
EduFlow: Análise do desempenho acadêmico de estudantes universitários utilizando técnicas de Visualização de Informação

Laura Ferreira Marquez



UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Laura Ferreira Marquez

**EduFlow: Análise do desempenho acadêmico de
estudantes universitários utilizando técnicas de
Visualização de Informação**

Dissertação de mestrado apresentada ao Programa de Pós-graduação da Faculdade de Computação da Universidade Federal de Uberlândia como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação.

Área de concentração: Ciência da Computação

Orientador: Prof. Dr. José Gustavo de Souza Paiva

Uberlândia
2025

Ficha Catalográfica Online do Sistema de Bibliotecas da UFU
com dados informados pelo(a) próprio(a) autor(a).

M357 Marquez, Laura Ferreira, 1990-
2025 EduFlow: Análise do desempenho acadêmico de estudantes universitários utilizando técnicas de Visualização de Informação [recurso eletrônico] / Laura Ferreira Marquez. - 2025.

Orientador: José Gustavo de Souza Paiva.
Dissertação (Mestrado) - Universidade Federal de Uberlândia,
Pós-graduação em Ciência da Computação.
Modo de acesso: Internet.
DOI <http://doi.org/10.14393/ufu.di.2025.466>
Inclui bibliografia.
Inclui ilustrações.

1. Computação. I. Paiva, José Gustavo de Souza,1979-, (Orient.).
II. Universidade Federal de Uberlândia. Pós-graduação em Ciência da Computação. III. Título.

CDU: 681.3

Bibliotecários responsáveis pela estrutura de acordo com o AACR2:

Gizele Cristine Nunes do Couto - CRB6/2091

Nelson Marcos Ferreira - CRB6/3074



UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Coordenação do Programa de Pós-Graduação em Computação
Av. João Naves de Ávila, 2121, Bloco 1A, Sala 243 - Bairro Santa Mônica, Uberlândia-MG,
CEP 38400-902

Telefone: (34) 3239-4470 - www.ppgco.facom.ufu.br - cpgfacom@ufu.br



ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Ciência da Computação			
Defesa de:	Dissertação, 16/2025, PPGCO			
Data:	28 de Julho de 2025	Hora de início:	14:05	Hora de encerramento:
Matrícula do Discente:	12222CCP012			
Nome do Discente:	Laura Ferreira Marquez			
Título do Trabalho:	EduFlow: Análise do desempenho acadêmico de estudantes universitários utilizando técnicas de Visualização de Informação			
Área de concentração:	Ciência da Computação			
Linha de pesquisa:	Ciência de Dados			
Projeto de Pesquisa de vinculação:	Demanda Universal FAPEMIG - Processo: APQ-00150-21.			

Reuniu-se por videoconferência, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Ciência da Computação, assim composta: Professores Doutores: Márcia Aparecida Fernandes - FACOM/UFU , Jean Roberto Ponciano - USP e José Gustavo de Souza Paiva - FACOM/UFU, orientador do candidato.

Os examinadores participaram desde as seguintes localidades: Os membros da banca participaram da cidade de Uberlândia e a aluna participou da cidade de Lisboa/Portugal.

Iniciando os trabalhos o presidente da mesa, Prof. Dr. José Gustavo de Souza Paiva , apresentou a Comissão Examinadora e o candidato, agradeceu a presença do público, e concedeu ao Discente a palavra para a exposição do seu trabalho. A duração da apresentação da Discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir o senhor presidente concedeu a palavra, pela ordem sucessivamente, aos examinadores, que passaram a arguir ao candidato. Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o candidato:

Aprovado

Esta defesa faz parte dos requisitos necessários à obtenção do título de Mestre.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.



Documento assinado eletronicamente por **Márcia Aparecida Fernandes, Professor(a) do Magistério Superior**, em 29/07/2025, às 16:35, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **José Gustavo de Souza Paiva, Professor(a) do Magistério Superior**, em 29/07/2025, às 19:05, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Jean Roberto Ponciano, Usuário Externo**, em 31/07/2025, às 10:38, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **6488730** e o código CRC **2521BFC7**.

Referência: Processo nº 23117.046334/2025-84

SEI nº 6488730

*Este trabalho é dedicado às crianças adultas que,
quando pequenas, sonharam em se tornar cientistas.*

Agradecimentos

Agradeço primeiramente à minha família, pelo apoio incondicional ao longo de toda a minha trajetória acadêmica. Aos meus orientadores, pelo incentivo, paciência e orientação técnica fundamentais para a realização deste trabalho. À Universidade Federal de Uberlândia e ao Programa de Pós-Graduação em Ciências da Computação pela estrutura e oportunidade de crescimento. Aos colegas e professores que, direta ou indiretamente, contribuíram com sugestões, críticas e apoio. Agradeço também à FAPEMIG - Projeto APQ-00150-21 pelo apoio financeiro ao desenvolvimento desta pesquisa.

“Sua vida pode ser dividida em dois períodos: antes de agora e a partir de agora.”
(Prof. Obvious Stating)

Resumo

Pesquisadores da área de Educação buscam compreender o desempenho acadêmico dos alunos em cursos de graduação, explorando a influência de diversos fatores, como frequência, notas, dentre outros aspectos. A análise de dados acadêmicos pode auxiliar na compreensão da relação entre componentes curriculares e o desempenho acadêmico dos alunos, propiciando a identificação de possíveis causas das deficiências dos alunos nesses componentes curriculares. Estratégias baseadas na visualização de desempenho acadêmico de alunos apresentam resultados satisfatórios quanto a análise de notas, faltas anteriores e predição de notas, características dos alunos, interações dos alunos em uma disciplina e seus padrões de progressão. No entanto, essas abordagens possuem algumas limitações, tais como a falta de correlações das informações para identificar as reais causas do fracasso ou sucesso do aluno no curso matriculado e a utilização de *layouts* muito básicos, que não exploram o potencial das informações para a tomada de decisão. Além disso, não provêem uma visão geral das correlações entre disciplinas e desempenho dos alunos. Por exemplo, alunos que fizeram a disciplina “A” primeiro, apresentam melhor aproveitamento na disciplina “B”, em quais disciplinas os alunos mais apresentaram absenteísmo e os possíveis motivos que levaram a esse fator. Diante do cenário apresentado, este trabalho propõe o **EduFlow**, um novo sistema de análise visual dos dados de desempenho acadêmico de alunos universitários, utilizando técnicas de *Visualização de Informação* interativas. Acredita-se que o uso de representações visuais que evidenciam ao usuário padrões e tendências a respeito dos fenômenos relacionados ao processo educacional permitem que os usuários compreendam melhor esses fenômenos e auxiliam na tomada de decisão por parte dos profissionais da educação, consequentemente melhorando o processo de aprendizado dos alunos. O EduFlow emprega *layouts* com foco na análise do desempenho acadêmico ao longo de um ou vários semestres, de forma a acompanhar sua evolução de desempenho, e na compreensão de como determinados momentos ou eventos impactam no aproveitamento e sucesso no curso.

Palavras-chave: Análise Visual, Desempenho Acadêmico, Layouts Interativos, Visualização de Dados Educacionais.

Abstract

Researchers in the field of Education seek to understand the academic performance of undergraduate students, exploring the influence of various factors such as attendance, grades, and other aspects. The analysis of academic data can assist in understanding the relationship between curricular components and students' academic performance, enabling the identification of potential causes of deficiencies in these curricular components. Strategies focusing on the visualization of academic performance present satisfactory results regarding the analysis of grades, past absences, grade prediction, student characteristics, student interactions within a module, and progression patterns. However, these approaches have some limitations, such as the lack of correlations between information to identify the real causes of student success or failure in their enrolled modules and the use of very basic layouts that do not fully explore the potential of the data for decision-making. Furthermore, they do not provide an overview of the correlations between subjects and student performance. For instance, students who take module "A" first may perform better in module "B", or insights on which modules have the highest absenteeism rates and the possible reasons behind this issue. Given this scenario, we propose **EduFlow**, a new system for visual analysis of university students' academic performance data, utilizing interactive *Information Visualization* techniques. We believe that the use of visual representations that highlight patterns and trends related to educational processes allows users to better understand these phenomena and supports decision-making by education professionals, consequently improving students' learning processes. EduFlow employs layouts focused on analyzing academic performance over one or multiple semesters to track performance evolution and understand how certain moments or events impact students' success and module achievements.

Keywords: Visual Analysis, Academic Performance, Interactive Layouts, Educational Data Visualization.

Listas de ilustrações

Figura 1 – Processo de <i>Visualização de Informação</i> adaptado de Card, Mackinlay e Shneiderman (1999)	35
Figura 2 – Comparação entre dois conjuntos de caminhos de notas do aluno. A cor vermelha representa o mesmo grupo de alunos que obtiveram C no Exame 3 e A no exame final; a cor azul representa os alunos que obtiveram C no Exame 3 e no exame final (DENG et al., 2019).	37
Figura 3 – Distribuição de notas para dois nós típicos (18 37.10 e 18 37.11) e dois nós <i>outliers</i> (18 37.12 e 18 37.13) na visualização GDV (Grade Distribution View) (GOULDEN et al., 2019).	38
Figura 4 – Visualização do <i>heatmap</i> do curso de Algoritmo e Programação I, comparando notas e faltas e dos alunos e em seguida um modelo de previsão com machine learning com o algoritmo <i>Random Forest</i> , para prever padrões de comportamento (ETEMADPOUR et al., 2019).	42
Figura 5 – Caminho geral das notas de exames selecionando o fluxo da nota C no Exame 3 para A no Exame final (DENG et al., 2019).	43
Figura 6 – Visualização do fluxo do aluno em toda a universidade, mostrando a progressão do aluno semestre a semestre. Cada nó folha corresponde a um único principal, e a espessura dos caminhos corresponde ao número de alunos progredindo através dos nós (RAJI et al., 2017).	45
Figura 7 – Detalhamento do ramo do gráfico radial que contém os cursos de Computação e Engenharia Elétrica da universidade. Esses cursos se separaram da maioria dos outros cursos no final do primeiro semestre (RAJI et al., 2017).	46

Figura 8 – Rede direcionada de movimento de aluno enquanto trabalhava na lição de casa (a), no teste intermediário (b) e no final (c) para outros componentes do curso na oferta de 2012 do MIT. A espessura das bordas codifica o número de movimentos do aluno e o tamanho dos nós codifica o tempo gasto em um componente do curso (EMMONS; LIGHT; BÖRNER, 2017)	47
Figura 9 – <i>Layout sunburst</i> mostrando padrões de navegação sequencial dos alunos em plataformas <i>on-line</i> (POON et al., 2017)	48
Figura 10 – A interface do usuário do DropoutVis inclui: <i>Temporal Circle View</i> , que fornece uma visão geral de diferentes padrões; <i>Windmill View</i> , que permite examinar o efeito de diferentes características sobre a evasão; <i>Feature Temporal Comparison View</i> , que apoia a exploração da associação entre diferentes períodos e a evasão; <i>Counterfactual Explanation View</i> , que oferece uma análise detalhada das razões para a evasão; e <i>Instance View</i> , que resume as regras de aprendizagem de indivíduos (ZHANG et al., 2023)	50
Figura 11 – Análise de explicações contrafactuals para o curso de Engenharia Industrial: A representa a seleção de alunos e seus respectivos contrafactuals; B exibe os contrafactuals, onde B1 mostra contrafactuals aleatórios, B2 mostra os mais viáveis, B3 mostra os mais factíveis e B4 mostra o contrafactual selecionado pelo usuário. Além disso, C mostra a distribuição dos alunos, e D mostra o impacto dos contrafactuals selecionados (B2–B4) (GARCIA-ZANABRIA et al., 2022)	52
Figura 12 – O painel de análise visual: (A) Visão de Seleção de Questões: Permite que os usuários selecionem questões. (B) Visão do aluno: Facilita a exploração do desempenho dos alunos em termos de diferentes métricas. (C) Visão de Resumo de Caminhos: Resume os caminhos e a informação de transição de <i>snapshots</i> de múltiplos alunos. (D) Visão de Sequência: Mostra as sequências completas de <i>snapshots</i> e eventos dos alunos (TSUNG et al., 2022)	54
Figura 13 – Arquitetura metodológica do sistema EduFlow. Estrutura da metodologia de análise visual, mostrando todas as etapas organizadas em componentes estruturados. Os usuários interagem com os <i>layouts</i> no componente Visualização, utilizando um conjunto de ferramentas de interação que proporciona a exploração de diferentes aspectos do EduFlow com foco na análise do desempenho acadêmico.	59
Figura 14 – Visões Analíticas do EduFlow	61
Figura 15 – Visão de Sequência de Execução, com alunos agrupados de acordo com o resultado de sua(s) tentativa(s) na disciplina.	63

Figura 16 – Interações na Visão de Sequência de execução. A: Exibição, no quadrado branco do primeiro nó, do número de alunos na categoria de desempenho dos alunos aprovados em GSI006. B: Exibição, na aresta, do percentual de alunos aprovados em GSI006 e que também foram aprovados em GSI015.	64
Figura 17 – Visão de Notas e Frequência, exibindo curvas de distribuição das notas (A) e faltas (B) em uma disciplina, juntamente com as linhas que representam médias gerais (roxa) e por status (amarela).	65
Figura 18 – Visão de Estágio. Distribuição proporcional de alunos que realizaram estágio (lado direito) e que não realizaram estágio (lado esquerdo). Cada barra horizontal corresponde a uma categoria de desempenho em uma disciplina, diferenciada por cores: verde escuro representa alunos aprovados na primeira tentativa, tons entre verde e vermelho representam aprovados em tentativas subsequentes, e vermelho representa alunos reprovados. O comprimento de cada barra é proporcional à quantidade relativa de alunos em cada grupo. A linha vertical central funciona como ponto de referência para comparação entre os dois grupos (com e sem estágio).	66
Figura 19 – Visão de Conteúdos. Cada gráfico de setor corresponde a um conteúdo, subdividido de acordo com a categoria de desempenho e tentativas dos alunos que compareceram à aula correspondente a esse conteúdo. Os conteúdos estão ordenados pela quantidade de aprovações. O conteúdo 13 está destacado em azul, mostrando a quantidade de alunos aprovados na primeira tentativa.	67
Figura 20 – Visão de Conteúdos, exibindo o número total de faltas em cada conteúdo de uma determinada disciplina.	68
Figura 21 – Fluxograma Curricular do Bacharelado em Sistemas de Informação. Fonte: https://facom.ufu.br/system/files/conteudo/fluxograma_curricular_do_bacharelado_em_si_0.pdf . Acesso em: 21 ago. 2025.	72
Figura 22 – Modelagem dos dados.	73
Figura 23 – Visão de Sequência de Execução para o fluxo GSI006 - GSI015.	76
Figura 24 – Exibição da proporção de alunos aprovados em GSI006 e que também foram aprovados em GSI015.	77
Figura 25 – A) Exibição da proporção de alunos reprovados em GSI006 e que também foram reprovados em GSI015. B) Exibição da proporção de alunos aprovados com duas tentativas em GSI006 e que foram aprovados diretamente em GSI015. C) Exibição da proporção de alunos aprovados com três tentativas em GSI006 e que foram aprovados diretamente em GSI015.	78

Figura 26 – Visão de Notas e Frequência da disciplina GSI006 com status reprovado, exibindo curvas de distribuição das notas (A) e faltas (B), juntamente com as linhas que representam médias gerais (roxa) e por status (amarela).	79
Figura 27 – Visão Estágio da disciplina GSI006.	80
Figura 28 – Visão de Conteúdos da disciplina GSI006: Gráfico de setores, com seleção dos conteúdos 11, 13, 15, 16, 18 e 19 em destaque.	81
Figura 29 – Visão de Conteúdos da disciplina GSI006: Faltas por Conteúdo.	82
Figura 30 – Visão de Sequência de Execução para o fluxo GSI015 - GSI006.	83
Figura 31 – Exibição da sequência de parte dos alunos reprovados em GSI015 e que também foram reprovados em GSI006.	84
Figura 32 – Visão de Notas e Frequência da disciplina GSI015 com status reprovado, exibindo curvas de distribuição das notas (A) e faltas (B), juntamente com as linhas que representam médias gerais (roxa) e por status (amarela).	85
Figura 33 – Visão Estágio da disciplina GSI015.	86
Figura 34 – Visão de Conteúdos da disciplina GSI015: Gráfico de setores, com seleção do conteúdo 11 em destaque.	86
Figura 35 – Visão de Conteúdos da disciplina GSI015: Faltas por Conteúdo.	87
Figura 36 – Visão de Sequência de Execução para o fluxo GSI002 - GSI006 - GSI016.	89
Figura 37 – Visão de Notas e Frequência da disciplina GSI016 com status reprovado, exibindo curvas de distribuição das notas (A) e faltas (B), juntamente com as linhas que representam médias gerais (roxa) e por status (amarela).	90
Figura 38 – Visão de Conteúdos da disciplina GSI016: Gráfico de setores.	91
Figura 39 – Visão de Conteúdos da disciplina GSI016: Faltas por Conteúdo.	91
Figura 40 – Visão de Sequência de Execução para o fluxo GSI002 - GSI016 - GSI006.	92
Figura 41 – Exibição da proporção de alunos aprovados em GSI016 e que também foram aprovados em GSI006 (29.41%). Exibição da proporção de alunos reprovados em GSI016 e que foram reprovados diretamente em GSI006 (69.23%).	93
Figura 42 – Visão de Notas e Frequência da disciplina GSI006 com status reprovado, exibindo curvas de distribuição das notas (A) e faltas (B), juntamente com as linhas que representam médias gerais (roxa) e por status (amarela).	94
Figura 43 – Visão de Conteúdos da disciplina GSI006: Gráfico de setores.	95
Figura 44 – Visão de Conteúdos da disciplina GSI006: Faltas por Conteúdo.	95
Figura 45 – Visão de Sequência de Execução para o fluxo GSI016 - GSI006 - GSI002.	96
Figura 46 – Visão de Sequência de Execução para o fluxo GSI006 - GSI002 - GSI016.	97

Lista de tabelas

Tabela 1 – Estrutura dos Dados.	61
Tabela 2 – Principais atributos utilizados nas visualizações	75

Sumário

1	INTRODUÇÃO	25
1.1	Objetivos e Desafios da Pesquisa	27
1.2	Estrutura do Documento	28
2	FUNDAMENTAÇÃO TEÓRICA	31
2.1	Conceitos Básicos	31
2.2	Análise de dados educacionais	32
2.3	Visualização de Informação	34
2.4	Considerações Finais	39
3	VISUALIZAÇÃO DE DADOS EDUCACIONAIS	41
3.1	Considerações Finais	55
4	DESCRIÇÃO DA PROPOSTA	57
4.1	Arquitetura do Sistema	57
4.2	Aspectos gerais do Design Visual	59
4.3	Estrutura dos Dados	60
4.4	Visões Analíticas	60
4.4.1	Visão de Sequência de Execução	62
4.4.2	Análise de uma disciplina específica	62
4.4.3	Visão de Conteúdos	67
4.5	Detalhes de Implementação	69
5	ANÁLISE DOS RESULTADOS	71
5.1	Contexto Curricular do Curso	72
5.2	Conjunto de Dados	73
5.2.1	Operações de Pré-processamento	74
5.3	Análises	76
5.3.1	Disciplinas GSI006 e GSI015	76

5.3.2	Disciplinas GSI002, GSI006 e GSI016	88
5.4	Discussão dos Resultados	99
6	CONCLUSÃO E TRABALHOS FUTUROS	101
	Referências	103

CAPÍTULO 1

Introdução

A análise do desempenho acadêmico de alunos do ensino superior é uma tarefa crucial e ao mesmo tempo desafiadora para profissionais da área de Educação. Educadores, coordenadores de programas e professores estão interessados em compreender como os alunos estão aprendendo conteúdos específicos, como determinados assuntos podem influenciar o aprendizado de outros tópicos, ou como o comportamento dos alunos em cada curso está relacionado ao seu desempenho nesse curso, dentre outras tarefas (ETEMADPOUR et al., 2019).

As instituições de ensino superior enfrentam desafios diversos no acompanhamento do desempenho de alunos, que vão além da simples avaliação de notas. Tarefas como identificação e acompanhamento de disciplinas com altas taxas de reprovação, análise das trajetórias dos alunos dentro do curso, tópicos mais desafiadores dentro de uma disciplina e análise da estrutura curricular dos cursos de graduação, dentre outras, são fundamentais para uma gestão educacional eficiente. A compreensão dessas dinâmicas permite que gestores e docentes tomem decisões mais assertivas sobre ajustes curriculares, estratégias de ensino e políticas de retenção. A evasão, por exemplo, é um dos problemas mais críticos, com causas complexas que envolvem desde fatores socioeconômicos até dificuldades com conteúdos disciplinares específicos (GONZÁLEZ-MORALES et al., 2025).

As universidades armazenam uma grande quantidade de informações sobre o desempenho dos alunos nas disciplinas que compõem os cursos de graduação, coletadas ao longo dos períodos letivos. Por exemplo, nos diários de classe dos docentes são registradas tanto as faltas ao longo do semestre quanto os conteúdos abordados em cada aula das disciplinas. Além disso, documentos como as matrizes curriculares e os planos de ensino oferecem detalhamentos importantes sobre a estrutura curricular, incluindo a organização sequencial dos componentes curriculares, os objetivos de aprendizagem, as competências previstas e as cargas horárias correspondentes. Informações relativas à realização de estágios também são registradas, normalmente por meio de sistemas acadêmicos institucionais, relatórios e formulários de acompanhamento, contendo dados como o período de realização, a carga horária cumprida e as disciplinas vinculadas. A análise integrada

desses dados pode auxiliar na compreensão da relação entre os componentes curriculares e o desempenho acadêmico dos alunos, permitindo a identificação de padrões de sucesso ou fracasso. Esse tipo de análise pode revelar, por exemplo, em quais disciplinas os alunos enfrentam maiores dificuldades, quais conteúdos estão mais associados à reprovação ou como a realização de estágio pode impactar o rendimento acadêmico. Para os gestores acadêmicos, essas informações representam uma base valiosa para avaliar a efetividade da estrutura curricular, planejar intervenções pedagógicas e propor ajustes nas políticas institucionais voltadas ao sucesso dos alunos.

Métodos analíticos e preditivos são comumente empregados na análise de informações relacionadas ao ensino superior, com o objetivo de auxiliar na compreensão do processo educacional e contribuir para a melhoria da eficácia do sistema como um todo. No entanto, segundo Almalawi et al. (2024), o principal desafio desses métodos está na utilização de informações relevantes nas análises. Muitos estudos baseiam-se majoritariamente em dados socioeconômicos e demográficos, como renda familiar, gênero, escolaridade dos pais e local de residência que, embora úteis para identificar padrões de vulnerabilidade, nem sempre representam com precisão a vivência acadêmica dos alunos. Tais dados pouco consideram o comportamento do aluno dentro das disciplinas, como frequência às aulas, desempenho em tópicos específicos, ordem de matrícula nas disciplinas e realização de estágios, que são aspectos diretamente ligados ao processo de aprendizagem. Além disso, a comunicação dos resultados obtidos por meio desses métodos frequentemente apresenta limitações, e as informações geradas podem ser de compreensão não-trivial pelo usuário. Além disso, esses métodos dependem de uma parametrização que nem sempre é simples de se obter.

Nesse sentido, abordagens que valorizem a interpretabilidade, a contextualização e a interatividade representam ferramentas importantes de análise. A *Visualização de Informação* é uma área que estuda a criação de técnicas visuais interativas por meio de computadores para auxiliar na análise e interpretação de grandes conjuntos de dados, facilitando a troca de informação entre humanos e computadores (FREITAS et al., 2001). Essa abordagem combina computação gráfica, interfaces homem-computador e mineração de dados, apresentando informações em *layouts* que permitem aos usuários utilizarem sua percepção visual para identificar padrões, tendências e valores atípicos nos dados (FREITAS et al., 2001) (HEER; BOSTOCK; OGIEVETSKY, 2010). O propósito da visualização é a percepção (*insight*), sendo que os principais objetivos dessa percepção são a descoberta, a tomada de decisões e o entendimento. No contexto educacional, técnicas de *Visualização de Informação* apresentam grande potencial para apoiar a análise do desempenho acadêmico dos alunos, ao oferecer representações visuais que facilitam a identificação de padrões, tendências e deficiências ao longo do tempo. Os *layouts* criados podem melhorar o processo analítico, reduzindo a carga cognitiva associada à sua interpretação e permitindo que gestores, professores e outros profissionais da educação

compreendam com mais clareza os fenômenos que afetam o processo de aprendizagem (WARE, 2004). Ao tornar os dados mais acessíveis e comprehensíveis, esses *layouts* promovem uma participação mais ativa desses profissionais no processo analítico, ampliando a possibilidade de tomada de decisões mais eficazes sobre o currículo e o desenvolvimento acadêmico dos alunos (KEIM, 2002).

1.1 Objetivos e Desafios da Pesquisa

Diante da necessidade de compreender o processo educacional acadêmico no ensino superior e a relação entre seus componentes, este trabalho propõe uma abordagem centrada tanto na análise e preparação dos dados acadêmicos quanto no desenvolvimento de representações visuais que favoreçam a interpretação de fenômenos complexos. Nesse contexto, apresenta-se o **EduFlow**, uma plataforma de análise dos dados de desempenho acadêmico de alunos universitários, utilizando técnicas interativas de *Visualização de Informação*. O EduFlow foi projetado para facilitar a compreensão das trajetórias acadêmicas dos alunos ao longo dos semestres e disciplinas cursadas, para analisar o impacto dessas sequências no sucesso ou fracasso desses alunos. A plataforma combina diferentes *layouts* interativos para relacionar informações associadas ao desempenho acadêmico, como notas, frequência, ordem de matrícula nas disciplinas e tópicos abordados em sala de aula. Além disso, o sistema possibilita a análise da relação entre participação em estágios e resultados acadêmicos, bem como a investigação de possíveis causas de reprovações associadas a conteúdos específicos.

A partir desse panorama, estabelecem-se os seguintes **objetivos específicos**:

- ❑ Prover um conjunto de dados capaz de revelar fenômenos relacionados ao desempenho do estudante em sequências de realização de disciplinas;
- ❑ Prover análises que permitam investigar o impacto de determinadas sequências de realização de disciplinas no desempenho dos alunos, além de análises que ilustrem o impacto no desempenho através do comparecimento ou não em determinados tópicos em uma disciplina;
- ❑ Projetar e implementar o sistema **EduFlow**, um sistema computacional de análise visual, que materializa os resultados das etapas anteriores em técnicas interativas de *Visualização de Informação* e orientadas ao apoio à decisão pedagógica.

Acredita-se que o uso de técnicas de *Visualização de Informação* para a análise de dados educacionais pode facilitar significativamente a compreensão de padrões e tendências presentes no processo de aprendizagem, permitindo que os fenômenos associados a esses dados sejam melhor compreendidos por profissionais da educação. Ao evidenciar visualmente comportamentos e correlações relevantes, tais como a relação entre frequência e

desempenho, ou entre disciplinas introdutórias e avançadas, o EduFlow contribui para a tomada de decisões mais fundamentadas, voltadas à melhoria do currículo, à definição de estratégias pedagógicas mais eficazes e à oferta de apoio direcionado aos alunos.

As principais contribuições desta pesquisa são:

- ❑ Uma abordagem de análise visual do desempenho acadêmico voltada à exploração da trajetória dos alunos, em termos da sequência de matrícula nas disciplinas e seu impacto no sucesso ou fracasso, bem como da análise de como seu comportamento durante os períodos letivos afeta o desempenho nessas disciplinas;
- ❑ Um sistema computacional que implementa a abordagem proposta de análise visual;
- ❑ Estudos de caso aplicados a um cenário real, demonstrando a eficácia da ferramenta na análise do desempenho acadêmico.

O trabalho descrito nesta dissertação resultou na submissão do seguinte artigo científico, atualmente em processo de revisão:

Marquez, L. F., Paiva, J. G. S. [*EduFlow: Analysis of University Students' Academic Performance Using Information Visualization Techniques*]. Journal of the Brazilian Computer Society. (under review)

1.2 Estrutura do Documento

Este trabalho está organizado conforme a seguinte estrutura:

- ❑ Capítulo 2 - Fundamentação Teórica: apresenta os conceitos básicos relacionados ao contexto educacional, incluindo definições fundamentais, abordagens de análise de dados educacionais e técnicas gerais de *Visualização de Informação*, que sustentam a proposta desenvolvida;
- ❑ Capítulo 3 - Visualização de Dados Educacionais: discute diferentes estratégias de visualização aplicadas ao contexto educacional, analisando suas contribuições, limitações e potenciais de uso para apoiar a compreensão do desempenho acadêmico;
- ❑ Capítulo 4 - Análise de Requisitos: descreve o processo de levantamento e definição dos requisitos para a construção do sistema EduFlow, com base em entrevistas com especialistas da área educacional e revisão da literatura;
- ❑ Capítulo 5 - Análise dos Resultados: detalha os experimentos realizados com o EduFlow, incluindo a descrição do conjunto de dados utilizado, os cenários analisados e as descobertas obtidas nas visualizações, além da avaliação de como os requisitos foram atendidos;

- ❑ Capítulo 6 - Conclusão e Trabalhos Futuros: apresenta as principais conclusões obtidas com a pesquisa, ressaltando as contribuições do sistema desenvolvido, e propõe direções para trabalhos futuros, visando a evolução e aplicação do EduFlow em diferentes contextos acadêmicos.

CAPÍTULO 2

Fundamentação Teórica

Um grande desafio para a área acadêmica é compreender e acompanhar o desempenho dos alunos, seja para detectar deficiências na estrutura curricular atual, seja na identificação de gargalos e represamentos, seja na compreensão de causas de evasão, dentre outras. A análise de dados acadêmicos tem se consolidado como uma importante ferramenta de apoio à gestão acadêmica, possibilitando uma compreensão mais aprofundada do desempenho de alunos e dos fatores que influenciam sua trajetória. Ao longo dos últimos anos, diferentes técnicas computacionais têm sido empregadas nesse contexto, incluindo métodos estatísticos, modelos preditivos e análises exploratórias (MORENO-MARCOS et al., 2018) (KUI et al., 2022) (ZHANG et al., 2023). Cada uma dessas abordagens oferece formas distintas de investigar os dados, com diferentes níveis de profundidade, complexidade e interpretabilidade.

Este capítulo apresenta a fundamentação teórica que embasa esta pesquisa. Inicialmente, são introduzidos os principais conceitos e terminologias utilizadas ao longo do trabalho. Em seguida, é abordado o contexto da análise de dados educacionais. Por fim, são explorados os fundamentos das técnicas de *Visualização de Informação*.

2.1 Conceitos Básicos

A seguir são apresentados os principais conceitos que fundamentam a análise de dados acadêmicos, com o objetivo de contextualizar a proposta deste estudo:

- ❑ **Aluno:** Indivíduo regularmente matriculado em um curso de graduação, que participa das atividades acadêmicas propostas com o objetivo de adquirir competências e concluir sua formação;
- ❑ **Curso:** Programa estruturado de estudos para formação de um profissional em determinada área de atuação, composto por diversas disciplinas, cada uma representando uma área específica do conhecimento;

- **Disciplina:** Componente curricular que aborda conteúdos específicos dentro do curso. Cada disciplina possui carga horária, objetivos e competências próprias;
- **Semestre acadêmico:** Unidade temporal que organiza o calendário acadêmico e define os períodos de realização das disciplinas ao longo do curso;
- **Turma:** Conjunto de alunos matriculados em uma mesma oferta de disciplina, no mesmo período letivo, compartilhando o mesmo cronograma de aulas, conteúdo programático e professor responsável;
- **Conteúdo:** Conjunto de assuntos a serem abordados dentro de uma disciplina, organizados segundo um plano de ensino;
- **Tópico:** Assunto componente de um conteúdo de disciplina, abordado em uma ou várias aulas;
- **Aula:** Sessão de aprendizagem agendada (como palestras, discussões ou atividades práticas) conduzida por um docente, para apresentação de um ou vários tópicos;
- **Frequência:** Indicador do nível de participação dos alunos em cada aula, confirmindo sua presença na mesma;
- **Falta:** Indicador de ausência de um aluno em uma aula ministrada;
- **Nota:** Avaliação numérica ou conceitual atribuída ao aluno, representando seu desempenho acadêmico na disciplina;
- **Aprovação:** Situação em que o aluno obtém desempenho e participação (nota e frequência) suficiente para adquirir a competência no conteúdo relacionado à disciplina;
- **Reprovação:** Situação em que o aluno não apresentou um nível mínimo necessário para dominar a competência no conteúdo relacionado à disciplina, seja por nota inferior à mínima exigida, seja por excesso de faltas, exigindo que o aluno refaça a disciplina;
- **Evasão:** Ato de abandono da disciplina ou do curso por parte do aluno, resultando na interrupção de sua trajetória acadêmica.

2.2 Análise de dados educacionais

De acordo com Deng et al. (2019) nos últimos anos, muito esforço tem sido dedicado ao desenvolvimento de estratégias de análise de aprendizagem para ajudar alunos, professores e profissionais da área de Educação a compreender os comportamentos de aprendizagem

e padrões de desempenho dos alunos. Os esforços iniciais foram focados em destacar os alunos em potencial em risco de fracasso acadêmico (ARNOLD; PISTILLI, 2012) (LIM et al., 2019).

Segundo Goulden et al. (2019), estudos adicionais sugerem que a taxa de retenção está relacionada ao desempenho do aluno (BEAN, 1987), e o primeiro ano é o mais crucial para a conclusão no programa (BAREFOOT et al., 2010). Para ajudar os alunos a terem sucesso em seus programas de curso e incentivá-los a voltar nos anos subsequentes, é fundamental que os instrutores entendam seus comportamentos de aprendizagem, além de identificar aqueles em risco o mais cedo possível. Também é importante que os alunos compreendam os hábitos de aprendizagem entre seus pares e busquem aprimorar suas estratégias de estudo. Essa compreensão pode estimular o engajamento acadêmico, promover o autocontrole sobre o processo de aprendizagem e contribuir para a construção de trajetórias mais consistentes ao longo do curso. Nesse contexto, diferentes estudos têm investigado como os alunos aprendem tópicos específicos, de que maneira determinados conteúdos influenciam o aprendizado de outros e como indicadores como notas e frequência podem refletir o desempenho acadêmico ao longo do curso (ETEMADPOUR et al., 2019). A análise dessas informações permite não apenas monitorar o progresso dos alunos, mas também identificar possíveis deficiências ou dificuldades recorrentes.

As universidades armazenam uma grande quantidade de informação a respeito do desempenho dos alunos nas disciplinas que compõem os cursos oferecidos, coletadas ao longo dos períodos letivos. A análise dessas informações pode auxiliar na compreensão da relação entre componentes curriculares e o desempenho acadêmico dos alunos, propiciando a descoberta das possíveis causas das deficiências dos alunos nesses componentes curriculares, que levam à reprovação. A mineração de dados educacionais extrai informações valiosas de dados brutos para obter melhores processos de aprendizagem nos cursos (DENG et al., 2019).

Nos últimos anos, muitas técnicas de *Machine Learning* foram amplamente adotadas para prever o desempenho dos alunos. A ideia é aprimorar a compreensão do processo de aprendizagem para focar na identificação, extração e avaliação de variáveis relacionadas a esse processo (EL-HALEES, 2009). Baradwaj e Pal (2011) utilizaram *Decision Trees* para prever o desempenho dos alunos no final do semestre, do curso de computação da Universidade Purvanchal, entre 2001 e 2010. Foram avaliados atributos como presença, testes em sala de aula, seminários e notas. Essa análise ajudou a identificar alunos que precisam de atenção especial para reduzir a taxa de reprovação e a tomada de medidas adequadas para as provas do próximo semestre. No entanto, essa análise não permite compreender as causas do baixo desempenho, os conteúdos com maior número de faltas e nem o caminho percorrido por alunos que foram reprovados.

Okubo et al. (2017) apresentam um método de previsão das notas finais de 108 alunos da Universidade Kyushu por meio de uma *Recurrent Neural Network* (RNN). O algoritmo

analisou dados de registro como a quantidade de relatórios e questionários respondidos por alunos, a quantidade de diários de bordo preenchidos e a quantidade de slides lidos no sistema. Com precisão superior a 90%, a partir dos resultados experimentais os autores concluíram que uma RNN é eficaz para a previsão antecipada das notas finais.

Moreno-Marcos et al. (2018) analisam um curso *on-line* sobre programação Java, da Universidade Carlos III de Madrid, para previsão de notas. Foram coletadas informações sobre notas anteriores dos alunos, a forma como alunos interagem no curso, resolução de exercícios e estrutura do curso. Quatro algoritmos – *Regression* (RG), *Support Vector Machines* (SVM), *Decision Trees* (DT) e *Random Forest* (RF) – foram considerados para fazer as análises de predição e, na maioria dos casos, é possível prever quem vai passar no curso, com um melhor aproveitamento do algoritmo *Random Forest*. No entanto, este trabalho não explora os principais motivos que levaram os alunos a um baixo desempenho, se podemos traçar um perfil desses alunos ou qual o melhor caminho de aprendizado a ser seguido por eles.

Dessa forma, a construção de sistemas baseados exclusivamente em métodos analíticos tradicionais ou técnicas de *Machine Learning* pode limitar a compreensão aprofundada sobre o desempenho acadêmico dos alunos. Tais abordagens, apesar de úteis em determinados contextos, nem sempre conseguem representar com precisão as múltiplas dimensões envolvidas na trajetória acadêmica, dificultando a identificação de gargalos curriculares, deficiências estruturais no curso ou padrões associados à evasão. Além disso, a interpretação dos resultados gerados por esses métodos muitas vezes exige conhecimento técnico especializado e depende de parametrizações complexas, que nem sempre refletem os objetivos pedagógicos ou as decisões práticas do ambiente educacional. Isso torna desafiadora a aplicação efetiva dessas análises por parte de educadores e gestores.

2.3 Visualização de Informação

A área de *Visualização de Informação* estuda a criação de representações gráficas interativas de dados abstratos com o intuito de auxiliar o processo de análise e compreensão de fenômenos que ocorrem em um conjunto de dados. Uma técnica de visualização é baseada numa representação visual e em mecanismos de interação que permitem ao usuário manipular essa representação de modo a melhor compreender o conjunto de dados ali representado, possibilitando ao mesmo observar características ou padrões no conjunto de dados (GERSHON; EICK, 1997).

O processo de visualização pode ser representado por uma sequência de mapeamentos ajustáveis de dados para uma representação visual, de modo a possibilitar interação do usuário com o espaço de informação e proporcionar a **cristalização do conhecimento**. A **Figura 1** mostra o processo completo de *Visualização de Informação*.

Na primeira etapa do processo, Dados Brutos são estruturados e transformados em

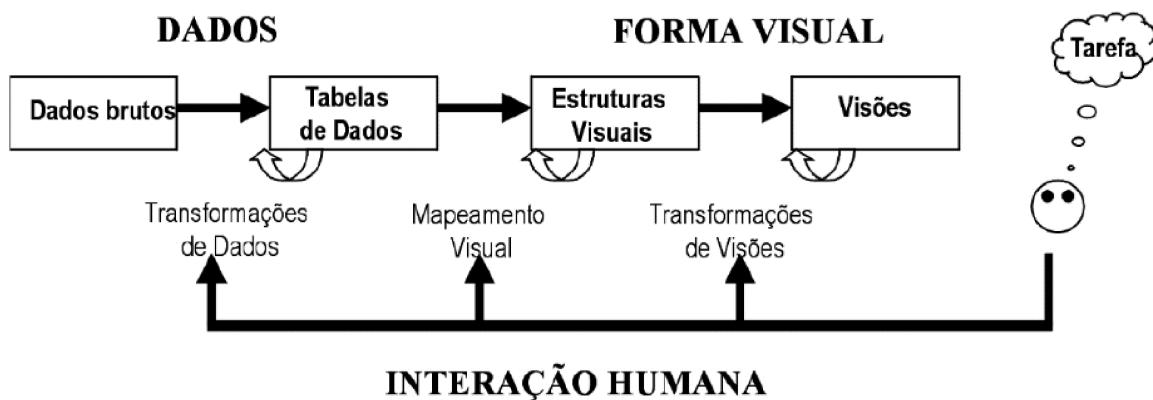


Figura 1 – Processo de *Visualização de Informação* adaptado de Card, Mackinlay e Shneiderman (1999)

Tabelas de Dados, que são descrições relacionais padronizadas que incluem metadados. Essa etapa é necessária em situações nas quais os Dados Brutos não apresentam uma padronização na representação das informações contidas, ou seu conteúdo não está organizado de maneira trivial. Novas transformações podem ser aplicadas, quer para agregar novos dados ao conjunto inicial, como, por exemplo, calcular grandezas estatísticas, quer para converter os dados originais para outros tipos ou reorganizar o conjunto de dados, classificando-o, por exemplo. As Tabelas de Dados são então mapeadas em Estruturas Visuais utilizando metáforas que estejam relacionadas com as tarefas a serem realizadas nesses dados, ao contexto nas quais estão inseridas, ou mesmo a convenções/costumes que auxiliam na compreensão pelo usuário. As Estruturas Visuais são então organizadas em *layouts* interativos, de acordo com um ou vários aspectos de análise, compondo Visões que serão exploradas pelos usuários na forma de tarefas de análise.

Utilizando técnicas de *Visualização de Informação*, dados podem ser analisados de forma que os usuários compreendam as informações e deduzam novos conhecimentos. Segundo Hansen, Borges e Holanda (2020), as técnicas de visualização trazem diversas vantagens para a análise de dados, tais como: i) compreender a informação mais rapidamente, devido ao fato de que a visão é o sentido humano com maior capacidade de capturar informações nos menores períodos de tempo; ii) visualizar grandes quantidades de dados de forma intuitiva e coesa; iii) descobrir valores atípicos e reconhecer padrões e relações; e iv) envolver o usuário por meio da veiculação de uma mensagem, que pode gerar impacto, funcionar como extensão da memória humana e também auxiliar no processo cognitivo (NASCIMENTO; FERREIRA, 2011).

Existem diversas técnicas aplicadas a análise de diferentes aspectos dos dados. Gráficos de barras, gráficos de linhas e gráficos de setores são exemplos de visualizações fundamentais que, como apontado por Card, Mackinlay e Shneiderman (1999), permitem

representar dados categóricos e quantitativos de forma intuitiva. Os gráficos de barras são eficazes para comparar valores entre diferentes categorias, com barras verticais ou horizontais cujo comprimento representa a magnitude dos dados. Gráficos de linhas são apropriados para representar variações temporais ou sequências ordenadas, evidenciando tendências, ciclos ou transições ao longo do tempo. Já os gráficos de setores são utilizados para ilustrar proporções dentro de um todo, permitindo observar a distribuição relativa entre categorias. Essas representações oferecem um ponto de partida acessível para o entendimento de padrões gerais e são particularmente úteis quando o número de variáveis é pequeno e as comparações são diretas. No entanto, à medida que a complexidade e a dimensionalidade dos dados aumentam, torna-se necessário adotar técnicas mais avançadas que permitam representar simultaneamente múltiplas variáveis.

A técnica de Coordenadas Paralelas (INSELBERG; DIMSDALE, 1990) é um exemplo no qual dados representados em um espaço n -dimensional são mapeados em uma estrutura bidimensional, que usa n eixos paralelos equidistantes (usualmente verticais). Uma linha representa cada instância interceptando cada eixo no local correspondente ao valor do atributo ao qual esse eixo corresponde, e a disposição das linhas cria padrões relacionados aos fenômenos observados nos dados. Linhas semelhantes podem identificar grupos e pontos de cruzamento identificam relações entre os atributos. A forma como os eixos são organizados nas coordenadas paralelas é decisivo para a análise de relação entre os atributos, sendo que analisar a relação entre eixos adjacentes é mais fácil que analisar a relação entre atributos distantes. Além disso, a visualização dos padrões pode apresentar alguns problemas com coleções muito grandes, já que a representação visual pode se tornar muito densa, reduzindo a representação a apenas um polígono de cor uniforme.

A **Figura 2** mostra um exemplo de coordenadas paralelas aplicado à base de dados de alunos do sistema **PerformanceVis** (DENG et al., 2019). Na Figura é possível observar que o grupo de alunos que obtiveram ‘C’ no Exame 3 e no exame final (classificados pela cor azul), tiveram pontuação menor nos exames 1 e 2 em comparação com o grupo dos alunos que obtiveram ‘C’ no Exame 3 e ‘A’ no exame final (classificados na cor vermelha). Isso sugere que o desempenho dos alunos nos Exames 1 e 2 pode afetar sua nota no exame final.

As técnicas de visualização hierárquica e relacional têm se mostrado fundamentais para representar e interpretar dados educacionais complexos. A visualização do tipo *sunburst*, conforme descrito por Kui et al. (2022), organiza dados hierárquicos em círculos concêntricos, nos quais o centro representa o nível mais alto da hierarquia e os segmentos radiais indicam subníveis. A largura dos segmentos é proporcional a uma métrica de interesse, como frequência ou quantidade, facilitando a análise de distribuição e a identificação de relações entre tópicos ou módulos educacionais, sendo especialmente eficaz para representar estruturas curriculares ou caminhos de navegação em ambientes de aprendizagem *on-line*. Por sua vez, os *heatmaps* representam dados quantitativos em matrizes

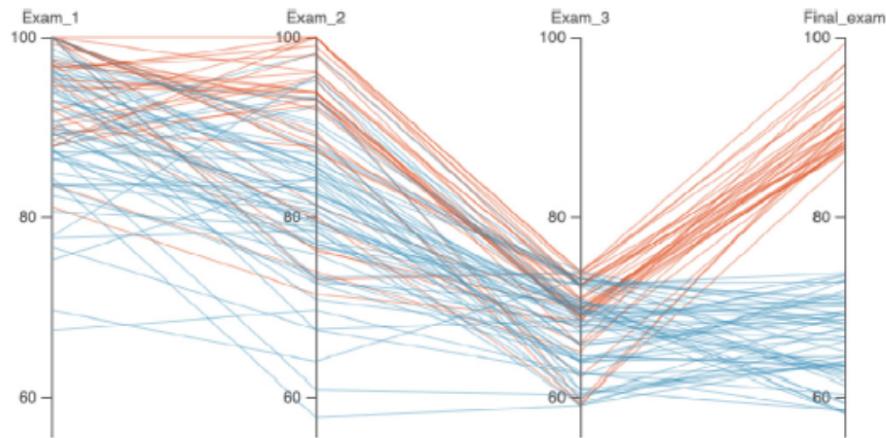


Figura 2 – Comparação entre dois conjuntos de caminhos de notas do aluno. A cor vermelha representa o mesmo grupo de alunos que obtiveram C no Exame 3 e A no exame final; a cor azul representa os alunos que obtiveram C no Exame 3 e no exame final (DENG et al., 2019).

bidimensionais, em que as células são preenchidas com cores graduadas de acordo com o valor que expressam. Essa técnica permite destacar rapidamente padrões, correlações, agrupamentos ou anomalias, sendo amplamente aplicada na comparação de desempenho entre alunos, na identificação de tópicos com maiores taxas de erro ou ausência, e na detecção de áreas críticas de engajamento. Já a visualização por grafos, como apontado por Emmons, Light e Börner (2017), representa entidades (como alunos, tópicos ou atividades) como nós, e suas relações (como interações, semelhanças de desempenho ou co-participações) como arestas. Essa abordagem é útil para explorar redes complexas em plataformas educacionais, como *Massive Open Online Course (MOOCs)*, permitindo identificar comunidades, padrões colaborativos, caminhos de navegação e nós centrais ou isolados, fornecendo informações relevantes para educadores. Por fim, a visualização em forma de árvore (*node-link*), como apontado por Raji et al. (2017), estrutura os dados em uma hierarquia explícita, onde os nós representam disciplinas ou eventos acadêmicos e as arestas denotam a progressão dos alunos. Essa técnica permite rastrear os caminhos percorridos pelos alunos ao longo do currículo, evidenciar taxas de transição entre disciplinas, identificar gargalos e desvios na trajetória acadêmica, além de revelar padrões de retenção ou evasão, oferecendo suporte à melhoria do planejamento pedagógico.

O Diagrama de Sankey (RIEHMANN; HANFLER; FROEHLICH, 2005) é uma técnica de visualização voltada à representação de fluxos quantitativos entre categorias distintas, permitindo a análise de transições, distribuições e proporções de forma intuitiva. Essa técnica utiliza nós para representar categorias discretas e fluxos conectores cuja espessura é proporcional à quantidade de elementos que transita entre essas categorias. Diagramas de Sankey interativos permitem manipulação dinâmica dos dados, reordenação dos nós e destaque de caminhos específicos, o que facilita a análise de grandes volumes de dados.

complexos e a identificação de padrões relevantes. Além disso, a integração de interatividade à visualização contribui significativamente para a percepção de relações entre variáveis, identificação de gargalos e interpretação de processos sequenciais. Essa abordagem tem sido amplamente utilizada em ambientes educacionais, para mapear trajetórias de aprendizagem e padrões de desempenho de alunos.

O sistema ***CCVis*** (GOULDEN et al., 2019) apresenta o *Grade Distribution View* (GDV), uma visualização baseada em diagrama de Sankey que tem como objetivo revelar a associação entre padrões de comportamento dos alunos definidos por sequências de cliques em conteúdos educacionais *on-line* e suas respectivas faixas de notas finais, conforme destacado na **Figura 3**. No lado esquerdo do Sankey, encontram-se os retângulos identificados por códigos como “18|37.xx”, que representam agrupamentos de alunos com comportamentos de navegação semelhantes. O prefixo numérico refere-se a uma combinação específica de curso e atividade, enquanto o sufixo distingue instâncias distintas dentro do mesmo contexto. Esses padrões podem ser classificados como típicos ou *outliers*: os típicos são exibidos em branco e indicam comportamentos cuja distribuição de notas está em conformidade com a média do grupo, enquanto os *outliers* são destacados em vermelho e representam padrões cujos desempenhos acadêmicos divergem significativamente da média, podendo indicar tanto comportamentos de risco quanto casos de desempenho excepcional. A pilha de retângulos abaixo da etiqueta “Choose node” funciona como um painel de seleção interativa, permitindo ao analista escolher até quatro padrões para análise detalhada. À medida que um ou mais padrões são selecionados, os fluxos correspondentes são destacados no Sankey, conectando os nós de comportamento à direita, onde estão representadas faixas de notas finais (por exemplo, 0.00–0.69 ou 0.90–0.99), com espessura proporcional ao número de alunos que seguiram determinado caminho. Essa representação permite observar como diferentes perfis de navegação influenciam os resultados acadêmicos, possibilitando a identificação de padrões comuns ou atípicos de desempenho.

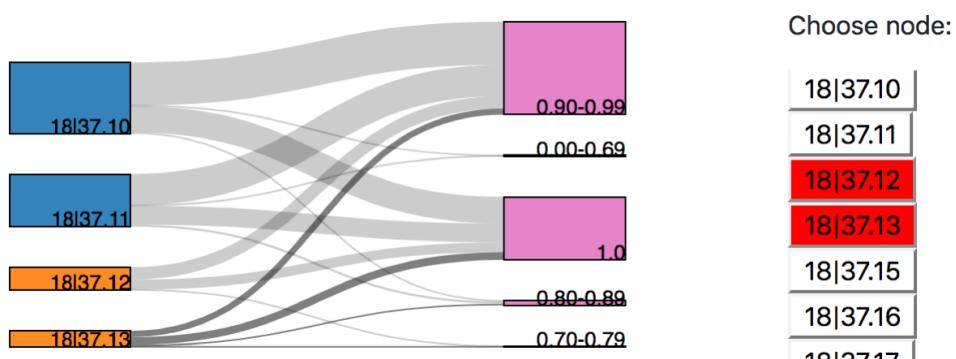


Figura 3 – Distribuição de notas para dois nós típicos (18|37.10 e 18|37.11) e dois nós *outliers* (18|37.12 e 18|37.13) na visualização GDV (Grade Distribution View) (GOULDEN et al., 2019).

2.4 Considerações Finais

A análise de dados educacionais, aliada às técnicas de *Visualização de Informação*, tem se mostrado uma abordagem promissora para ampliar a compreensão sobre o desempenho acadêmico dos alunos e os fatores que influenciam sua trajetória ao longo do curso. A fundamentação teórica apresentada neste capítulo destaca não apenas os conceitos essenciais relacionados ao contexto educacional, mas também a relevância de métodos visuais interativos como ferramentas complementares a análises estatísticas e modelos preditivos. Ao representar graficamente padrões, relações e progressões acadêmicas, as técnicas de visualização possibilitam uma leitura mais intuitiva e acessível dos dados, promovendo *insights* que podem subsidiar decisões pedagógicas e estratégias de intervenção mais eficazes. Dessa forma, a integração entre análise de dados e visualização interativa representa um avanço significativo na construção de sistemas de apoio à gestão acadêmica, com potencial para impactar positivamente a qualidade do ensino e a permanência dos alunos.

CAPÍTULO 3

Visualização de Dados Educacionais

Diversas pesquisas vêm aplicando técnicas de visualização de dados no contexto educacional com o objetivo de apoiar tarefas analíticas como a identificação de padrões de desempenho, o monitoramento da evasão dos alunos, a análise do progresso acadêmico e a avaliação da efetividade curricular (OLSSON; MOZELIUS; COLLIN, 2015; KOUROUNIOTIS, 2023). Essas abordagens visuais permitem a exploração interativa de grandes volumes de dados acadêmicos, facilitando a compreensão de fatores que influenciam o sucesso ou fracasso dos alunos. Neste capítulo, são apresentadas e discutidas propostas relevantes da literatura que utilizam estratégias de visualização aplicadas à educação, destacando suas contribuições, limitações e oportunidades de aprimoramento, com foco especial em como essas técnicas podem apoiar o processo de tomada de decisão em ambientes educacionais.

Etemadpour et al. (2019) apresentam uma ferramenta de análise visual para análise de desempenho acadêmico utilizando informações sobre frequência e notas dos alunos. No trabalho são utilizados vários modelos de *Machine Learning* para prever tendências de dados dos alunos ao longo de vários semestres, que são comparados com dados reais para avaliação dos modelos. Foram considerados onze diferentes modelos de *Machine Learning* no sistema, cujos resultados da predição de cada modelo são mostrados em *heatmaps*, conforme a **Figura 4**, que mostra uma comparação entre os dados reais e dados preditos por um modelo preditivo baseado em *Random Forest*. Esse *layout* utiliza um gráfico bidimensional cujo eixo horizontal representa o número de faltas dos alunos e o eixo vertical corresponde às suas notas finais no curso de *Algorithm and Programming I*. Cada célula da matriz representa uma combinação específica de número de ausências e nota obtida, colorido de acordo com a densidade de alunos que compartilham essas características. Quanto maior a intensidade de cor da célula, menor o número de alunos naquela interseção; quanto menor a intensidade, maior é a densidade de alunos com aquela característica. A Figura também apresenta círculos sobrepostos a determinadas células, utilizados para enfatizar áreas de interesse na análise. O círculo verde do gráfico com dados reais (parte superior da figura) indica que alunos com até 3 faltas conseguiram mais de

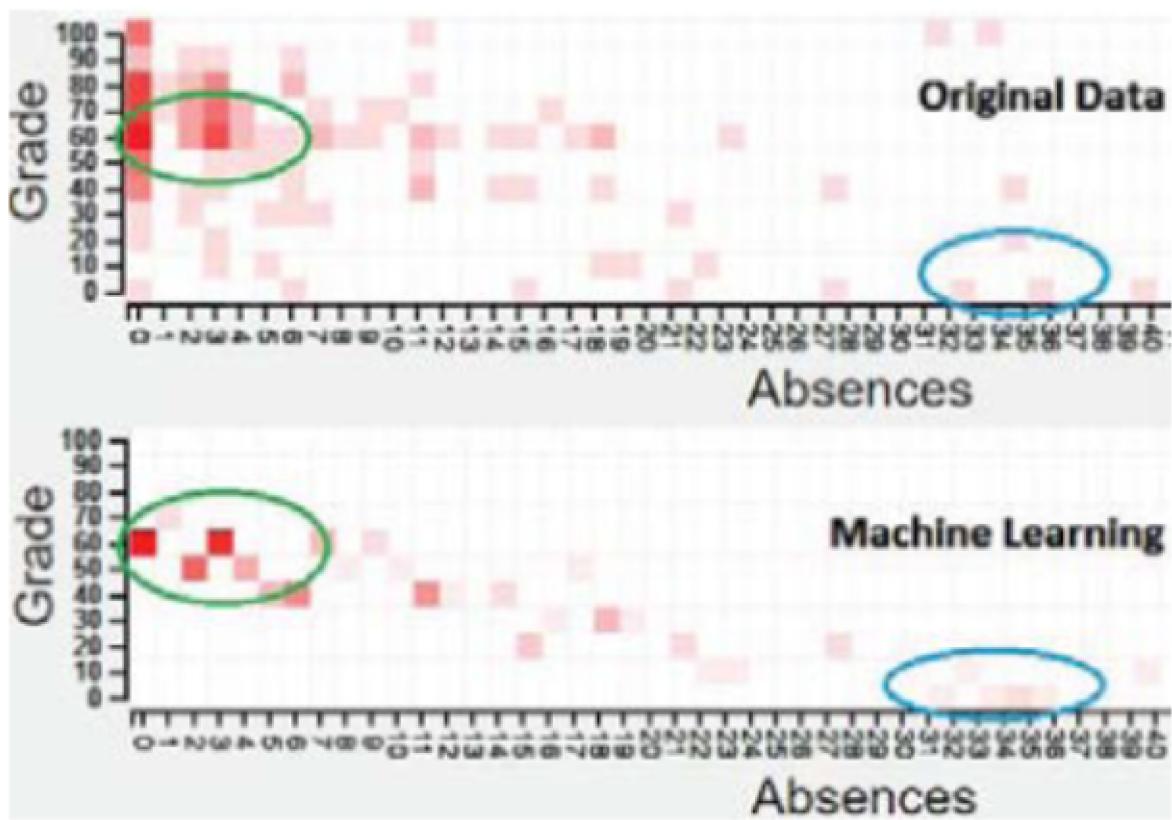


Figura 4 – Visualização do *heatmap* do curso de Algoritmo e Programação I, comparando notas e faltas e dos alunos e em seguida um modelo de previsão com machine learning com o algoritmo *Random Forest*, para prever padrões de comportamento (ETEMADPOUR et al., 2019).

60% da nota. Já no círculo azul observa-se que uma grande concentração de alunos com muitas faltas, apresentam médias muito baixas, até 20%, o que sugere uma forte correlação entre frequência e desempenho. Esse padrão também é parcialmente reproduzido no *heatmap* gerado pelo modelo *Random Forest* (parte inferior da figura), sugerindo que o modelo conseguiu capturar parte da relação entre absenteísmo e rendimento acadêmico. A correspondência parcial entre as distribuições revela que o número de faltas, embora relevante, não é o único fator determinante no desempenho, apontando para a necessidade de incorporar outras variáveis no modelo preditivo. Os usuários podem interagir com o sistema para explorar diferentes modelos preditivos, comparando os resultados reais e previstos a fim de avaliar a acurácia dos modelos. Apesar de ser útil na avaliação do desempenho dos alunos, o sistema se limita a exibir a relação entre notas e faltas em uma disciplina específica, bem como a avaliar o desempenho dos modelos de predição nessas disciplinas.

Deng et al. (2019) desenvolveram o sistema **PerformanceVis**, uma ferramenta interativa de análise visual voltada para a compreensão do desempenho dos alunos em um curso

de Química. A proposta se apoia em três categorias principais de dados: (i) desempenho acadêmico, (ii) características dos alunos e (iii) comportamento de aprendizagem. A primeira inclui as notas finais do curso e as pontuações obtidas em tarefas, testes e atividades de laboratório. A segunda abrange atributos demográficos e acadêmicos dos alunos, como etnia, pontuação geral e classificação acadêmica. Já a terceira diz respeito à forma como os alunos interagem com as avaliações, incluindo número de tentativas em questões, uso de dicas e desistências. Uma das principais visualizações do sistema é representada por um diagrama de Sankey bidimensional, conforme ilustrado na **Figura 5**.

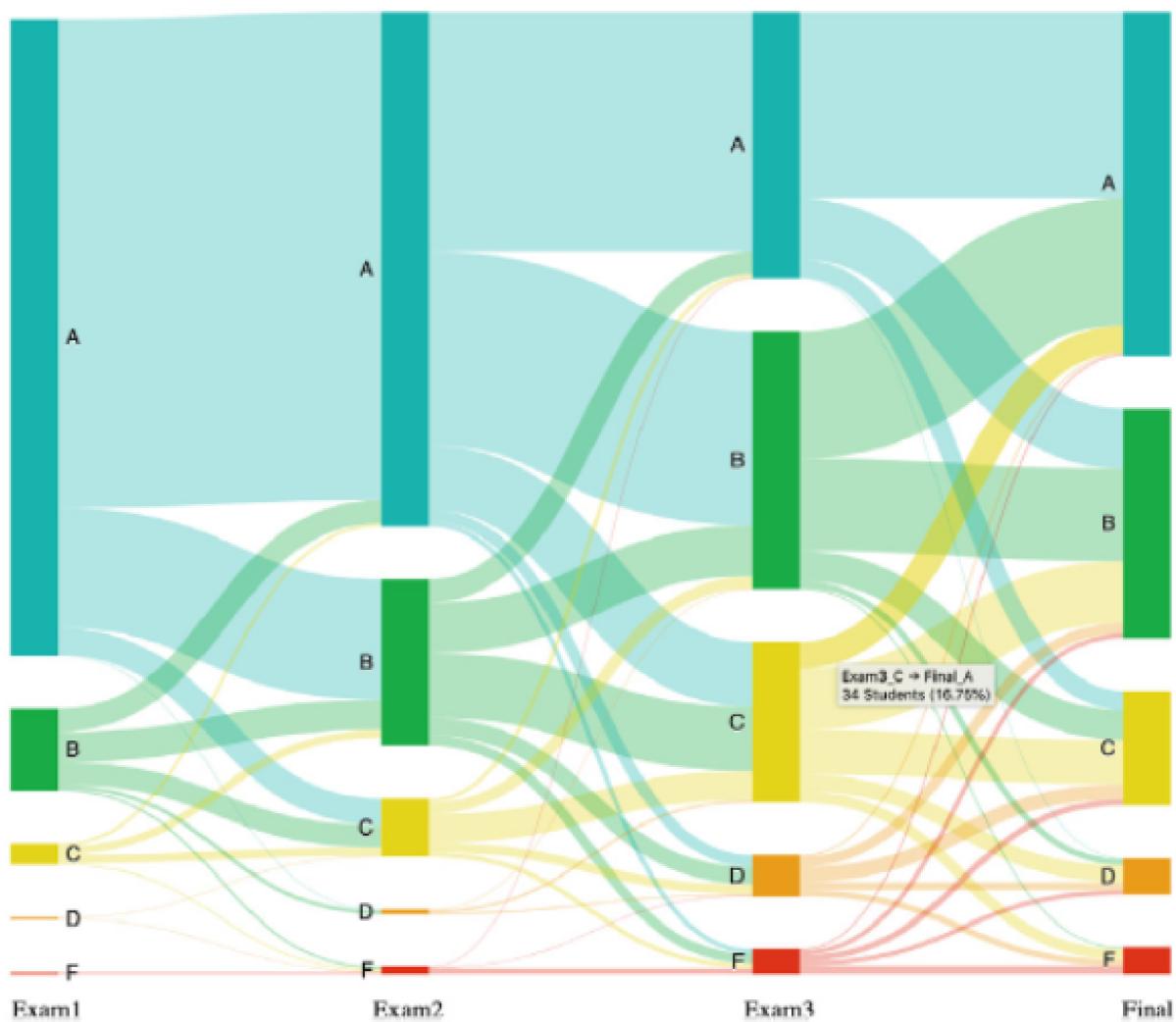


Figura 5 – Caminho geral das notas de exames selecionando o fluxo da nota C no Exame 3 para A no Exame final (DENG et al., 2019).

Esse diagrama mostra a evolução das notas de 949 alunos ao longo de quatro exames distintos. No eixo horizontal, têm-se os momentos avaliativos (*Exam1* a *Final*), enquanto o eixo vertical representa as faixas de nota categorizadas de A a F. Os nós indicam as distribuições das notas em cada exame, e os fluxos entre os nós representam a transição

dos alunos de uma nota para outra ao longo do tempo. A espessura de cada fluxo é proporcional à quantidade de alunos que seguem aquele caminho, permitindo identificar padrões recorrentes de melhora ou piora no desempenho. Os usuários podem encontrar informações mais detalhadas sobre cada fluxo, incluindo o número exato e a proporção de alunos que caem no fluxo, passando o mouse sobre o fluxo. A ferramenta oferece uma visão geral do desempenho de todos os alunos e uma visão das tendências gerais das notas ao longo do curso. Além disso, os usuários podem identificar grupos específicos de alunos que diminuíram ou melhoraram ao longo do tempo com a lista de distribuições de notas de exames entre todos os alunos. Com isso, é possível identificar, por exemplo, que um grupo de alunos que obteve nota C no terceiro exame conseguiu atingir nota A no exame final. A visualização, portanto, facilita a detecção de tendências e mudanças significativas no desempenho dos alunos ao longo do curso. Entretanto, o PerformanceVis não permite observar um panorama das correlações entre disciplinas e desempenho dos alunos, por exemplo, alunos que fizeram a disciplina “A” primeiro, apresentam melhor aproveitamento na disciplina “B”.

Raji et al. (2017) propõem o **eCamp**, um sistema de análise visual desenvolvido para representar fluxos estudantis em larga escala dentro de uma universidade. A interface permite explorar a progressão dos alunos entre cursos e departamentos, integrando dados de graduação, histórico acadêmico e classificações curriculares. A **Figura 6** apresenta uma visão geral do sistema baseada em uma árvore radial no estilo Sankey. Nessa visualização, os nós folhas representam os cursos de graduação (*majors*), enquanto os ramos e suas respectivas espessuras indicam os caminhos percorridos por grupos de alunos ao longo dos semestres. Quanto mais espesso o caminho, maior o número de alunos que o seguiu. Essa estrutura permite observar de forma agregada os fluxos de entrada e saída entre diferentes áreas acadêmicas, revelando tendências de migração, evasão e conclusão de curso. As cores dos nós indicam departamentos ou áreas temáticas, facilitando a identificação de padrões interdisciplinares. A disposição radial temporal permite que se observe a progressão dos alunos desde a entrada até a formatura ou abandono. Já a **Figura 7** destaca um subconjunto da árvore, focando especificamente nos fluxos relacionados aos cursos de Ciência da Computação e Engenharia Elétrica. Nela, observa-se que alguns ramos originados nesses cursos se dispersam em direções distintas ainda no primeiro ou segundo semestre, evidenciando um desvio precoce de alunos para outras áreas, como Psicologia ou cursos interdisciplinares. Por exemplo, o ramo inferior evidencia uma trajetória que se interrompe após o primeiro nó, indicando possível evasão; outro ramo mostra transição para outro curso após o segundo semestre, e um terceiro segue até um curso não especificado. Essa configuração revela como o sistema permite identificar rotas alternativas, momentos críticos de decisão dos alunos, e áreas de maior ou menor retenção. Entretanto, o sistema não analisa os percentuais de ausência por conteúdo das disciplinas nem sua correlação com o desempenho acadêmico, o que o impede de analisar em



Figura 6 – Visualização do fluxo do aluno em toda a universidade, mostrando a progressão do aluno semestre a semestre. Cada nó folha corresponde a um único principal, e a espessura dos caminhos corresponde ao número de alunos progredindo através dos nós (RAJI et al., 2017).

quais componentes curriculares a presença é um fator crítico para o sucesso. Além disso, o sistema não oferece uma visão geral dos perfis de alunos mais propensos à aprovação ou reprovação, tampouco fornece *insights* sobre quais disciplinas deveriam ser cursadas primeiro para otimizar o desempenho acadêmico ao longo do programa. Seguindo uma linha semelhante, Gutierrez-Pachas et al. (2022) utilizam visualizações baseadas em redes complexas, diagramas de fluxo de progressão e gráficos estatísticos para ilustrar tendências de desempenho e evasão, bem como o impacto de mudanças no desenho curricular sobre o sucesso acadêmico. Os autores destacam como diferentes fatores, como origem

socioeconômica, equilíbrio de gênero e sequências de conclusão de disciplinas afetam a permanência e a progressão dos alunos. O sistema permite explorar as relações entre estruturas curriculares e o sucesso discente, identificar padrões de risco de evasão e analisar a influência da sequência de disciplinas sobre o desempenho acadêmico.



Figura 7 – Detalhamento do ramo do gráfico radial que contém os cursos de Computação e Engenharia Elétrica da universidade. Esses cursos se separaram da maioria dos outros cursos no final do primeiro semestre (RAJI et al., 2017).

Alguns estudos têm como foco específico a análise de plataformas de ensino *on-line*. Emmons, Light e Börner (2017) consideram a importância da análise de dados educacionais para melhorar plataformas de cursos *on-line* e, consequentemente, reduzir a evasão de alunos nesses cursos. O estudo apresenta o uso de análise de dados e técnicas de visualização com o intuito de auxiliar alunos, professores, desenvolvedores e pesquisadores das plataformas a compreenderem um grande volume de dados acerca do processo de aprendizado dos alunos e interação dos alunos com o material disponibilizado pelo curso. A **Figura 8** mostra um exemplo de análise utilizando o *layout* proposto, na qual examina as interações em um curso, que podem ser entre alunos, entre materiais de um curso e entre os diferentes cursos de uma plataforma de aprendizado *on-line*. Na Figura é apresentada uma visualização em forma de grafo direcionado que representa os movimentos dos alunos entre diferentes componentes do curso durante três momentos distintos: ao realizar a lição de casa (a), no exame intermediário (b) e no exame final (c). Cada nó colorido corresponde a um componente de aprendizagem, como fórum de discussão, laboratório, vídeo de aula, perguntas de aula, livro didático, tutoriais e wiki. A cor de cada nó indica o tipo do recurso, enquanto o tamanho do nó codifica o tempo total que os alunos dedicaram àquele componente. As setas representam os movimentos dos alunos a partir de

um ponto de partida (lição de casa, exame intermediário ou exame final) em direção aos demais recursos consultados. A espessura das arestas indica a frequência de transições realizadas pelos alunos entre os componentes, ou seja, quanto mais espessa a seta, maior o número de alunos que fez aquele caminho de navegação. Esse *layout* permite observar, por exemplo, que o livro didático (*Book*, identificado na figura como o círculo roxo) foi acessado com mais frequência durante os exames do que durante as lições de casa, sugerindo seu uso como material de apoio na preparação para avaliações. Esse tipo de visualização permite que instrutores ajustem os recursos pedagógicos com base na forma como os alunos interagem com os diferentes conteúdos ao longo do curso. Os usuários podem interagir com o sistema para explorar como os diferentes recursos são utilizados ao longo do curso, o que auxilia os instrutores na otimização dos materiais e os alunos no aprimoramento de suas estratégias de estudo.

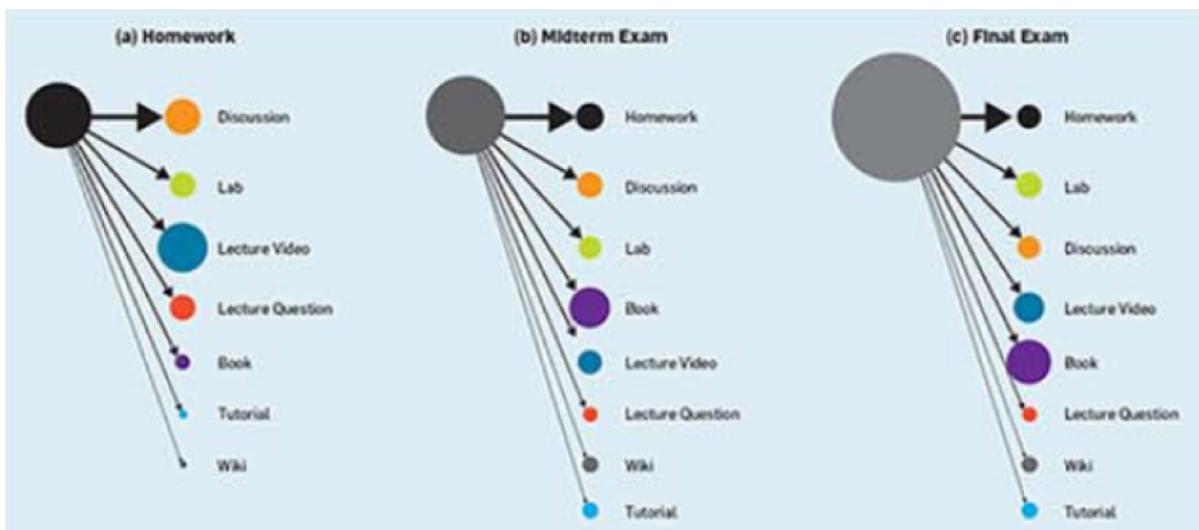


Figura 8 – Rede direcionada de movimento de aluno enquanto trabalhava na lição de casa (a), no teste intermediário (b) e no final (c) para outros componentes do curso na oferta de 2012 do MIT. A espessura das bordas codifica o número de movimentos do aluno e o tamanho dos nós codifica o tempo gasto em um componente do curso (EMMONS; LIGHT; BÖRNER, 2017).

Kui et al. (2022) exploram diversas técnicas de análise visual aplicadas à educação *online*, categorizando-as em quatro tipos principais: análise do comportamento de aprendizagem, análise de conteúdo educacional, análise de interações entre alunos e predição/recomendação. Os *layouts* utilizados incluem *sunburst*, diagramas de Sankey, *heatmaps* e visualizações baseadas em linhas do tempo. Os usuários podem explorar interativamente os dados para detectar áreas problemáticas, prever o desempenho dos alunos e otimizar os conteúdos educacionais. O *layout sunburst* (**Figura 9**) é utilizado para representar a estrutura hierárquica de sequências de navegação dos alunos em plataformas de

e-learning. Cada camada concêntrica representa uma etapa da navegação e cada segmento mostra um recurso acessado (exemplo: vídeo, fórum, exercício). A largura dos segmentos indica a frequência com que aquela sequência foi seguida por alunos. Esse tipo de visualização permite identificar os caminhos mais comuns percorridos pelos alunos durante a navegação, revelando padrões recorrentes ou desvios que possam indicar dificuldades de usabilidade ou comportamento não ideal de aprendizagem. A análise permite que os professores ajustem a organização dos conteúdos e intervenham em pontos de dispersão.



Figura 9 – *Layout sunburst* mostrando padrões de navegação sequencial dos alunos em plataformas *on-line* (POON et al., 2017).

O sistema **DropoutVis** (ZHANG et al., 2023) propõe uma abordagem interativa de análise visual para investigar padrões de evasão em ambientes de aprendizagem *on-line*, integrando diferentes fontes de informação para identificar os fatores que influenciam o abandono dos cursos. O sistema se baseia em três componentes visuais principais: (1) a *Trajectory View*, que apresenta as trajetórias temporais dos alunos ao longo do curso, permitindo observar a evolução do comportamento de participação e desempenho; (2) a *Counterfactual Explanation View*, que oferece explicações contrafactualis — isto é, cenários alternativos em que, ao modificar uma ou mais variáveis do histórico do aluno, poderia-

se evitar a evasão — e (3) a *Group Comparison View*, que facilita a comparação entre grupos de alunos (evadidos e não evadidos), destacando as diferenças nas interações com os recursos do curso. Essas visualizações são organizadas de forma coordenada, permitindo uma exploração fluida e aprofundada dos dados. Os contrafactuals, por exemplo, são gerados com base em modelos preditivos e exibidos de maneira visual, indicando quais mudanças hipotéticas no comportamento ou desempenho poderiam levar à retenção do aluno. A **Figura 10** apresenta uma interface composta por cinco visões interativas, cada uma projetada para revelar aspectos distintos da evasão escolar em ambientes de aprendizagem *on-line*, descritas a seguir:

- *Temporal Circle View*: apresenta padrões temporais agregados de evasão ao longo do tempo. Cada círculo representa um ponto no tempo (por exemplo, uma semana ou uma fase do curso), e o tamanho do círculo é proporcional ao número de evasões registradas naquele período. As cores indicam diferentes padrões de evasão, agrupados por similaridade temporal, por exemplo, círculos em azul podem representar evasões precoces, enquanto tons mais quentes (como laranja ou vermelho) indicam evasões em fases posteriores. Essa visão permite ao usuário identificar períodos críticos em que a taxa de evasão se intensifica;
- *Windmill View*: explora o impacto de diferentes variáveis (ou características dos alunos) sobre a evasão. Cada “pá” do moinho representa uma característica, como frequência, desempenho em avaliações ou nível de participação. A intensidade da cor representa a correlação da variável com a evasão (tons escuros indicam forte associação), enquanto o comprimento da pá é proporcional à influência da característica no modelo preditivo. Essa visualização permite identificar quais fatores mais contribuem para o risco de evasão;
- *Feature Temporal Comparison View*: compara a variação de uma ou mais características ao longo do tempo, relacionando esses padrões com os momentos de evasão. Linhas coloridas representam diferentes turmas, perfis ou grupos de alunos. Os *layouts* temporais apresentam métricas como tempo de login, participação em fóruns ou pontuação média semanal. Áreas sombreadas ou barras verticais podem indicar os momentos em que os alunos evadiram. Essa visão oferece suporte à análise de quais comportamentos precedem a evasão;
- *Counterfactual Explanation View*: apresenta explicações contrafactuals, ou seja, simulações do que precisaria mudar no comportamento ou características de um aluno para que a evasão fosse evitada. Cada bloco representa uma característica modificável (por exemplo, aumentar o tempo de engajamento semanal), com cores que indicam se o efeito da mudança é positivo (verde) ou negativo (vermelho). Os números associados a cada bloco mostram o impacto estimado daquela mudança no risco de evasão, permitindo análises do tipo “e se”;

- *Instance View*: resume as regras de aprendizagem geradas pelo sistema para alunos individuais. Cada instância (aluno) é representada por um retângulo contendo suas principais características comportamentais e de desempenho. As regras são extraídas de modelos explicativos e os elementos gráficos (como ícones ou cores) indicam o nível de risco, variando de verde (baixo risco) a vermelho (alto risco), com a possibilidade de comparar instâncias para identificar padrões comuns.



Figura 10 – A interface do usuário do DropoutVis inclui: *Temporal Circle View*, que fornece uma visão geral de diferentes padrões; *Windmill View*, que permite examinar o efeito de diferentes características sobre a evasão; *Feature Temporal Comparison View*, que apoia a exploração da associação entre diferentes períodos e a evasão; *Counterfactual Explanation View*, que oferece uma análise detalhada das razões para a evasão; e *Instance View*, que resume as regras de aprendizagem de indivíduos (ZHANG et al., 2023).

Garcia-Zanabria et al. (2022) desenvolveram o **SDA-Vis**, um sistema de visualização para análise de evasão estudantil para explorar e identificar os fatores que contribuem para a evasão e propor soluções para preveni-la. A ferramenta utiliza dados acadêmicos, socioeconômicos e demográficos dos alunos para criar cenários contrafactuals, proporcionando *insights* sobre as ações que poderiam reduzir o risco de evasão. A **Figura 11** ilustra a interface do sistema SDA-Vis aplicada à análise de evasão no curso de Engenharia Industrial, com foco na geração e avaliação de explicações contrafactuals. A visualização é dividida em cinco componentes interativos:

- *Student Projection View* (A): representa os alunos classificados como potenciais evadidos com base em modelos preditivos. O eixo horizontal mostra a probabilidade de evasão, enquanto o eixo vertical indica a distância euclidiana (*feasibility*) entre o aluno original e seus contrafactuals. Cada ponto verde representa um aluno e os pontos vermelhos correspondem aos contrafactuals gerados. A linha cinza demarca o limiar de decisão do modelo (por exemplo, 0,5);
- *Counterfactual Exploration View* (B): esta área é subdividida em quatro blocos principais (B1 a B4) apresentando diferentes conjuntos de contrafactuals para cada aluno. O bloco B1 exibe contrafactuals aleatórios, o B2 mostra os mais viáveis (menor número de alterações nas variáveis), o B3 apresenta os mais factíveis (mais próximos dos dados reais de alunos não evadidos) e o B4 permite ao usuário selecionar manualmente o contrafactual com base em sua expertise. Cada linha nesta visualização corresponde a um contrafactual e cada coluna a uma variável. As células coloridas indicam as variáveis que precisam ser alteradas, com destaque em vermelho para chamar atenção às mudanças mais impactantes. Setas apontam as direções de mudança de valores, como, por exemplo, aumento da nota média ou redução da taxa de pobreza;
- *Feature Distribution Bars View* (C): fornece histogramas que representam a distribuição de características demográficas ou acadêmicas da população de alunos selecionada. No exemplo da figura, visualiza-se a proporção de alunos por gênero e a distribuição etária, facilitando a comparação de perfis de evasão e contrafactuals por grupos populacionais;
- *Impact View* (D): quantifica o efeito de cada contrafactual selecionado (B2 a B4) sobre um grupo específico de alunos. Cada linha representa uma variável e cada ponto vermelho mostra o número de alunos que seriam beneficiados se determinada variável fosse modificada conforme o contrafactual sugerido. A espessura das linhas codifica a magnitude do impacto. Essa visualização permite avaliar qual alteração tem maior potencial de reduzir a evasão dentro de um grupo-alvo.

Essa composição visual integrada permite que usuários identifiquem, comparem e priorizem contrafactuals com base em critérios quantitativos (viabilidade, factibilidade, impacto) e qualitativos (escolha manual com base na experiência), possibilitando intervenções mais eficazes e direcionadas para reduzir a evasão.

Outra proposta, com escopo mais específico, é apresentada por Tsung et al. (2022), denominada **BlockLens**, uma ferramenta interativa de análise visual voltada à compreensão dos comportamentos de programação de alunos em ambientes baseados em blocos, como o *Scratch*. O sistema é projetado para oferecer suporte à análise de trajetórias de resolução de problemas e identificação de padrões comuns e pontos críticos no processo

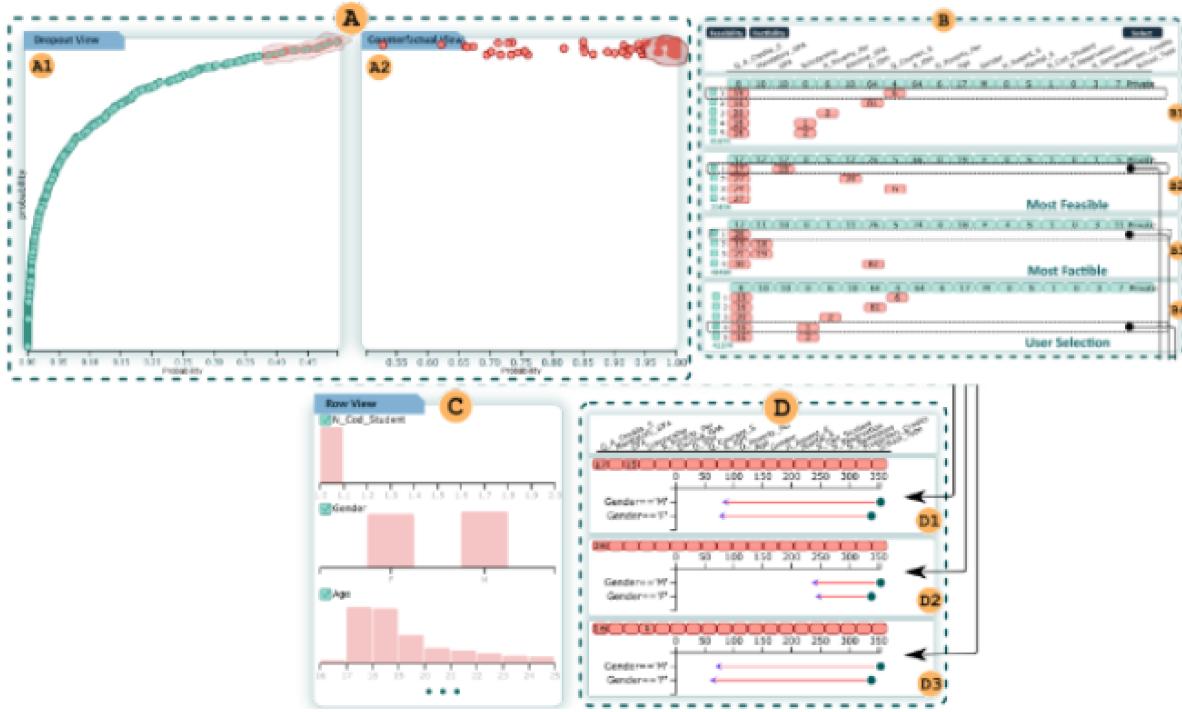


Figura 11 – Análise de explicações contrafactuals para o curso de Engenharia Industrial:
A representa a seleção de alunos e seus respectivos contrafactuals; B exibe os contrafactuals, onde B1 mostra contrafactuals aleatórios, B2 mostra os mais viáveis, B3 mostra os mais factíveis e B4 mostra o contrafactual selecionado pelo usuário. Além disso, C mostra a distribuição dos alunos, e D mostra o impacto dos contrafactuals selecionados (B2–B4) (GARCIA-ZANABRIA et al., 2022).

de programação. A interface é composta por quatro componentes principais, ilustrados na **Figura 12**:

- *Visão de Seleção de Questões* (A): exibe estatísticas-chave de desempenho, como taxa de sucesso, número médio de passos e tempo médio gasto, para cada questão. Essas métricas são visualizadas por meio de representações visuais compactas, como barras de progresso e ícones de relógio, permitindo aos instrutores escolher rapidamente quais questões analisar mais a fundo;
- *Visão do Aluno* (B): organiza os alunos com base na quantidade de *checkpoints* (momentos associados ao sucesso) e *warning signs* (momentos associados ao fracasso) identificados nas sequências de código. O gráfico principal é um plano cartesiano onde o eixo x representa a quantidade de checkpoints e o eixo y a quantidade de *warning signs*. Cada ponto representa um aluno: os pontos verdes indicam alunos que acertaram a questão, enquanto os pontos laranja representam aqueles que erraram. O tamanho do ponto corresponde ao número de passos realizados por cada

aluno. Abaixo, dois histogramas mostram a distribuição do tempo total e da contagem de passos para alunos com sucesso e sem sucesso, facilitando a detecção de padrões como desistência precoce ou excesso de tentativas;

- *Visão de Resumo de Caminhos* (C): apresenta uma visualização inspirada em diagramas de Sankey, onde cada nó representa um *snapshot* de código e as conexões entre nós representam transições entre estados consecutivos no processo de codificação. *Snapshots* frequentes e significativos são destacados com suas estruturas completas, enquanto os menos relevantes aparecem de forma compacta. A coloração dos retângulos indica sua função: verde para *checkpoints*, laranja para *warning signs* e cinza para demais estados. A largura das conexões reflete a frequência de transição entre *snapshots*, e barras de distribuição de passos mostram em quais estágios do processo esses *snapshots* ocorrem mais frequentemente;
- *Visão de Sequência* (D): detalha a trajetória completa de um aluno selecionado, exibindo a sequência de *snapshots* e eventos associados, como inserção ou remoção de blocos. Ao contrário da Visão de Caminhos, todos os *snapshots* são exibidos de forma expandida, permitindo uma análise granular das ações individuais dos alunos. Essa combinação de visualizações permite aos educadores identificar estratégias comuns de resolução de problemas, reconhecer padrões de erro recorrentes e tomar decisões pedagógicas informadas com base em dados comportamentais concretos. Embora seja eficaz na análise de estratégias de resolução de problemas, o sistema analisa o comportamento dos alunos em tarefas isoladas, não considerando o histórico de desempenho ao longo do tempo. Portanto, não permite inferir evolução individual ou mudanças no estilo de resolução de problemas, o que seria crucial para intervenções pedagógicas mais eficazes.

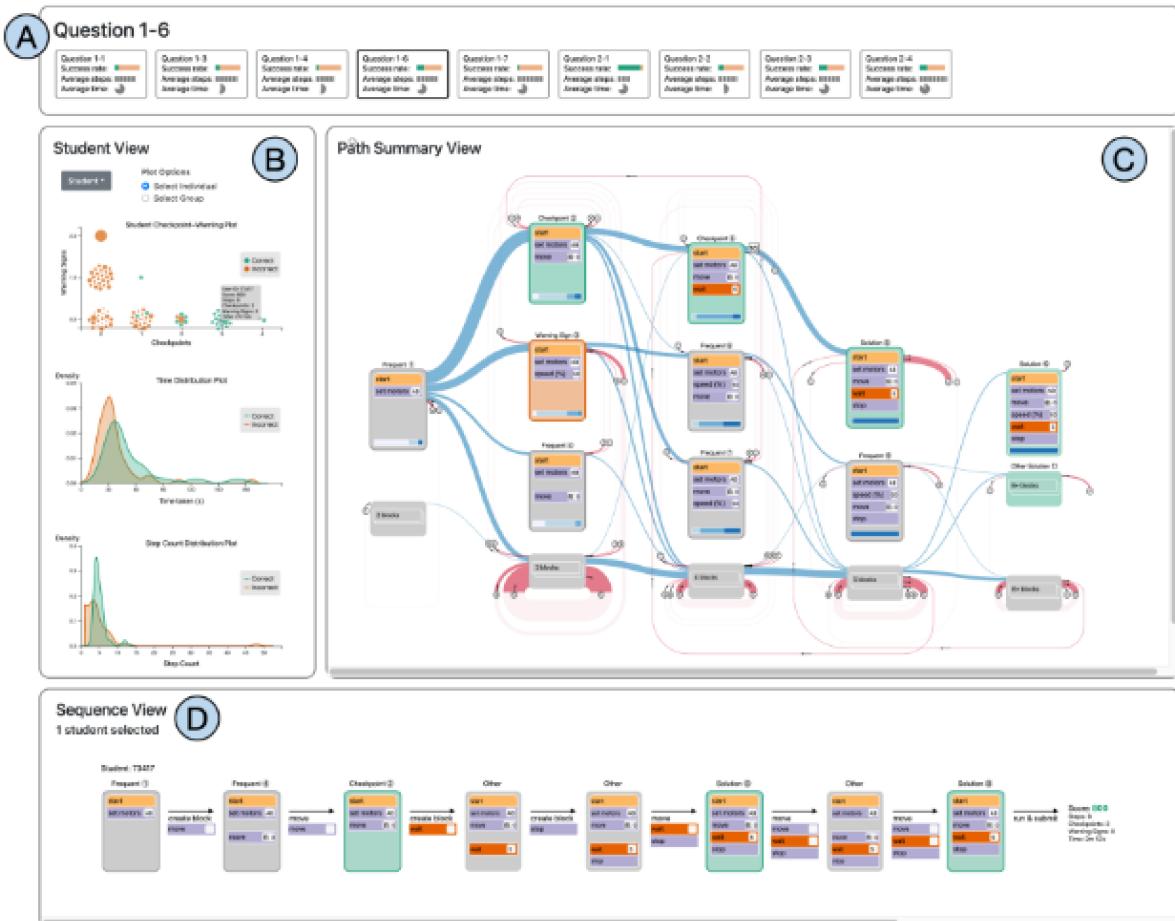


Figura 12 – O painel de análise visual: (A) Visão de Seleção de Questões: Permite que os usuários selezionem questões. (B) Visão do aluno: Facilita a exploração do desempenho dos alunos em termos de diferentes métricas. (C) Visão de Resumo de Caminhos: Resume os caminhos e a informação de transição de *snapshots* de múltiplos alunos. (D) Visão de Sequência: Mostra as sequências completas de *snapshots* e eventos dos alunos (TSUNG et al., 2022).

3.1 Considerações Finais

Conforme apresentado neste capítulo, embora existam diversas estratégias visuais para análise de dados acadêmicos que oferecem *insights* valiosos sobre o comportamento dos alunos e seu impacto no desempenho, essas abordagens apresentam algumas limitações. De modo geral, observa-se a ausência de análises de correlação que permitam identificar as causas reais do sucesso ou fracasso dos alunos nas disciplinas em que estão matriculados. O uso de dados externos, como indicadores socioeconômicos e demográficos, tem motivado muitos trabalhos científicos, pois permite identificar padrões de vulnerabilidade estudantil, compreender fatores de evasão relacionados a contexto social e familiar, além de subsidiar políticas institucionais voltadas ao apoio acadêmico. Tais informações oferecem um panorama mais amplo das condições que influenciam a trajetória dos alunos, contribuindo para análises de risco e para a definição de estratégias de inclusão. Entretanto, quando utilizados isoladamente, esses dados podem resultar em *layouts* cujos padrões ressaltados não estão necessariamente relacionados à vida acadêmica dos alunos, ou ao comportamento desses alunos nas disciplinas do curso, o que pode comprometer o potencial das informações para fins de tomada de decisão. Além disso, nenhuma dessas abordagens investiga o impacto da sequência de disciplinas no desempenho acadêmico, tampouco quais conteúdos podem ser cruciais para o sucesso dos alunos em uma disciplina específica. A proposta deste trabalho é preencher as lacunas citadas provendo formas de análise do desempenho acadêmico dos alunos ao longo de diferentes sequências de disciplinas cursadas ao longo do curso, ressaltando quais dificuldades em cada disciplina foram observadas, e como elas se relacionam com esse desempenho, permitindo a compreensão dessas relações e auxiliando a criação de estratégias que auxiliem no processo educacional.

CAPÍTULO 4

Descrição da Proposta

Este capítulo apresenta a proposta desenvolvida no contexto deste trabalho, detalhando sua concepção e implementação. Inicialmente, são descritos os requisitos funcionais que orientaram o desenvolvimento do sistema, com base em revisão da literatura e validações com especialistas. Em seguida, são discutidos os principais aspectos da arquitetura do sistema e da estrutura dos dados utilizada, estabelecendo um padrão genérico que permite sua adaptação a diferentes instituições. Por fim, é apresentada uma descrição detalhada do EduFlow, incluindo suas visões analíticas, os *layouts* interativos empregados e os recursos de implementação utilizados para oferecer suporte à análise do desempenho acadêmico de alunos universitários.

4.1 Arquitetura do Sistema

O sistema EduFlow foi desenvolvido com o objetivo de apoiar a análise do desempenho acadêmico de alunos universitários por meio de representações visuais interativas. Para que o sistema atendesse de forma eficaz às necessidades dos usuários, foi conduzido um processo sistemático de levantamento de requisitos, apresentado neste capítulo.

De forma a validar a importância das análises visuais dos dados, foram requisitados pareceres dos profissionais da educação com experiência profissional no acompanhamento das atividades acadêmicas, para identificar as necessidades existentes no processo de análise de desempenho dos alunos. Além disso, para desenvolver a estratégia de análise, foi levado em conta a investigação realizada pela revisão da literatura e as lições aprendidas a partir desta pesquisa. Estudos anteriores, como os de Deng et al. (2019) e Etemadpour et al. (2019) mostram que, embora existam ferramentas para analisar o desempenho acadêmico de alunos, elas apresentam limitações significativas, discutidas no Cap. 3. Após o estudo realizado, os seguintes requisitos foram identificados:

- ❑ **R1 - Análise do impacto da sequência de cumprimento de disciplinas no desempenho acadêmico:** o sistema deve permitir aos usuários examinar de que

forma a conclusão de determinadas disciplinas influencia o desempenho em disciplinas subsequentes e o sucesso acadêmico geral. Conforme evidenciado nos estudos de Raji et al. (2017) e Etemadpour et al. (2019), uma das principais limitações das ferramentas de visualização existentes é a ausência de correlação entre o desempenho em diferentes disciplinas ao longo do tempo. Assim, o sistema deve possibilitar a compreensão do impacto causado por determinada sequência de disciplinas no desempenho dos alunos;

- **R2 - Análise do desempenho dos alunos em disciplinas individuais:** compreender o desempenho dos alunos em determinadas disciplinas é fundamental para a identificação e compreensão de desafios acadêmicos, especialmente ao se analisar resultados específicos de grupos de alunos que seguiram trajetórias semelhantes. A revisão da literatura evidenciou a necessidade de ferramentas que forneçam uma visão mais integrada do desempenho em disciplinas individuais, indo além de indicadores simples de aprovação ou reprovação (DENG et al., 2019);
- **R3 - Análise da relação entre estágios e desempenho acadêmico:** identificar como experiências acadêmicas externas, como os estágios, impactam o desempenho dos alunos é um aspecto relevante da análise educacional. Estudos como os de Baradwaj e Pal (2011) e Santana et al. (1996) destacam a importância de compreender os fatores que contribuem para o sucesso ou a dificuldade dos alunos na conclusão de sequências de disciplinas. Nesse contexto, a análise da relação entre a participação em atividades de estágio e os resultados acadêmicos pode oferecer *insights* valiosos para orientar políticas institucionais, visando à criação de estratégias mais eficazes de inserção profissional e à construção de uma estrutura curricular em que as atividades de estágio estejam melhor alinhadas aos objetivos educacionais do curso. Na revisão da literatura não foram encontrados sistemas que fazem essa análise;
- **R4 - Análise dos tópicos abordados nas disciplinas:** a análise do engajamento dos alunos com tópicos específicos dentro de uma disciplina auxilia os educadores na identificação de conteúdos que representam maior dificuldade. Estudos anteriores, como os de Freitas et al. (2001) e Heer, Bostock e Ogievetsky (2010), ressaltam a necessidade de ferramentas visuais que vão além das métricas gerais de desempenho, permitindo explorar como tópicos individuais influenciam nas taxas de aprovação e reprovação. A análise do comportamento dos alunos em diferentes tópicos contribui para uma compreensão mais profunda da estrutura interna de cada disciplina, possibilitando avaliar se a sequência e a distribuição atual dos conteúdos favorecem a aprendizagem efetiva, e se ajustes podem ser necessários para melhorar a compreensão e os resultados. Em outras palavras, a análise dos tópicos das disciplinas não apenas evidencia os pontos de maior dificuldade para os alunos, como também

orienta possíveis aprimoramentos no desenho curricular, de forma a alinhar melhor o conteúdo às necessidades de aprendizagem dos alunos.

A **Figura 13** mostra um diagrama da estratégia de análise visual empregada no EduFlow. Cada um dos seus componentes é detalhado nas próximas seções.

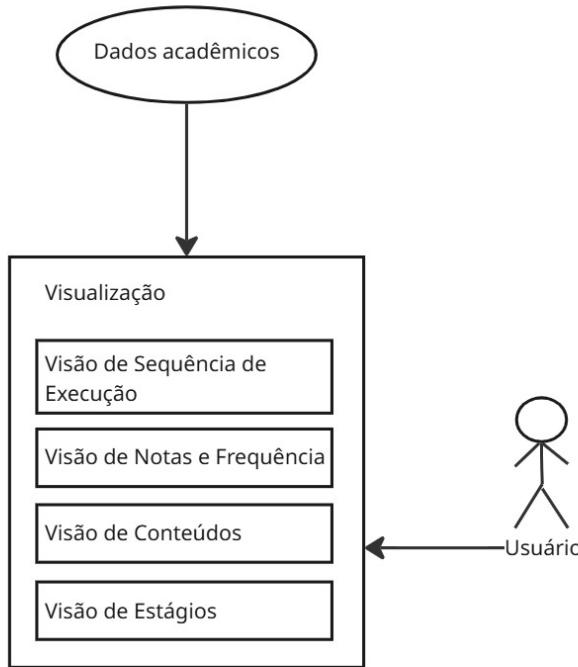


Figura 13 – Arquitetura metodológica do sistema EduFlow. Estrutura da metodologia de análise visual, mostrando todas as etapas organizadas em componentes estruturados. Os usuários interagem com os *layouts* no componente Visualização, utilizando um conjunto de ferramentas de interação que proporciona a exploração de diferentes aspectos do EduFlow com foco na análise do desempenho acadêmico.

4.2 Aspectos gerais do Design Visual

O EduFlow é um sistema de análise visual interativa desenvolvido para apoiar a compreensão do desempenho acadêmico de alunos universitários. Seu funcionamento baseia-se na utilização de informações relevantes sobre o percurso dos alunos, como: histórico de aprovações e reprovações por disciplina, sequência e período em que as disciplinas foram cursadas, registros de frequência e notas, desempenho por conteúdos ministrados e envolvimento em atividades complementares, como estágios. Os alunos são classificados

em três grupos principais: aprovados, reprovados e aprovados após múltiplas reprovações. Considera-se um aluno desistente de uma disciplina aquele para o qual não há registros com situação final igual a “aprovado” naquela disciplina, independentemente do número de tentativas. Essa definição visa identificar padrões de abandono ou interrupção definitiva no cumprimento da disciplina. Para representar essas categorias de forma visualmente intuitiva, o sistema adota uma codificação cromática padronizada: a cor verde representa alunos aprovados diretamente, a cor vermelha indica alunos reprovados, e tons intermediários, interpolados entre verde e vermelho, denotam o número de tentativas até a aprovação. Além disso, o sistema permite a análise do comportamento do aluno ao longo da disciplina, conceito que, neste contexto, refere-se à distribuição de presenças e ausências em cada tópico ministrado durante o semestre em que esteve matriculado. Essa definição do comportamento permite caracterizar o grau de engajamento do aluno com os conteúdos da disciplina.

A interface do EduFlow é composta por quatro visões principais, cada uma composta por *layouts* interativos que possibilitam a exploração de aspectos específicos do desempenho dos alunos. As seções subsequentes descrevem essas visões e discutem como cada uma contribui para o atendimento aos requisitos definidos na etapa de análise.

4.3 Estrutura dos Dados

Após a definição das tarefas analíticas e das visões correspondentes, um conjunto de informações necessárias para cada uma dessas visões foi definido. Essas informações são utilizadas pelos módulos que constroem os *layouts* do sistema, e incluem informações de matrícula dos alunos, disciplinas cursadas, seu comportamento (notas e faltas), além de informações sobre a estrutura das disciplinas em um curso. Essa padronização garante que dados sobre quaisquer curso de quaisquer instituições sejam utilizados no sistema. A **Tabela 1** apresenta a lista das informações, com uma breve descrição sobre o conteúdo esperado em cada um.

4.4 Visões Analíticas

A análise de sequências de disciplinas concentra-se nas sequências de disciplinas cumpridas pelos alunos, permitindo compreender como o desempenho em uma disciplina pode influenciar o resultado em disciplinas subsequentes. Inspirado pelos métodos de análise de sequências de eventos, o EduFlow permite que o usuário analise o desempenho dos alunos que cumpriram uma sequência específica de disciplinas selecionadas. A **Figura 14** mostra as Visões Analíticas do EduFlow, as quais serão detalhadas nas sessões a seguir.

Tabela 1 – Estrutura dos Dados.

Visão / Atributo	Descrição
Dados Comuns a Todas as Visões	
Disciplina	Identificação da disciplina cursada (código e nome)
Matrícula	Identificador único do aluno
Situação final	Situação final do aluno na disciplina (aprovado, reprovado etc.)
Ano, Período	Ano e período letivo em que a disciplina foi cursada
Visão de Notas e Frequência	
Média final	Nota final obtida na disciplina
Faltas	Número total de faltas na disciplina
Visão de Estágio	
Data das aulas	Datas das aulas da disciplina cursada
Início do estágio	Data de início do estágio realizado
Fim do estágio	Data de conclusão do estágio realizado
Visão de Conteúdos	
Aula	Identificador único da aula
Conteúdo	Descrição do conteúdo abordado na aula
Faltas (por aula)	Quantidade de faltas do aluno em uma aula específica

Fonte: Elaborado pela autora.

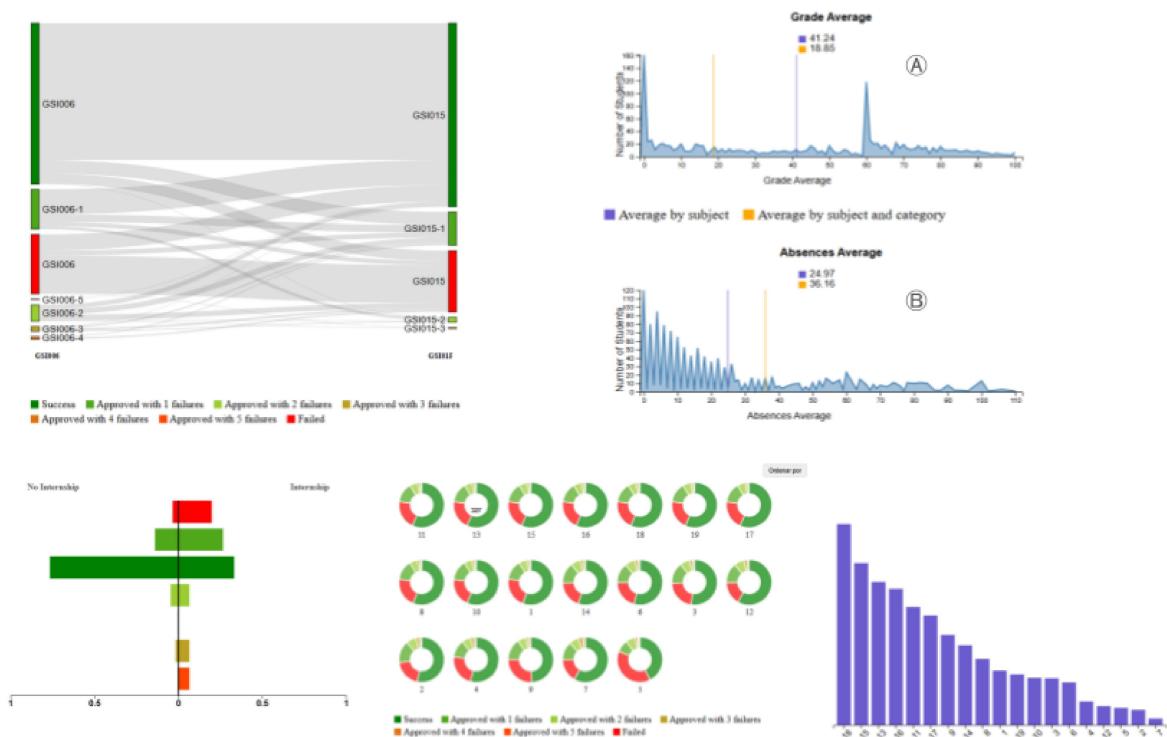


Figura 14 – Visões Analíticas do EduFlow

4.4.1 Visão de Sequência de Execução

Essa análise é representada pela **Visão de Sequência de Execução**, mostrada na **Figura 15**. Nessa visão, o progresso dos alunos, considerando uma sequência específica de disciplinas, é exibido em termos de proporção de aprovações e reprovações em uma ou diversas tentativas. O *layout* dessa análise utiliza um diagrama Sankey, no qual a altura de cada nó representa a quantidade de alunos daquela categoria de desempenho da disciplina e a espessura de cada aresta representa a quantidade de alunos que transitam entre categorias de desempenho em diferentes disciplinas. O *layout* utiliza nomes diferenciados para cada quantidade de reprovação, sendo o código da disciplina seguido da quantidade de reprovações acumuladas. Dessa forma, considerando a disciplina GSI006 como exemplo, os nós serão rotulados como GSI006, GSI006-1, GSI006-2, etc. para primeira tentativa, segunda tentativa (1 reprovação) e terceira tentativa (2 reprovações), etc., respectivamente. O usuário pode, assim, analisar como alunos reprovados e/ou aprovados em uma disciplina específica se comportam, em termos de desempenho, em outras disciplinas. Além disso, permite identificar como cada sequência de cumprimento de disciplinas escolhida pelos alunos impactou no seu desempenho nessas disciplinas, auxiliando na identificação das sequências mais interessantes dentre as existentes.

Ao posicionar o cursor sobre um nó (**Figura 16A**), o usuário pode visualizar detalhes sobre a quantidade de alunos naquela categoria de desempenho. Ao posicionar o cursor sobre uma aresta (**Figura 16B**), são exibidas informações sobre o percentual de alunos que transitaram entre as categorias conectadas. No exemplo mostrado, 85.06% dos alunos que foram aprovados em GSI006, também foram aprovados em GSI015. Além disso, ao clicar em um nó, é possível verificar uma análise mais detalhada da disciplina selecionada, através das demais Visões do EduFlow. Essas interações tornam a análise mais personalizável, oferecendo *insights* claros para diferentes análises.

Essa visualização dá suporte ao requisito R1 ao oferecer uma visão geral das taxas de sucesso e reprovação em diferentes sequências de disciplinas, permitindo ao usuário analisar como a conclusão de uma disciplina influencia o desempenho em disciplinas subsequentes.

4.4.2 Análise de uma disciplina específica

Ao selecionar um nó específico de uma disciplina na Visão de Sequência de Execução, é iniciada uma análise detalhada dessa disciplina. Esta seção explora duas visões complementares: **Visão de Notas e Frequência** e a **Visão de Estágio**.

A Visão de Notas e Frequência (**Figura 17**) apresenta uma análise do desempenho e do engajamento dos alunos em uma disciplina específica, utilizando dois gráficos distintos: um para notas e outro para faltas. Essas visualizações permitem compreender como o comportamento dos alunos, em termos de notas e frequência, estão associados ao sucesso

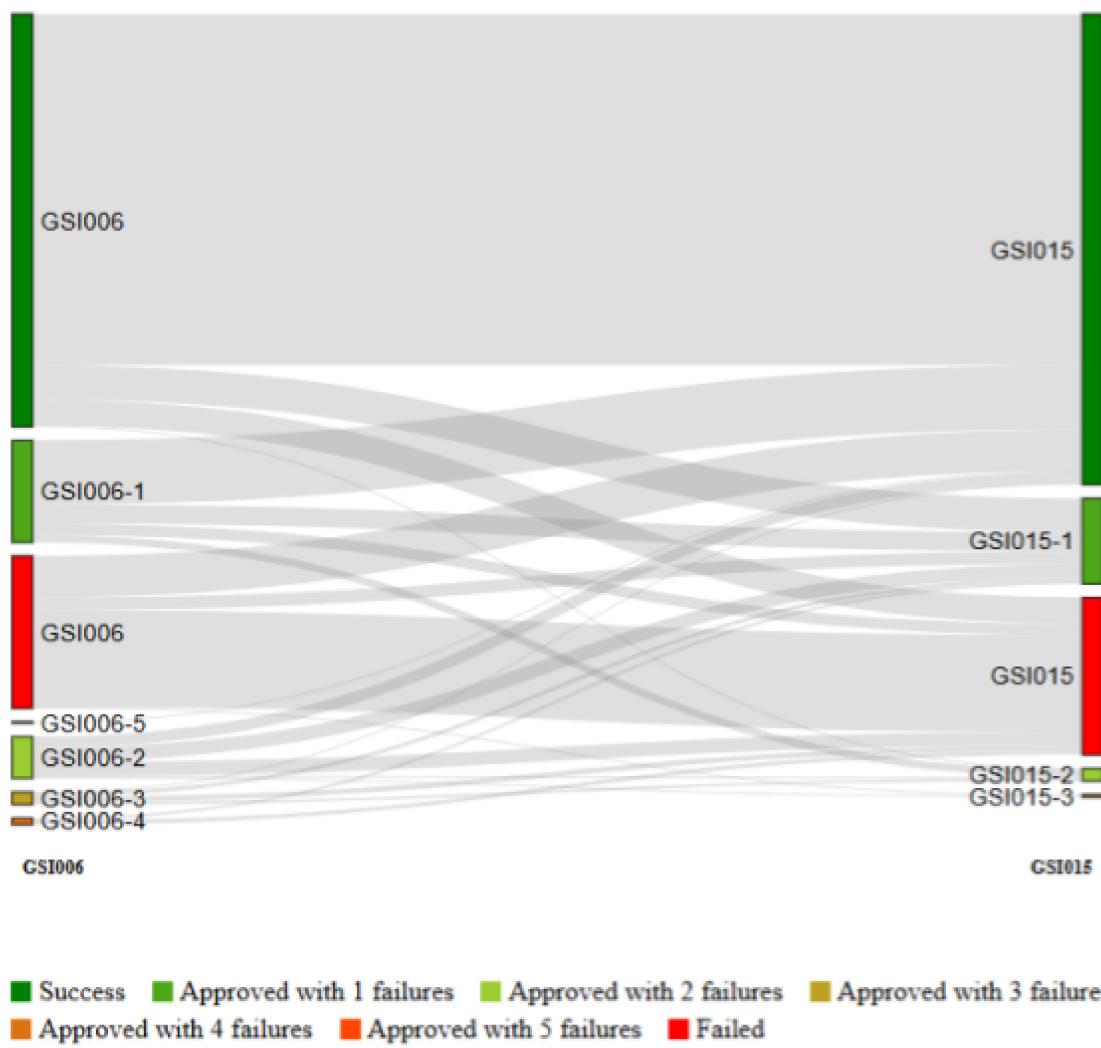


Figura 15 – Visão de Sequência de Execução, com alunos agrupados de acordo com o resultado de sua(s) tentativa(s) na disciplina.

ou insucesso na disciplina. O eixo x de cada gráfico representa a distribuição de notas ou de faltas dos alunos na disciplina, enquanto o eixo y mostra o número de alunos em cada faixa de notas ou frequência. A área azul no gráfico indica a quantidade de alunos em diferentes intervalos de notas ou faltas. A linha roxa representa a média geral da disciplina para a métrica analisada (notas ou faltas), enquanto a linha amarela indica a média da categoria de desempenho (nó) da disciplina selecionada. Essas linhas foram incluídas para facilitar a identificação de tendências gerais e comparações específicas por categorias de alunos da disciplina. A ideia dessas visualizações é auxiliar na compreensão de como a variação nas notas e na frequência pode impactar o sucesso acadêmico.

A Visão de Notas e Frequência está alinhada ao requisito R2, pois possibilita a identificação de comportamentos acadêmicos característicos, como notas baixas e altos índices de faltas, em grupos específicos de alunos que seguiram uma determinada sequência de

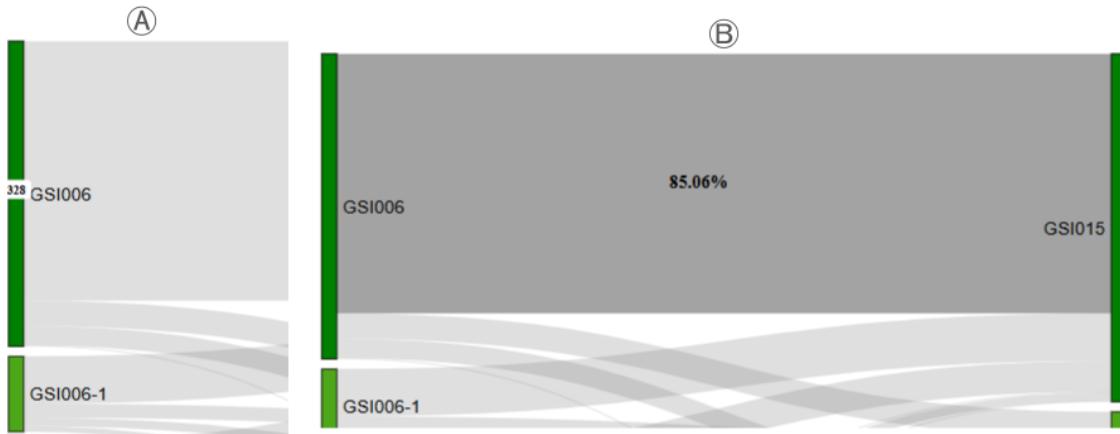


Figura 16 – Interações na Visão de Sequência de execução. A: Exibição, no quadrado branco do primeiro nó, do número de alunos na categoria de desempenho dos alunos aprovados em GSI006. B: Exibição, na aresta, do percentual de alunos aprovados em GSI006 e que também foram aprovados em GSI015.

disciplinas e obtiveram um mesmo resultado na disciplina selecionada. Ao permitir a comparação visual entre o desempenho médio desses grupos e a média geral da turma, o sistema favorece uma compreensão mais integrada e contextualizada das dificuldades enfrentadas pelos alunos. Essa abordagem vai além de classificações simples de aprovação ou reprovação, oferecendo *insights* sobre como indicadores acadêmicos, como frequência e notas, podem ajudar a explicar padrões de desempenho em nível de disciplina.

A Visão de Estágio (**Figura 18**) utiliza um gráfico de pirâmide para prover uma análise comparativa da proporção de alunos que realizaram ou não o estágio enquanto cursavam a disciplina selecionada. Esse *layout* foi projetado para explorar a influência da experiência prática proporcionada pelo estágio no desempenho acadêmico pela investigação de possíveis correlações entre a prática profissional e o sucesso ou dificuldades enfrentadas na disciplina. O eixo x no *layout* diferencia os dois grupos: “Sem Estágio” (à esquerda) e “Com Estágio” (à direita), enquanto o eixo y apresenta as categorias de desempenho dos alunos. As barras são espelhadas horizontalmente em direção ao centro, representando a proporção de alunos em cada categoria de desempenho, considerando ambas as situações. Neste exemplo, nota-se que a grande maioria dos reprovados na disciplina (barra vermelha) faziam estágio simultaneamente. A Visão de Estágio está alinhada ao requisito R3, pois permite a análise de como a participação em estágios se relaciona

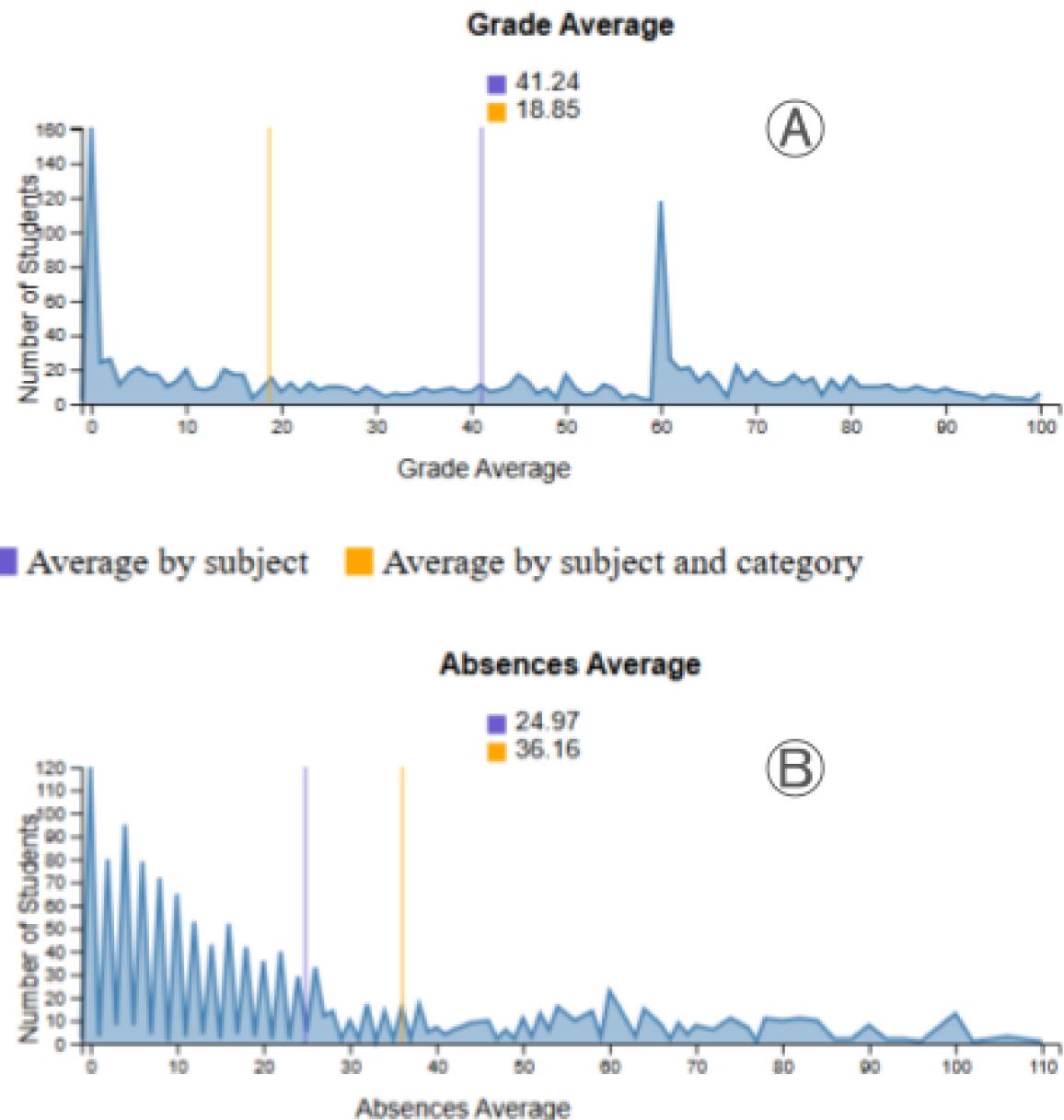


Figura 17 – Visão de Notas e Frequência, exibindo curvas de distribuição das notas (A) e faltas (B) em uma disciplina, juntamente com as linhas que representam médias gerais (roxa) e por status (amarela).

com o desempenho dos alunos na disciplina selecionada. Ao comparar a distribuição de desempenho entre alunos que realizaram estágio e aqueles que não realizaram, o sistema contribui para discussões sobre como a prática profissional se alinha ao sucesso acadêmico. Esses *insights* podem orientar a criação de estratégias de melhorias nas políticas de estágio e no desenho curricular, garantindo que essas experiências apoiem de forma eficaz os objetivos educacionais do curso.

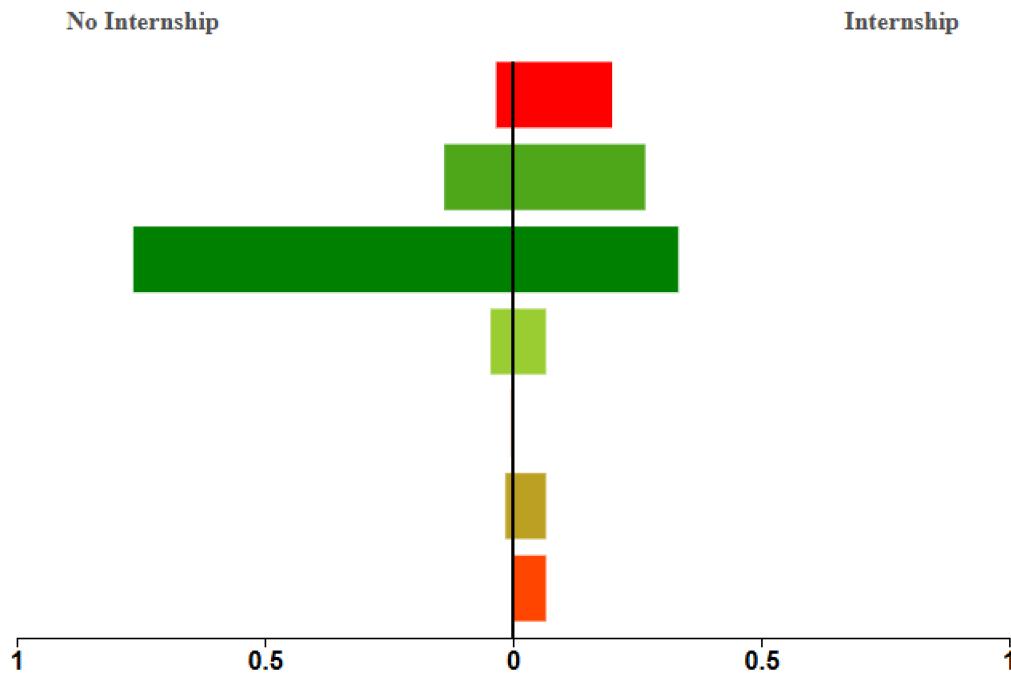


Figura 18 – Visão de Estágio. Distribuição proporcional de alunos que realizaram estágio (lado direito) e que não realizaram estágio (lado esquerdo). Cada barra horizontal corresponde a uma categoria de desempenho em uma disciplina, diferenciada por cores: verde escuro representa alunos aprovados na primeira tentativa, tons entre verde e vermelho representam aprovados em tentativas subsequentes, e vermelho representa alunos reprovados. O comprimento de cada barra é proporcional à quantidade relativa de alunos em cada grupo. A linha vertical central funciona como ponto de referência para comparação entre os dois grupos (com e sem estágio).

4.4.3 Visão de Conteúdos

A **Visão de Conteúdos** fornece uma análise detalhada em relação à disciplina selecionada na Visão de Sequência de Execução. Essa visão é composta por dois *layouts* principais: o primeiro destaca as correlações entre os conteúdos e o desempenho geral dos alunos, enquanto o segundo explora a frequência de participação dos alunos em cada conteúdo.

No *layout* apresentado na **Figura 19**, cada gráfico de setores representa um conteúdo da disciplina, com o tamanho das fatias coloridas indicando a distribuição proporcional dos alunos que compareceram à aula relacionada a determinado conteúdo, para cada categoria de desempenho. Quanto maior a fatia de um setor, maior a proporção de alunos



Figura 19 – Visão de Conteúdos. Cada gráfico de setor corresponde a um conteúdo, subdividido de acordo com a categoria de desempenho e tentativas dos alunos que compareceram à aula correspondente a esse conteúdo. Os conteúdos estão ordenados pela quantidade de aprovações. O conteúdo 13 está destacado em azul, mostrando a quantidade de alunos aprovados na primeira tentativa.

em uma determinada categoria que compareceu à aula associada ao conteúdo analisado. Ao posicionar o cursor sobre uma fatia de um setor, o usuário pode visualizar detalhes sobre a quantidade de alunos naquela categoria de desempenho, conforme destacado no quadrado azul. Esse esquema visual permite ao usuário compreender como a presença ou

ausência nas aulas correspondentes a um conteúdo se relacionam com seu desempenho na disciplina como um todo. Além disso, o *layout* permite ordenar, usando o botão destacado em rosa, os gráficos de setores de forma decrescente em relação à quantidade de alunos que compareceram às aulas de cada conteúdo ou em relação à quantidade em determinada categoria de desempenho. Ao identificar gráficos com maior proporção de reprovações (setores vermelhos), o usuário pode investigar, em fontes externas, quais fatores contribuíram para que esses conteúdos tenham predominância de alunos com baixo desempenho na disciplina. O segundo *layout* (**Figura 20**) utiliza um gráfico de barras para ilustrar o número de faltas por conteúdo, ordenados em ordem decrescente. Cada barra representa um conteúdo específico, cuja altura indica índices de ausências.

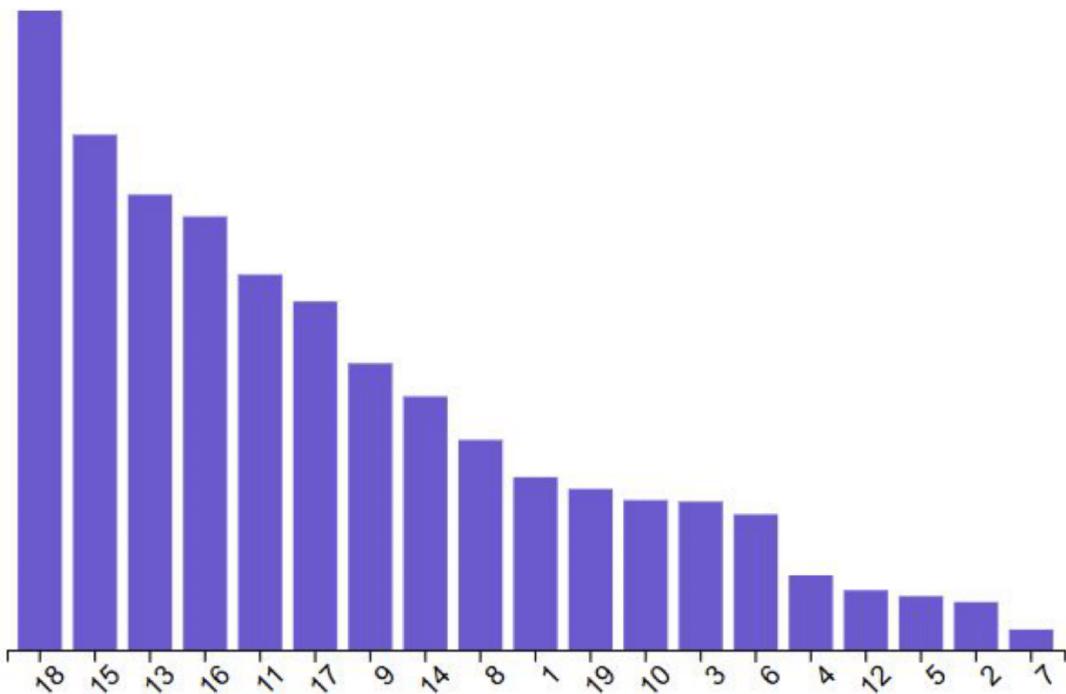


Figura 20 – Visão de Conteúdos, exibindo o número total de faltas em cada conteúdo de uma determinada disciplina.

Essas visualizações atendem especificamente ao requisito R4, ao permitir que os educadores explorem como tópicos individuais impactam o sucesso acadêmico de forma geral. A combinação desses dois *layouts* da Visão de Conteúdos oferece uma visão integrada do desempenho acadêmico e do engajamento dos alunos, permitindo ao usuário explorar a relação entre presença e dificuldade percebida em cada conteúdo.

4.5 Detalhes de Implementação

O sistema EduFlow foi implementado como uma aplicação *web* baseada na arquitetura cliente-servidor, estruturada de forma modular para facilitar a manutenção e a expansão de funcionalidades. O armazenamento dos dados é realizado em um banco de dados *PostgreSQL*¹, executado em um ambiente isolado via *Docker*², o que facilita a replicação do ambiente e garante portabilidade. A *API* responsável por fornecer os dados ao *frontend* foi implementada em *JavaScript*, permitindo comunicação assíncrona entre cliente e servidor por meio de requisições *HTTP*. A camada de visualização foi inteiramente desenvolvida com a biblioteca *D3.js*³, permitindo a construção de *layouts* interativos e personalizados, como diagramas de Sankey, gráficos de linha, barras e setores. Esses componentes visuais são atualizados dinamicamente com base nos filtros selecionados pelo usuário, como disciplinas ou alunos, oferecendo uma experiência exploratória e responsiva. Para isso, foram utilizados mecanismos avançados de interação, como filtros múltiplos, ordenações dinâmicas e legendas contextuais, todos gerados via *JavaScript*. Por fim, recursos adicionais de *CSS* foram incorporados para estilização da interface e aprimoramento da experiência do usuário, garantindo clareza na apresentação dos dados e coerência visual entre os diferentes tipos de *layout*.

¹ <https://www.postgresql.org/>

² <https://www.docker.com/>

³ <https://d3js.org/>

CAPÍTULO 5

Análise dos Resultados

Este capítulo explora a utilização da plataforma EduFlow, aplicando os *layouts* em um conjunto de cenários educacionais escolhidos com base em percepções prévias da coordenação do curso e de docentes, que apontaram históricos de sucesso ou dificuldades significativas em determinadas sequências de disciplinas. O objetivo é investigar como os *layouts* auxiliam na identificação de situações relacionadas ao sucesso ou fracasso dos alunos, bem como na compreensão de possíveis razões que explicam esses resultados de desempenho.

A análise foi conduzida em duas etapas complementares. Primeiramente, os *layouts* foram utilizados para fornecer uma visão geral do desempenho dos alunos na realização de sequências específicas de disciplinas. Foram consideradas sequências originalmente definidas na grade curricular do curso associadas a disciplinas com alto represamento. Foram consideradas também sequências popularmente consideradas pelos alunos que são diferentes das sequências originais, de forma a analisar o impacto desses caminhos alternativos. Em seguida, foram examinadas categorias específicas de desempenho, buscando detalhes que expliquem os comportamentos observados, de modo a confirmar ou refutar hipóteses levantadas pela coordenação.

Ao longo da análise, são indicados explicitamente quais requisitos funcionais do sistema foram atendidos em cada caso, demonstrando como os *layouts* auxiliam na identificação de padrões de desempenho, frequência, evasão, ordem de realização de disciplinas e correlações com atividades extracurriculares como estágios.

As seções seguintes detalham o conjunto de dados utilizado, bem como o contexto no qual esses dados estão inseridos. São discutidos também os resultados obtidos pelas análises.

5.1 Contexto Curricular do Curso

O curso de Bacharelado em Sistemas de Informação da Universidade Federal de Uberlândia (UFU) segue uma matriz curricular definida em seu Projeto Pedagógico de Curso¹, e é estruturada em oito semestres, composta por 42 componentes curriculares sequenciais que abordam tanto fundamentos teóricos quanto aplicações práticas da computação, além do Estágio Supervisionado, como ilustrado na **Figura 21**. A organização das disciplinas segue uma sequência pedagógica recomendada, na qual certos componentes devem ser cursados em uma ordem específica para garantir a progressão lógica do aprendizado. A execução sequencial ocorre quando dois componentes são organizados no fluxo curricular de forma que um seja pré-requisito do outro ou, mesmo sem exigência formal, sua realização em ordem específica seja recomendada devido à progressão lógica dos conteúdos. Esse alinhamento curricular é projetado para promover a construção de uma base de conhecimento interconectada.

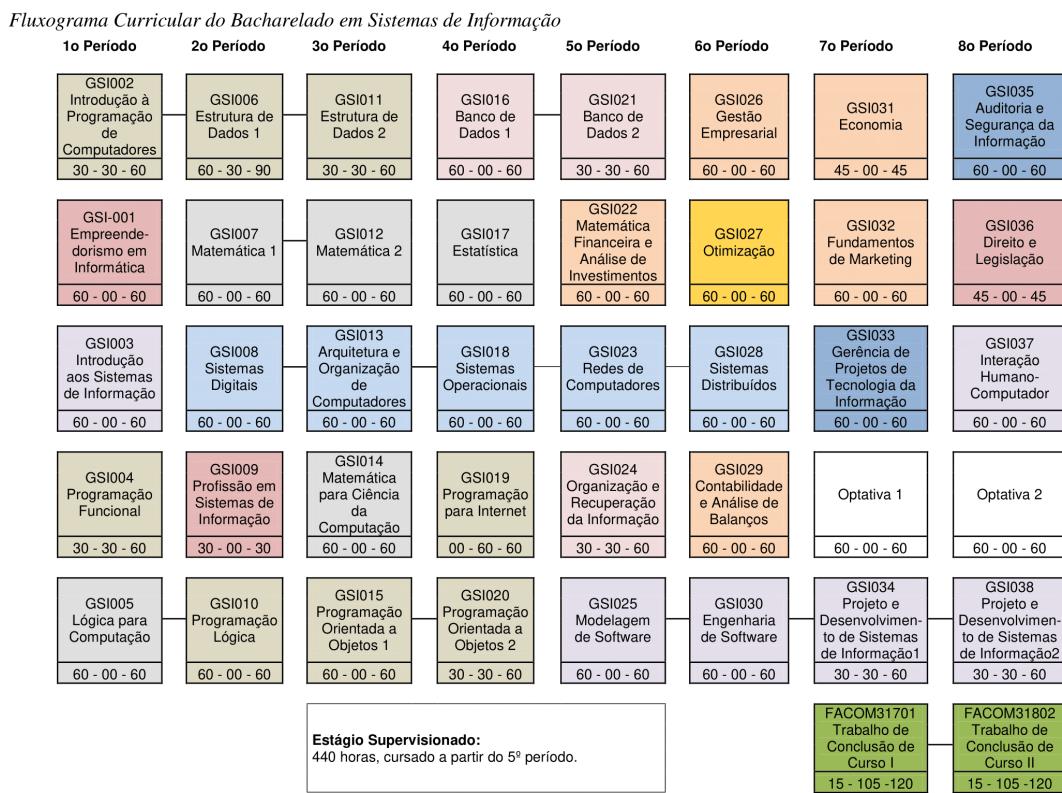


Figura 21 – Fluxograma Curricular do Bacharelado em Sistemas de Informação.

Fonte: https://facom.ufu.br/system/files/conteudo/fluxograma_curricular_do_bacharelado_em_si_0.pdf. Acesso em: 21 ago. 2025.

¹ <https://facom.ufu.br/graduacao/bsi-santamonica/projeto-pedagogico>. Acesso em: 21 ago. 2025.

5.2 Conjunto de Dados

O conjunto de dados utilizado nos experimentos de avaliação do sistema consiste em informações detalhadas de desempenho de 3.199 alunos matriculados no curso de Sistemas de Informação da UFU, entre 2009 e 2019. A base de dados dos alunos, disponibilizada devidamente anonimizada pela UFU, integra diferentes fontes de informação, incluindo:

- ❑ Histórico acadêmico dos alunos (notas finais, faltas, situação da disciplina, período de realização etc.);
- ❑ Frequência em sala de aula (presença/faltas por aula);
- ❑ Participação em estágios supervisionados;
- ❑ Conteúdos ministrados em cada aula por disciplina.

A **Figura 22** exibe a modelagem do conjunto de dados do curso de Sistemas de Informação da UFU antes do processamento.

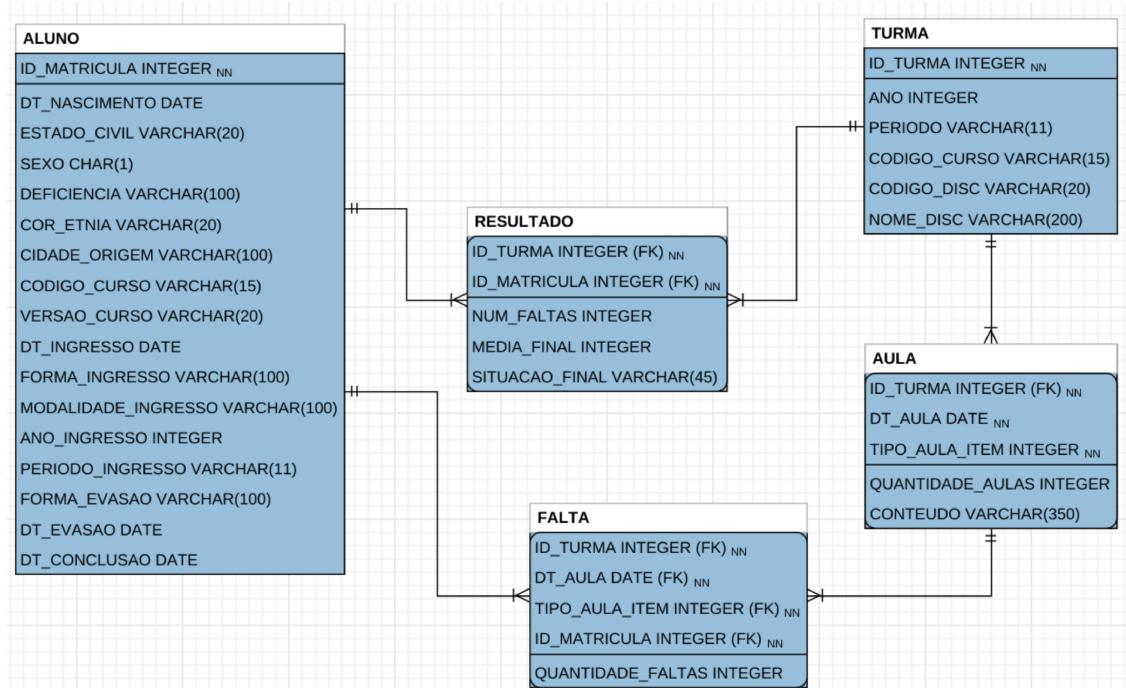


Figura 22 – Modelagem dos dados.

5.2.1 Operações de Pré-processamento

Para garantir a adequação do conjunto de dados à análise no EduFlow, foi realizada uma etapa de pré-processamento, com o objetivo de assegurar a consistência e a integridade das informações. Essa etapa foi estruturada em três tipos principais de operações:

- **Preenchimento de Dados Ausentes:** Valores ausentes em colunas-chave, como nota final e número de faltas, foram tratados por meio da inserção de um valor sentinela igual a -1. Essa escolha visa manter a integridade das instâncias da base de dados, evitando o descarte de registros que poderiam ser úteis para outras análises. Além disso, como o valor -1 está fora do intervalo válido desses atributos, ele permite a identificação explícita de dados ausentes e sua exclusão automática em cálculos estatísticos, como médias e desvios. Essa abordagem foi escolhida em detrimento de estratégias como preenchimento por média, que poderiam introduzir viés em distribuições de desempenho e mascarar comportamentos reais dos alunos, sobretudo em cenários com heterogeneidade de trajetórias;
- **Criação de Novos Atributos:** Alguns atributos adicionais foram derivados a partir dos dados brutos, ou a partir de consultas ao *website* do curso² com o intuito de padronizar a identificação de disciplinas, alunos e turmas. Foram definidos códigos unificados de disciplina, períodos e situações finais (aprovado, reprovado). Essas transformações facilitaram a criação de filtros e a agregação de dados para visualização no sistema;
- **Criação de Tópicos por Disciplina:** Com base nas informações presentes nas tabelas, foram organizados os tópicos ministrados em cada disciplina, a partir da descrição textual do conteúdo e da data da aula. Considerando que as descrições dos conteúdos ministrados foram inseridas de forma livre por diferentes docentes, observou-se uma variação na forma como os tópicos eram registrados. Para lidar com essa heterogeneidade, foi elaborada uma lista padronizada de tópicos por disciplina, tomando como base os conteúdos previstos na estrutura curricular de cada módulo. Em seguida, as descrições informadas pelos professores foram associadas manualmente aos tópicos dessa lista padronizada. Nos casos em que não havia informações suficientes, os tópicos ausentes foram inferidos com base no histórico de semestres anteriores, respeitando a ordem usual de progressão dos conteúdos. Como resultado desse processo, foram definidos identificadores e nomes únicos para os tópicos (*idt_aula* e *dsc_programa*), substituindo os registros textuais originais do campo *conteudo*. Essa reformulação favoreceu uma exploração mais precisa do envolvimento dos alunos com os diferentes conteúdos, possibilitando examinar de

² <https://facom.ufu.br/graduacao/bsi-santamonica/projeto-pedagogico>. Acesso em: 21 ago. 2025.

forma mais clara a relação entre presença em aula, desempenho e os tópicos abordados.

As informações estão organizadas em tabelas relacionais e cada tabela contribui com diferentes aspectos da modelagem acadêmica, permitindo análises integradas sobre desempenho, engajamento e progressão curricular. Os atributos finais, após o pré-processamento dos dados, são exibidos na **Tabela 2**.

Tabela 2 – Principais atributos utilizados nas visualizações do EduFlow.

Atributo	Descrição
<code>id_matricula</code>	Identificador único do aluno
<code>cod_disc, nome_disc</code>	Identificação da disciplina cursada
<code>media_final, num_faltas</code>	Nota final e número total de faltas na disciplina
<code>data_evasao, dt_conclusao</code>	Datas de evasão e de conclusão do curso
<code>estagio.datamatricula</code>	Data da matrícula do estágio realizado
<code>estagio.datafinalizacao</code>	Data do fim do estágio realizado
<code>id_turma, dt_aula</code>	Identificação da turma e data da aula ministrada
<code>quantidade_faltas</code>	Quantidade de faltas do aluno na aula específica
<code>situacao_final</code>	Situação final do aluno na disciplina, como aprovado ou reprovado, por nota ou por falta
<code>ano_ingresso, periodo_ingresso</code>	Ano e período letivo em que o aluno ingressou no curso
<code>conteudo</code>	Conteúdo ou tópico abordado na aula
<code>idt_aula, dsc_programa</code>	Identificadores únicos para os tópicos

Fonte: Elaborado pela autora.

A **Tabela 2** apresenta os principais atributos gerados ou refinados ao longo do processo de pré-processamento, os quais constituem a base de dados estruturada utilizada pelo EduFlow. Esses atributos foram definidos com o objetivo de fornecer suporte às visualizações interativas propostas pelo sistema. A correspondência entre esses atributos e os requisitos de análise definidos na **Tabela 1** evidencia como o processo de transformação e padronização dos dados foi conduzido para viabilizar a exploração visual nas diferentes visões do sistema. O pré-processamento dos dados educacionais, incluindo extração, limpeza, estruturação e categorização de informações como notas, frequência e situação final dos alunos, foi implementada utilizando a linguagem *Python*³. Para isso, foram empregadas bibliotecas amplamente utilizadas na área de ciência de dados, como o *Pandas*⁴, para manipulação e transformação de estruturas tabulares, e o *NumPy*⁵, para operações vetoriais e numéricas eficientes. Essas ferramentas possibilitaram um tratamento eficiente dos dados, assegurando a integridade e a consistência das informações utilizadas nas etapas seguintes da análise visual. Ao assegurar que todos os atributos necessários estejam devidamente normalizados, identificáveis e consistentes, o pré-processamento garante que

³ <https://www.python.org/>

⁴ <https://pandas.pydata.org/>

⁵ <https://numpy.org/>

os dados possam ser integrados de forma eficiente às visões do EduFlow, promovendo análises mais precisas, reutilizáveis e compatíveis com múltiplos contextos acadêmicos.

5.3 Análises

5.3.1 Disciplinas GSI006 e GSI015

A **Figura 23** ilustra a Visão de Sequência de Execução, mostrando a relação entre as disciplinas **Estrutura de Dados 1 (GSI006)** e **Programação Orientada a Objetos 1 (GSI015)**, bem como o desempenho geral de 579 alunos que seguiram essa sequência.

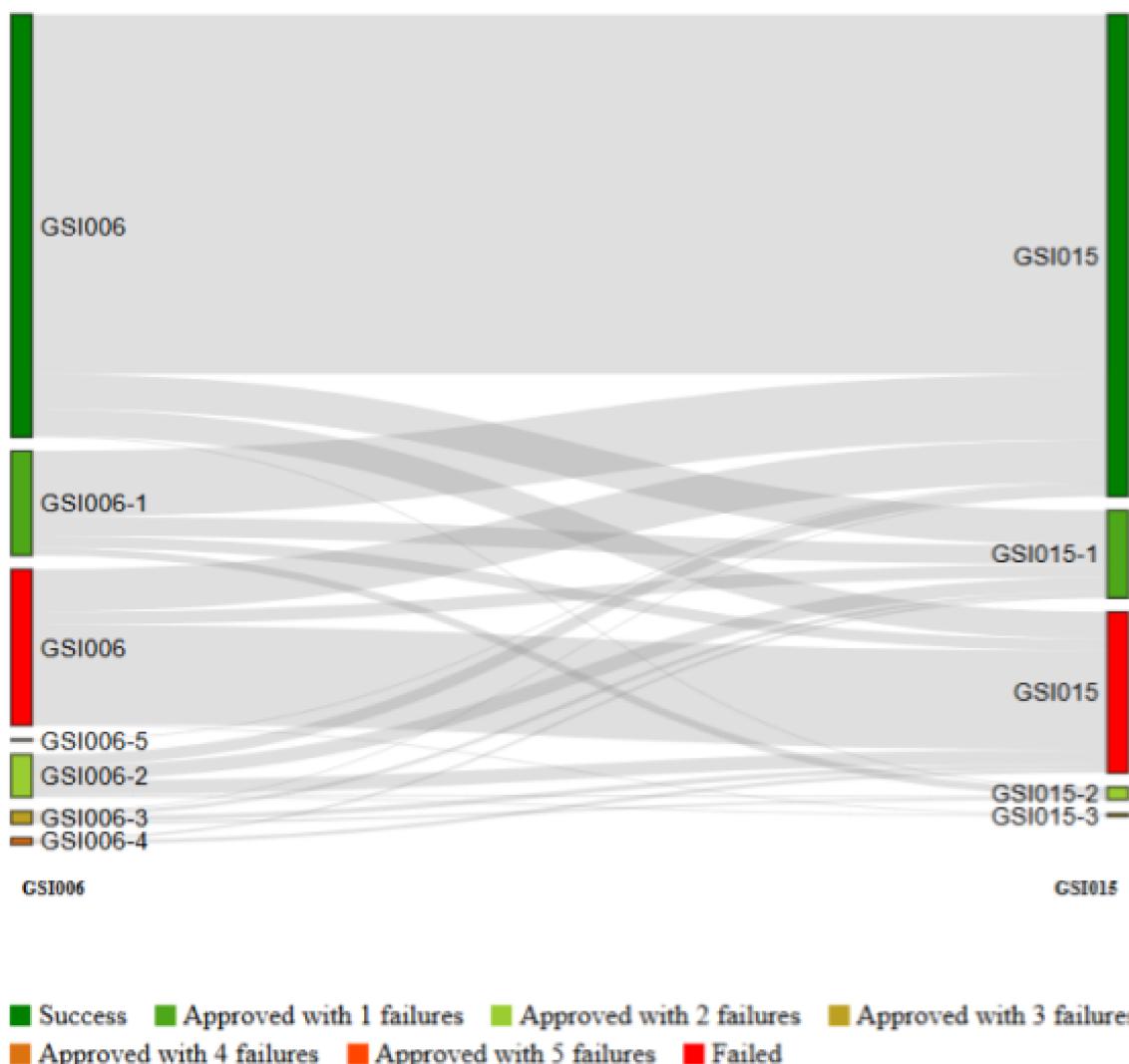


Figura 23 – Visão de Sequência de Execução para o fluxo GSI006 - GSI015.

Historicamente, essas disciplinas apresentam elevados índices de reprovação e, conforme mencionado anteriormente, estão organizadas em um fluxograma curricular que

sugere sua realização em sequência. GSI006 é ofertada no segundo período e GSI015, no terceiro, de modo que, no fluxo padrão do curso, os alunos geralmente cursam uma no semestre imediatamente anterior à outra. Apesar de não haver uma relação formal de pré-requisito entre elas, os conceitos aprendidos em GSI006 são frequentemente aplicados e aprofundados em GSI015. Essa continuidade sugere que dificuldades enfrentadas em GSI006 podem refletir diretamente no desempenho em GSI015, tornando a relação entre as disciplinas um ponto relevante para análise.

A altura dos nós e a espessura das arestas permitem uma avaliação quantitativa e detalhada das trajetórias acadêmicas dos alunos. O *layout* sugere que o sucesso em GSI006 influencia diretamente o desempenho em GSI015, uma vez que alunos que obtêm êxito na primeira disciplina tendem a alcançar melhores resultados na segunda.

Ao interagir com a visualização (marcação azul na **Figura 24**), observa-se que a maioria dos alunos (85,06%) que foram aprovados em GSI006 também foram aprovados em GSI015. Por outro lado, a altura dos nós vermelhos indica que uma proporção menor de alunos foi reprovada na disciplina, sendo que 63,64% dos alunos que reprovaram em GSI006 também foram reprovados em GSI015 (**Figura 25A**). Adicionalmente, 62,96% dos alunos com apenas uma reprovação anterior (GSI006-1) foram aprovados diretamente em GSI015 (**Figura 25B**), enquanto apenas 27,27% dos alunos obtiveram aprovação após duas reprovações anteriores (**Figura 25C**).

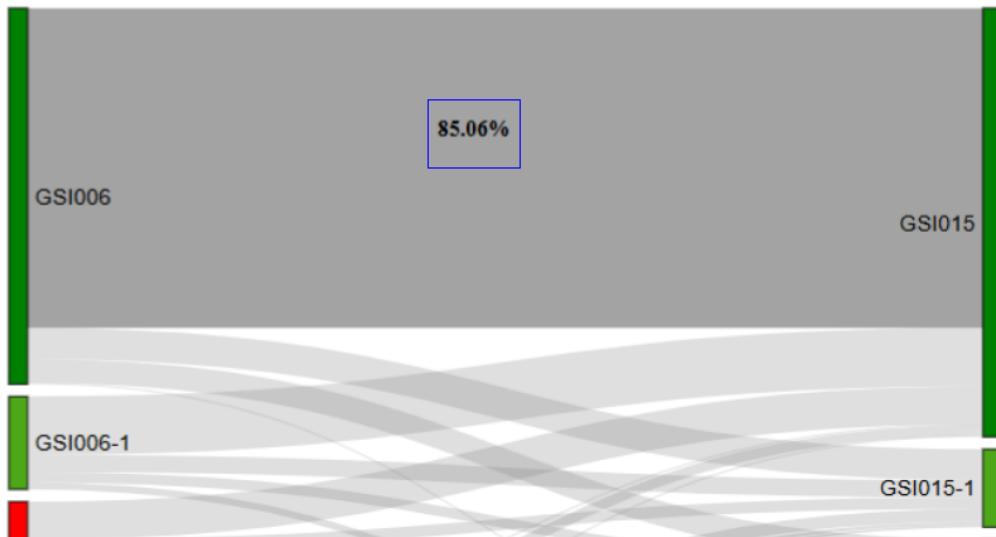


Figura 24 – Exibição da proporção de alunos aprovados em GSI006 e que também foram aprovados em GSI015.

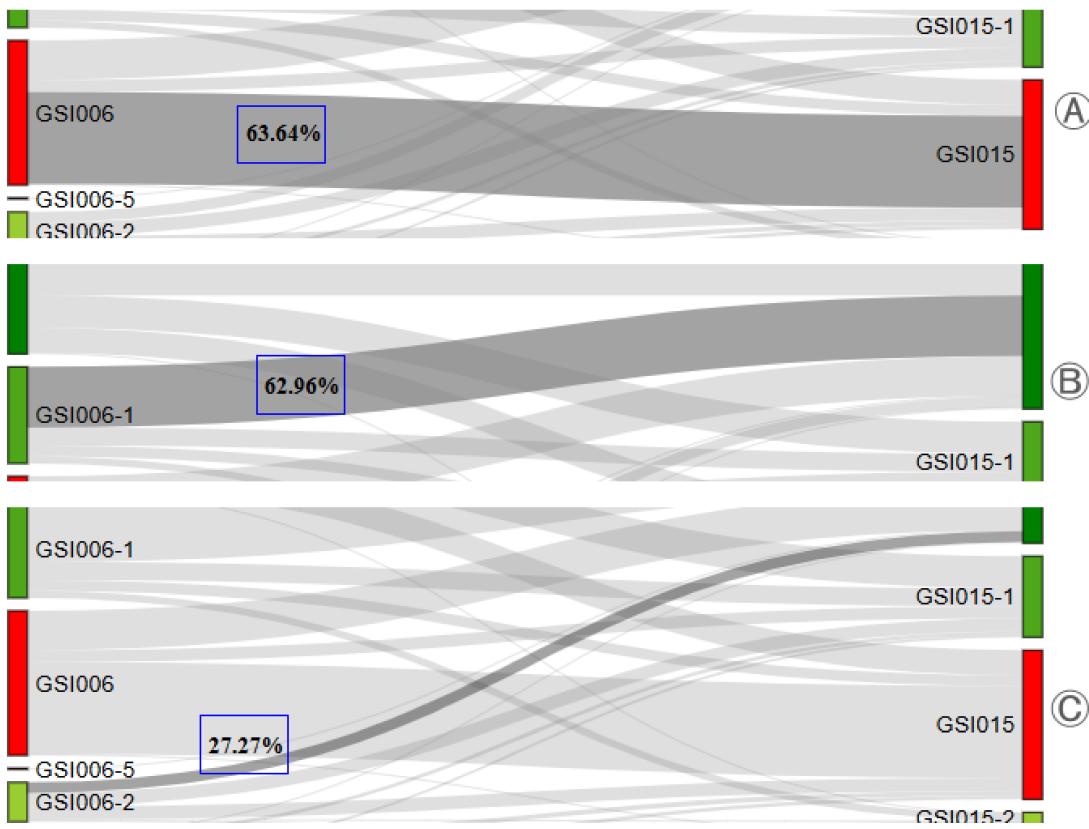


Figura 25 – A) Exibição da proporção de alunos reprovados em GSI006 e que também foram reprovados em GSI015. B) Exibição da proporção de alunos aprovados com duas tentativas em GSI006 e que foram aprovados diretamente em GSI015. C) Exibição da proporção de alunos aprovados com três tentativas em GSI006 e que foram aprovados diretamente em GSI15.

Ao selecionar o nó que representa a proporção de alunos reprovados em GSI006, o sistema fornece análises detalhadas relacionadas à disciplina como um todo, incluindo informações específicas sobre a categoria de reprovação. Essas análises específicas são particularmente apresentadas na Visão de Notas e Frequência, que destacam como o desempenho acadêmico e a frequência impactam os resultados dos alunos. A **Figura 26** analisa a relação entre desempenho acadêmico e frequência em GSI006, destacando os alunos reprovados em comparação com o desempenho geral nessa disciplina. A **Figura 26A** apresenta a distribuição das notas, na qual se observa que a média dos alunos reprovados é substancialmente inferior à média geral. Ambas as médias estão abaixo do mínimo exigido para aprovação. Já na **Figura 26B**, que exibe a distribuição de faltas, nota-se que a média de faltas entre os alunos reprovados é significativamente superior à média geral. Esse padrão sugere que a frequência é um fator crucial para a aprendizagem e o desempenho acadêmico, indicando que alunos com maior assiduidade tendem a obter

melhores resultados.

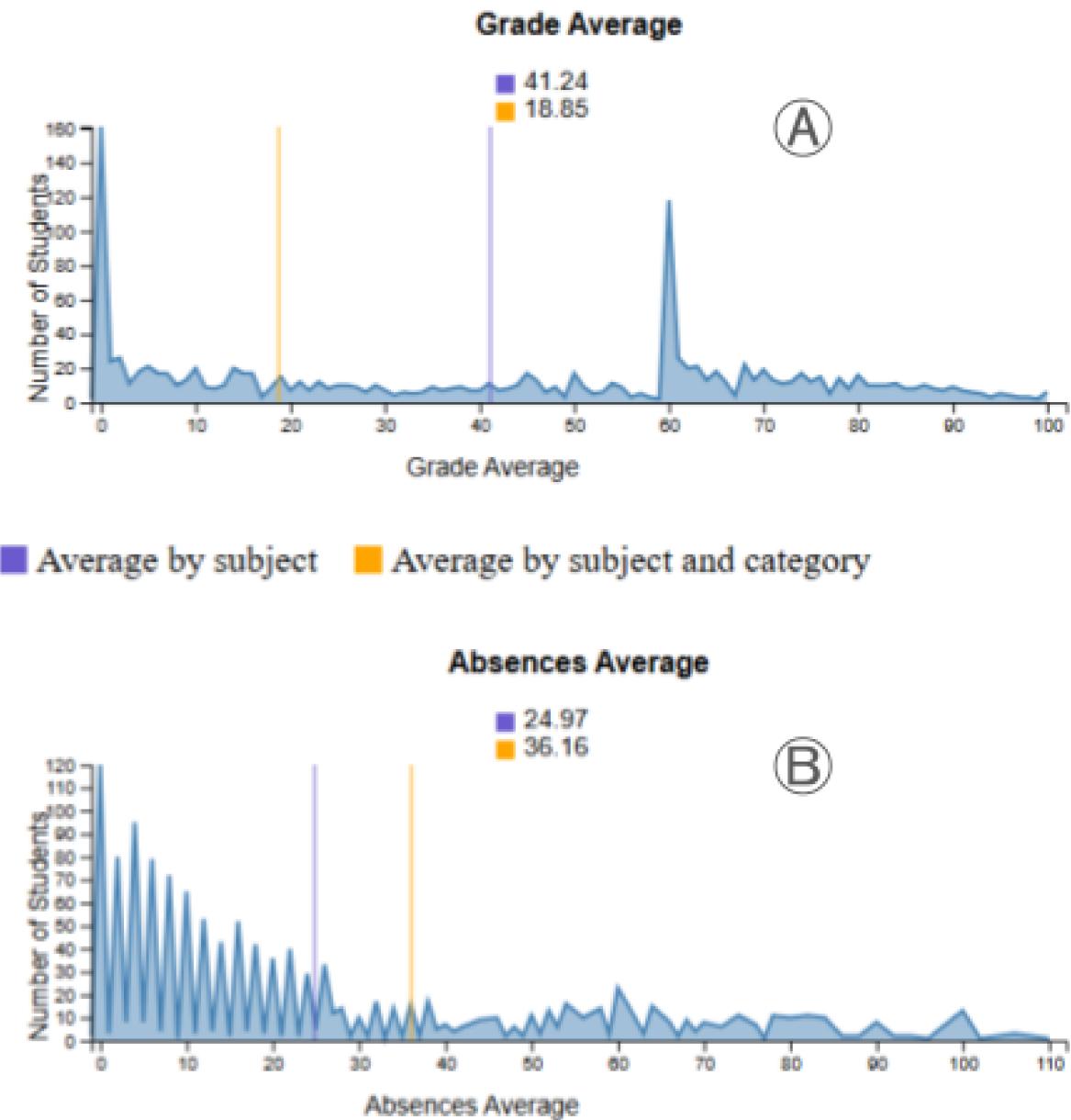


Figura 26 – Visão de Notas e Frequência da disciplina GSI006 com status reprovado, exibindo curvas de distribuição das notas (A) e faltas (B), juntamente com as linhas que representam médias gerais (roxa) e por status (amarela).

Na Visão de Estágio (**Figura 27**), observa-se que a maior parte dos alunos aprovados diretamente pertence ao grupo que não realizou estágio durante o curso da disciplina. Por outro lado, um número significativo de alunos que realizaram estágios e que foram reprovados, ou que conseguiram aprovação apenas após múltiplas tentativas, pode ser observado. Esse padrão sugere que o envolvimento em estágio, embora enriquecedor em termos de experiência prática, pode estar associado a dificuldades no desempenho

acadêmico em disciplinas como GSI006, possivelmente devido a desafios relacionados à gestão de tempo ou à sobrecarga de atividades. No entanto, é importante destacar que também existe uma proporção de alunos que conseguiram conciliar a realização de estágios com um bom desempenho acadêmico, obtendo aprovação na disciplina. Esse fato indica que, embora desafiador, é possível para alguns alunos gerenciar com sucesso as demandas de estágio e os requisitos acadêmicos.

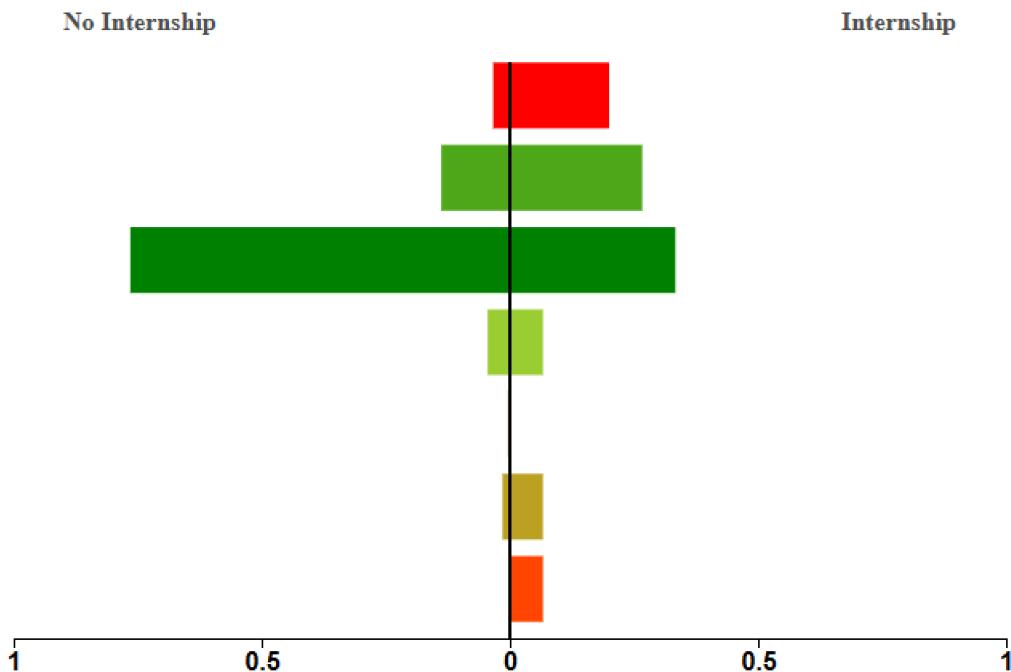


Figura 27 – Visão Estágio da disciplina GSI006.

A análise dos conteúdos abordados na disciplina GSI006 é apresentada pela Visão de Conteúdos na **Figura 28** e na **Figura 29**. Observa-se que a maioria dos conteúdos apresentam distribuições semelhantes entre as categorias de desempenho, com predominância da cor verde, indicando uma maior proporção de alunos aprovados. No entanto, alguns tópicos revelam padrões distintos, com predominância de alunos que estiveram presentes nessas aulas e foram reprovados pelo menos uma vez na disciplina. Ao ordenar os tópicos pela quantidade de alunos reprovados em GSI006, nota-se que os conteúdos representados pelos números 11, 13, 15, 16, 18 e 19 (destacados em azul) apresentam uma faixa maior de alunos reprovados (120 alunos), em relação à maioria dos conteúdos. Além disso, quase metade dos alunos que compareceram às aulas relacionadas a esses tópicos foram reprovados pelo menos uma vez na disciplina. Essa conclusão é possível ao observar que apenas

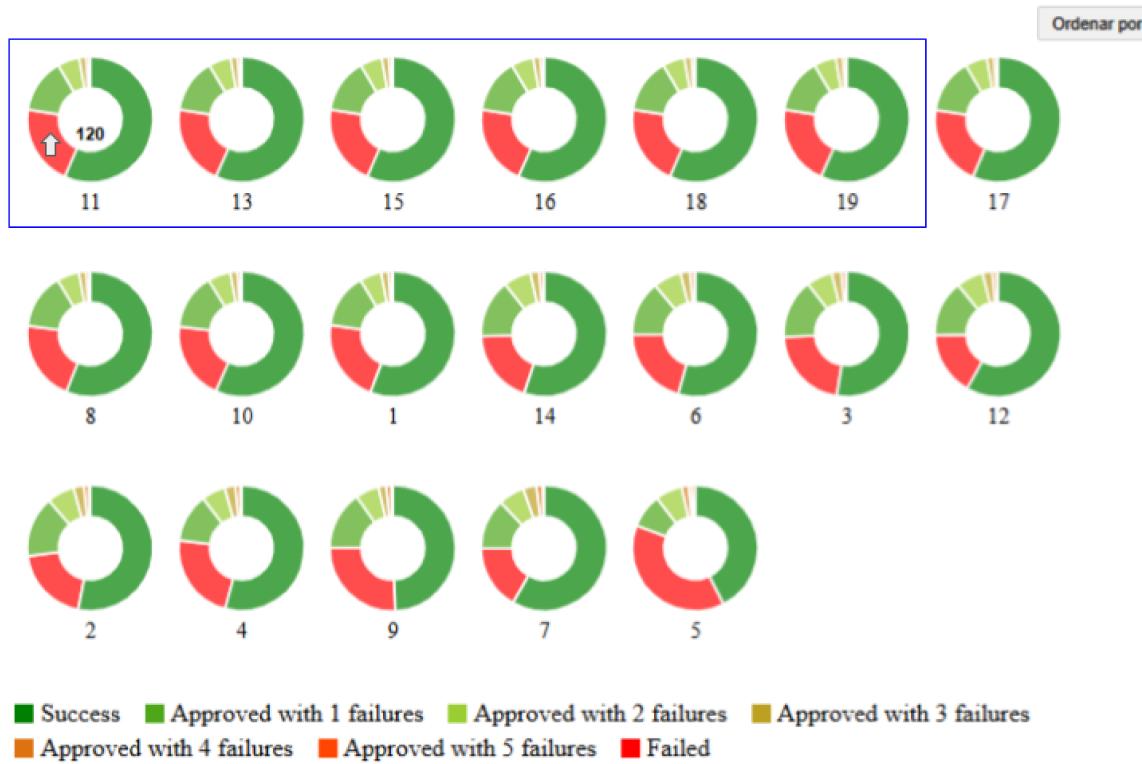


Figura 28 – Visão de Conteúdos da disciplina GSI006: Gráfico de setores, com seleção dos conteúdos 11, 13, 15, 16, 18 e 19 em destaque.

a fatia verde escura, indicada na legenda como *Success*, representa os alunos aprovados diretamente, ou seja, sem reprovações anteriores. Todas as demais fatias, em diferentes tonalidades, correspondem a alunos que obtiveram aprovação somente após uma ou mais tentativas ou que foram reprovados. Ao analisar a Visão de Conteúdos (**Figura 29**), esses mesmos tópicos apresentam maior quantidade de faltas do que “Algoritmos de busca” (tópico 7), por exemplo, que por sua vez apresenta menor quantidade de alunos reprovados (50 alunos), sugerindo que a ausência pode ser um fator crítico para o desempenho satisfatório. Por outro lado, o conteúdo introdutório “Introdução, Execução de Programas - ARGC e ARGV” (tópico 1) registra um índice de faltas não tão alto (**Figura 29**), quando comparado aos demais conteúdos. Apesar disso, esse tópico apresenta uma quantidade significativa de alunos (117 alunos) reprovados na disciplina. É importante destacar que essas aulas introdutórias não envolvem conteúdo técnico. Elas frequentemente registram alta participação por serem momentos em que o professor apresenta o planejamento do curso, o que é essencial para contextualizar a disciplina e alinhar expectativas. O *layout* auxilia a verificar se os alunos estão comparecendo a essas aulas iniciais, oferecendo um indicador de engajamento que, embora não esteja diretamente relacionado ao desempenho

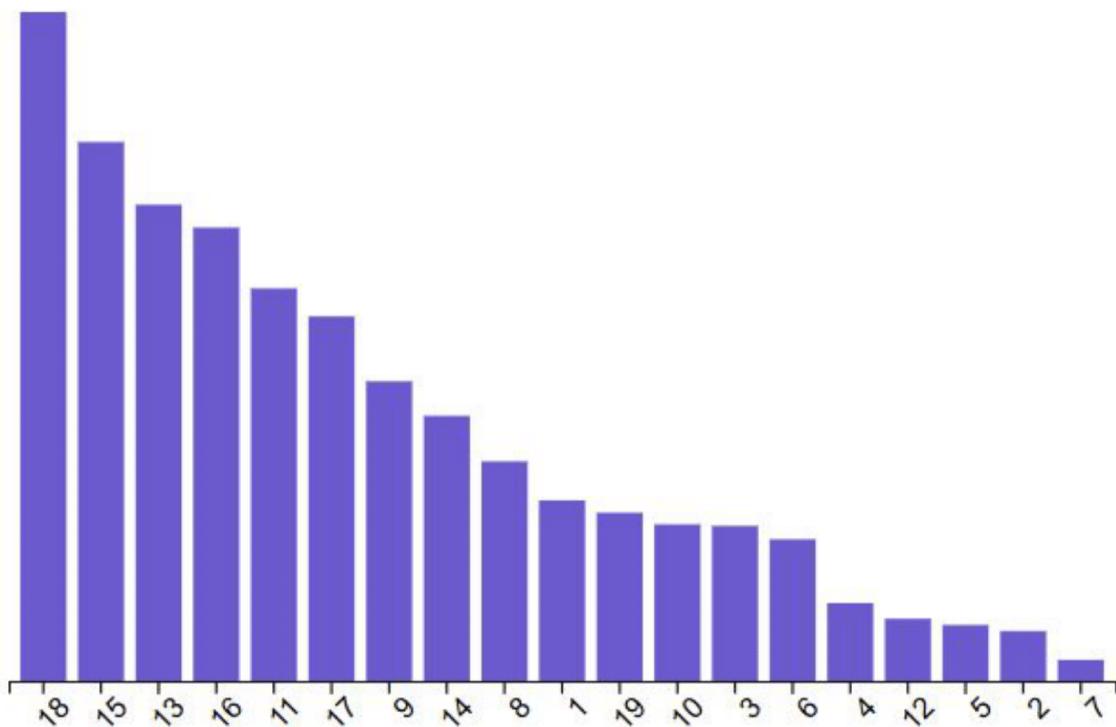


Figura 29 – Visão de Conteúdos da disciplina GSI006: Faltas por Conteúdo.

acadêmico, é relevante para a organização e integração dos alunos.

Ao analisar o desempenho de 158 alunos que cursaram a disciplina GSI015 antes de GSI006, conforme apresentado na Visão de Sequência de Execução (**Figura 30**), observa-se um aumento significativo nos índices de reprovação, sugerindo que a sequência proposta na grade curricular, que recomenda cursar GSI006 antes de GSI015, é mais propícia para o sucesso acadêmico. O *layout* ressalta o fato de que o domínio dos conceitos fundamentais de GSI006 é essencial para a assimilação dos conteúdos mais avançados abordados em GSI015. Nota-se que os nós verdes em GSI006 são consideravelmente menores quando os alunos cursam GSI015 primeiro, em comparação com o *layout* da **Figura 23**, o que sugere que, mesmo após obterem aprovação em GSI015, muitos alunos enfrentam dificuldades para alcançar um desempenho satisfatório em GSI006. Apenas 27.66% dos alunos aprovados em GSI015 obtêm o mesmo êxito em GSI006, conforme destacado ao posicionar o cursor na aresta correspondente.

Além disso, os nós vermelhos, que simbolizam reprovações, evidenciam que a reprovação inicial em GSI015 está frequentemente associada a um desempenho insatisfatório posterior em GSI006. De fato, 78.57% dos alunos reprovados em GSI015 também reproduziram em GSI006 (**Figura 31**). O *layout* nesse caso realça como a estrutura curricular escolar está adequada no sentido de contribuir com o bom desempenho acadêmico dos

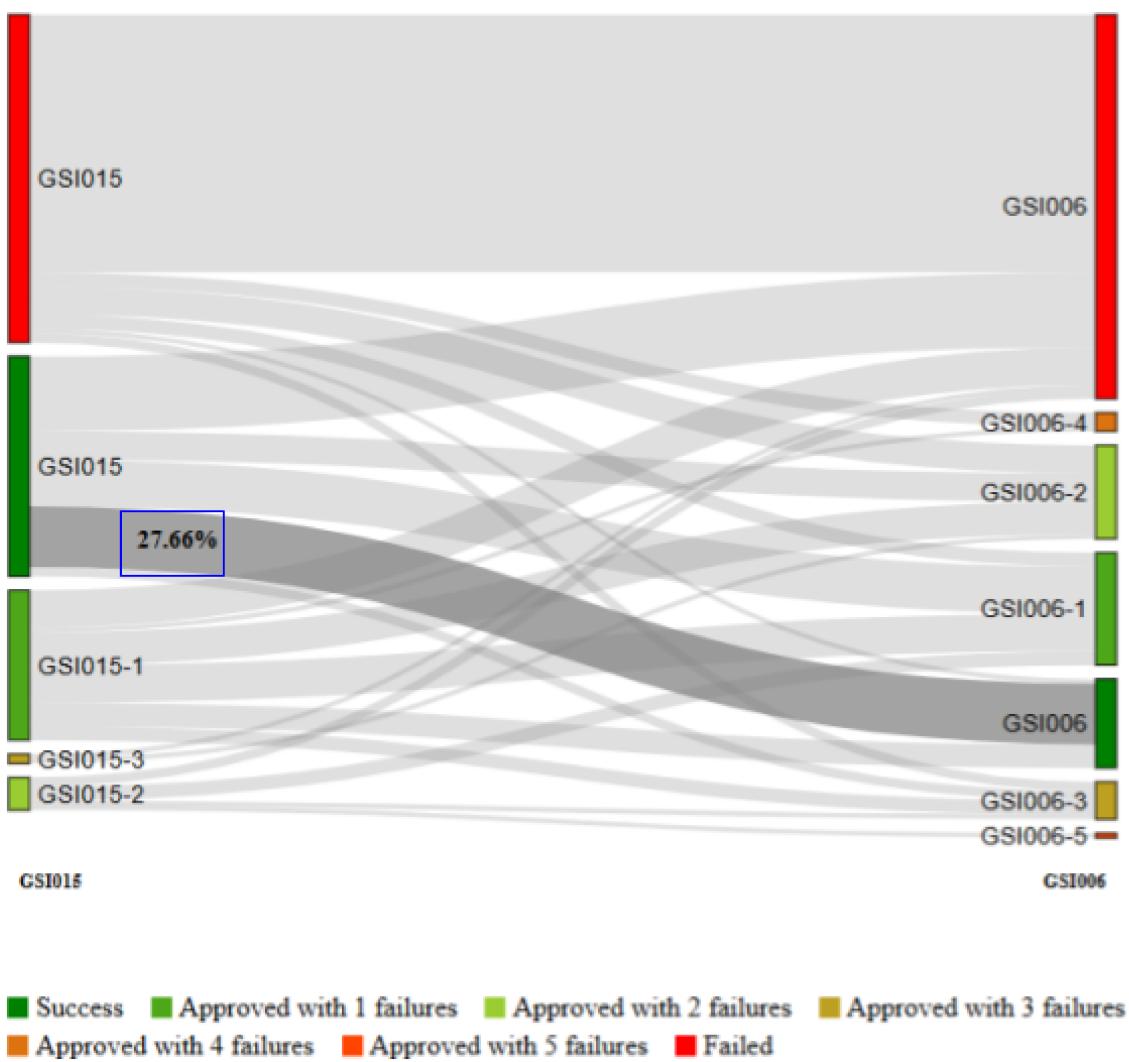


Figura 30 – Visão de Sequência de Execução para o fluxo GSI015 - GSI006.

alunos, garantindo que conceitos fundamentais sejam dominados antes de avançar para conteúdos mais complexos.

Ao clicar no nó vermelho em GSI015, são apresentadas análises complementares em relação à disciplina, como a Visão de Notas e Frequência, a Visão de Estágio e a Visão de Conteúdos.

A Visão de Notas e Frequência da disciplina GSI015 é apresentada na **Figura 32**. Semelhante à análise feita com a disciplina GSI006, o *layout* revela que alunos reprovados em GSI015 apresentam médias de notas mais baixas que a média geral considerando todos os alunos da disciplina, enquanto a média de faltas é maior. Estratégias voltadas para incentivar a assiduidade podem ser especialmente eficazes na redução das taxas de reprovação. Os dados revelam que alunos com maior presença nas aulas têm maior probabilidade de alcançar melhores resultados.

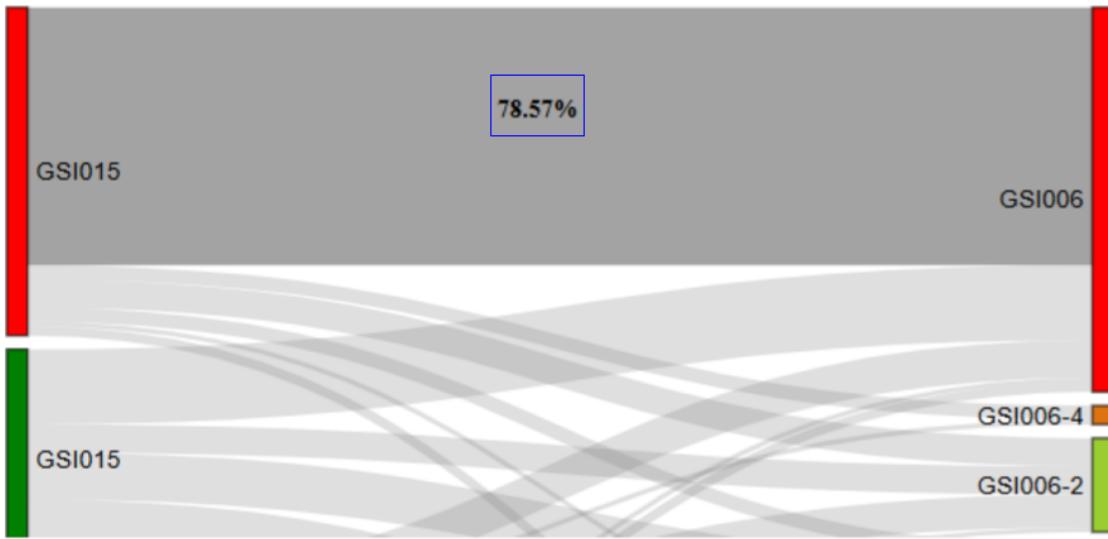


Figura 31 – Exibição da sequência de parte dos alunos reprovados em GSI015 e que também foram reprovados em GSI006.

Na Visão de Estágio (**Figura 33**), observa-se que a proporção de alunos reprovados na disciplina que participaram do estágio é superior àqueles que não realizaram estágio durante o mesmo período. Além disso, a proporção dos alunos aprovados diretamente é maior entre aqueles que não realizaram o estágio, sugerindo que as mesmas dificuldades relacionadas à gestão de tempo e sobrecarga de atividades impactam negativamente o desempenho acadêmico em GSI015. No entanto, é importante ressaltar que os padrões observados em GSI015 diferem no que diz respeito à dispersão dos dados. A distribuição apresentada é mais heterogênea entre as categorias de desempenho. Esse comportamento sugere que outros fatores, além da realização de estágio, podem influenciar no desempenho dos alunos nesta disciplina.

A análise dos conteúdos abordados na disciplina GSI015 é apresentada na Visão de Conteúdos, na **Figura 34** e **Figura 35**. Na **Figura 34**, os conteúdos estão ordenados de forma decrescente em relação à quantidade de alunos na categoria “reprovados”. Cada um dos nove primeiros tópicos inclui, individualmente, 70 alunos reprovados, como ilustrado ao se posicionar o cursor sobre essa categoria de desempenho do conteúdo 11, “Tratamento de exceções”. Ao analisar os cinco tópicos com maiores índices de faltas, apresentados na **Figura 35** (tópicos 11, 3, 13, 8 e 15), evidencia-se que todos esses tópicos correspondem aos conteúdos com os maiores números de reprovações. Por outro lado, o conteúdo 10, “Uso de bibliotecas em linguagens orientadas a objetos”, apresenta o menor índice de faltas e também o menor número de reprovações. Esses resultados reforçam a hipótese de

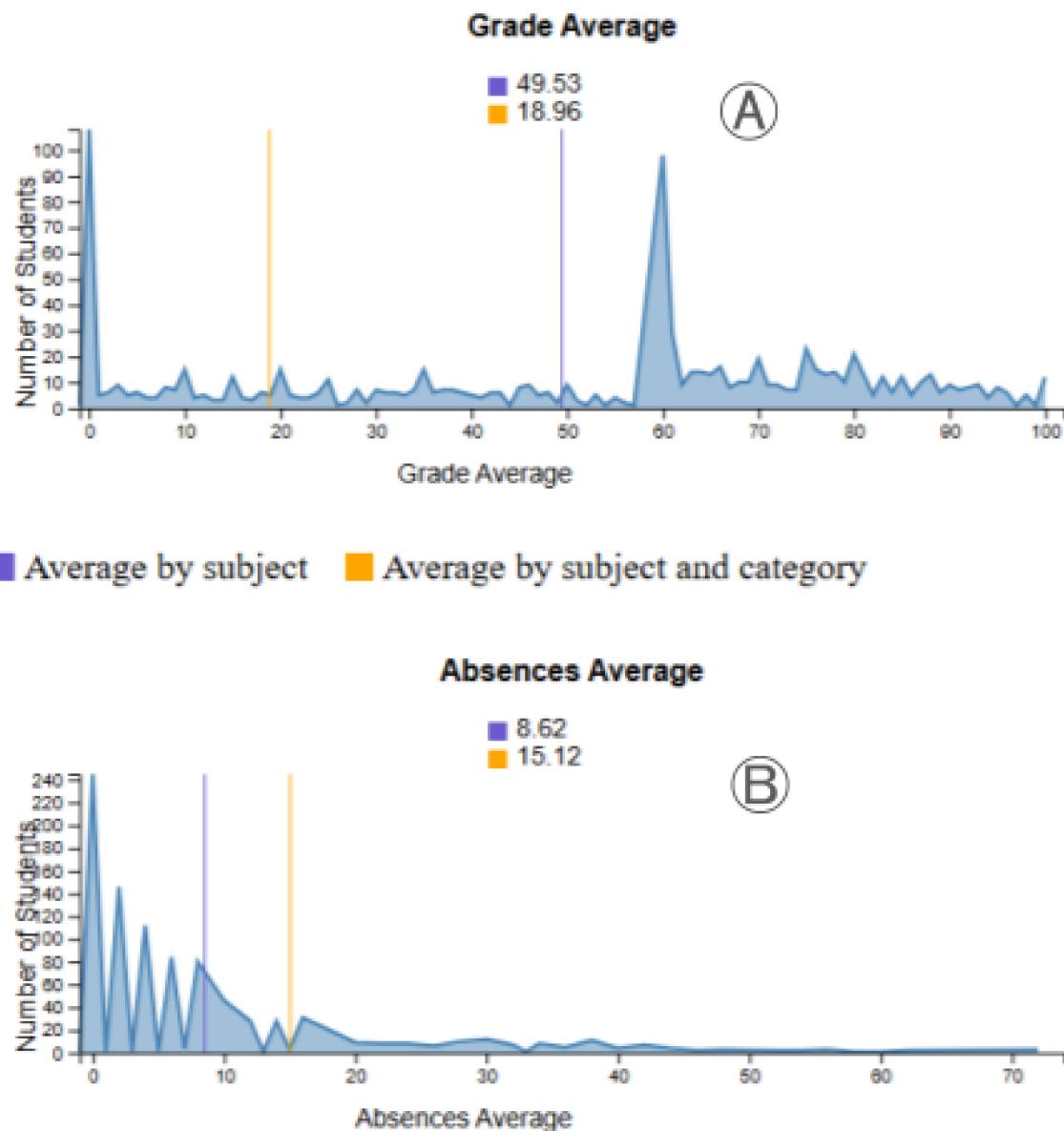


Figura 32 – Visão de Notas e Frequência da disciplina GSI015 com status reprovado, exibindo curvas de distribuição das notas (A) e faltas (B), juntamente com as linhas que representam médias gerais (roxa) e por status (amarela).

que a presença em sala é um fator determinante para o desempenho acadêmico, conforme já identificado anteriormente na análise da disciplina GSI006.

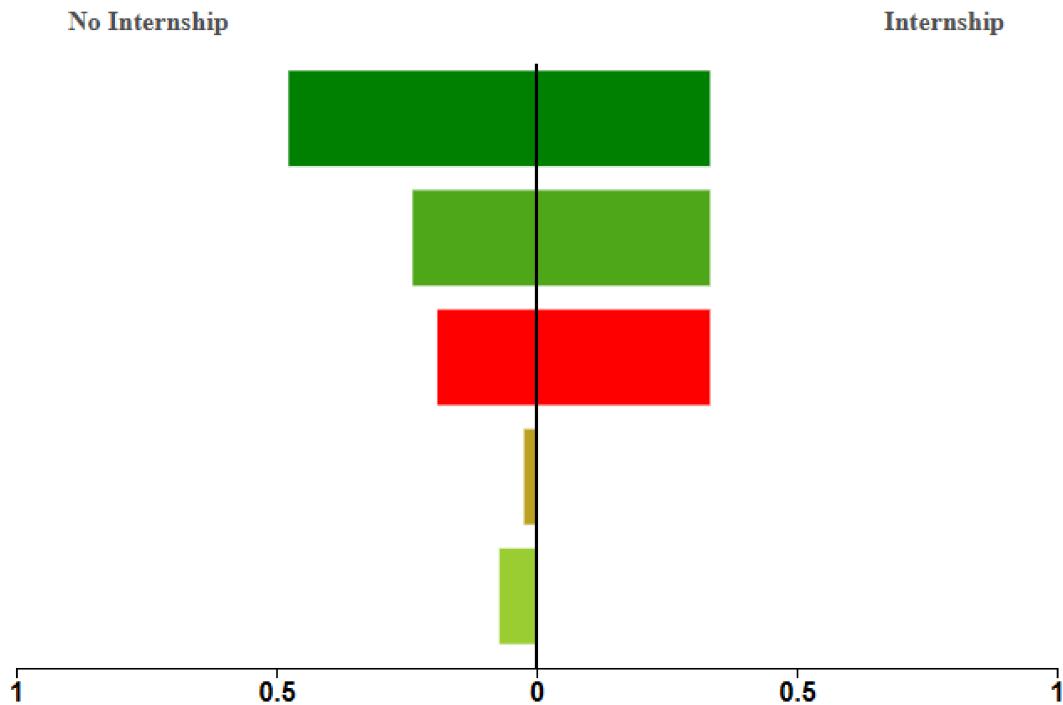


Figura 33 – Visão Estágio da disciplina GSI015.

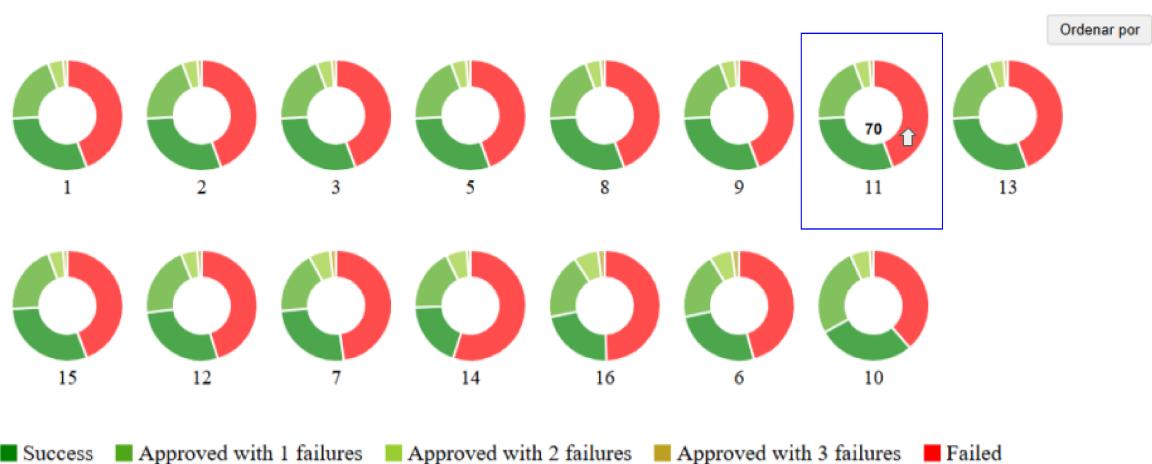


Figura 34 – Visão de Conteúdos da disciplina GSI015: Gráfico de setores, com seleção do conteúdo 11 em destaque.

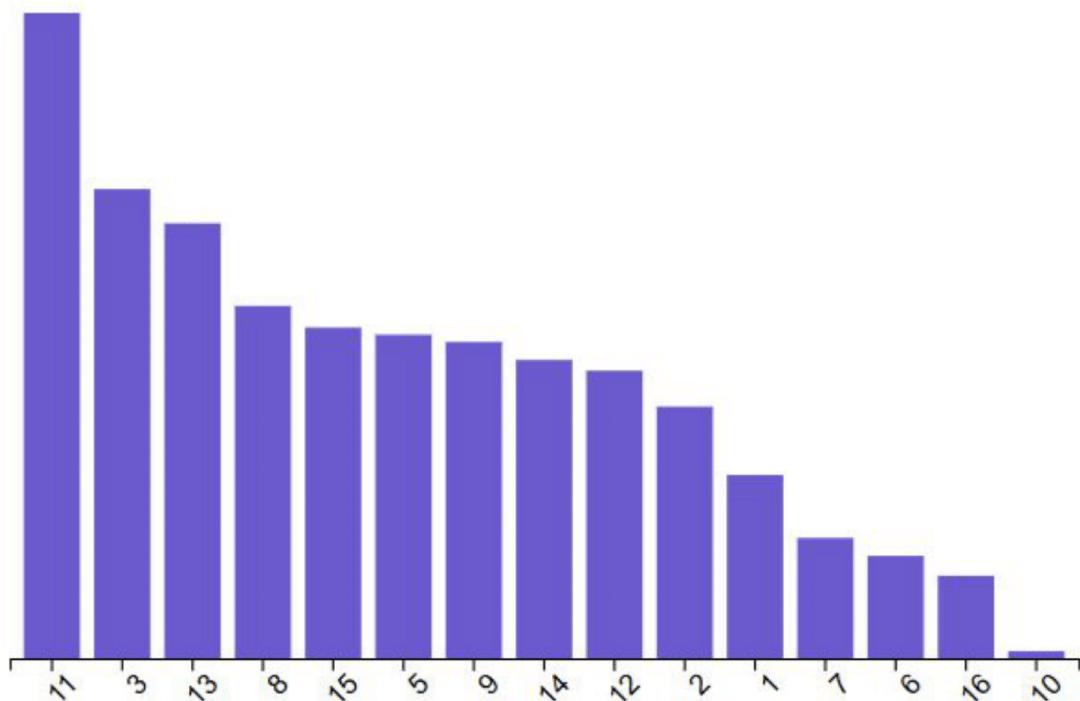


Figura 35 – Visão de Conteúdos da disciplina GSI015: Faltas por Conteúdo.

5.3.2 Disciplinas GSI002, GSI006 e GSI016

O estudo de caso apresentado nesta seção explora o uso do EduFlow para analisar a relação entre as disciplinas **Introdução à Programação de Computadores (GSI002)**, **Estrutura de Dados 1 (GSI006)** e **Banco de Dados 1 (GSI016)**. A disciplina GSI002 é oferecida no primeiro semestre e introduz os fundamentos da lógica de programação e algoritmos, formando a base essencial para os componentes curriculares subsequentes. GSI006, no segundo semestre, aprofunda os conceitos de programação com foco em estruturas de dados mais complexas, utilizadas tanto no desenvolvimento de algoritmos quanto na manipulação eficiente de informações. Já GSI016, oferecida no quarto semestre, aplica os conceitos aprendidos em GSI006 no contexto de bancos de dados, com ênfase em modelagem, consultas SQL e gestão de dados. Essa estrutura sequencial facilita a consolidação de conhecimentos e prepara os alunos para conteúdos mais avançados a serem aprendidos posteriormente no curso. No entanto, apesar dessa lógica curricular, muitos alunos não percebem a forte relação entre essas disciplinas no que diz respeito aos conceitos aprendidos. Essa falta de percepção leva, em diversos casos, à tentativa de cursá-las fora da ordem recomendada, o que pode contribuir para dificuldades no aprendizado e, consequentemente, resultar em reprovações.

A Visão de Sequência de Execução (**Figura 36**) apresenta 482 alunos que seguiram a ordem recomendada. Observa-se que esses alunos apresentam menores taxas de reprovação e demonstram uma progressão mais consistente entre as disciplinas. Uma grande proporção dos alunos que foram aprovados em GSI002 também obteve sucesso em GSI006 (69,54%) e, entre os que foram aprovados em GSI006, 75,09% foram aprovados em GSI016 na primeira tentativa. Esses resultados indicam que o domínio dos conceitos fundamentais nas disciplinas iniciais contribui para o sucesso nas disciplinas subsequentes. Por outro lado, a reprovação em disciplinas anteriores está associada à reprovação nas seguintes. Por exemplo, 74,71% dos alunos reprovados em GSI006 também foram reprovados em GSI016. O *layout* evidencia claramente esse padrão de transição, com nós vermelhos (reprovações) e arestas entre eles indicando fortes correlações negativas.

Ao selecionar o nó correspondente aos alunos reprovados em GSI016, tornam-se disponíveis análises detalhadas sobre a disciplina. A **Figura 37** apresenta a Visão de Notas e Frequência, que fornece uma análise detalhada da relação entre o desempenho acadêmico e a assiduidade dos alunos, comparando alunos que foram reprovados com os alunos das outras categorias de desempenho da disciplina. Na **Figura 37A**, observa-se que a média de notas dos alunos reprovados é substancialmente inferior à média geral. Ambas as médias encontram-se abaixo do mínimo exigido para aprovação, o que evidencia não apenas o desempenho inferior dos alunos reprovados em comparação com os demais, mas também um desempenho geral insatisfatório da turma como um todo na disciplina. Essa informação é crucial para uma análise mais detalhada, pois permite identificar lacunas no processo de aprendizagem, orientar estratégias pedagógicas focadas em recuperação e

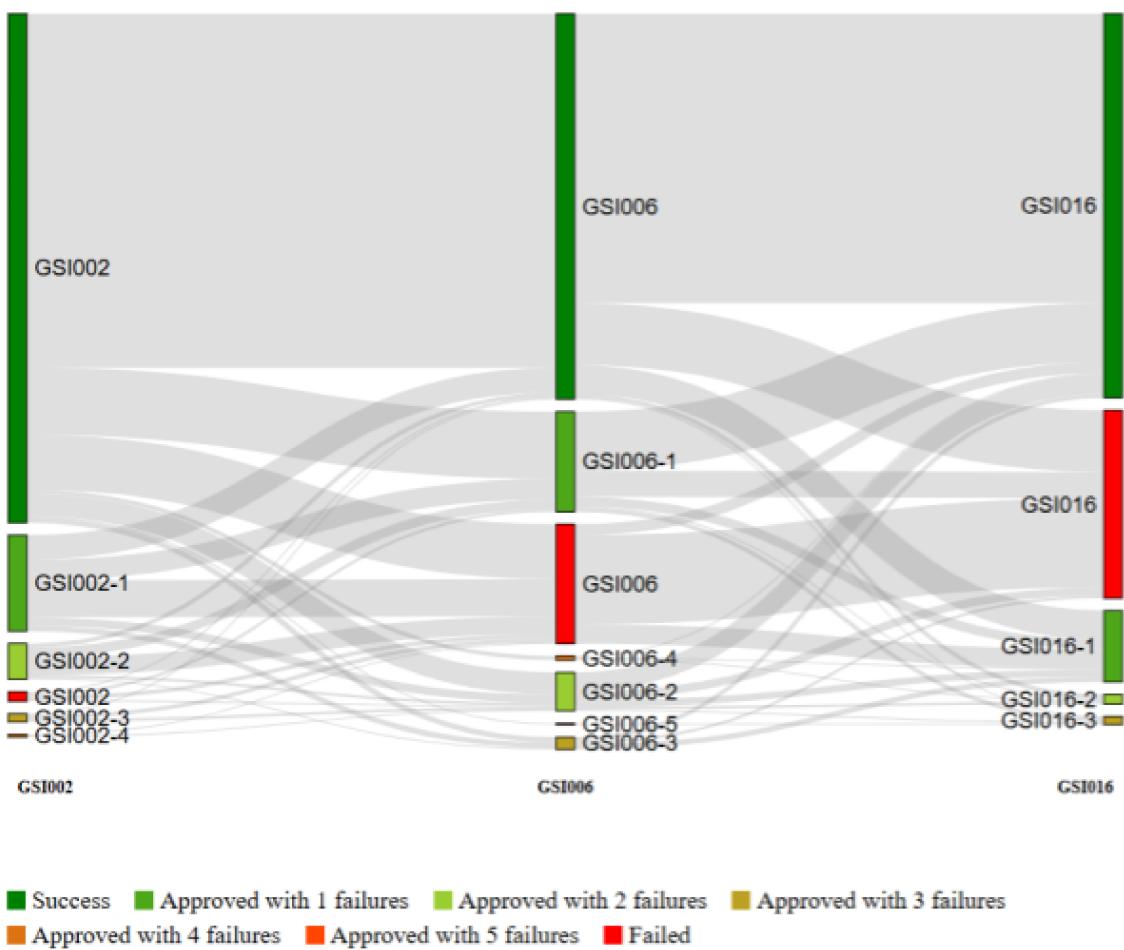


Figura 36 – Visão de Sequência de Execução para o fluxo GSI002 - GSI006 - GSI016.

revisar o planejamento didático de conteúdos com alto índice de reprovação. Já a **Figura 37B** mostra que a média de faltas dos alunos reprovados é maior do que a média geral. Isso sugere que a assiduidade é um fator determinante para o aprendizado e desempenho acadêmico, indicando que alunos com maior presença tendem a obter melhores resultados. Esses dados sinalizam a necessidade de ações preventivas voltadas à melhoria da presença em sala de aula, bem como de políticas institucionais que identifiquem e atuem sobre os fatores associados à evasão por faltas.

A Visão de Conteúdos (**Figura 38** e **Figura 39**) da disciplina GSI016 oferece *insights* adicionais. Na **Figura 38**, os tópicos estão ordenados em ordem decrescente com base no número de alunos na categoria “reprovados”, destacada em vermelho. Cada um dos quatro primeiros tópicos inclui, individualmente, 137 alunos reprovados. Observa-se que a maioria dos tópicos apresenta distribuições semelhantes entre as categorias de desempenho, com predominância da cor verde, o que indica maiores taxas de aprovação. Ao

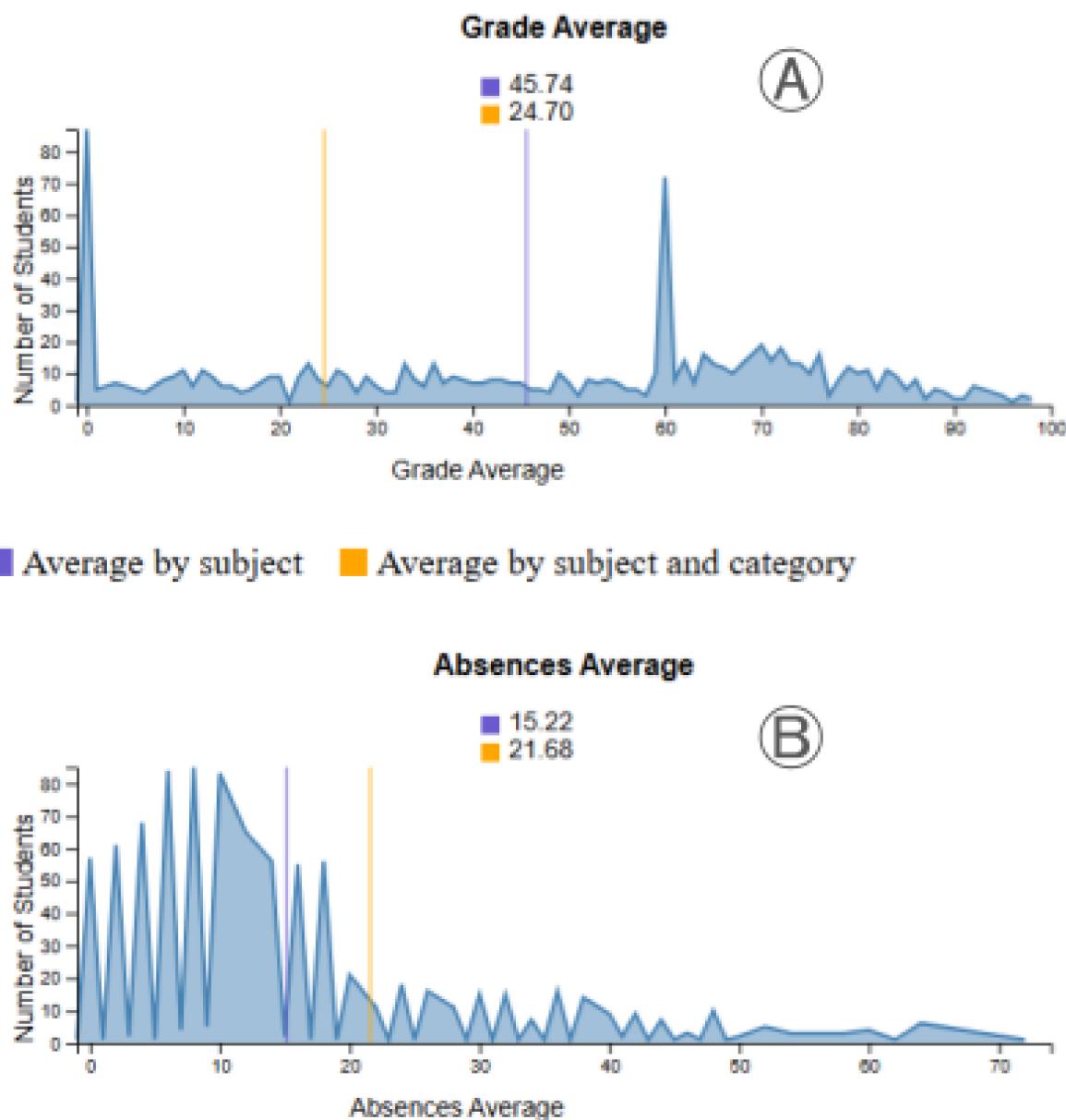


Figura 37 – Visão de Notas e Frequência da disciplina GSI016 com status reprovado, exibindo curvas de distribuição das notas (A) e faltas (B), juntamente com as linhas que representam médias gerais (roxa) e por status (amarela).

analisar os três tópicos com as maiores índices de faltas na **Figura 39** (tópicos 4, 3 e 2), observa-se que esses tópicos também correspondem àqueles com as maiores taxas de reprovação. Esses resultados reforçam a hipótese de que a frequência às aulas é um fator determinante no desempenho acadêmico.

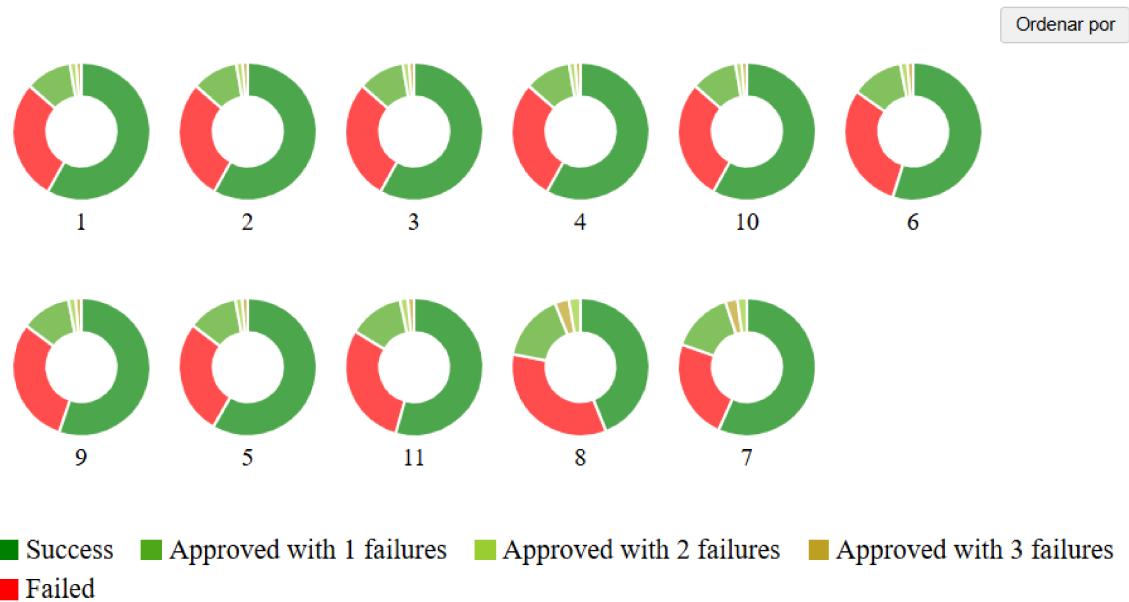


Figura 38 – Visão de Conteúdos da disciplina GSI016: Gráfico de setores.

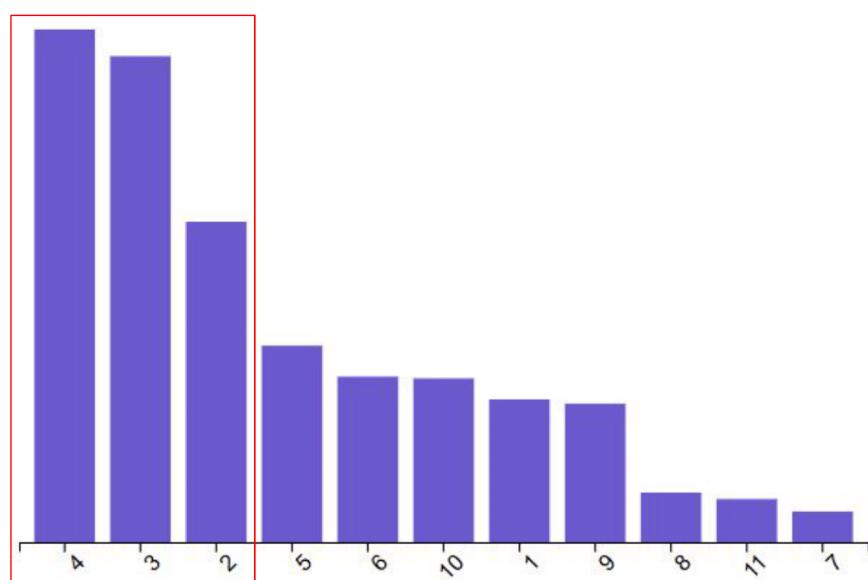


Figura 39 – Visão de Conteúdos da disciplina GSI016: Faltas por Conteúdo.

Em contraste, a **Figura 40** destaca os desafios enfrentados por um grupo de 61 alunos que não seguiram a sequência recomendada, evidenciando o impacto negativo de cursar disciplinas avançadas sem a devida preparação proporcionada pelos componentes curriculares anteriores, conforme apresentado na **Figura 36**. Os nós vermelhos referentes às

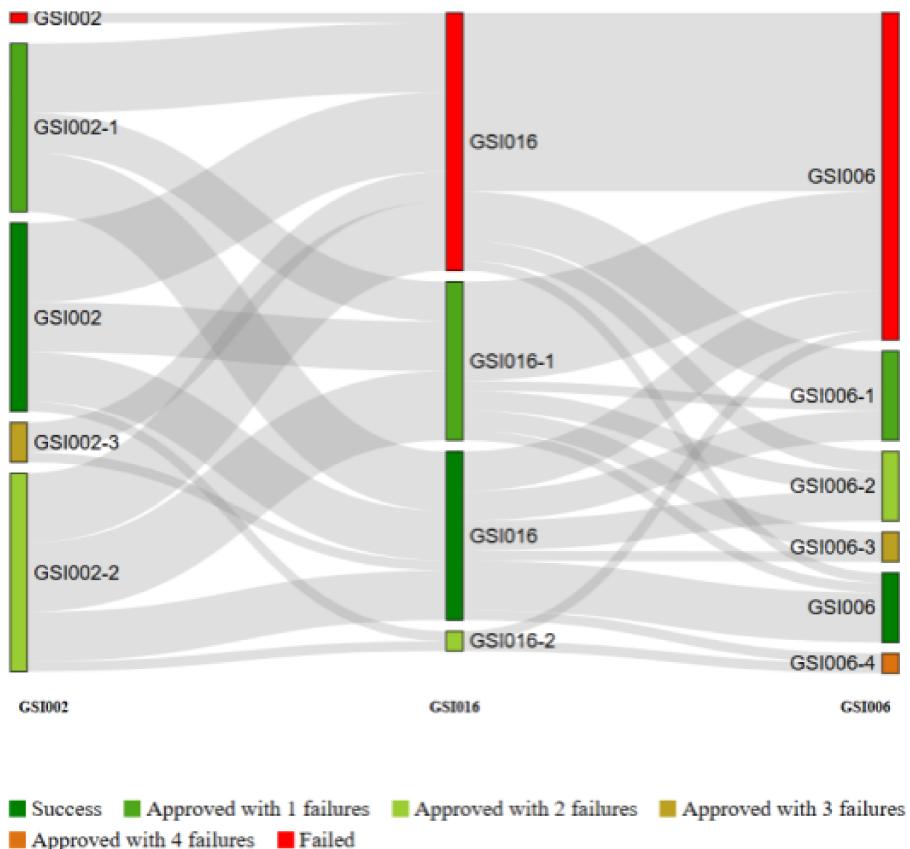


Figura 40 – Visão de Sequência de Execução para o fluxo GSI002 - GSI016 - GSI006.

disciplinas GSI016 e GSI006 aparecem mais altos e visualmente mais destacados. Além disso, observa-se que as barras representando as disciplinas aparecem mais segregadas e fragmentadas, sugerindo um cenário acadêmico mais problemático, com maior variabilidade no desempenho dos alunos. Por outro lado, os nós verdes referentes às disciplinas GSI016 e GSI006 são mais baixos, indicando que uma proporção menor de alunos obteve aprovação direta nessas disciplinas. Alunos que cursaram GSI016 antes de concluir GSI006 apresentaram uma taxa de aprovação direta de apenas 29,41%, enquanto 69,23% deles foram reprovados na disciplina (**Figura 41**). Esse padrão reforça a importância de seguir o fluxo de disciplinas recomendado na matriz curricular.

Ao selecionar o nó vermelho em GSI006, são exibidas análises complementares. A Visão de Notas e Frequência revela que os alunos reprovados em GSI006 apresentam médias de notas inferiores à média geral da disciplina (**Figura 42A**). Já na **Figura 42B**, que apresenta a distribuição de faltas, observa-se que o número médio de ausências entre os

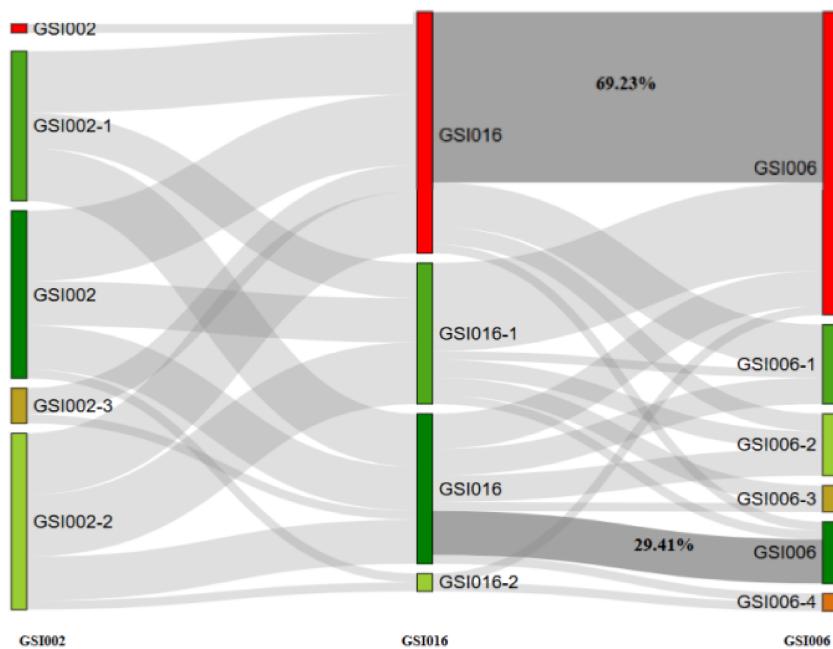


Figura 41 – Exibição da proporção de alunos aprovados em GSI016 e que também foram aprovados em GSI006 (29.41%). Exibição da proporção de alunos reprovados em GSI016 e que foram reprovados diretamente em GSI006 (69.23%).

alunos reprovados é superior à média geral. Os dados, mais uma vez, sugerem que alunos com maiores taxas de frequência tendem a alcançar melhores resultados acadêmicos.

A Visão de Conteúdos (**Figura 43** e **Figura 44**) ilustra que a maioria dos tópicos apresenta distribuições semelhantes entre as categorias de desempenho, com uma predominância da cor vermelha, indicando uma alta concentração de alunos reprovados na disciplina.

Na **Figura 43**, os conteúdos da disciplina GSI006 estão ordenados de forma decrescente em relação à quantidade de alunos na categoria “reprovados”. Observa-se uma predominância da cor vermelha nos *layouts*, o que evidencia de forma clara e direta a concentração de reprovações. Essa característica visual facilita a identificação dos conteúdos mais desafiadores para os alunos. Um aspecto particularmente relevante é que, em alguns tópicos, como os conteúdos 3, 4, 6 e 2, praticamente não há registros de alunos aprovados diretamente. Isso sugere que esses conteúdos podem estar associados a barreiras significativas no processo de aprendizagem, exigindo múltiplas tentativas até que os alunos obtenham aprovação. Ao comparar com a **Figura 38**, percebe-se como os padrões visuais variam entre diferentes disciplinas, refletindo a influência da sequência em que são cursadas. Cada um dos nove primeiros tópicos em GSI006 inclui, individualmente, 33 alunos reprovados. Além disso, ao analisar os três conteúdos com maiores índices de faltas, apresentados na **Figura 44** (tópicos 15, 18 e 13), nota-se que esses tópicos pertencem aos conteúdos com a maior quantidade de alunos reprovados na disciplina. Esses resultados

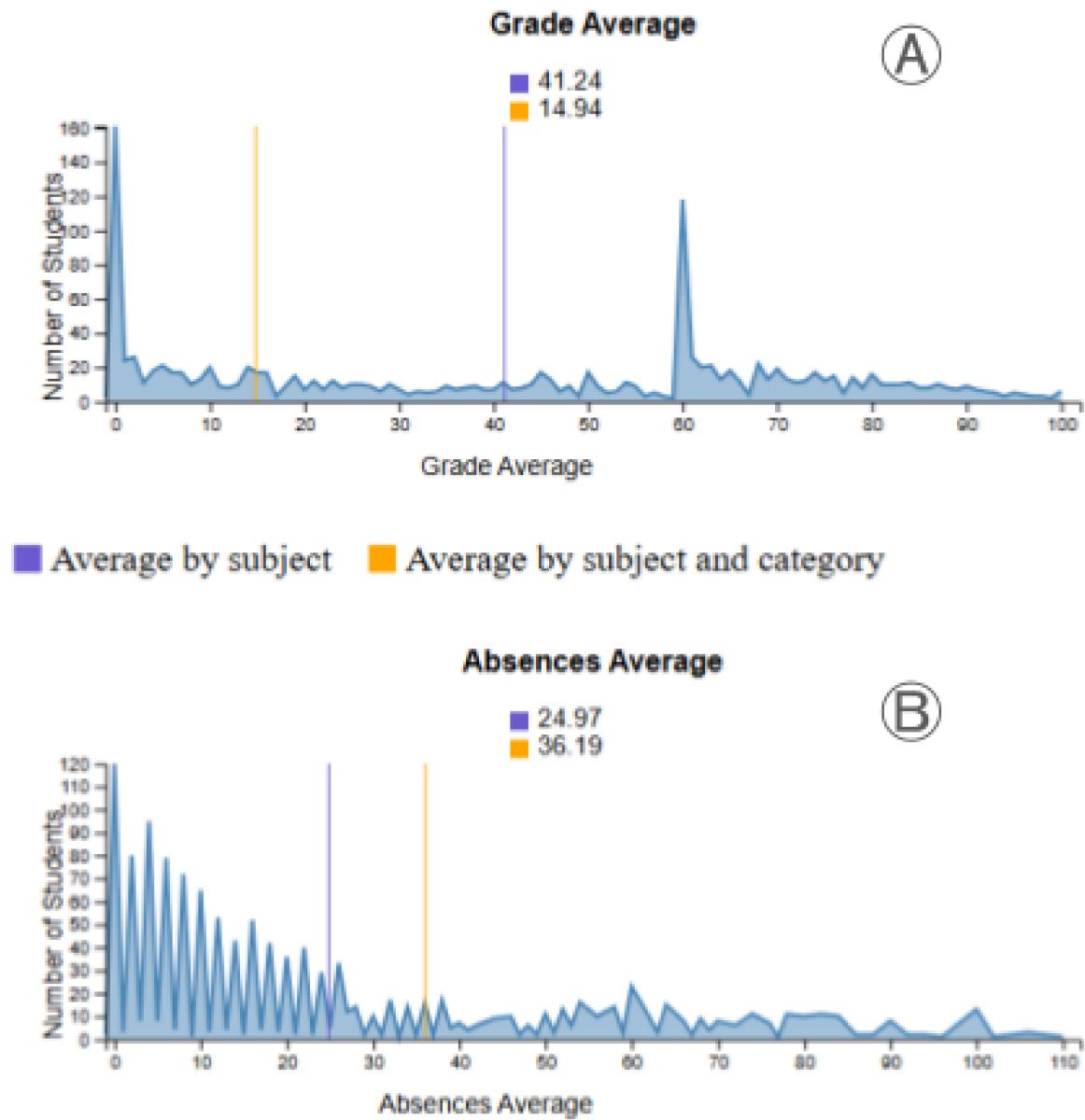


Figura 42 – Visão de Notas e Frequência da disciplina GSI006 com status reprovado, exibindo curvas de distribuição das notas (A) e faltas (B), juntamente com as linhas que representam médias gerais (roxa) e por status (amarela).

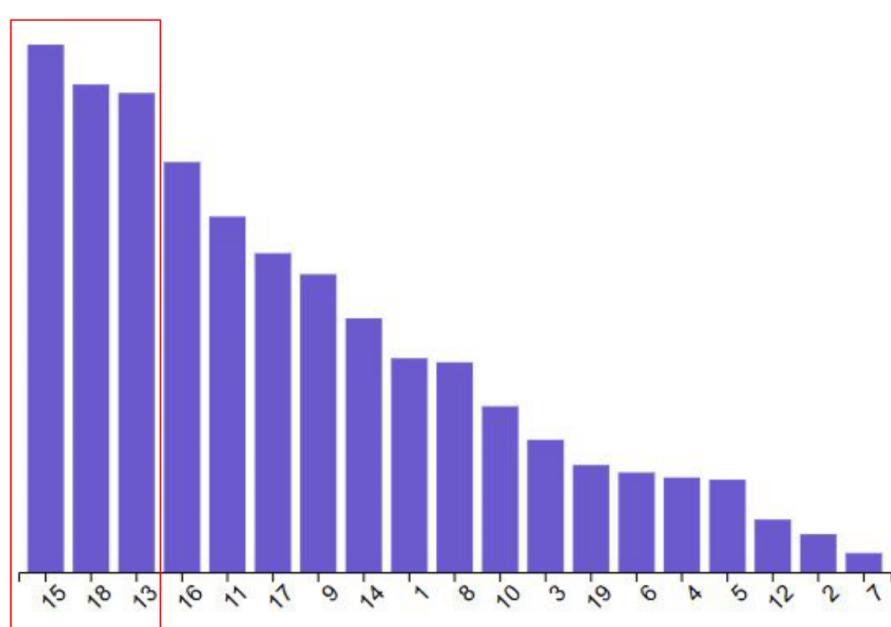
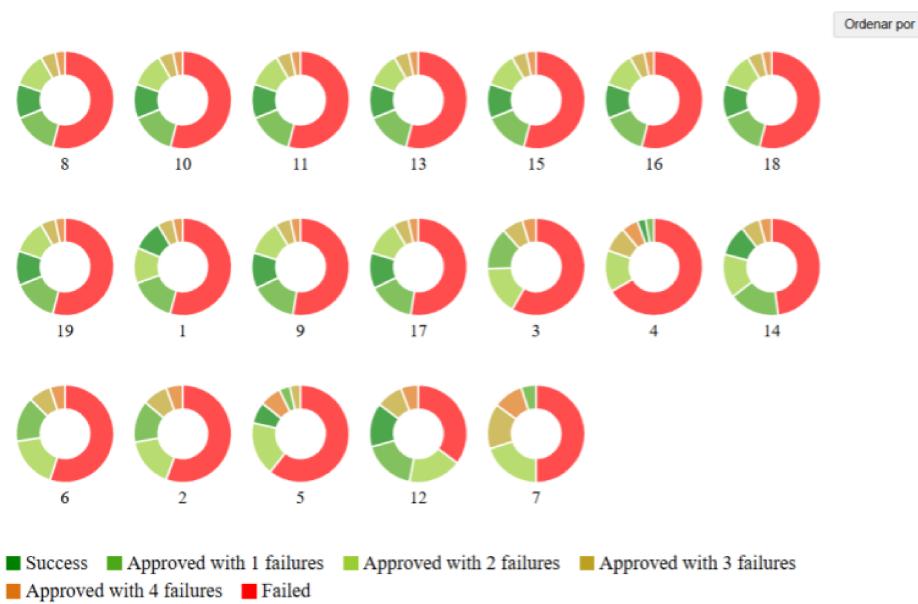


Figura 44 – Visão de Conteúdos da disciplina GSI006: Faltas por Conteúdo.

sugerem que a ausência também contribui de forma significativa para o baixo desempenho observado na Visão de Conteúdos. O *layout* desempenha um papel fundamental ao permitir a visualização desses padrões de maneira intuitiva, facilitando a compreensão das dificuldades enfrentadas pelos alunos. A variação desses padrões entre disciplinas diferentes sugere que a organização curricular e a sequência de conteúdos podem ter um impacto significativo no desempenho acadêmico.

A **Figura 45** apresenta a análise da sequência GSI016 - GSI006 - GSI002, explorando um cenário alternativo em que os alunos optaram por não seguir a ordem recomendada. Embora apenas dois alunos tenham seguido essa sequência, o *layout* ainda pode ser utilizado para sugerir possíveis cenários adversos. O *layout* mostra que o primeiro aluno (destacado em azul), que foi reprovado em GSI016, enfrentou dificuldades semelhantes em GSI006, sendo novamente reprovado. No entanto, esse aluno obteve aprovação direta em GSI002. Por outro lado, o segundo aluno, que foi aprovado em GSI016, necessitou de três tentativas (“GSI006-2”) para ser aprovado em GSI006 e de duas tentativas (“GSI002-1”) para ser aprovado em GSI002, o que sugere que a ausência de uma preparação adequada nas disciplinas introdutórias comprometeu seu desempenho. Apesar dos desafios enfrentados em GSI016 e GSI006, ambos os alunos conseguiram, ao final, obter aprovação em GSI002.

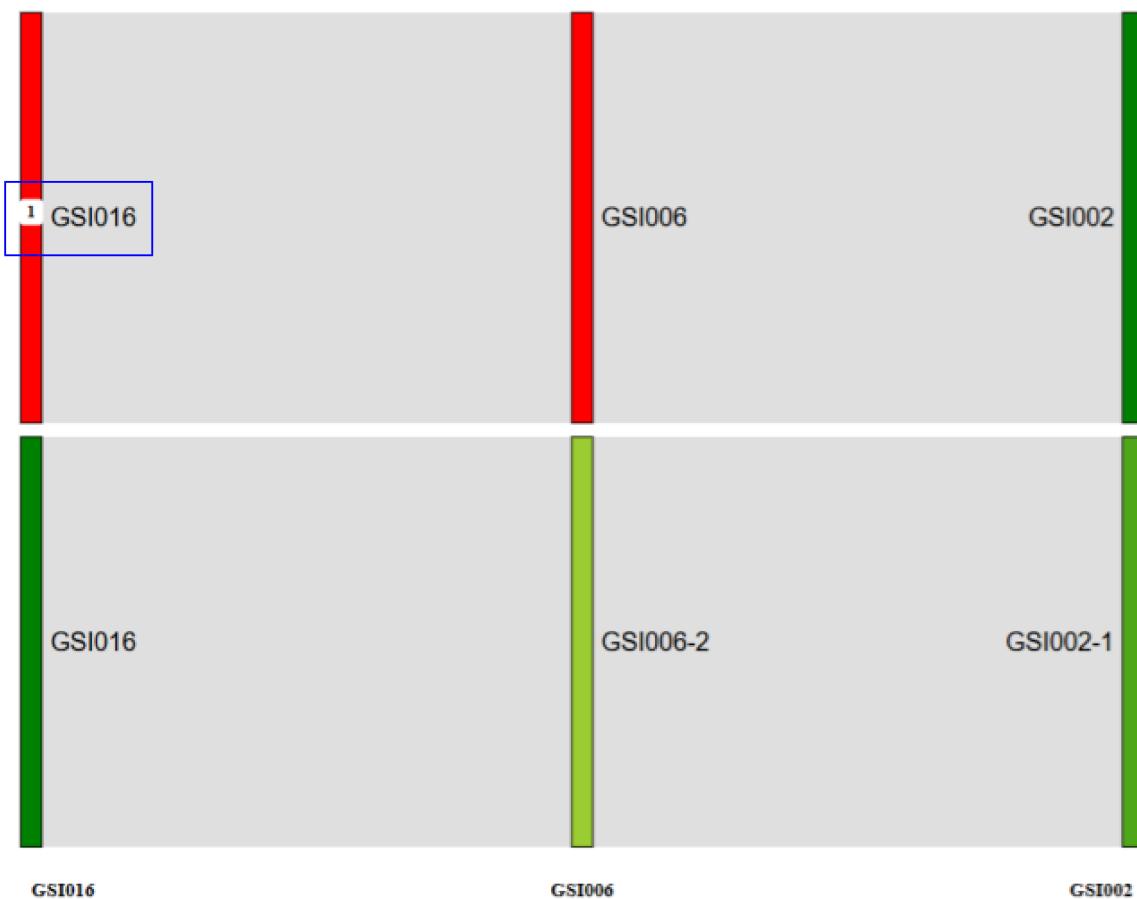


Figura 45 – Visão de Sequência de Execução para o fluxo GSI016 - GSI006 - GSI002.

O *layout* fornece ainda alguns *insights* adicionais: é importante destacar, através dos nós vermelho e verde claro, que nenhum dos dois alunos obteve aprovação na primeira tentativa em GSI006 antes de cursar GSI002, evidenciando a relevância de um accompa-

nhamento mais rigoroso na ordem das disciplinas. Embora a análise seja limitada pelo pequeno número de alunos, ela fornece um panorama inicial que reforça a importância de manter a sequência estipulada pela coordenação, muitas vezes planejada para promover maior preparo e sucesso acadêmico.

A **Figura 46** apresenta a análise da sequência GSI006 - GSI002 - GSI016, avaliando o terceiro caso com cinco alunos que optaram por não seguir a sequência recomendada. Observa-se que nenhum aluno conseguiu aprovação em GSI006 na primeira tentativa antes de cursar GSI002. Em relação a GSI016, os dados mostram que 40% dos alunos alcançaram aprovação na primeira tentativa, enquanto 20% obtiveram êxito na segunda tentativa. No entanto, os demais 40% enfrentaram reprovações, conforme a aresta destacada.

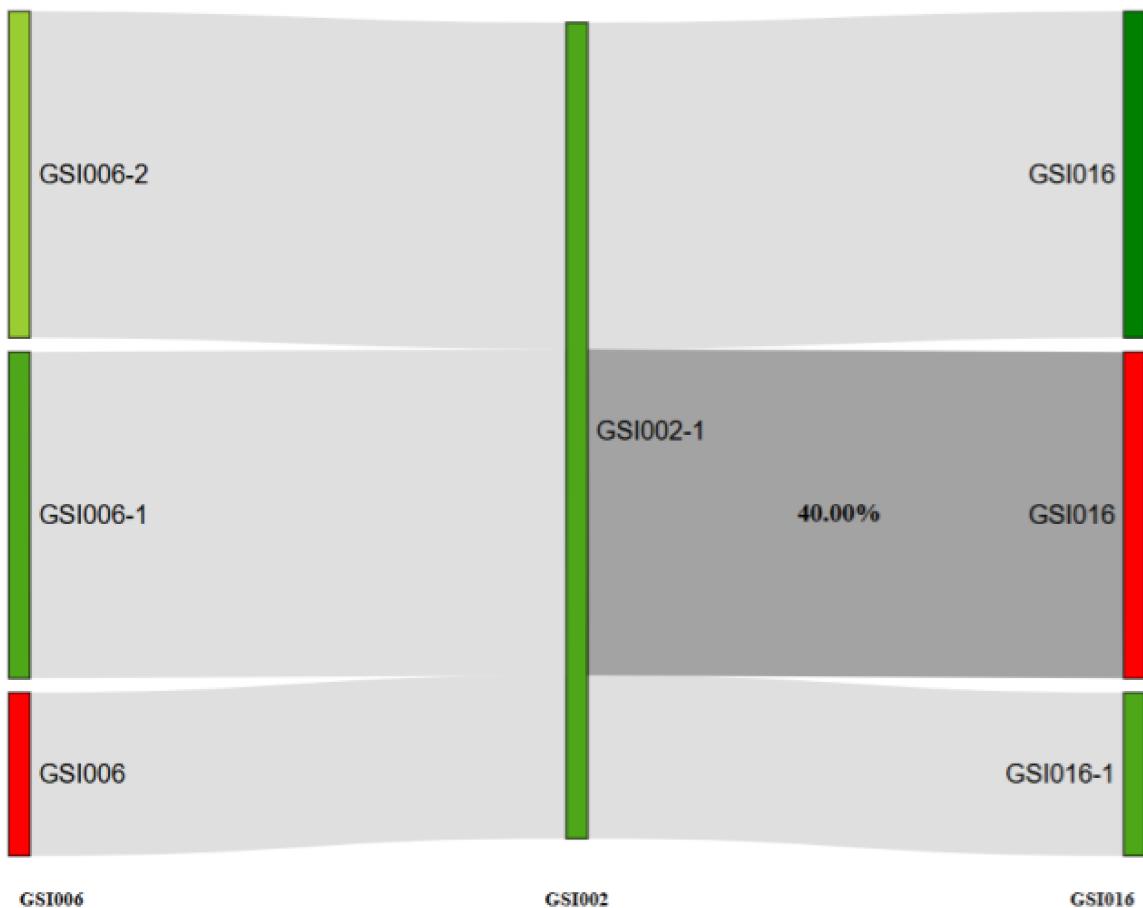


Figura 46 – Visão de Sequência de Execução para o fluxo GSI006 - GSI002 - GSI016.

Ao clicar no nó correspondente à proporção de alunos reprovados em GSI016 (**Figura 46**), o sistema disponibiliza uma análise detalhada da disciplina. A Visão de Notas e Frequência da disciplina GSI016 é ilustrada na **Figura 47**. Na **Figura 47A**, a média de notas dos alunos reprovados é substancialmente inferior à média geral. Ambas as médias são

inferiores ao mínimo exigido para aprovação. Já na **Figura 47B**, a média de faltas dos alunos reprovados é consideravelmente maior do que a média geral. Esse padrão sugere novamente que a assiduidade é um fator determinante para o aprendizado e desempenho acadêmico, indicando que alunos com maior presença tendem a obter melhores resultados.

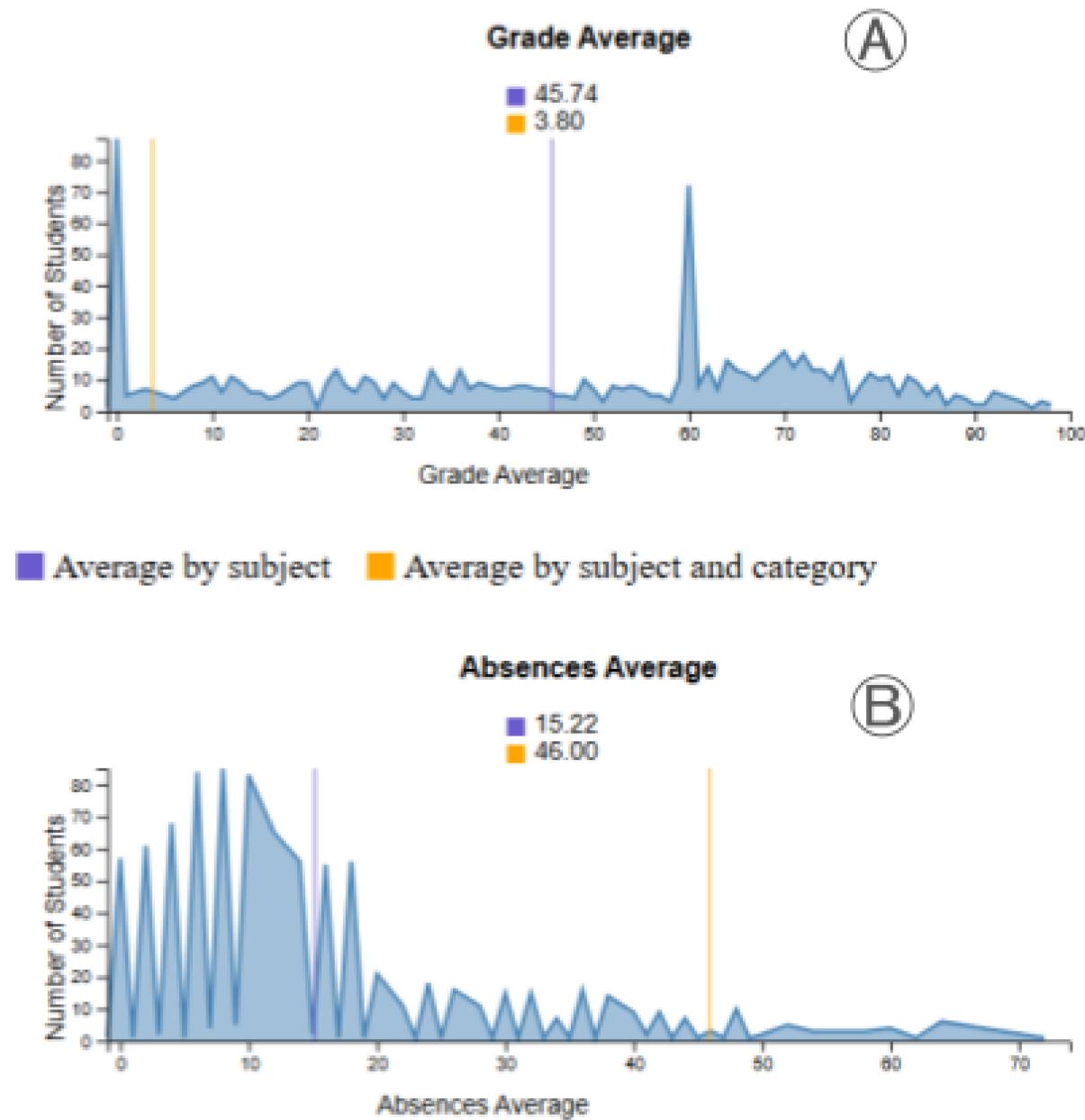


Figura 47 – Visão de Notas e Frequência da disciplina GSI016 com status reprovado, exibindo curvas de distribuição das notas (A) e faltas (B), juntamente com as linhas que representam médias gerais (roxa) e por status (amarela).

5.4 Discussão dos Resultados

A aplicação do EduFlow aos dados do curso de Sistemas de Informação da UFU permitiu identificar padrões relevantes no desempenho acadêmico dos alunos, revelando o potencial dos *layouts* visuais como ferramentas de apoio à análise educacional. De maneira geral, os *layouts* facilitaram a detecção de correlações entre a sequência de disciplinas, a assiduidade e os resultados obtidos, permitindo interpretações mais aprofundadas do que as viabilizadas por análises estatísticas convencionais.

A Visão de Sequência de Execução fornece pistas importantes sobre o impacto da ordem em que os componentes curriculares são cursados. Por meio da representação das transições entre disciplinas, foi possível verificar que alunos que seguem a sequência recomendada tendem a apresentar menores taxas de reprovação, reforçando a importância do alinhamento pedagógico proposto pela matriz curricular. Os *layouts* permitiram observar diretamente que cursar disciplinas avançadas antes das introdutórias está frequentemente associado a um desempenho acadêmico inferior.

A Visão de Notas e Frequência evidencia que alunos com maiores taxas de presença tendem a obter médias superiores, sugerindo que a assiduidade é um fator determinante para o sucesso acadêmico. Esse padrão foi observado de maneira consistente em diferentes disciplinas e contextos analisados, indicando que políticas institucionais que promovam o engajamento em sala de aula podem contribuir para a redução das reprovações.

Além disso, os *layouts* da Visão de Conteúdos revelaram tópicos específicos que concentram um número elevado de reprovações, mesmo entre alunos que estavam presentes nas respectivas aulas. Essa pista visual indicou que certos conteúdos podem representar maiores desafios conceituais e, portanto, demandar abordagens didáticas diferenciadas. A ordenação por número de reprovações ou por faltas permitiu destacar esses casos de maneira imediata e acessível.

A Visão de Estágio sugere uma possível relação entre a realização de estágio e a redução no desempenho acadêmico, uma vez que a maioria dos alunos aprovados diretamente não realizou estágio, enquanto há uma concentração significativa de alunos com reprovações ou aprovações tardias entre aqueles que estagiaram. Esse padrão sugere que o envolvimento em estágio, embora enriquecedor em termos de experiência prática, pode estar associado a dificuldades no desempenho acadêmico, possivelmente devido a desafios relacionados à gestão de tempo ou à sobrecarga de atividades. No entanto, é importante destacar que também existe uma proporção de alunos que conseguiram conciliar a realização de estágios com um bom desempenho acadêmico, obtendo aprovação na disciplina, indicando que, embora desafiadora, essa conciliação é viável em alguns casos.

Em síntese, os *layouts* fornecidos pelo EduFlow se mostraram eficazes em destacar padrões críticos de desempenho e em orientar investigações mais específicas, consolidando seu papel como uma ferramenta promissora para a análise visual de dados educacionais.

CAPÍTULO 6

Conclusão e Trabalhos Futuros

Esta dissertação apresentou o EduFlow, um sistema interativo de análise visual do desempenho acadêmico de alunos universitários. O sistema foi desenvolvido com o objetivo de permitir a exploração das trajetórias acadêmicas e desempenho dos alunos ao longo de diferentes disciplinas e períodos letivos. Para isso, o EduFlow integra quatro visualizações coordenadas: a Visão de Sequência de Execução, que mostra o encadeamento das disciplinas cursadas e o impacto dessas trajetórias em termos de aprovação e reprovação; a Visão de Notas e Frequência, que detalha a distribuição estatística das notas e presenças por categoria de desempenho; a Visão de Conteúdos, que relaciona tópicos abordados nas aulas com o desempenho dos alunos; e a Visão de Estágios, que permite correlacionar experiências extracurriculares com os resultados acadêmicos. A coordenação entre os *layouts* permite ao usuário filtrar e explorar diferentes aspectos do desempenho estudantil de forma interativa e integrada.

Os *layouts* desenvolvidos se mostraram eficazes na identificação de padrões relevantes. A Visão de Sequência de Execução evidenciou possíveis impactos da ordem em que as disciplinas são cursadas nas taxas de aprovação, sugerindo que desvios da sequência recomendada podem aumentar a reprovação. A Visão de Notas e Frequência revelou que alunos com maior assiduidade tendem a apresentar melhores notas, enquanto os reprovados acumulam mais faltas e têm médias significativamente abaixo do mínimo exigido para aprovação. Já a Visão de Conteúdos permitiu identificar tópicos específicos que concentram um número elevado de alunos reprovados mesmo comparecendo regularmente às respectivas aulas, o que levanta hipóteses sobre dificuldades conceituais ou metodológicas relacionadas a esses conteúdos. A Visão de Estágio permitiu analisar a relação entre a prática de estágio e o desempenho nas disciplinas, sugerindo inclusive que a realização do estágio pode estar associado a um desempenho inferior dos alunos, provavelmente devido a uma possível sobrecarga ou dificuldade na gestão do tempo. Ainda assim, uma parcela dos alunos conseguiu conciliar as duas atividades com sucesso, o que demonstra que, embora desafiador, o equilíbrio entre estágio e desempenho acadêmico é viável. A coordenação entre as visualizações possibilitou uma análise complementar e multifacetada,

ampliando a capacidade de interpretação dos dados e auxiliando na tomada de decisão pedagógica e curricular.

A análise realizada pelo sistema depende de conjuntos de dados bem estruturados e completos, o que significa que dados ausentes ou inconsistentes podem afetar a interpretação dos *insights* gerados. A precisão da análise está diretamente relacionada à qualidade e completude dos dados de entrada. Dessa forma, quaisquer falhas na coleta ou estruturação dos dados serão refletidas nas visualizações, impactando os resultados apresentados. De fato, o sistema pode ser utilizado para evidenciar essas inconsistências, tornando-se uma ferramenta útil para identificar lacunas, erros ou informações ausentes em bases de dados acadêmicas. A escalabilidade também pode representar uma limitação potencial do EduFlow, especialmente ao lidar com conjuntos de dados maiores que envolvam um número elevado de disciplinas e interdependências. Embora o sistema seja capaz de processar esses dados, a Visão de Sequência de Execução pode tornar-se de difícil navegação quando muitas disciplinas são selecionadas simultaneamente, o que pode comprometer a capacidade do usuário de interpretar as transições e os padrões de desempenho.

Como trabalhos futuros, destacam-se as seguintes frentes de pesquisa e desenvolvimento:

- **Melhoria da usabilidade em visualizações densas:** a incorporação de mecanismos interativos avançados, como técnicas de *zoom*, foco e filtros progressivos, pode facilitar a navegação e reduzir a sobrecarga visual nas visualizações com muitos elementos, permitindo uma exploração mais criteriosa dos dados;
- **Aplicação do sistema em outros contextos curriculares:** A validação do EduFlow em cursos de diferentes áreas do conhecimento e instituições permite avaliar sua adaptabilidade a outras estruturas curriculares, bem como sua eficácia em contextos educacionais diversos. Essa aplicação é importante pois pode revelar novas necessidades analíticas e promover melhorias contínuas no sistema;
- **Integração com ambientes de aprendizagem virtual:** Com o avanço do ensino remoto e híbrido, é relevante explorar a integração do EduFlow com sistemas de gestão de aprendizagem (LMS), ampliando a análise para além das disciplinas presenciais, incorporando dados de atividades *on-line* e interações em ambientes virtuais.

Essas direções futuras visam não apenas ampliar o escopo e a aplicabilidade do EduFlow, mas também consolidá-lo como uma ferramenta robusta de apoio à análise educacional baseada em dados.

Referências

- ALMALAWI, A. et al. Educational Data Mining for Student Performance Prediction: A Systematic Literature Review. **AI**, MDPI, v. 8, n. 12, p. 187, 2024. DOI: [10.3390/ai8120187](https://doi.org/10.3390/ai8120187). Disponível em: <<https://doi.org/10.3390/ai8120187>>.
- ARNOLD, K. E.; PISTILLI, M. D. Course signals at Purdue: Using learning analytics to increase student success. In: PROCEEDINGS of International Conference on Learning Analytics and Knowledge. [S.l.: s.n.], 2012. P. 267–270.
- BARADWAJ, B.; PAL, S. Mining educational data to analyze student's performance. In: 2. v. 6, p. 63–69.
- BAREFOOT, B. O. et al. **Achieving and Sustaining Institutional Excellence for the First Year of College**. [S.l.]: John Wiley & Sons, 2010.
- BEAN, J. P. **Increasing student retention: Effective programs and practices for reducing the dropout rate**. [S.l.: s.n.], 1987.
- CARD, S. K.; MACKINLAY, J. D.; SHNEIDERMAN, B. Information Visualization. In: READINGS in Information Visualization - Using Visualization to Think. San Francisco: Morgan Kaufmann Publ., 1999. P. 1–34.
- DENG, H. et al. PerformanceVis: Visual analytics of student performance data from an introductory chemistry course. **Vis. Informatics**, v. 3, p. 166–176, 2019.
- EMMONS, S. R.; LIGHT, R. P.; BÖRNER, K. MOOC visual analytics: Empowering students, teachers, researchers, and platform developers of massively open online courses. **Journal of the Association for Information Science and Technology**, v. 68, n. 10, p. 2350–2363, 2017.
- ETEMADPOUR, Ronak et al. Academic Performance Analysis Supported by a Web-Based Visual Analytics Tool. In: PROCEEDINGS of the 19th IEEE International Conference on Advanced Learning Technologies (ICALT). Maceió, Brazil: IEEE, 2019. P. 158. DOI: [10.1109/ICALT.2019.00000](https://doi.org/10.1109/ICALT.2019.00000).

- FREITAS, C. M. D. S. et al. Introdução à Visualização de Informações. **Revista de Informática Teórica e Aplicada**, Porto Alegre, RS, v. 8, n. 2, p. 143–158, 2001.
- GARCIA-ZANABRIA, G. et al. SDA-Vis: A Visualization System for Student Dropout Analysis Based on Counterfactual Exploration. **Applied Sciences**, v. 12, p. 5785, 2022. DOI: 10.3390/app12125785.
- GERSHON, N.; EICK, S. G. Information Visualization. **IEEE Computer Graphics and Applications**, v. 17, n. 4, p. 29–31, jul. 1997. DOI: 10.1109/MCG.1997.595265.
- GONZÁLEZ-MORALES, Oscar et al. Dropping out of higher education: Analysis of variables that explain it. **Personality and Individual Differences**, v. 219, p. 112241, 2025. DOI: 10.1016/j.paid.2024.112241. Disponível em: <<https://doi.org/10.1016/j.paid.2024.112241>>.
- GOULDEN, M. C. et al. CCVis: Visual Analytics of Student Online Learning Behaviors Using Course Clickstream Data. **Visualization and Data Analysis**, 2019.
- GUTIERREZ-PACHAS, Daniel A. et al. How Do Curricular Design Changes Impact Computer Science Programs?: A Case Study at San Pablo Catholic University in Peru. **Education Sciences**, v. 12, n. 4, p. 242, 2022. DOI: 10.3390/educsci12040242. Disponível em: <<https://doi.org/10.3390/educsci12040242>>.
- EL-HALEES, Alaa. **Mining students data to analyze e-Learning behavior: A Case Study**. [S.l.], 2009.
- HANSEN, L.; BORGES, V. R. P.; HOLANDA, M. A Literature Study of Visual Analysis in an Educational Context. In: 2020 IEEE Frontiers in Education Conference (FIE). [S.l.: s.n.], 2020. P. 1–8. DOI: 10.1109/FIE44824.2020.9274178.
- HEER, J.; BOSTOCK, M.; OGIEVETSKY, V. A Tour Through the Visualization Zoo. **Communications of the ACM**, ACM Press, v. 5, n. 6, 2010. DOI: 10.1145/1743546.1743567.
- INSELBERG, A.; DIMSDALE, B. Parallel coordinates: a tool for visualizing multidimensional geometry. In: PROCEEDINGS of the 1st conference on Visualization'90. [S.l.]: IEEE Computer Society Press, 1990. P. 361–378.
- KEIM, D. Information visualization and visual data mining. **IEEE Transactions on Visualization and Computer Graphics**, IEEE, v. 8, n. 1, p. 1–8, 2002. DOI: 10.1109/TVCG.2002.938985.
- KOUROUNIOTIS, S. Seeing Data: Using Visualization to Reveal Insights and Make Decisions. **EDUCAUSE Review**, EDUCAUSE, 2023. Disponível em: <<https://er.educause.edu/articles/2023/2/seeing-data-using-visualization-to-reveal-insights-and-make-decisions>>.

- KUI, Xiaoyan et al. A Survey of Visual Analytics Techniques for Online Education. **Visual Informatics**, Elsevier, v. 6, p. 67–77, 2022. DOI: 10.1016/j.visinf.2022.07.004.
- LIM, L. et al. Exploring students' sensemaking of learning analytics dashboards: does frame of reference make a difference? In: PROCEEDINGS of International Conference on Learning Analytics and Knowledge. [S.l.: s.n.], 2019. P. 250–259.
- MORENO-MARCOS, P. M. et al. Analysing the predictive power for anticipating assignment grades in a massive open online course. **Behav. Inf. Technol.**, v. 38, 10–11, p. 1021–1036, 2018.
- NASCIMENTO, H. A.; FERREIRA, C. B. Uma introdução à visualização de informações. **Visualidades**, v. 9, n. 2, 2011.
- OKUBO, F. et al. A neural network approach for students' performance prediction. In: PROCEEDINGS of International Conference on Learning Analytics and Knowledge. [S.l.: s.n.], 2017. P. 598–599.
- OLSSON, Marie; MOZELIUS, Peter; COLLIN, Jonas. Visualisation and Gamification of e-Learning and Programming Education. **Electronic Journal of e-Learning**, v. 13, n. 6, p. 441–454, 2015. Disponível em: <<https://files.eric.ed.gov/fulltext/EJ1087309.pdf>>.
- POON, L. K. et al. Learning Analytics for Monitoring Students' Participation Online: Visualizing Navigational Patterns on Learning Management System. In: PROCEEDINGS of the International Conference on Blended Learning. [S.l.]: Springer, Cham, 2017. P. 166–176.
- RAJI, M. et al. Visual progression analysis of student records data. In: 2017 IEEE Visualization in Data Science (VDS). Phoenix, AZ: [s.n.], 2017. P. 31–38.
- RIEHMANN, Patrick; HANFLER, Michael; FROEHLICH, Bernd. Interactive Sankey diagrams. In: IEEE Symposium on Information Visualization. [S.l.]: IEEE, 2005. P. 233–240.
- SANTANA, A. P. et al. **Evasão escolar em escolas públicas municipais rurais localizadas em Montes Claros**. 1996. F. 20. Undergraduate Thesis – Universidade Estadual de Montes Claros, Montes Claros, Brazil.
- TSUNG, S. et al. BlockLens: Visual Analytics of Student Coding Behaviors in Block-Based Programming Environments. In: PROCEEDINGS of the Ninth ACM Conference on Learning @ Scale (L@S '22). [S.l.: s.n.], 2022. P. 1–5. DOI: 10.1145/3491140.3528298.
- WARE, C. **Information Visualization, Perception for Design**. 2nd. [S.l.]: Morgan Kaufmann Publishers, 2004.

ZHANG, Huijie et al. Visual analytics of potential dropout behavior patterns in online learning based on counterfactual explanation. **Journal of Visualization**, The Visualization Society of Japan, v. 26, p. 723–741, 2023. DOI: [10.1007/s12650-022-00899-8](https://doi.org/10.1007/s12650-022-00899-8).