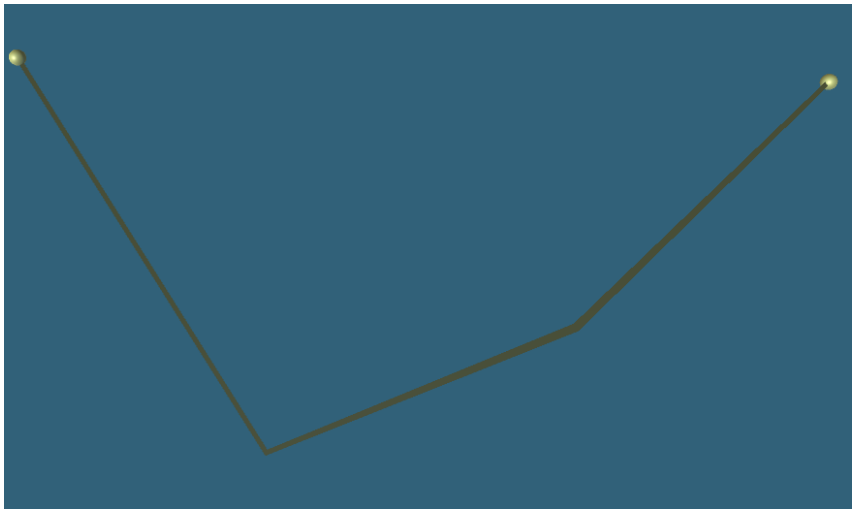
Já utilizando uma das estratégias de filtragem para seleção de trajetória, a Figura 11 apresenta uma trajetória composta por quatro instâncias temporais consecutivas do estudante 100107. Cada esfera colorida representa um semestre letivo e as linhas conectando as esferas indicam a evolução temporal. Essa visualização permite observar o progresso acadêmico individual ao longo do tempo. Em alguns cenários, apresentar todas as instâncias temporais pode levar a uma visualização complexa e confusa, especialmente quando se comparam muitas trajetórias simultaneamente. O excesso de informação (esferas e linhas) pode dificultar a identificação de padrões e tendências gerais. Em face disso, o protótipo oferece a opção de esconder as esferas nos tempos intermediários, exibindo somente a esfera inicial e final das trajetórias. A Figura 12 ilustra a mesma trajetória do estudante 100107 sem a exibição das instâncias temporais intermediárias.

Figura 11 – Trajetória simples de um estudante ao longo de quatro semestres.



Fonte: Elaborada pelo autor

Figura 12 – Trajetória simples de um estudante ao longo de quatro semestres exibindo somente as instancias temporais inicial e final.

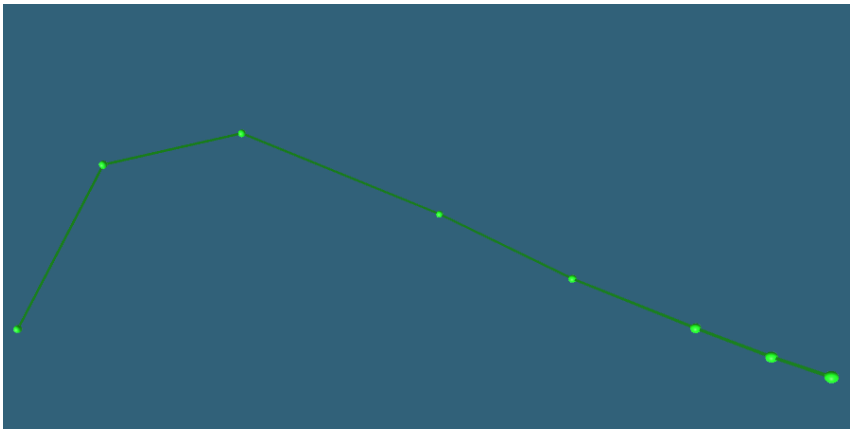


Fonte: Elaborada pelo autor

4.5.3 Trajetória Média de Classe Sumarizada

A Figura 13 exibe a trajetória média da classe **Formado4**, que representa estudantes que concluíram o curso em quatro anos. Essa trajetória é composta pela média das instâncias de todos os estudantes da classe em cada semestre letivo, conforme descrito na Seção 3.5. A trajetória média pode ser utilizada como referência para comparações com trajetórias individuais ou de outras classes.

Figura 13 – Trajetória média da classe Formado4.



Fonte: Elaborada pelo autor

4.6 Considerações Finais

Este capítulo apresentou a implementação prática do método proposto para visualização de trajetórias acadêmicas, utilizando dados reais fornecidos pela Universidade Federal de São Carlos. O conjunto de dados foi detalhado, incluindo as transformações realizadas para adequação à modelagem adotada. A representação das instâncias temporais, tanto individuais quanto de grupos, foi realizada por meio de vetores de características organizados a partir das disciplinas obrigatórias do curso.

A protótipo desenvolvido contempla todas as etapas propostas na Capítulo 3 em uma ferramenta funcional e interativa. O sistema implementa diferentes estratégias de consulta — incluindo filtros relacionais, buscas por similaridade métrica e métrico-temporal, bem como abordagens híbridas — além de recursos para a criação e sumarização de grupos de estudantes.

O próximo capítulo apresentará os cenários exploratórios realizados a partir da ferramenta desenvolvida, ilustrando o potencial analítico da abordagem proposta e os tipos de interpretações que podem ser extraídas a partir da visualização das trajetórias no contexto educacional.

Capítulo 5

Cenários Exploratórios e Resultados

Este capítulo apresenta a aplicação e validação das técnicas de representação e análise visual de trajetórias acadêmicas propostas no Capítulo 3.

5.1 Considerações iniciais

Neste capítulo, serão apresentados diversos cenários exploratórios para demonstrar a capacidade do método em gerar *insights* relevantes sobre o desempenho e o percurso dos estudantes, utilizando os dados históricos do curso de Bacharelado em Ciência da Computação da Universidade Federal de São Carlos, conforme detalhado no Capítulo 4.

A ênfase dos exemplos está em evidenciar a capacidade da técnica de representar, projetar e analisar visualmente a evolução acadêmica dos estudantes (representadas por dados complexos), segundo a sua similaridade. A análise de trajetórias pode ser realizada de estudantes individuais ou em grupo sumarizados, filtradas por meio de diferentes estratégias de consultas e projetadas em um ambiente visual onde é possível avaliar a similaridade.

Como o conjunto de dados utilizado já possui os desfechos acadêmicos definidos para todos os estudantes, é possível validar as inferências obtidas pela análise visual, comparando com o que de fato aconteceu e assim fortalecendo as evidências da eficácia da abordagem.

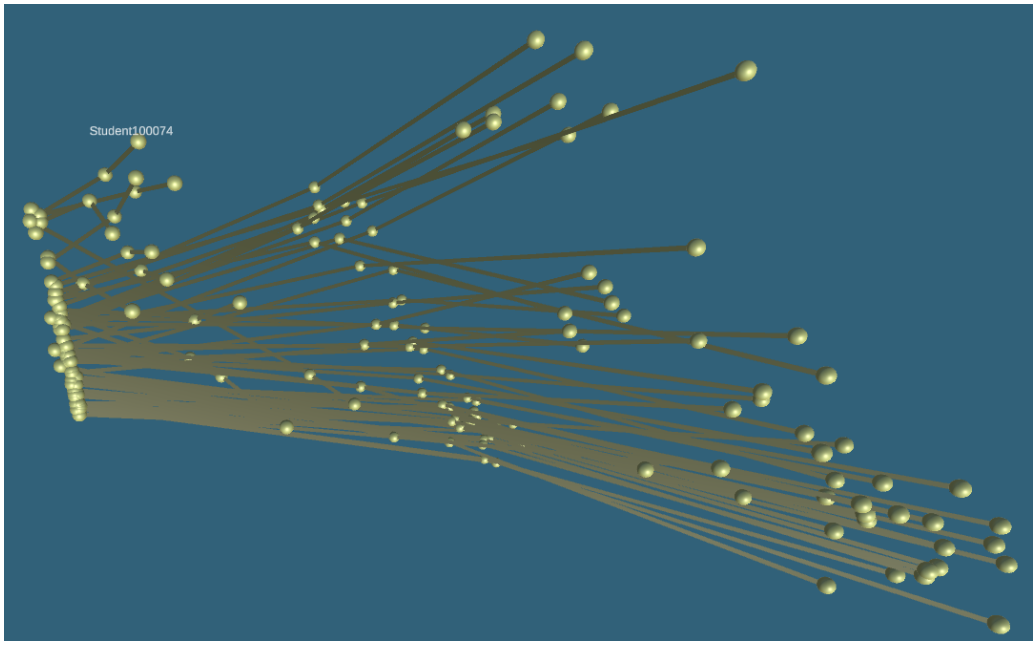
As Figuras apresentadas ao longo deste capítulo são capturas de tela do protótipo desenvolvido, mas o foco da discussão reside na interpretação dos padrões e informações possíveis de serem reveladas pela aplicação das técnicas.

5.2 Cenário 1 — Análise Geral de uma Turma e Identificação de Estudante Atípico

Este primeiro cenário busca ilustrar o uso da técnica de visualização de trajetórias para explorar o comportamento coletivo de uma turma ao longo dos semestres iniciais do curso, bem como identificar estudantes cujas trajetórias se desviam do padrão predominante.

Inicialmente, foi realizado um filtro para selecionar todas as trajetórias dos estudantes da turma de 1996[1] nos três primeiros semestres letivos. O resultado é exibido na Figura 14 onde é possível observar a configuração espacial das trajetórias e identificar agrupamentos visuais que indicam padrões de progressão similares. Trata-se de uma análise geral do comportamento da turma, adequada para um primeiro nível de exploração das trajetórias acadêmicas.

Figura 14 – Visão geral da turma de 1996 nos três primeiros semestres do curso.



Fonte: Elaborada pelo autor

Durante a exploração, observou-se que o estudante 100074 apresentou uma trajetória com deslocamento significativo em relação ao núcleo principal da turma, caracterizando um comportamento atípico.

Para investigar a possível evolução futura desse estudante, aplicou-se uma consulta por similaridade de trajetória (conforme Seção 3.8.2), considerando seus dados até o terceiro semestre como trajetória-alvo e buscando em todo o banco de dados histórico. A consulta

[1] Os dados apresentados neste artigo foram anonimizados para proteger a privacidade dos estudantes. O ano de 1996 é utilizado como um rótulo genérico e não corresponde ao ano real que representa a turma dos estudantes.

retornou as k trajetórias mais semelhantes, permitindo comparar a trajetória do estudante com a de outros que seguiram percursos similares nos mesmos períodos.

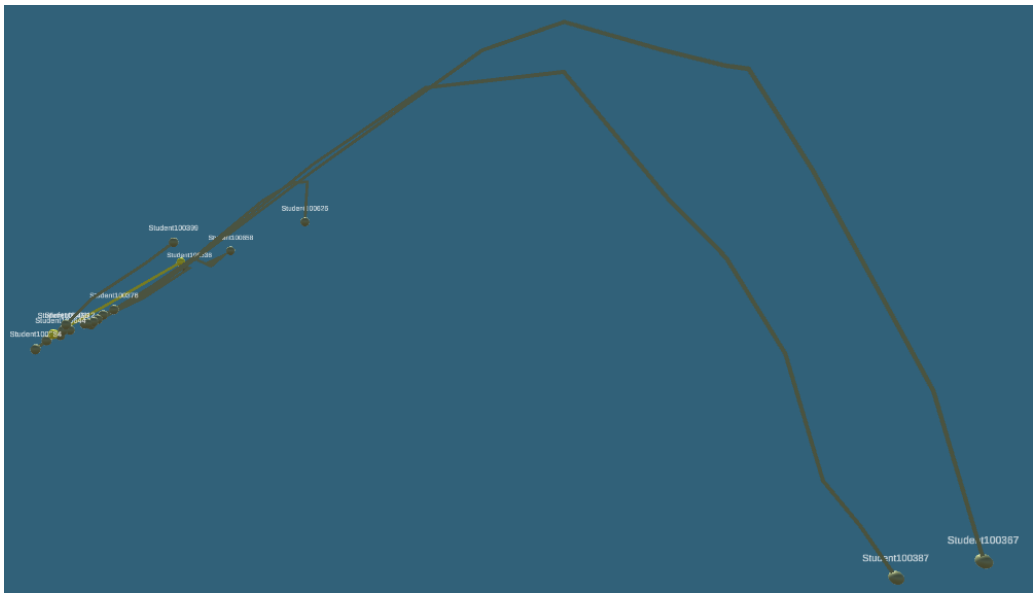
O resultado é apresentado na Figura 15 onde a inspeção visual mostrou que a maior parte das trajetórias similares pertencia a estudantes que, posteriormente, se evadiram do curso. Essa observação sugere que o estudante analisado também pode estar em risco de evasão. Como os dados utilizados são históricos, foi possível verificar que o estudante, de fato, se evadiu no oitavo semestre, validando a inferência realizada a partir da análise visual. Estatisticamente, das 11 trajetórias retornadas na consulta, 9 (82%) são de estudantes com desfecho de evasão.

Figura 15 – Trajetória do estudante 100074 até o 3º semestre e as 5-NN tratórias

Fonte: Elaborada pelo autor

A Figura 16 ilustra o mesmo cenário com *zoom* reduzido para abranger todas as trajetórias retornadas pela consulta, onde apenas as duas trajetórias mais longas e afastadas são de estudantes formados.

Figura 16 – Trajetória do estudante 100074 até o 3º semestre e as 5-NN tratórias com *zoom* reduzido



Fonte: Elaborada pelo autor

Este cenário demonstra como a técnica pode ser utilizada para:

- ❏ Analisar o comportamento coletivo de uma turma;
- ❏ Identificar estudantes cujas trajetórias se afastam do padrão da turma;
- ❏ Utilizar consultas por similaridade para inferir tendências futuras de estudantes ativos comparando-os com estudantes que já possuem desfecho conhecido;
- ❏ Validar inferências com base em desfechos históricos reais.

5.3 Cenário 2 — Criação de Classes Personalizadas para Validação de Hipóteses

Um dos diferenciais da abordagem proposta é a possibilidade de criação de grupos personalizados de estudantes com base em critérios definidos pelo analista, permitindo a modelagem e a análise de trajetórias médias desses grupos. Neste cenário, essa funcionalidade é utilizada para investigar uma hipótese específica: *estudantes que reprovam na disciplina CAP (Construção de Algoritmos e Programação) nos primeiros semestres apresentam maior risco de evasão.*

Para testar essa hipótese, foi criada uma classe personalizada denominada **CAP<6**, composta por estudantes cuja nota final em CAP foi inferior a 6.0, caracterizando reprovação. Com base nos estudantes classificados nesta classe, foi calculada a trajetória

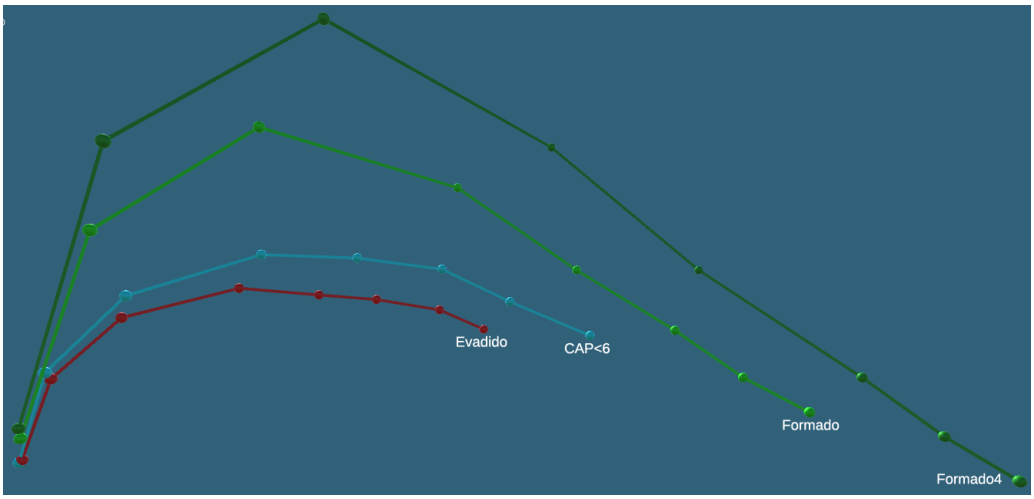
média do grupo por meio da agregação das instâncias temporais de cada semestre letivo (conforme o método descrito na Seção 3.5).

Em seguida, para acrescentar informações ao contexto exploratório, foram incluídas na visualização as trajetórias sumarizadas das seguintes classes:

- **Formado**: trajetória média de todos os estudantes que concluíram o curso;
- **Formado4**: trajetória média de formandos em 4 anos, considerada referência de desempenho ótimo;
- **Evadido**: trajetória média de todos os estudantes que se evadiram do curso;
- **CAP<6**: trajetória média dos estudantes reprovados em CAP.

A Figura 17 apresenta a visualização das quatro trajetórias sumarizadas. Observa-se que a trajetória da classe **CAP<6** se posiciona mais próxima da classe **Evadido** do que das classes **Formado** ou **Formado4**, reforçando visualmente a hipótese investigada.

Figura 17 – Trajetórias sumarizadas das classes Formado, Formado4, CAP<6 e Evadido



Fonte: Elaborada pelo autor

Além disso, foi realizada uma análise estatística dos dados: apenas 33,5% dos estudantes que reprovaram em CAP concluíram o curso com êxito, enquanto 66,5% se evadiram. Essa evidência numérica, associada à configuração espacial da visualização, sugere que a reprovação em CAP pode ser um indicador relevante de risco acadêmico precoce.

Continuando a análise exploratória, foram incluídas ao contexto as trajetórias dos estudantes 100230 e 100237, utilizando seus dados até o quarto semestre do curso, com o objetivo de observar o alinhamento de cada estudante com os diferentes perfis médios. A Figura 18 ilustra o resultado onde é possível identificar que, no período analisado, a trajetória do estudante 100230 se posicionava entre as classes **Formado** e **Formado4**, enquanto a do estudante 100237 encontrava-se entre as trajetórias **CAP<6** e **Evadido**.

Essa disposição sugere que o primeiro estudante mantinha um bom ritmo de progressão acadêmica, ao passo que o segundo apresentava sinais de risco de evasão.

Figura 18 – Trajetórias sumarizadas das classes Formado, Formado4, Cap<6 e Evadido

Fonte: Elaborada pelo autor

Os dados históricos confirmaram essa tendência: o estudante 100230 concluiu o curso em 6 anos, enquanto o estudante 100237 se evadiu no 5º semestre.

Este cenário exemplifica a capacidade do método em:

- ❏ Criar classes personalizadas baseadas em atributos de interesse do analista;
- ❏ Sumarizar trajetórias de grupos e compará-las visualmente com outras classes;
- ❏ Validar hipóteses por meio da proximidade entre trajetórias médias;
- ❏ Combinar análise visual e estatística para fundamentar conclusões;
- ❏ Explorar casos individuais em relação a perfis de grupo.

As informações obtidas neste cenário, baseadas em evidências extraídas das trajetórias visuais, podem ser utilizadas para embasar decisões institucionais, como ações de apoio pedagógico em disciplinas-chave, e para a construção de estratégias de acompanhamento acadêmico.

5.4 Cenário 3 — Consulta por Similaridade de Trajetória com Comparação a Classes Sumarizadas

O objetivo deste cenário é realizar uma consulta exploratória e analisar a trajetória individual de estudantes, comparando com as trajetórias vizinhas e de classes sumarizadas.

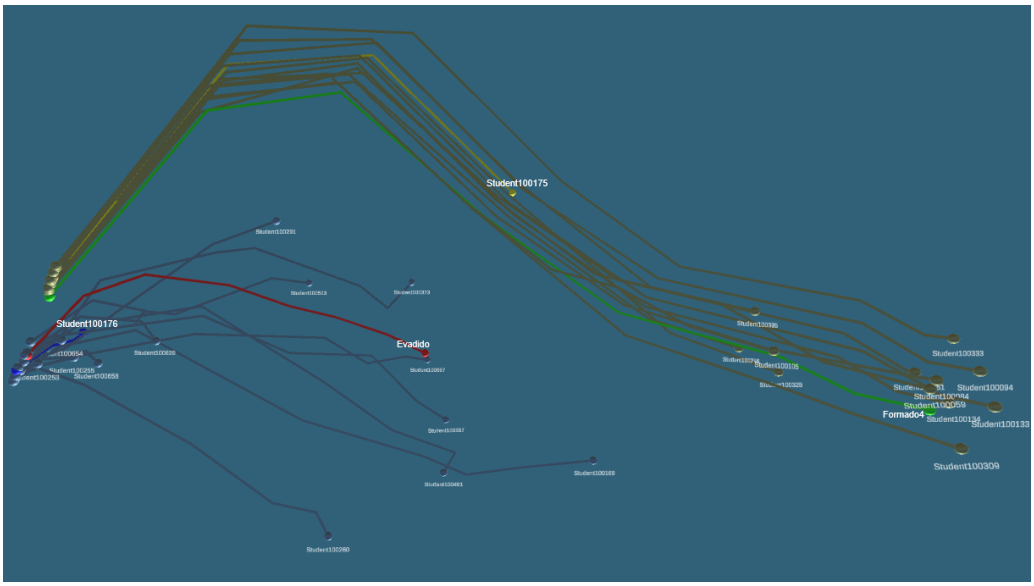
A proposta é demonstrar como a abordagem permite avaliar possíveis desfechos futuros com base na trajetória percorrida até um ponto intermediário do curso.

A exploração se inicia selecionando as trajetórias dos estudantes **100175** e **100176** considerando os dados da trajetória destes estudantes até o quarto semestre do curso. Em seguida, foi aplicada uma consulta por similaridade de trajetória, utilizando a função de distância métrica para recuperar as **5-NN** trajetórias mais próximas de cada um dos estudantes.

Para adicionar mais informações ao contexto exploratório, foram incluídas na visualização as trajetórias médias das classes **Formado4** (formados em 4 anos) e **Evadido** (estudantes que abandonaram o curso), representando os perfis sucesso e insucesso acadêmico, respectivamente.

A Figura 19 apresenta o resultado da consulta, onde as trajetórias dos estudantes-alvo 100175 e 100176 são destacadas nas cores **amarela** e **azul**, enquanto suas trajetórias vizinhas aparecem em **dourado** e **azul claro**, respectivamente. As trajetórias sumarizadas das classes **Formado4** e **Evadido** são representadas pelas cores **verde** e **vermelha**. Ressalta-se que as trajetórias-alvo foram consideradas até a conclusão do quarto semestre do curso.

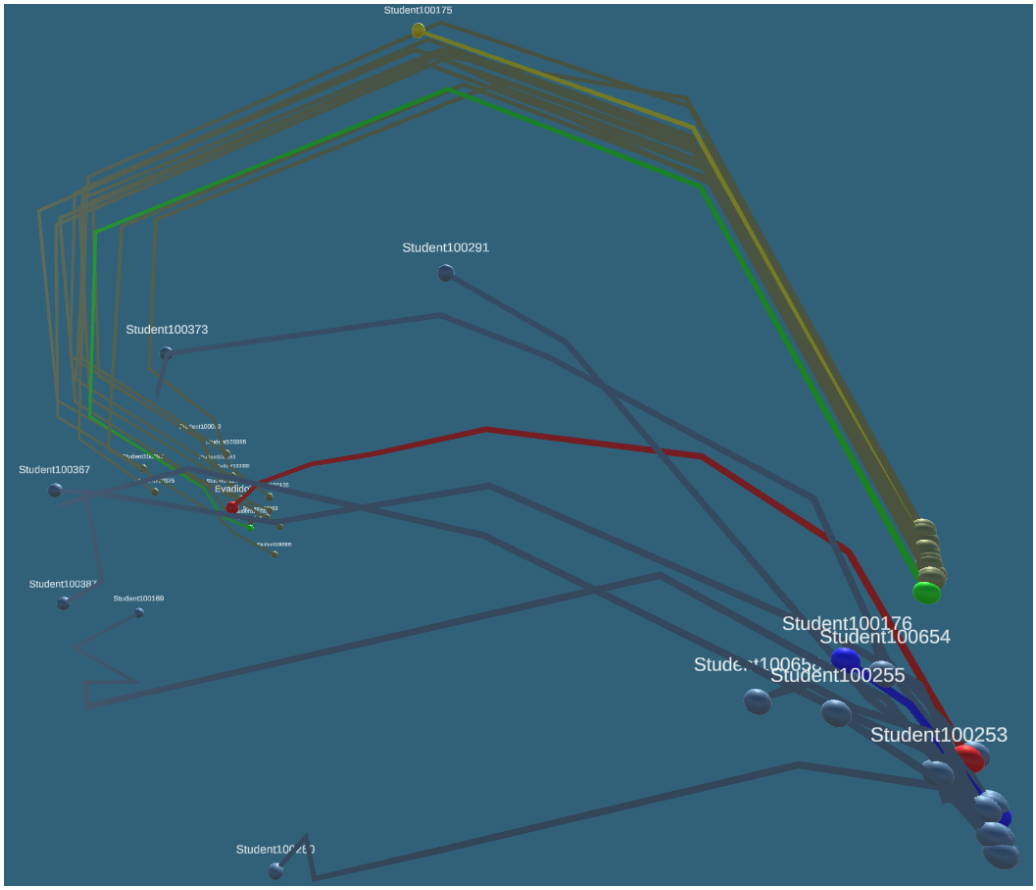
Figura 19 – Análise comparativa da trajetória de 2 estudantes-alvo no início do quinto semestre com a tendência futura dos vizinhos e classes sumarizadas.



Fonte: Elaborada pelo autor

A Figura 20 ilustra o mesmo cenário em ângulo rotacionado e *zoom* ampliado para melhor análise da trajetória do estudante 100176.

Figura 20 – Análise do cenário por outro ângulo.



Fonte: Elaborada pelo autor

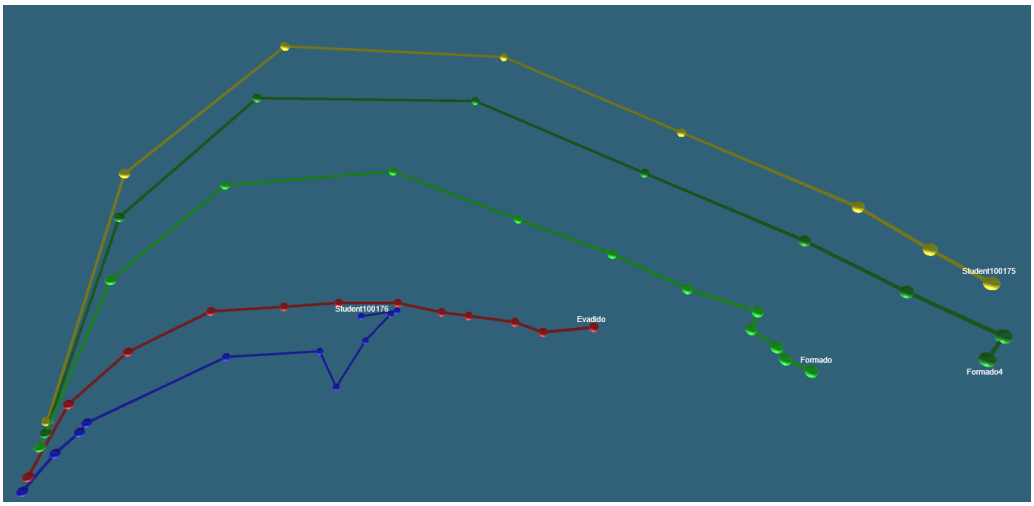
A inspeção visual revela que a trajetória do estudante **100175** e as trajetórias de seus vizinhos estão espacialmente próximas da trajetória da classe **Formado4**, sugerindo uma tendência de sucesso acadêmico. Em contrapartida, o estudante **100176** apresenta uma trajetória mais dispersa e próxima da classe **Evadido**, assim como a maioria de suas vizinhanças.

Uma análise quantitativa das trajetórias retornadas revela: entre as 12 trajetórias visualizadas, que são as mais semelhantes à do estudante 100175, 100% correspondem a estudantes que concluíram o curso com êxito. Já no caso do estudante 100176, 9 (69%) das 13 trajetórias similares visualizadas pertencem a estudantes que abandonaram o curso.

Ao consultar os dados históricos, confirma-se que o estudante **100175 concluiu o curso em 8 semestres**, enquanto o estudante **100176 se evadiu no 12º semestre**. A Figura 21 ilustra a trajetória completa dos estudantes. É importante notar que, embora a visualização da trajetória do estudante 100176 aparente conter 11 esferas, ela representa 12 instâncias temporais no total. Essa aparente discrepância visual ocorre devido à sobreposição da esfera referente ao 10º semestre com a do 11º, indicando que o estudante não apresentou variação em suas características relevantes entre esses dois períodos, possivelmente devido a situações como trancamento ou não cursar disciplinas que alterassem

seu vetor de características.

Figura 21 – Trajetória completa dos estudantes 100175 e 100176 e das classes sumarizadas.



Fonte: Elaborada pelo autor

Este cenário ilustra a aplicabilidade da técnica de consulta por similaridade de trajetória na comparação de perfis acadêmicos e na identificação de tendências de evolução. A proximidade de trajetórias pode indicar similaridades de comportamento ao longo do tempo, contribuindo para:

- ❏ Antecipação de prováveis desfechos acadêmicos com base em padrões históricos;
- ❏ Identificação de trajetórias alinhadas a perfis de sucesso ou risco;
- ❏ Apoio a ações de monitoramento e intervenção individualizada.

A análise reforça a utilidade da abordagem proposta como instrumento exploratório de apoio à gestão acadêmica e à pesquisa em dados educacionais.

5.5 Cenário 4 — Comparação entre Representações Métrica e Métrico-Temporal

Este cenário tem como objetivo demonstrar como o componente temporal pode influenciar na busca por similaridade de trajetória. Para isso, foram comparados os resultados obtidos ao aplicar a consulta por similaridade de trajetória utilizando duas formas distintas de representação: a **representação métrica** e a **representação métrico-temporal**, conforme discutido na Seção 3.4.1.

A representação métrica leva em conta exclusivamente a distância entre os dados que compõem as instâncias temporais, ignorando a posição temporal das mesmas no cálculo

de similaridade. Já a representação métrico-temporal incorpora a posição temporal como parte do cálculo de distância, de forma que o tempo passa a influenciar a medida de similaridade entre trajetórias. Ambas as abordagens são válidas e úteis em diferentes contextos analíticos e a escolha entre elas depende do objetivo da análise e das características dos dados.

Para o exemplo, foi selecionado o estudante **100509**, que apresentava percurso acadêmico fora do esperado nos semestres iniciais com o objetivo de tentar identificar seu provável desfecho ao compará-lo com trajetórias similares. A trajetória do 100509 até o **3º semestre** foi utilizada como base para a busca por estudantes com trajetórias similares.

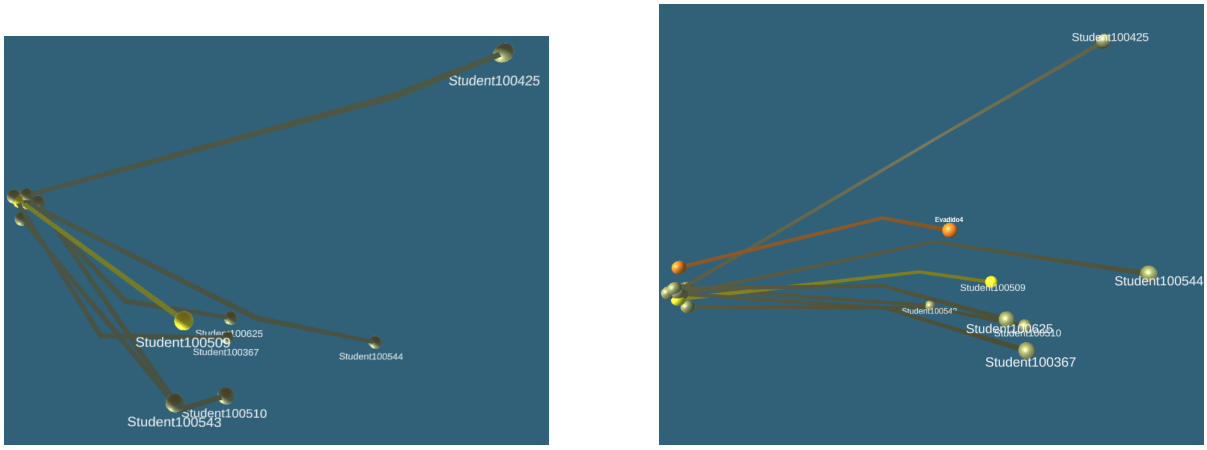
A consulta foi realizada em duas versões:

1. Utilizando apenas a **representação métrica**;

2. Utilizando a **representação métrico-temporal**, com pesos $\alpha = 0.042$ e $\beta = 0.125$. Estes pesos foram calculados conforme proposto em Bueno et al. (2009). Para isso, foram inicialmente determinadas as *dimensões intrínsecas* de cada componente do espaço: o componente métrico (representado por vetores de características com notas em disciplinas obrigatórias) e o componente temporal (representado pela sequência de semestres letivos). A dimensão intrínseca do componente métrico foi estimada utilizando a técnica do *Distance Exponent* (TRAINA; TRAINA; FALOUTSOS, 2000), resultando em um valor aproximado de 2,5. Já para o componente temporal, cuja representação é unidimensional, adotou-se o valor 1 como sua dimensão intrínseca. Com as dimensões intrínsecas estimadas, aplicou-se o processo de normalização para obter os pesos α e β. A normalização foi feita dividindo a dimensão intrínseca pela maior distância observada entre as instâncias do respectivo componente. Para o componente métrico, encontrou-se a maior distância euclidiana observada entre pares de instâncias, igual a 59,47, resultando em $\alpha = \frac{2,5}{59,47} \approx 0,042$. Para o componente temporal, a maior distância entre semestres foi considerada como 8, resultando em $\beta = \frac{1}{8} = 0,125$. Esses valores normalizados foram então utilizados na equação da representação métrico-temporal para garantir uma contribuição balanceada entre os dois componentes.

Em ambas as versões, foram retornadas as **4-NN** trajetórias mais similares ao estudante-alvo. A Figura 22 ilustra os resultados lado a lado.

Figura 22 – Comparação entre resultados da consulta por similaridade métrica (à esquerda) e métrico-temporal (à direita) para o estudante 100509.



Fonte: Elaborada pelo autor

A inspeção visual do resultado mostra que a **classe sumarizada "Evadido4"** foi recuperada na consulta com representação métrico-temporal (trajetória laranja na visualização), mas não apareceu na consulta puramente métrica. A diferença obtida é devida à influência do componente temporal considerado no espaço métrico-temporal e a presença dessa classe sugere uma possível associação do estudante-alvo com o padrão de evasão no quarto semestre, o que de fato se confirmou ao verificar os dados históricos: o estudante se evade no quarto semestre.

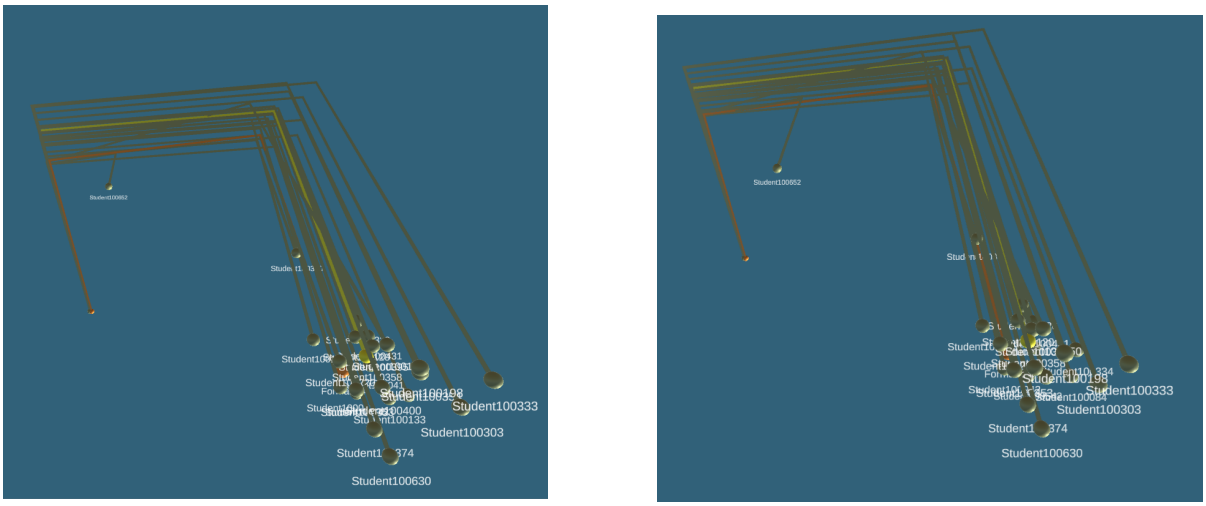
Este cenário demonstra que diferentes estratégias de modelagem e cálculo de similaridade podem produzir resultados distintos e complementares. Enquanto a abordagem métrica pura valoriza exclusivamente o conteúdo das instâncias, a abordagem métrico-temporal agrega contexto e ritmo evolutivo à análise.

Ambas as representações são compatíveis com o método proposto e sua escolha pode ser ajustada dinamicamente no sistema, conforme a necessidade do analista. A flexibilidade na definição de métricas de distância permite explorar a complexidade dos dados educacionais sob múltiplas perspectivas.

Durante a execução deste cenário, foi possível constatar que a utilização da abordagem métrico-temporal, no contexto deste trabalho, em que os estudantes são modelados por vetores de características com notas em disciplinas obrigatórias acumulativas ao longo do tempo, tende a produzir resultados similares àqueles obtidos com a distância somente métrica. A diferença entre as representações métrica e métrico-temporal se manifesta principalmente em casos atípicos, como estudantes com desempenho irregular ou que cursam disciplinas fora do período regular, situações em que o componente temporal exerce maior influência no cálculo da distância. Para trajetórias regulares, entretanto, a seleção das trajetórias em uma consulta por similares apresenta resultado bem próximo ou até idêntico, como é ilustrado na Figura 23, onde foi realizada uma consulta 10-

NN nas duas abordagens para a trajetória-alvo do estudante 100358 (Formado4) nos 4 primeiros semestres. Foram recuperadas 25 trajetórias idênticas, sendo a diferença observada apenas na posição espacial dos elementos na visualização, o métrico-temporal aproxima as distâncias devido à ponderação aplicada aos componentes métrico e temporal na abordagem métrico-temporal. Em outros contextos e modelagens de estudantes, pode ser que os resultados sejam diferentes.

Figura 23 – Comparação entre resultados da consulta por similaridade métrica (à esquerda) e métrico-temporal (à direita) para o estudante 100358.



Fonte: Elaborada pelo autor

5.6 Cenário 5 — Aplicação da Consulta Híbrida: Reprovação em CAP e Trajetórias Similares

Este cenário tem como objetivo demonstrar o uso da **consulta híbrida**, que combina filtros relacionais sobre atributos dos estudantes com a consulta por similaridade de trajetória. A proposta permite conduzir análises mais refinadas ao restringir o universo de busca com base em critérios específicos e, em seguida, explorar visualmente as trajetórias dos estudantes mais similares ao alvo definido.

A motivação deste cenário é investigar um caso específico com base na hipótese levantada na seção 5.3, de que estudantes que reprovam na disciplina Construção de Algoritmos e Programação nos primeiros semestres do curso apresentam maior risco de evasão. Considerando essa hipótese, busca-se avaliar o caso do estudante **100058**, que teve reprovação em CAP no início de sua trajetória acadêmica, levantando-se a possibilidade de que ele também estivesse em situação de risco.

Entretanto, em vez de avaliar o desfecho exclusivamente com base nesse evento isolado, a técnica proposta permite realizar uma análise contextualizada, observando a similaridade da trajetória do estudante com a de outros que também reprovaram em CAP. Dessa

forma, é possível verificar se o padrão de evolução do estudante 100058 se aproxima mais de perfis de evasão ou de sucesso acadêmico.

Para criar o contexto exploratório, o procedimento adotado foi o seguinte:

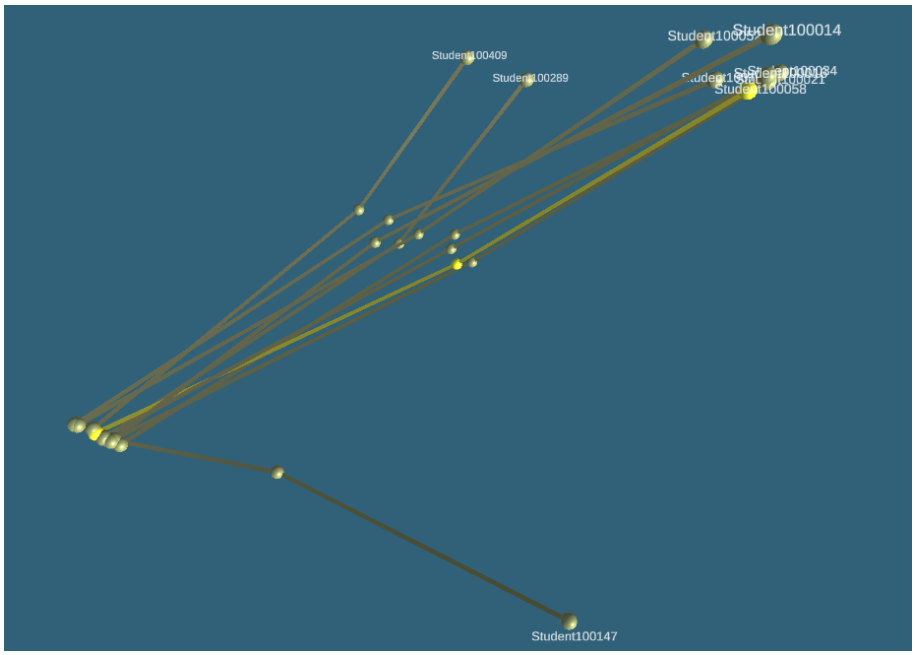
1. Inicialmente, foi aplicado um **filtro relacional** para selecionar todos os estudantes que **reprovaram em CAP**.

2. Em seguida, sobre o conjunto filtrado, foi aplicada uma **consulta por similaridade de trajetória**, selecionando como alvo a trajetória do estudante 100058, utilizando seus dados até o **3º semestre**.

3. A consulta foi realizada com a configuração de **5-NN** (cinco vizinhos mais próximos) utilizando a representação métrica.

A Figura 24 apresenta a visualização resultante da consulta híbrida. A trajetória do estudante-alvo é destacada na cor **amarela**, e as demais trajetórias pertencem a estudantes que também reprovaram em CAP.

Figura 24 – Consulta híbrida: comparação entre o estudante 100058 (em amarelo) e estudantes reprovados em CAP com trajetórias similares.



Fonte: Elaborada pelo autor

A visualização mostra que, apesar da reprovação em CAP, o estudante 100058 apresenta uma trajetória similar à de diversos estudantes que **obtiveram sucesso na conclusão do curso**. Entre os estudantes mais similares ao 100058 recuperados, apenas um (representado pela trajetória mais dispersa) apresentou desfecho de evasão. Os demais

alcançaram a formatura. Isso sugere que, embora a reprovação em CAP seja estatisticamente associada a maior risco de evasão (como discutido no Cenário 2), o fato da trajetória do estudante-alvo analisado estar mais próxima da trajetória de estudantes formados, ameniza a hipótese do risco de evasão. Ao verificar os dados históricos do banco de dados, confirmou-se que o estudante 100058 efetivamente concluiu o curso com êxito.

Este cenário evidencia o potencial da consulta híbrida para análises mais específicas, ao permitir:

- Restringir a base de análise a perfis de interesse (por exemplo, estudantes com um evento comum, como reprovação);
- Explorar visualmente variações dentro desse grupo, revelando padrões individuais e coletivos;
- Apoiar decisões informadas, considerando o histórico completo do estudante e não apenas eventos isolados.

As estratégias de filtragem permitem a exploração visual de trajetórias com diferentes contextos analíticos, como identificar estudantes em risco ou investigar o impacto de eventos específicos no desempenho acadêmico.

5.7 Cenário 6: Identificação de Estudantes com um perfil específico

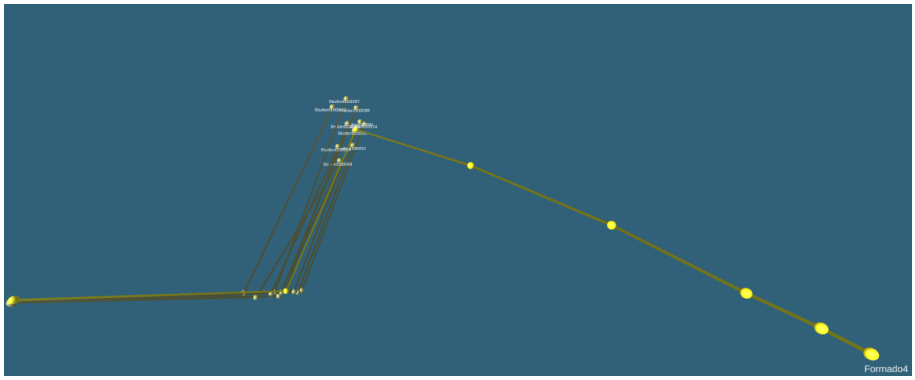
Este cenário tem como objetivo destacar a flexibilidade da técnica proposta, que permite a utilização de trajetórias sumarizadas de grupos de estudantes como elemento central em consultas por similaridade. Ao invés de selecionar um ou mais estudantes individuais como referência, é possível adotar um grupo representativo como alvo de comparação, ampliando as possibilidades de análise e inferência.

Neste exemplo, considera-se uma situação hipotética em que se deseja identificar um perfil específico de estudantes de graduação com potencial para seguirem na carreira acadêmica, ingressando em programas de pós-graduação. Como não há, neste estudo, dados diretos sobre ex-estudantes que efetivamente ingressaram na pós-graduação, utilizou-se como aproximação a trajetória sumarizada da classe **Formado4** — composta por estudantes que concluíram o curso no tempo mínimo regular de 4 anos —, assumindo que esta trajetória representa um perfil de excelência acadêmica e, portanto, um grupo com provável aptidão para continuidade nos estudos.

A classe **Formado4** é composta por estudantes com desempenho acadêmico consistente ao longo de toda a graduação. Com base nessa trajetória média, foi realizada uma consulta por similaridade métrica 5-NN para identificar trajetórias de estudantes ativos

cujas trajetórias, até o terceiro semestre, se aproximam do padrão representado pela classe **Formado4**. O resultado é exibido na Figura 25, onde é possível identificar estudantes com percursos semelhantes à trajetória da classe **Formado4**, o que indica que eles também compartilham características associadas ao bom desempenho acadêmico. Dos 10 estudantes retornados, 50% concluíram em 4 anos, 30% em 5 anos e 20% em 6 anos.

Figura 25 – Consulta por similaridade com uma trajetória sumarizada como alvo.



Fonte: Elaborada pelo autor

Considerando o cenário fictício de identificação de perfis com maior chance de sucesso na pós-graduação, essa abordagem poderia ser utilizada para:

- ❏ Identificar perfis e indicá-los para programas de *iniciação científica*;
- ❏ Estimular a participação em eventos acadêmicos e publicações;
- ❏ Direcionar convites para mentorias, monitorias ou estágios de pesquisa;
- ❏ Fortalecer políticas de incentivo à carreira acadêmica e aumentar o número de egressos que ingressam em programas de pós-graduação;
- ❏ Oferecer suporte personalizado para manter o alto desempenho desses estudantes.

Este cenário evidencia que a técnica proposta não está restrita à análise de situações de risco, como evasão, mas pode ser aplicada de forma proativa e estratégica para identificar perfis específicos, como mostrado no cenário hipotético onde a técnica seria utilizada para **potencializar trajetórias de sucesso**. Ao permitir a seleção de trajetórias sumarizadas como referência em uma consulta por similaridade de trajetórias, aumentam-se as possibilidades de análises e criação de contextos exploratórios, contribuindo para a mineração de informações que podem ser usadas na gestão acadêmica e planejamento institucional.

5.8 Cenário 7: Detecção de Risco Acadêmico em Semestre Específico

Neste cenário, demonstra-se a aplicação da técnica de consulta por similaridade pontual de trajetória para a identificação de potenciais situações de risco acadêmico em momentos específicos da graduação. Essa estratégia permite avaliar se determinado estado do estudante em um semestre específico pode estar associado a padrões de evasão, baixo rendimento ou desvio de trajetória.

Considera-se o caso de um estudante que tenha terminado o **terceiro semestre** do curso e que apresenta um histórico irregular em sua trajetória acadêmica. Com o intuito de investigar sinais de risco, foi selecionada a instância temporal correspondente ao **segundo semestre letivo** do estudante — ponto que, hipoteticamente, pode ser considerado crítico para a continuidade acadêmica — e realizada uma consulta por similaridade pontual, buscando as 5 trajetórias mais próximas a essa instância na representação métrica.

A Figura 26 apresenta o resultado da consulta, evidenciando que, das cinco trajetórias recuperadas, **quatro pertencem a estudantes que se evadiram** do curso nos semestres seguintes. Essa proximidade no espaço métrico indica que o estudante-alvo apresentava, já no segundo semestre, características comuns a estudantes que não concluíram a graduação, o que poderia ter servido de alerta para ações de apoio ou acompanhamento.

Figura 26 – Consulta por similaridade pontual: análise da trajetória do estudante 100510 no segundo semestre.

Fonte: Elaborada pelo autor

Os dados históricos confirmam que o estudante em questão **se evadiu ao final do quarto semestre letivo**, validando a evidência levantada no segundo semestre.

Esse cenário destaca uma das principais vantagens da abordagem de similaridade pontual: a **capacidade de realizar diagnósticos situacionais** com base em instantes específicos do percurso acadêmico. Isso permite, por exemplo:

- ❏ Antecipar situações de risco a partir de pontos críticos da trajetória;

- ❏ Identificar padrões de evasão vinculados a momentos estruturantes do curso;
- ❏ Realizar análises focadas em disciplinas ou semestres chave;
- ❏ Apoiar decisões sobre quando e como intervir no processo formativo dos estudantes.

Portanto, a análise pontual representa uma estratégia de análise temporal específica de trajetórias, permitindo que gestores e coordenadores atuem de forma mais **oportuna e direcionada**.

5.9 Considerações finais

A análise dos cenários apresentados neste capítulo busca demonstrar a aplicabilidade e a flexibilidade do método proposto para a exploração visual de trajetórias acadêmicas. Ao modelar os estudantes como dados complexos em evolução temporal, representá-los em espaços métricos e projetá-los, torna-se possível investigar, de forma visual e interativa, comportamentos individuais e coletivos ao longo da formação acadêmica, baseando-se em similaridade.

Os resultados dos cenários apresentados buscaram evidenciar a capacidade das técnicas propostas em gerar visualizações para o reconhecimento de padrões espaciais entre trajetórias, para ampliar a percepção analítica sobre os dados educacionais e favorecer o surgimento de *insights* exploratórios que dificilmente seriam alcançados por meio de relatórios tabulares, estatísticas agregadas, visualizações estáticas ou gráficos bidimensionais tradicionais, que não capturam a complexa evolução temporal e multivariada do estudante.

As diferentes estratégias para selecionar as trajetórias a serem visualizadas, compostas pelas consultas por similaridade de trajetória, similaridade pontual e abordagem híbrida, mostraram-se eficazes na construção de contextos analíticos adaptados aos objetivos de investigação. Isso permitiu a criação de cenários personalizados, tais como:

- ❏ Identificação de estudantes com maior risco de evasão, por meio da comparação com trajetórias históricas semelhantes;
- ❏ Estimativa de desfechos prováveis com base em padrões anteriores;
- ❏ Validação de hipóteses pedagógicas, como a relação entre o desempenho em disciplinas-chave e o sucesso na graduação;
- ❏ Exploração de grupos sumarizados, como as classes personalizadas, viabilizando comparações e análises baseadas em comportamento médio.

Capítulo 6

Conclusões

As técnicas de visualização de dados desempenham um papel fundamental na análise de informações, permitindo que analistas de dados extraiam conhecimento a partir da representação visual e da interpretação intuitiva das informações exibidas. A visualização de dados educacionais pode ser uma ferramenta essencial para entender a evolução acadêmica dos estudantes ao longo do tempo.

Este trabalho apresentou um método inovador para a modelagem, análise e visualização de trajetórias acadêmicas, tratando os estudantes como dados complexos em evolução temporal. A proposta baseou-se na representação das instâncias temporais em espaço métrico, posteriormente projetado em um espaço multidimensional que possibilita a visualização gráfica das trajetórias.

A abordagem é genérica, flexível e adaptável a diferentes domínios educacionais e tipos de dados, viabilizando a análise visual de trajetórias tanto de indivíduos quanto de grupos de estudantes. O desenvolvimento do protótipo STVis foi desenvolvido para demonstrar a viabilidade prática da proposta, integrando mecanismos de consulta relacional, busca por similaridade e sumarização de grupos.

Os resultados apresentados ao longo dos cenários exploratórios (Capítulo 5) forneceram evidências que comprovam a hipótese central deste trabalho. A exploração visual de trajetórias acadêmicas, modeladas a partir de dados complexos em evolução temporal, demonstrou que pode ser capaz de gerar *insights* sobre a evolução acadêmica dos estudantes. Por meio da análise de similaridade da evolução temporal entre estudantes e grupos, foi possível identificar padrões de desempenho e tendências de progressão que seriam de difícil percepção em abordagens analíticas convencionais, que comparam poucos atributos e não consideram o aspecto temporal dos estudantes. Cenários como a detecção precoce de estudantes em risco de evasão e a validação de hipóteses educacionais ilustraram como

a visualização interativa, aliada a mecanismos de filtragem e sumarização, amplia o potencial de interpretação e auxilia diretamente a tomada de decisão por parte de gestores, coordenadores e pesquisadores da área educacional. A comprovação dessas descobertas, comparando-as com os desfechos acadêmicos reais dos estudantes, disponíveis nos dados históricos, demonstra a viabilidade da abordagem proposta e reafirma o valor da visualização de dados complexos para a compreensão de fenômenos educacionais. Também foi possível constatar o comportamento da função de distância métrico-temporal frente à modelagem adotada nesta pesquisa. A abordagem proposta permite incorporar tanto a similaridade entre perfis acadêmicos quanto a proximidade temporal.

Com isso, conclui-se que as técnicas apresentadas conseguem representar as trajetórias acadêmicas de maneira eficaz, sendo capazes de mapear as diversas nuances inerentes ao percurso formativo dos estudantes ao longo de seu período acadêmico, permitindo criar oportunidades para investigação, intervenção e melhoria dos processos educacionais, a partir da exploração visual.

6.1 Contribuições do Trabalho

Este trabalho propôs uma abordagem para a análise visual do percurso formativo de estudantes, modelando-os como dados complexos em evolução temporal e representando suas trajetórias acadêmicas em espaço métrico. As principais contribuições desta dissertação foram:

- uma nova abordagem para modelar estudantes como objetos de dados complexos em evolução temporal representadas em espaço métrico. Esta modelagem permite capturar a natureza dinâmica das trajetórias acadêmicas e oferece flexibilidade para incorporar diferentes tipos de dados, desde que uma função de distância possa ser definida;
- a definição e implementação de um conjunto de estratégias de filtragem para selecionar subconjuntos relevantes de trajetórias acadêmicas para visualização e análise. As estratégias desenvolvidas incluem consultas baseadas em atributos relacionais, consultas baseadas em similaridade de trajetória e consultas híbridas. As diferentes estratégias permitem distintos tipos de análise, como análise de turmas, classes sumarizadas e identificação de estudantes semelhantes;
- a aplicação de uma técnica de sumarização de trajetórias que calcula trajetórias médias representativas de grupos de estudantes. A técnica facilita a identificação de tendências e padrões gerais de comportamento acadêmico. A sumarização auxilia na redução da complexidade visual e destaca informações importantes, permitindo comparar o desempenho de diferentes grupos de estudantes e personalizar os grupos para análise específica;

- a implementação de um protótipo de ferramenta de visualização interativa que possibilita a exploração de trajetórias acadêmicas em um ambiente 3D. O protótipo implementa funcionalidades para importação de dados, filtragem nos dados a serem visualizados através de consultas e visualização 3D. A ferramenta inclui recursos de interação, como rotação e zoom, que facilitam a análise dos resultados pelo usuário;

- a realização de um estudo de caso com dados reais de estudantes do curso de graduação em Ciência da Computação da UFSCar. O estudo de caso demonstra a aplicabilidade prática das técnicas propostas e fornece *insights* sobre os padrões de desempenho dos estudantes. O estudo valida a abordagem em um contexto do mundo real e ilustra seu potencial para apoiar a tomada de decisões educacionais.

6.2 Limitações e Trabalhos Futuros

Embora os cenários exploratórios apresentados tenham evidenciado o potencial da abordagem proposta, algumas limitações devem ser consideradas.

A principal limitação está relacionada à aplicação da proposta em um conjunto de dados relativamente pequeno e modelagem dos estudantes apenas com dados numéricos estruturados. Considerando que a principal característica da proposta é ser genérica, possibilitando trabalhar com dados heterogêneos, não estruturados, não convencionais (desde que possam ser representados em espaços métricos), o estudo de caso utilizou apenas notas em disciplinas obrigatórias e informações temporais. Essa limitação decorre dos dados disponíveis para validação da proposta e não compromete a generalidade da abordagem, mas restringe o escopo da validação empírica. Investigações futuras poderão explorar representações mais complexas e diversificadas.

Em segundo lugar, a qualidade da representação das trajetórias está diretamente associada à qualidade dos dados utilizados. Informações ausentes, inconsistentes ou desatualizadas podem comprometer a fidelidade das representações e, consequentemente, a validade das descobertas. Além disso, a técnica depende da definição de uma função de distância adequada para representar a similaridade entre instâncias temporais. Neste estudo, utilizou-se métricas como a distância euclidiana e a distância métrico-temporal, embora outras métricas possam ser mais apropriadas para determinados tipos de dados ou contextos específicos.

Outro aspecto relevante diz respeito ao processo de sumarização das trajetórias de grupos. A estratégia adotada, baseada na média aritmética, não lida explicitamente com a presença de *outliers* dentro de um grupo, o que pode impactar a representatividade das trajetórias médias geradas.

Apesar dessas limitações, o trabalho abre possibilidades para pesquisas futuras. Entre as principais perspectivas, destacam-se:

- **Ampliação das representações dos estudantes**: investigar outras formas de representar as instâncias temporais, incorporando dados complementares além de notas, como histórico socioeconômico, participação em atividades extracurriculares e desempenho em avaliações externas, visando enriquecer as análises e proporcionar uma visão mais holística do percurso acadêmico.

- **Adoção de novas técnicas de projeção multidimensional**: explorar métodos como t-SNE, UMAP ou PCA, que oferecem diferentes propriedades de preservação local e global da estrutura dos dados, possibilitando comparações e refinamentos das visualizações.

- **Aplicação em diferentes contextos educacionais**: testar o método em outros níveis de ensino ou contextos institucionais, como cursos técnicos, programas de pós-graduação ou mesmo em instituições com diferentes estruturas curriculares.

- **Validação em tempo real**: aplicar a abordagem com dados de estudantes ativos, com métodos de acompanhamento e validação dos resultados.

- **Exploração de novas técnicas de interação**: implementar recursos como filtros dinâmicos e interativos, por exemplo, ao selecionar uma instância temporal, disparar automaticamente uma nova consulta por similaridade para ampliar o potencial analítico e a fluidez da exploração.

- **Investigação de novos métodos de agregação**: estudar alternativas à média aritmética para a sumarização de grupos, como o uso de medidas para identificar *outliers* (mediana, quartis) ou a consideração de estatísticas adicionais (desvio padrão), com o objetivo de aprimorar a representatividade das trajetórias médias.

Referências

AIGNER, W. et al. **Visualization of Time-Oriented Data**. 1st. ed. [S.l.]: Springer Publishing Company, Incorporated, 2011. ISBN 0857290789.

ALHAZMI, E.; SHENEAMER, A. Early predicting of students performance in higher education. **IEEE Access**, v. 11, p. 27579–27589, 2023. Disponível em: <https://doi.org/10.1109/ACCESS.2023.3250702>.

ARANTES, A. S. **Consultas por similaridade complexas em gerenciadores relacionais**. Tese (Tese de Doutorado em Ciências de Computação e Matemática Computacional) — Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2005. Disponível em: <https://doi.org/10.11606/T.55.2005.tde-13112014-165634>.

ARGIENTO, R.; FILIPPI-MAZZOLA, E.; PACI, L. **Model-based clustering of categorical data based on the Hamming distance**. 2024. Disponível em: <https://arxiv.org/abs/2212.04746>.

ATLASBR. **Atlas Brasil**. 2021. Disponível em: <http://www.atlasbrasil.org.br/>.

BARIONI, M. C. N. et al. Data visualization in rdbms. In: **Proceedings of the IASTED International Conference on Information Systems and Databases (ISDB 2002)**. Tokyo, Japan: [s.n.], 2002. p. 264–269.

BELUSSI, A.; FALOUTSOS, C. Self-spacial join selectivity estimation using fractal concepts. **ACM Trans. Inf. Syst.**, Association for Computing Machinery, New York, NY, USA, v. 16, n. 2, p. 161–201, abr. 1998. ISSN 1046-8188. Disponível em: <https://doi.org/10.1145/279339.279342>.

BODILY, R.; VERBERT, K. Trends and issues in student-facing learning analytics reporting systems research. In: **Proceedings of the Seventh International Learning Analytics & Knowledge Conference**. New York, NY, USA: Association for Computing Machinery, 2017. (LAK '17), p. 309–318. ISBN 9781450348706. Disponível em: <https://doi.org/10.1145/3027385.3027403>.

BÖHLEN, M. H. et al. The 3vdm approach: A case study with clickstream data. In: **Visual Data Mining: Theory, Techniques and Tools for Visual Analytics**. Springer, 2008, (Lecture Notes in Computer Science, v. 4404). p. 13–29. Disponível em: <https://doi.org/10.1007/978-3-540-71080-6_2>.

BöHM, C.; BERCHTOLD, S.; KEIM, D. A. Searching in high-dimensional spaces: Index structures for improving the performance of multimedia databases. **ACM Comput. Surv.**, Association for Computing Machinery, New York, NY, USA, v. 33, n. 3, p. 322–373, set. 2001. ISSN 0360-0300. Disponível em: <https://doi.org/10.1145/502807.502809>.

BRIN, S. Near neighbor search in large metric spaces. In: **Proceedings of the 21th International Conference on Very Large Data Bases**. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1995. (VLDB '95), p. 574–584. ISBN 1558603794.

BUENO, R. et al. Time-aware similarity search: A metric-temporal representation for complex data. In: MAMOULIS, N. et al. (Ed.). **Advances in Spatial and Temporal Databases**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. (Lecture Notes in Computer Science, v. 5644), p. 302–319. ISBN 978-3-642-02982-0. Disponível em: <https://doi.org/10.1007/978-3-642-02982-0_20>.

BUENO, R. et al. Metric data analysis enhanced through temporal visualization. In: **2010 14th International Conference Information Visualisation**. London, UK: IEEE Computer Society, 2010. p. 116–121. Disponível em: <https://doi.org/10.1109/IV.2010.26>.

CHÁVEZ, E. et al. Searching in metric spaces. **ACM computing surveys (CSUR)**, ACM New York, NY, USA, v. 33, n. 3, p. 273–321, 2001.

CHEN, L. et al. Indexing metric spaces for exact similarity search. **ACM Comput. Surv.**, Association for Computing Machinery, New York, NY, USA, v. 55, n. 6, dez. 2022. ISSN 0360-0300. Disponível em: <https://doi.org/10.1145/3534963>.

CHEN, M. et al. **Foundations of data visualization**. Springer, 2020. Disponível em: <https://doi.org/10.1007/978-3-030-34444-3>.

CIACCIA, P.; PATELLA, M.; ZEZULA, P. M-tree: An efficient access method for similarity search in metric spaces. In: **Proceedings of the 23rd International Conference on Very Large Data Bases**. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1997. (VLDB '97), p. 426–435. ISBN 1558604707.

DIAS, R. L. et al. Visual-interactive k-ndn method (VIK): A novel approach to visualize and interact with content-based image retrieval systems regarding similarity and diversity. In: **2017 21st International Conference Information Visualisation (IV)**. London, UK: [s.n.], 2017. p. 72–77. Disponível em: <https://doi.org/10.1109/iV.2017.41>.

DIMARA, E.; STASKO, J. A critical reflection on visualization research: Where do decision making tasks hide? **IEEE Transactions on Visualization and Computer Graphics**, v. 28, n. 1, p. 1128–1138, 2022. Disponível em: <https://doi.org/10.1109/TVCG.2021.3114813>.

FADEL, S. G. et al. Loch: A neighborhood-based multidimensional projection technique for high-dimensional sparse spaces. **Neurocomputing**, v. 150, p. 546–556, 2015. ISSN 0925-2312. Special Issue on Information Processing and Machine Learning for Applications of Engineering Solving Complex Machine Learning Problems with Ensemble Methods Visual Analytics using Multidimensional Projections. Disponível em: <https://doi.org/10.1016/j.neucom.2014.07.071>.

FALOUTSOS, C.; LIN, K.-I. Fastmap: A fast algorithm for indexing, data-mining and visualization of traditional and multimedia datasets. In: **Proceedings of the 1995 ACM SIGMOD International Conference on Management of Data**. New York, NY, USA: Association for Computing Machinery, 1995. (SIGMOD '95), p. 163–174. ISBN 0897917316. Disponível em: <https://doi.org/10.1145/223784.223812>.

FLEXA, C. et al. Polygonal coordinate system: Visualizing high-dimensional data using geometric dr, and a deterministic version of t-sne. **Expert systems with applications**, Elsevier Ltd, New York, v. 175, p. 114741, 2021. ISSN 0957-4174.

GUERRA, J. et al. Trac: Visualizing students academic trajectories. In: **Transforming Learning with Meaningful Technologies**. Cham: Springer International Publishing, 2019. (Lecture Notes in Computer Science, v. 11722), p. 765–768. ISBN 978-3-030-29735-0. Disponível em: <https://doi.org/10.1007/978-3-030-29736-7_84>.

GUERRA LAURA E ARCINIEGAS, S. Academic management through the visualization of information. In: **2019 14th Iberian Conference on Information Systems and Technologies (CISTI)**. Coimbra, Portugal: [s.n.], 2019. p. 1–5. Disponível em: <https://doi.org/10.23919/CISTI.2019.8760770>.

HARRIS, R. **Information Graphics: A Comprehensive Illustrated Reference**. Oxford University Press, 1999. (BusinessPro collection). ISBN 9780195135329. Disponível em: <https://books.google.com.br/books?id=LT1RXREvkGIC>.

HASUGIAN, P. M. et al. Review of high-dimensional and complex data visualization. In: **2023 International Conference of Computer Science and Information Technology (ICOSNIKOM)**. [s.n.], 2023. p. 1–7. Disponível em: <https://doi.org/10.1109/ICoSNIKOM60230.2023.10364377>.

HEGDE, V.; PRAGEETH, P. P. Higher education student dropout prediction and analysis through educational data mining. In: **2018 2nd International Conference on Inventive Systems and Control (ICISC)**. Coimbatore, India: [s.n.], 2018. p. 694–699. Disponível em: <https://doi.org/10.1109/ICISC.2018.8398887>.

HUANG, J. et al. T-distributed stochastic neighbor embedding echo state network with state matrix dimensionality reduction for time series prediction. **Engineering Applications of Artificial Intelligence**, v. 122, p. 106055, 2023. ISSN 0952-1976. Disponível em: <https://doi.org/10.1016/j.engappai.2023.106055>.

ILIINSKY, N.; STEELE, J. **Designing Data Visualizations: Representing Informational Relationships**. [S.l.]: "O'Reilly Media, Inc.", 2011. (Designing Data Visualizations). ISBN 9781449312282.

JAYARATNA, S. et al. Analysing the learning paths of computer science students via trajectory analytics. In: **2020 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)**. [s.n.], 2020. p. 273–280. Disponível em: <https://doi.org/10.1109/TALE48869.2020.9368342>.

JOIA, P. et al. Local affine multidimensional projection. **IEEE Transactions on Visualization and Computer Graphics**, v. 17, n. 12, p. 2563–2571, 2011.

JUNIOR, R. D. M.; BUENO, R. Visualization of trajectory-based queries in images database. In: **2021 25th International Conference Information Visualisation (IV)**. Sydney, Australia: [s.n.], 2021. p. 328–333. Disponível em: <https://doi.org/10.1109/IV53921.2021.00059>.

KEIM, D. Designing pixel-oriented visualization techniques: theory and applications. **IEEE Transactions on Visualization and Computer Graphics**, v. 6, n. 1, p. 59–78, 2000.

KEIM, D. Information visualization and visual data mining. **IEEE Transactions on Visualization and Computer Graphics**, v. 8, n. 1, p. 1–8, 2002.

KEIM, D.; KRIEGEL, H.-P.; ANKERST, M. Recursive pattern: a technique for visualizing very large amounts of data. In: **Proceedings Visualization '95**. Atlanta, GA, USA: [s.n.], 1995. p. 279–286. Disponível em: <https://doi.org/10.1109/VISUAL.1995.485140>.

KEIM, D. et al. Challenges in visual data analysis. In: **Tenth International Conference on Information Visualisation (IV'06)**. London, UK: [s.n.], 2006. p. 9–16. Disponível em: <https://doi.org/10.1109/IV.2006.31>.

KIM, M. C.; ZHU, Y.; CHEN, C. How are they different? a quantitative domain comparison of information visualization and data visualization (2000-2014). **Scientometrics**, v. 107, p. 123–165, 01 2016. Disponível em: <https://doi.org/10.1007/s11192-015-1830-0>.

KOSTIC, Z. et al. Exploring mid-air hand interaction in data visualization. **IEEE Transactions on Visualization and Computer Graphics**, v. 30, n. 9, p. 6347–6364, 2024. Disponível em: <https://doi.org/10.1109/TVCG.2023.3332647>.

KUNZAYILA, N. Uso da visualização de informações para análise de dados educacionais. In: PALETTA, F. C.; BANDEIRA, P. M. (Ed.). **Anais do I Simpósio Organização e Representação da Informação**. São Paulo: ECA-USP, 2022. p. 79. ISBN 978-65-88640-64-7. Disponível em: <https://toi.eca.usp.br/vitoi/libraryFiles/downloadPublic/43>.

LI, F. Vr interactive game design based on unity3d engine. In: **2020 International Conference on Robots Intelligent System (ICRIS)**. [S.l.: s.n.], 2020. p. 142–145.

LI, K. On integrating information visualization techniques into data mining: A review. **CoRR**, abs/1503.00202, 2015. Disponível em: <https://doi.org/10.48550/arXiv.1503.00202>.

LIMA, E. L. **Espaços Métricos**. 2. ed. Rio de Janeiro: Projeto Euclides / Instituto de Matemática Pura e Aplicada, 1977. (Coleção Projeto Euclides: 4).

LIU, S. et al. Visualizing high-dimensional data: Advances in the past decade. **IEEE Transactions on Visualization and Computer Graphics**, v. 23, n. 3, p. 1249–1268, 2017. Disponível em: <https://doi.org/10.1109/TVCG.2016.2640960>.

LOPES, M. A. D. S.; NETO, A. D. D.; MARTINS, A. D. M. Parallel t-sne applied to data visualization in smart cities. **IEEE Access**, v. 8, p. 11482–11490, 2020. Disponível em: <https://doi.org/10.1109/ACCESS.2020.2964413>.

MAATEN, L. van der. Accelerating t-sne using tree-based algorithms. **Journal of Machine Learning Research**, v. 15, n. 93, p. 3221–3245, 2014. Disponível em: <https://dl.acm.org/doi/10.5555/2627435.2697068>.

MAATEN, L. van der; HINTON, G. Visualizing data using t-sne. **Journal of Machine Learning Research**, v. 9, n. 86, p. 2579–2605, 2008. Disponível em: <http://jmlr.org/papers/v9/vandermaaten08a.html>.

MANNING, C. D.; RAGHAVAN, P.; SCHüTZE, H. **Introduction to Information Retrieval**. Cambridge: Cambridge University Press, 2008. ISBN 9780521865715.

MILLER, J. et al. Ens-t-sne: Embedding neighborhoods simultaneously t-sne. In: **2024 IEEE 17th Pacific Visualization Conference (PacificVis)**. [s.n.], 2024. p. 222–231. Disponível em: <https://doi.org/10.1109/PacificVis60374.2024.00032>.

NASCIMENTO, H. A. D.; FERREIRA, C. B. Uma introdução à visualização de informações. **Visualidades**, v. 9, n. 2, 2011.

NAVARRO, G. Searching in metric spaces by spatial approximation. In: **6th International Symposium on String Processing and Information Retrieval. 5th International Workshop on Groupware (Cat. No.PR00268)**. [s.n.], 1999. p. 141–148. Disponível em: <https://doi.org/10.1109/SPIRE.1999.796589>.

NEVES, T. T. d. A. T. **Projeções multidimensionais para a análise de fluxos de dados**. Tese (Tese de Doutorado) — Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2016. Disponível em: <https://doi.org/10.11606/T.55.2017.tde-12012017-150124>.

NEVES, T. T. d. A. T. et al. Análise visual utilizando projeções multidimensionais. **Revista de Informática Teórica e Aplicada**, v. 22, n. 2, p. 258–288, 2015. Disponível em: <https://doi.org/10.22456/2175-2745.56498>.

ORTEGA, M. P. P.; BRAVO, F.; PEÑA, L. I. Analítica del aprendizaje, visualización de trayectoria académica (learning analytics, dashboard for academic trajectory). In: **Latin American Conference on Learning Analytics**. [s.n.], 2019. Disponível em: <https://api.semanticscholar.org/CorpusID:245635414>.

ORTIGOSSA, E. S.; DIAS, F. F.; NASCIMENTO, D. C. d. Getting over high-dimensionality: How multidimensional projection methods can assist data science. **Applied Sciences**, v. 12, n. 13, 2022. ISSN 2076-3417. Disponível em: <https://doi.org/10.3390/app12136799>.

PAIVA, C. E.; BUENO, R. Delimitation of regions of interest in similarity queries visualization. In: **2019 23rd International Conference Information Visualisation (IV)**. Paris, France: [s.n.], 2019. p. 31–36. ISSN 2375-0138. Disponível em: <https://doi.org/10.1109/IV.2019.00015>.

PAIVA, C. E.; MALAQUIAS, R. D.; BUENO, R. Visualization of similarity queries with trajectory estimation in complex data. In: **2020 24th International Conference Information Visualisation (IV)**. Melbourne, Australia: [s.n.], 2020. p. 92–97. Disponível em: <https://doi.org/10.1109/IV51561.2020.00025>.

PAULOVICH, F. V. et al. Least square projection: A fast high-precision multidimensional projection technique and its application to document mapping. **IEEE Transactions on Visualization and Computer Graphics**, v. 14, n. 3, p. 564–575, 2008. Disponível em: <https://doi.org/10.1109/TVCG.2007.70443>.

PUDIL, P.; HOVOVICOVA, J. Novel methods for subset selection with respect to problem knowledge. **IEEE Intelligent Systems and their Applications**, v. 13, n. 2, p. 66–74, 1998.

RAUT, A.; HAJARE, S. Transforming education with data mining: Opportunities, applications, and challenges. In: **2025 International Conference on Inventive Computation Technologies (ICICT)**. [s.n.], 2025. p. 1–6. Disponível em: <https://doi.org/10.1109/ICICT64420.2025.11005098>.

ROMERO, C.; VENTURA, S. Educational data mining: A review of the state of the art. **IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)**, v. 40, n. 6, p. 601–618, 2010. Disponível em: <https://doi.org/10.1109/TSMCC.2010.2053532>.

SAKTHEESWARAN, A.; SRINIVASAN, A.; STASKO, J. Touch? speech? or touch and speech? investigating multimodal interaction for visual network exploration and analysis. **IEEE Transactions on Visualization and Computer Graphics**, Institute of Electrical and Electronics Engineers (IEEE), v. 26, n. 6, p. 2168–2179, jun. 2020. ISSN 2160-9306. Disponível em: <http://dx.doi.org/10.1109/TVCG.2020.2970512>.

SANTOS, R. S. S. dos; PONTI, M. A.; RODRIGUES, K. R. da H. The use of digital reports to support the visualization and identification of university dropout data. In: **Human Interface and the Management of Information**. Berlin, Heidelberg: Springer-Verlag, 2022. p. 308–323. ISBN 978-3-031-06423-4. Disponível em: <https://doi.org/10.1007/978-3-031-06424-1_23>.

SEDLMAIR, M.; MEYER, M.; MUNZNER, T. Design study methodology: Reflections from the trenches and the stacks. **IEEE Transactions on Visualization and Computer Graphics**, v. 18, n. 12, p. 2431–2440, 2012. Disponível em: <https://doi.org/10.1109/TVCG.2012.213>.

SHNEIDERMAN, B. The eyes have it: a task by data type taxonomy for information visualizations. In: **Proceedings 1996 IEEE Symposium on Visual Languages**. [S.l.: s.n.], 1996. p. 336–343.

SIEMENS, G.; LONG, P. Penetrating the fog: Analytics in learning and education. **EDUCAUSE Review**, v. 5, p. 30–32, 01 2011. Disponível em: <https://doi.org/10.17471/2499-4324/195>.

SILAHTAROĞLU, G.; ALAYOGLU, N. Using or not using business intelligence and big data for strategic management: An empirical study based on interviews with executives in various sectors. **Procedia - Social and Behavioral Sciences**, v. 235, p. 208–215, 2016. ISSN 1877-0428. 12th International Strategic Management Conference, ISMC 2016, 28-30 October 2016, Antalya, Turkey. Disponível em: <https://doi.org/10.1016/j.sbspro.2016.11.016>.

SMELSER, K.; MILLER, J.; KOBOUROV, S. “normalized stress” is not normalized: How to interpret stress correctly. In: **2024 IEEE Evaluation and Beyond - Methodological Approaches for Visualization (BELIV)**. [s.n.], 2024. p. 41–50. Disponível em: <https://doi.org/10.1109/BELIV64461.2024.00010>.

TRAINA, C.; TRAINA, A.; FALOUTSOS, C. Distance exponent: A new concept for selectivity estimation in metric trees. In: **Proceedings of 16th International Conference on Data Engineering (Cat. No.00CB37073)**. [S.l.: s.n.], 2000. p. 195–195.

TRIMM, D.; RHEINGANS, P.; DESJARDINS, M. Visualizing student histories using clustering and composition. **IEEE Transactions on Visualization and Computer Graphics**, v. 18, n. 12, p. 2809–2818, 2012.

TUFTE, E. **The Visual Display of Quantitative Information**. Graphics Press, 2001. ISBN 9781930824133. Disponível em: <https://books.google.com.br/books?id=qmjNngEACAAJ>.

UNIVERSIDADE FEDERAL DE SÃO CARLOS. DEPARTAMENTO DE COMPUTAÇÃO. **BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO — PROJETO PEDAGÓGICO** —. São Carlos: [s.n.], 2018. Documento interno. AUTORIA: NÚCLEO DOCENTE ESTRUTURANTE (NDE) COORDENAÇÃO DE CURSO. Disponível em: <https://www.prograd.ufscar.br/pt-br/assets/arquivos/cursos/cursos-oferecidos/ciencias-da-computacao/PPCBCC_2019.pdf>.

VERMA, B. K.; SRIVASTAVA, N.; BHARTI, A. K. Comparative analysis of student performance prediction using machine learning techniques. In: **2024 International Conference on Information Science and Communications Technologies (ICISCT)**. [S.l.: s.n.], 2024. p. 332–337.

WALKER, J.; BORGO, R.; JONES, M. W. Timenotes: A study on effective chart visualization and interaction techniques for time-series data. **IEEE Transactions on Visualization and Computer Graphics**, v. 22, n. 1, p. 549–558, 2016. Disponível em: <https://doi.org/10.1109/TVCG.2015.2467751>.

WARD, M. O.; GRINSTEIN, G.; KEIM, D. **Interactive Data Visualization: Foundations, Techniques, and Applications**. 2nd. ed. USA: A. K. Peters, Ltd., 2015. ISBN 1482257378.

WILKE, C. O. **Fundamentals of data visualization: a primer on making informative and compelling figures**. [S.l.]: O’Reilly Media, 2019.

XIA, J. et al. Interactive visual cluster analysis by contrastive dimensionality reduction. **IEEE Transactions on Visualization and Computer Graphics**, v. 29, n. 1, p. 734–744, 2023. Disponível em: <https://doi.org/10.1109/TVCG.2022.3209423>.

ZHANG, G. et al. Towards a better understanding of the role of visualization in online learning: A review. **Visual Informatics**, v. 6, n. 4, p. 22–33, 2022. ISSN 2468-502X. Disponível em: <https://doi.org/10.1016/j.visinf.2022.09.002>.

ZHANG, T. et al. Bertscore: Evaluating text generation with bert. In: **International Conference on Learning Representations (ICLR)**. [s.n.], 2020. Disponível em: <https://doi.org/10.48550/arXiv.1904.09675>.

ZHUANG, M.; CONCANNON, D.; MANLEY, E. A framework for evaluating dashboards in healthcare. **IEEE Transactions on Visualization and Computer Graphics**, v. 28, n. 4, p. 1715–1731, 2022.