



UNIVERSIDADE FEDERAL DO CEARÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO DO CAMPUS DE QUIXADÁ
MESTRADO ACADÊMICO EM COMPUTAÇÃO

ANTONIO GILDO DE FREITAS FILHO

ANÁLISE DA TRAJETÓRIA ACADÊMICA ESTUDANTIL UTILIZANDO
MINERAÇÃO DE PROCESSOS COM ÊNFASE NA EVASÃO DO CURSO DE CIÊNCIA
DA COMPUTAÇÃO DA UFC EM QUIXADÁ

QUIXADÁ

2025

ANTONIO GILDO DE FREITAS FILHO

ANÁLISE DA TRAJETÓRIA ACADÊMICA ESTUDANTIL UTILIZANDO MINERAÇÃO
DE PROCESSOS COM ÊNFASE NA EVASÃO DO CURSO DE CIÊNCIA DA
COMPUTAÇÃO DA UFC EM QUIXADÁ

Dissertação apresentada ao Curso de Mestrado Acadêmico em Computação do Programa de Pós-Graduação em Computação do Campus de Quixadá da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Computação. Área de Concentração: Algoritmos e Teoria da Computação.

Orientador: Prof. Dr. Davi Romero de Vasconcelos.

Coorientador: Prof. Dr. Paulo de Tarso Guerra Oliveira.

QUIXADÁ

2025

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

- F936a Freitas Filho, Antonio Gildo de.
Análise da trajetória acadêmica estudantil utilizando mineração de processos com ênfase na evasão do curso de Ciência da Computação da UFC em Quixadá / Antonio Gildo de Freitas Filho. – 2025.
107 f. : il. color.
- Dissertação (mestrado) – Universidade Federal do Ceará, Campus de Quixadá, Programa de Pós-Graduação em Computação, Quixadá, 2025.
Orientação: Prof. Dr. Davi Romero de Vasconcelos.
Coorientação: Prof. Dr. Paulo de Tarso Guerra.
1. Evasão. 2. Mineração de Processos. 3. Análise de Conformidade. 4. Redes de Petri. 5. Educação Superior. I. Título.

CDD 005

ANTONIO GILDO DE FREITAS FILHO

ANÁLISE DA TRAJETÓRIA ACADÊMICA ESTUDANTIL UTILIZANDO MINERAÇÃO
DE PROCESSOS COM ÊNFASE NA EVASÃO DO CURSO DE CIÊNCIA DA
COMPUTAÇÃO DA UFC EM QUIXADÁ

Dissertação apresentada ao Curso de Mestrado Acadêmico em Computação do Programa de Pós-Graduação em Computação do Campus de Quixadá da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Computação. Área de Concentração: Algoritmos e Teoria da Computação.

Aprovada em: 30/04/2025.

BANCA EXAMINADORA

Prof. Dr. Davi Romero de Vasconcelos (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Paulo de Tarso Guerra
Oliveira (Coorientador)
Universidade Coorientador (UFC)

Prof. Dr. Edward Hermann Haeusler
Pontifícia Universidade Católica do Rio de Janeiro
(PUC-RIO)

Prof. Dr. Criston Pereira de Souza
Universidade Federal do Ceará (UFC)

AGRADECIMENTOS

A Deus, por me conceder saúde, força e sabedoria para chegar até aqui. Pela presença constante, mesmo nos momentos de incerteza, e por iluminar meu caminho com fé e esperança.

Aos meus pais, Francisca Lucia da Costa de Freitas e Antonio Gildo de Freitas, por todo amor, dedicação e incentivo incondicional. Vocês são meu alicerce e inspiração. Tudo o que sou e conquistei tem raízes no esforço e exemplo que recebi de vocês.

À minha família, pelo apoio contínuo, pelas palavras de encorajamento e por compreenderem minha ausência em tantos momentos importantes. Cada gesto de carinho foi essencial nessa trajetória.

Ao meu orientador, Prof. Dr. Davi Romero de Vasconcelos, por sua orientação firme, generosa e precisa. Agradeço pela confiança depositada em mim, pela disponibilidade e por me conduzir com seriedade e respeito ao longo deste percurso acadêmico.

Ao meu coorientador, Prof. Dr. Paulo de Tarso Guerra Oliveira, por suas contribuições valiosas, pela escuta atenta e pelas observações criteriosas que enriqueceram este trabalho.

A todas as pessoas que, direta ou indiretamente, contribuíram para a realização desta dissertação, deixo aqui minha mais sincera gratidão.

RESUMO

A gestão das instituições de ensino superior enfrenta diversos desafios, sendo a evasão estudantil um dos mais complexos e preocupantes. O abandono dos cursos por parte dos discentes acarreta prejuízos individuais, sociais e institucionais. Nesse contexto, a aplicação de técnicas de Mineração de Processos (*Process Mining*) surge como uma ferramenta analítica promissora para compreender as trajetórias acadêmicas que levam à evasão. Este trabalho propõe a modelagem da estrutura curricular do curso de Ciência da Computação da Universidade Federal do Ceará (UFC), Campus de Quixadá, utilizando *logs* de eventos extraídos dos históricos acadêmicos dos estudantes. Os dados incluem informações sobre componentes curriculares cursados, *status* (matriculado, aprovado, reprovado, reprovado por falta, trancado, suprimido ou trancamento total), identificação do discente anonimizado, ano e semestre letivo. A estrutura curricular é formalizada por meio de redes de Petri, permitindo representar a progressão acadêmica e simular diferentes trajetórias. Posteriormente, aplicam-se técnicas de verificação de conformidade com a biblioteca *PM4PY*, como *token replay* e alinhamento (*alignments*), para avaliar o grau de aderência entre os percursos reais e o modelo ideal, identificar desvios, gargalos e padrões de retenção. A análise revelou componentes curriculares e períodos letivos com maior incidência de reprovações e desistências, evidenciando pontos críticos que impactam a permanência dos estudantes. Os resultados indicam que a modelagem baseada em mineração de processos pode subsidiar a tomada de decisões acadêmicas, contribuindo para a redução da evasão e o aprimoramento das políticas institucionais de acompanhamento e apoio ao discente.

Palavras-chave: evasão; mineração de processos; análise de conformidade; redes de petri; educação superior.

ABSTRACT

Higher education institutions face several challenges, among which student dropout is one of the most complex and concerning. Course abandonment by students leads to individual, social, and institutional losses. In this context, the application of Process Mining techniques emerges as a promising analytical tool to understand the academic trajectories that lead to dropout. This study proposes the modeling of the curriculum structure of the Computer Science program at the Federal University of Ceará (UFC), Campus of Quixadá, using event logs extracted from students' academic records. The dataset includes information about the courses taken, their status (enrolled, approved, failed, failed due to absence, withdrawal, canceled, or full withdrawal), anonymized student IDs, and academic terms. The curriculum structure is formalized through Petri nets, enabling the representation of academic progression and the simulation of different trajectories. Subsequently, conformance checking techniques, such as token replay and alignments, are applied using the PM4PY library to assess the alignment between real academic paths and the ideal model, identifying deviations, bottlenecks, and retention patterns. The analysis revealed specific courses and semesters with high rates of failure and dropout, highlighting critical points that affect student retention. The results indicate that curriculum modeling based on process mining can support academic decision-making, contributing to reducing dropout rates and improving institutional policies for student support and retention.

Keywords: dropout; process mining; conformance checking; petri nets; higher education.

LISTA DE FIGURAS

Figura 1 – Coortes de 2014, 2015 e 2016 do curso de Ciência da Computação no Brasil	19
Figura 2 – Coortes de 2014, 2015 e 2016 do curso de Ciência da Computação na UFC- Quixadá	20
Figura 3 – Play-In, Play-Out e Replay	23
Figura 4 – Elementos básicos de uma rede de Petri	25
Figura 5 – Exemplo de uma rede de Petri e seus elementos	27
Figura 6 – Exemplo de uma rede de Petri	29
Figura 7 – Exemplo de uma rede de Petri	29
Figura 8 – Exemplo de uma rede de Petri	30
Figura 9 – Exemplo de uma rede de Petri	30
Figura 10 – Exemplo de uma modelagem de processo	34
Figura 11 – Passo um: <i>Replay</i> do trace $\sigma_1 = \langle a, c, d, e, h \rangle$ em modelo PN_1	37
Figura 12 – Passo dois: <i>Replay</i> do trace $\sigma_1 = \langle \emptyset, c, d, e, h \rangle$ em modelo PN_1	38
Figura 13 – Passo três: <i>Replay</i> do trace $\sigma_1 = \langle \emptyset, \emptyset, d, e, h \rangle$ em modelo PN_1	38
Figura 14 – Passo quatro: <i>Replay</i> do trace $\sigma_1 = \langle \emptyset, \emptyset, \emptyset, e, h \rangle$ em modelo PN_1	38
Figura 15 – Passo cinco: <i>Replay</i> do trace $\sigma_1 = \langle \emptyset, \emptyset, \emptyset, \emptyset, h \rangle$ em modelo PN_1	39
Figura 16 – Passo seis: <i>Replay</i> do trace $\sigma_1 = \langle \emptyset, \emptyset, \emptyset, \emptyset, \emptyset \rangle$ em modelo PN_1	39
Figura 17 – Passo um: <i>Replay</i> do trace $\sigma_2 = \langle a, d, c, e, h \rangle$ em modelo PN_2	41
Figura 18 – Passo um: <i>Replay</i> do trace $\sigma_2 = \langle \emptyset, d, c, e, h \rangle$ em modelo PN_2	41
Figura 19 – Passo um: <i>Replay</i> do trace $\sigma_2 = \langle \emptyset, \emptyset, c, e, h \rangle$ em modelo PN_2	41
Figura 20 – Passo um: <i>Replay</i> do trace $\sigma_2 = \langle \emptyset, \emptyset, \emptyset, e, h \rangle$ em modelo PN_2	42
Figura 21 – Passo um: <i>Replay</i> do trace $\sigma_2 = \langle \emptyset, \emptyset, \emptyset, \emptyset, h \rangle$ em modelo PN_2	42
Figura 22 – Passo um: <i>Replay</i> do trace $\sigma_2 = \langle \emptyset, \emptyset, \emptyset, \emptyset, \emptyset \rangle$ em modelo PN_2	43
Figura 23 – Alinhamento do trace $\sigma = \langle a, c, d, e, h \rangle$	44
Figura 24 – Alinhamento do trace $\sigma = \langle a, c, d, e, h \rangle$	44
Figura 25 – Estrutura curricular do curso de Ciência da Computação da UFC do Campus de Quixadá.	61
Figura 26 – Transição dos semestres.	62
Figura 27 – Modelagem das disciplinas	62
Figura 28 – Modelagem dos estado de aprovação ou não aprovação nos componentes curriculares.	63

Figura 29 – Modelagem do pré-requisito das disciplinas.	64
Figura 30 – Modelo reduzido da estrutura curricular do curso de Ciência da Computação	64
Figura 31 – Modelo completo da estrutura curricular do curso de Ciência da Computação	65
Figura 32 – Modelo reduzido da estrutura curricular para simulação	66
Figura 33 – Modelo reduzido da estrutura curricular para simulação	68
Figura 34 – Modelo reduzido da estrutura curricular para simulação	68
Figura 35 – Modelo reduzido da estrutura curricular para simulação	69
Figura 36 – Modelo reduzido da estrutura curricular para simulação	69
Figura 37 – Modelo reduzido da estrutura curricular para simulação	69
Figura 38 – Modelo reduzido da estrutura curricular para simulação	70
Figura 39 – Modelo reduzido da estrutura curricular para simulação	70
Figura 40 – Modelo reduzido da estrutura curricular para simulação	71
Figura 41 – Modelo reduzido da estrutura curricular para simulação	72
Figura 42 – Modelo reduzido da estrutura curricular para simulação	73
Figura 43 – Modelo reduzido da estrutura curricular para simulação	73
Figura 44 – Modelo reduzido da estrutura curricular para simulação	74
Figura 45 – Modelo reduzido da estrutura curricular para simulação	75
Figura 46 – Modelo reduzido da estrutura curricular para simulação	75
Figura 47 – Modelo reduzido da estrutura curricular para simulação	76
Figura 48 – Modelo reduzido da estrutura curricular para simulação	77
Figura 49 – Modelo reduzido da estrutura curricular para simulação	78
Figura 50 – Modelo reduzido da estrutura curricular para simulação	78
Figura 51 – Modelo reduzido da estrutura curricular para simulação	79
Figura 52 – Modelo reduzido da estrutura curricular para simulação	79
Figura 53 – Modelo reduzido da estrutura curricular para simulação	79
Figura 54 – Modelo reduzido da estrutura curricular para simulação	80
Figura 55 – Dispersão do <i>trace fitness</i> dos discentes	85
Figura 56 – Frequência das transições por disciplinas	86
Figura 57 – Distribuição de <i>tokens</i> remanescentes por disciplinas	87
Figura 58 – <i>Tokens</i> remanescentes ao final do Semestre	89
Figura 59 – Custo de alinhamento dos discentes formados	92
Figura 60 – <i>Fitness</i> de alinhamento dos discentes formados	93

Figura 61 – Custo e <i>Fitness</i> de alinhamento dos discentes não formados	94
Figura 62 – Frequência de componentes curriculares cursados fora do semestre ideal . .	95
Figura 63 – Frequência de componentes curriculares cursadas antes do semestre ideal .	96
Figura 64 – Matriz curricular dos discentes evadidos	97
Figura 65 – Estrutura curricular do curso de Ciência da Computação da UFC do Campus de Quixadá.	106

LISTA DE TABELAS

Tabela 1 – Um exemplo de um <i>log</i> de eventos, descrevendo o comportamento relacionado ao processo de solicitação de compensação para ingresso de concerto.	15
Tabela 2 – Exemplo de um <i>log</i> de eventos.	32
Tabela 3 – Simplificação do <i>log</i> de eventos da Tabela 2.	33
Tabela 4 – Conjunto de dados brutos (as seções exibem diferentes colunas de uma única tabela, cujas linhas compartilham a mesma posição e referência de matrícula)	53
Tabela 5 – Dados brutos acadêmicos dos discentes do curso de Ciência da Computação	54
Tabela 6 – Conjunto de dados após limpeza	56
Tabela 7 – Conjunto de dados após concatenação das colunas código e descrição	57
Tabela 8 – Conjunto de dados após inserção das atividades de semestre e formado	58
Tabela 9 – <i>Log</i> de eventos do curso de Ciência da Computação	59
Tabela 10 – <i>Log</i> de eventos do discente 999001	60
Tabela 11 – <i>Trace</i> de eventos do discente 1001	67
Tabela 12 – <i>Trace</i> de eventos do discente 1002	72
Tabela 13 – <i>Trace</i> de eventos do discente 1003	77
Tabela 14 – Estatísticas dos alunos formados	82
Tabela 15 – Estatísticas dos alunos evadidos	83
Tabela 16 – Resultado do alinhamento do 1º semestre de um discente.	91

LISTA DE QUADROS

Quadro 1 – Quadro comparativo das abordagens dos trabalhos relacionados e deste trabalho.	51
---	----

SUMÁRIO

1	INTRODUÇÃO	13
1.1	Objetivos	15
1.1.1	<i>Objetivo Geral</i>	15
1.1.2	<i>Objetivos Específicos</i>	16
1.2	Organização	16
2	FUNDAMENTAÇÃO TEÓRICA	17
2.1	Evasão no ensino superior	17
2.1.1	<i>Causas da evasão</i>	20
2.2	Modelagem de processos	22
2.2.1	<i>Redes de Petri</i>	24
2.2.2	<i>Workflow nets</i>	30
2.3	Mineração de processos	31
2.4	CrITÉRIOS de avaliação de processos	33
2.4.1	<i>Verificação de conformidade por token replay</i>	35
2.4.2	<i>Verificação de conformidade por alinhamento</i>	43
3	TRABALHOS RELACIONADOS	47
3.1	Análise da trajetória de aprendizagem do aluno no ensino a distância por meio da mineração de processos	47
3.2	Identificação de trajetórias de aprendizagem em um curso de graduação e sua relação com a evasão escolar	48
3.3	Mineração de dados educacionais para predição da evasão em cursos de graduação presenciais no ensino superior	49
3.4	Análise de dados pré-universitários para predizer a evasão de alunos ingressantes em uma instituição de ensino superior	49
3.5	Modelo de predição de evasão escolar com base em dados de autoavaliação de cursos de graduação	50
3.6	Quadro comparativo	50
4	REPRESENTAÇÃO DA ESTRUTURA CURRICULAR COM REDE DE PETRI	52
4.1	Preparação dos dados e criação do <i>log</i> de eventos	52

4.1.1	<i>Limpeza dos dados</i>	55
4.1.2	<i>Construção do log de eventos</i>	56
4.1.3	<i>Ferramentas utilizadas</i>	59
4.2	Modelo formal da estrutura curricular do Curso de Ciência da Computação em rede de Petri	59
4.2.1	<i>Simulação preliminar com token replay</i>	66
4.2.1.1	<i>Cenário discente ideal</i>	66
4.2.1.2	<i>Cenário discente com reprovações</i>	71
4.2.1.3	<i>Cenário de reprovações em todos os componentes de um semestre</i>	76
5	EXPERIMENTOS	81
5.1	Análise de conformidade utilizando a técnica de <i>token replay</i>	81
5.1.1	<i>Resultados para discentes formados</i>	82
5.1.2	<i>Resultados para discentes evadidos</i>	83
5.1.3	<i>Comparação entre grupos</i>	84
5.1.4	<i>Distribuição do trace fitness e padrões de progressão acadêmica</i>	84
5.1.5	<i>Fatores estruturais que influenciam a evasão</i>	85
5.1.6	<i>Análise dos pontos críticos nas trajetórias acadêmicas</i>	87
5.1.7	<i>Distribuição de tokens remanescentes por número de semestre letivo até a evasão</i>	89
5.2	Análise de conformidade utilizando a técnica de alinhamentos (<i>Alignments</i>)	90
5.2.1	<i>Alinhamento dos discentes formados</i>	92
5.2.2	<i>Alinhamento dos discentes não formados</i>	93
5.2.2.1	<i>Frequência de componentes curriculares cursados fora do semestre ideal</i>	94
5.2.2.2	<i>Frequência de componentes curriculares cursados antes do semestre ideal</i>	96
5.2.2.3	<i>Matriz curricular empírica dos discentes evadidos</i>	97
5.3	Considerações finais	98
6	CONCLUSÕES E TRABALHOS FUTUROS	100
	REFERÊNCIAS	102
7	ESTRUTURA CURRICULAR DO CURSO DE CIÊNCIA DA COMPUTAÇÃO	106

1 INTRODUÇÃO

A gestão de instituições de ensino superior federal enfrenta desafios significativos, entre os quais a evasão destaca-se como um dos mais complexos. A evasão é um indicador de alta complexidade para o acompanhamento das políticas de ensino superior no Brasil.

De acordo com o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP, 2017), a evasão concretiza-se com a saída antecipada, antes do ano de conclusão do ciclo do curso, por desistências, representando, portanto, uma condição final de insucesso quanto ao objetivo de promover o aluno a uma situação superior à de ingresso. Esse fenômeno não apenas compromete o progresso acadêmico dos estudantes, mas também impacta negativamente na qualidade e eficácia das instituições de ensino superior.

De acordo com INEP (2017), a taxa acumulada de evasão é calculada pela relação entre o número de estudantes que desistiram do curso e o número total de ingressantes no ano. Nos cursos presenciais ofertados pelas Instituições de Ensino Superior (IES) no Brasil, essa taxa é notavelmente alta. De acordo com os indicadores de fluxo da educação superior no Ciclo 2018-2022, divulgados por INEP (2023), a evasão acumulada atinge 52,61%, evidenciando um problema crítico para a gestão acadêmica e o planejamento de políticas educacionais no país.

A elevada taxa de evasão nas instituições de ensino superior reforça a necessidade de ferramentas mais avançadas para monitorar e combater esse fenômeno. Neste contexto, a mineração de dados educacionais, conforme descrito por Romero e Ventura (2010), surge como uma abordagem promissora. Trata-se da aplicação de técnicas de mineração de dados e aprendizado de máquina ao campo educacional, visando extrair padrões e gerar conhecimento a partir dos dados.

Diversos estudos, como em Oliveira e Medeiros (2024) e Jesus *et al.* (2021), utilizam essas técnicas para prever o risco de evasão, destacando o potencial da análise preditiva no enfrentamento da evasão acadêmica. Essas pesquisas utilizam dados de autoavaliação e históricos acadêmicos para treinar algoritmos de aprendizagem de máquina, proporcionam uma base sólida para a tomada de decisões mais eficazes por parte das instituições, conforme também apontado por Silva *et al.* (2021), que busca identificar as razões que levam os discentes a desistirem dos cursos.

Diversos autores, como ??, ?? e Silva *et al.* (2016), apontam que a estrutura curricular dos cursos de graduação é um dos fatores que contribuem para a evasão estudantil. Essas estruturas são compostas por componentes curriculares obrigatórios e eletivos, distribuídos

ao longo de vários semestres. Essa organização pode ser interpretada como um *processo*, no qual as atividades representam etapas que conduzem o estudante de um estado inicial de conhecimento a um estado de conhecimento mais avançado, até atingir um objetivo final Harte *et al.* (1994).

Nesse contexto, é possível compreender o percurso do estudante como uma trajetória educacional, ou seja, um caminho formativo estruturado com base no currículo do curso. Essa trajetória reflete a sequência cronológica de disciplinas cursadas por um discente e pode ser observada em três situações distintas: permanência (continuidade no curso), conclusão (finalização com sucesso) e desistência (abandono antes do término). Essa visão permite analisar o processo educacional não apenas pela estrutura curricular ideal, mas também pelas diferentes formas de progresso e desvio que ocorrem ao longo do percurso acadêmico.

Compreender a trajetória acadêmica como um *processo* permite que técnicas de mineração de processos (*process mining*) sejam aplicadas. De acordo com Aalst *et al.* (2011), as técnicas de mineração de processos são capazes de extrair conhecimento dos *logs* de eventos frequentemente disponíveis nos sistemas de informação atuais. Essas técnicas fornecem novos meios para descobrir, monitorar e melhorar processos em diversos domínios.

O *log* de eventos é o ponto de partida para a mineração de processos. Cada evento refere-se a uma atividade e está relacionado a um caso específico. Os eventos pertencentes a um caso são ordenados e podem ser vistos como uma execução do processo Aalst (2012). A Tabela 1 apresenta o *log* de eventos do processo de solicitação de compensação para ingresso de concerto. Sua estrutura básica possui algumas propriedades, a saber: *Case Id* que corresponde a identificação de forma única cada instância do processo; *Timestamp* registra o momento exato em que a atividade foi realizada; *Activity* refere-se à tarefa realizada dentro do processo; e *Resource* refere-se ao recurso associado à execução da atividade, o qual pode representar, por exemplo, o responsável pela ação, como ilustrado na tabela, entre outros tipos de recursos possíveis.

Tais registros podem ser convertido em estruturas que permitam uma visualização e análise mais intuitiva do processo educacional, como as redes de Petri. Segundo Murata (1989), as redes de Petri são ferramentas formais adequadas para modelar e analisar sistemas nos quais ocorrem atividades simultâneas, concorrentes ou paralelas, como é o caso da progressão acadêmica dos estudantes em uma estrutura curricular.

A partir da modelagem da estrutura curricular por meio de redes de Petri, é possível aplicar técnicas de análise de conformidade com base no *Replay*. De acordo com Aalst (2016), o *Replay* permite comparar o comportamento esperado, conforme definido no modelo, com o

Tabela 1 – Um exemplo de um *log* de eventos, descrevendo o comportamento relacionado ao processo de solicitação de compensação para ingresso de concerto.

<i>Case Id</i>	<i>Timestamp</i>	<i>Activity</i>	<i>Resource</i>
...			
12373	30-7-2019 11.02	Register request	Barbara
12373	30-7-2019 11.12	Register request	Barbara
12374	30-7-2019 11.32	Register request	Jan
12374	30-7-2019 11.44	Register request	Jan
12373	30-7-2019 12.12	Check ticket	Hajo
12374	30-7-2019 14.16	Examine casually	Jorge
12375	30-7-2019 14.32	Register request	Josep
...			

Fonte: Aalst (2016)

comportamento real registrado no *log* de eventos. Essa técnica consiste em reexecutar o processo modelado com os dados reais dos estudantes, identificando como os discentes percorrem as disciplinas ao longo do tempo e detectando desvios, gargalos ou padrões recorrentes que podem estar associados à evasão.

O trabalho de Unger *et al.* (2022) aplica mineração de processos em um *log* de eventos para compreender o comportamento dos estudantes no ambiente virtual de aprendizagem (AVA) com o objetivo de identificar problemas na interação do discente nas atividades propostas e auxiliar na tomada de decisão visando à melhoria dos instrumentos utilizados no ensino. Este trabalho concentra-se na compreensão das atividades executadas no AVA e não considera o impacto do currículo na trajetória acadêmica ao longo do curso

Portanto, modelar as estruturas curriculares como um processo permite a aplicação de técnicas de mineração de processos, possibilitando a extração de informações que contribuam para a redução da evasão no curso de graduação do campus da UFC em Quixadá, proposta que não foi abordada em Unger *et al.* (2022).

1.1 Objetivos

1.1.1 Objetivo Geral

Realizar uma análise da trajetória acadêmica dos estudantes do curso de Ciência da Computação da Universidade Federal do Ceará (UFC), *campus* de Quixadá, por meio da modelagem da estrutura curricular com redes de Petri e da aplicação de técnicas de Mineração de Processos, com ênfase na identificação de pontos críticos associados à evasão estudantil.

1.1.2 *Objetivos Específicos*

São objetivos específicos desse trabalho:

- Modelar formalmente a estrutura curricular do curso de Ciência da Computação da UFC, *campus* de Quixadá, por meio de redes de Petri.
- Aplicar a técnica de *token replay* para verificar a conformidade entre os dados reais dos estudantes e o modelo acadêmico.
- Aplicar a técnica de alinhamento (*alignments*) para identificar e analisar inconformidades nas trajetórias acadêmicas dos discentes.
- Identificar e analisar pontos críticos na trajetória estudantil que possam estar associados à evasão.

1.2 **Organização**

Os demais capítulos deste trabalho estão organizados da seguinte forma: o Capítulo 2 explora os principais conceitos abordados neste estudo. O Capítulo 3 apresenta uma revisão dos trabalhos relacionados ao tema. O Capítulo 4 descreve a modelagem formal da estrutura curricular do curso de Ciência da Computação, por meio de redes de Petri, incluindo a construção do modelo e a simulação preliminar de diferentes trajetórias acadêmicas. O Capítulo 5 apresenta os experimentos realizados com base nas técnicas de *token replay* e alinhamento (*alignments*), destacando padrões de progressão, retenção e evasão. Por fim, o Capítulo 6 expõe as considerações finais do estudo e apresenta sugestões para trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, são apresentados os conceitos que fundamentam este trabalho, com foco na análise da evasão no ensino superior a partir de abordagens baseadas em modelagem e mineração de processos. Na Seção 2.1, discute-se o fenômeno da evasão acadêmica, suas definições institucionais e metodologias de acompanhamento longitudinal, com destaque para os indicadores de permanência, conclusão e desistência utilizados pelo INEP. A Seção 2.2 introduz as redes de Petri como formalismo matemático capaz de representar a dinâmica dos processos educacionais. Na Seção 2.3, apresentam-se os três principais tipos de análise em mineração de processos: descoberta, verificação de conformidade e aprimoramento de modelos, com o *log* de eventos formalizado como elemento central dessas abordagens. Por fim, a Seção 2.4 discute os critérios de avaliação dos modelos de processo — adequação, precisão, generalização e simplicidade — além das técnicas de verificação de conformidade, como *token replay* e alinhamento (*alignment*), acompanhadas de fórmulas e exemplos práticos que permitem mensurar o grau de aderência entre a execução real e o modelo idealizado.

2.1 Evasão no ensino superior

De acordo com Oliveira (2023), a evasão escolar é um problema antigo e multifacetado que afeta todos os níveis de ensino, especialmente o ensino superior. Compreender as razões que levam os discentes a abandonarem seus cursos tem sido o foco de diversos estudos na literatura, como evidenciam Vasconcelos e Andrade (2018), Carvalho *et al.* (2019), Sousa *et al.* (2022), reforçando o caráter persistente e desafiador dessa problemática.

Embora existam múltiplas definições para evasão, este trabalho adota o conceito proposto pela Comissão Especial do Ministério da Educação (MEC), conforme ANDIFES *et al.* (1996), que a define como o desligamento do estudante do curso superior em decorrência de abandono, desistência formal, transferência, reopção de curso ou exclusão por normas institucionais.

Esse fenômeno é acompanhado pelo Censo da Educação Superior, principal instrumento de coleta de dados sobre o ensino superior no Brasil. Realizado anualmente pelo INEP, o Censo tem como objetivo fornecer informações estatísticas confiáveis que possibilitem conhecer, monitorar e avaliar o sistema de educação superior. Os dados levantados permitem o cálculo de indicadores fundamentais para a formulação e implementação de políticas públicas, além

de subsidiar a atuação de gestores institucionais, autoridades governamentais, pesquisadores e organismos nacionais e internacionais.

Por meio do Censo da Educação Superior, torna-se possível acompanhar o fluxo acadêmico dos discentes com base em sua situação de vínculo com o curso. Anualmente, o Censo registra o último vínculo acadêmico de cada aluno no ano de referência, classificando-o em uma das seguintes condições: matrícula ativa (cursando ou trancado), formado, desvinculado por desistência ou transferência, ou falecido. Essas informações permitem caracterizar a posição do estudante em três dimensões analíticas fundamentais: permanência, conclusão e desistência.

A permanência refere-se aos discentes que mantêm vínculo ativo com o curso, seja por matrícula regular ou trancada; a conclusão abrange aqueles que finalizaram com êxito todos os requisitos curriculares; e a desistência compreende os que encerraram o vínculo sem completar a formação, seja por evasão, abandono ou transferência. Essas categorias são mutuamente exclusivas e coletivamente exaustivas, constituindo a base dos principais indicadores educacionais utilizados na análise da trajetória acadêmica.

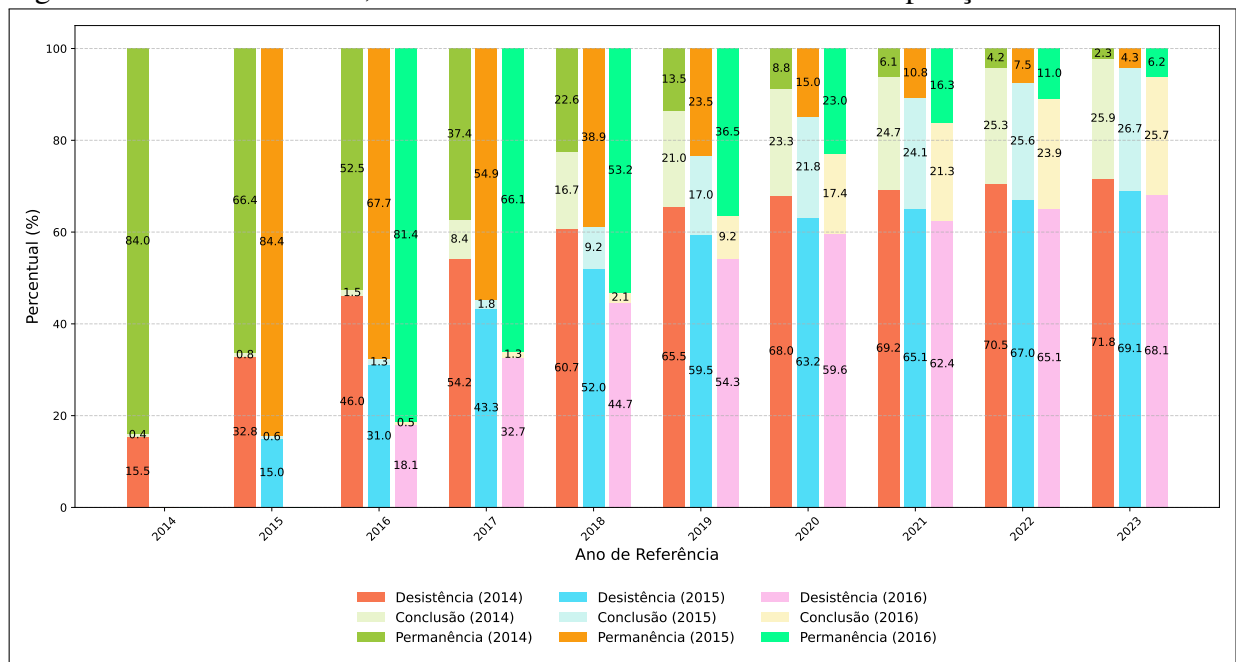
Para fins de acompanhamento longitudinal, INEP (2017) define *coortes* de ingressantes como o conjunto de alunos que iniciou um curso em determinado ano (T), cujos vínculos são registrados nas bases do Censo nos quatro primeiros anos após o ingresso, com monitoramento estendido por até dez anos. Esse modelo permite identificar, ao longo do tempo, o status de cada estudante e sua movimentação no sistema.

No cálculo dos indicadores, observa-se que, independentemente do ano de referência (t), a soma da Taxa de Permanência Acumulada (TAP), da Taxa de Conclusão Acumulada (TCA) e da Taxa de Desistência Acumulada (TDA) será sempre igual a 100%. Isso ocorre porque os três indicadores contemplam, de forma abrangente e exclusiva, todas as possíveis situações de vínculo dos discentes com o curso, permitindo uma análise precisa do fluxo acadêmico ao longo da trajetória formativa.

A Figura 1 apresenta as *coortes* do curso de Ciência da Computação no Brasil nos anos de 2014, 2015 e 2016, divulgados por INEP (2023). Dentre elas, apenas a *coorte* de 2014 possui o ciclo completo de acompanhamento divulgado, enquanto as demais ainda não contam com informações referentes aos anos finais do período de análise.

A partir dos dados apresentados, é possível compreender a gravidade do fenômeno da evasão. Os discentes que ingressaram em 2014 apresentam, ao final do ciclo (2023), uma taxa de desistência acumulada de aproximadamente 71,8%, enquanto a taxa de conclusão atinge cerca

Figura 1 – *Coortes* de 2014, 2015 e 2016 do curso de Ciência da Computação no Brasil



Fonte: elaborada pelo autor.

de 25,9%. Isso significa que, para cada aluno que conclui o curso, quase três o abandonam ao longo do percurso. Embora os ciclos das *coortes* de 2015 e 2016 ainda não estejam oficialmente encerrados, os percentuais observados indicam trajetória semelhante, com taxas de desistência de 69,1% e 65,1%, respectivamente.

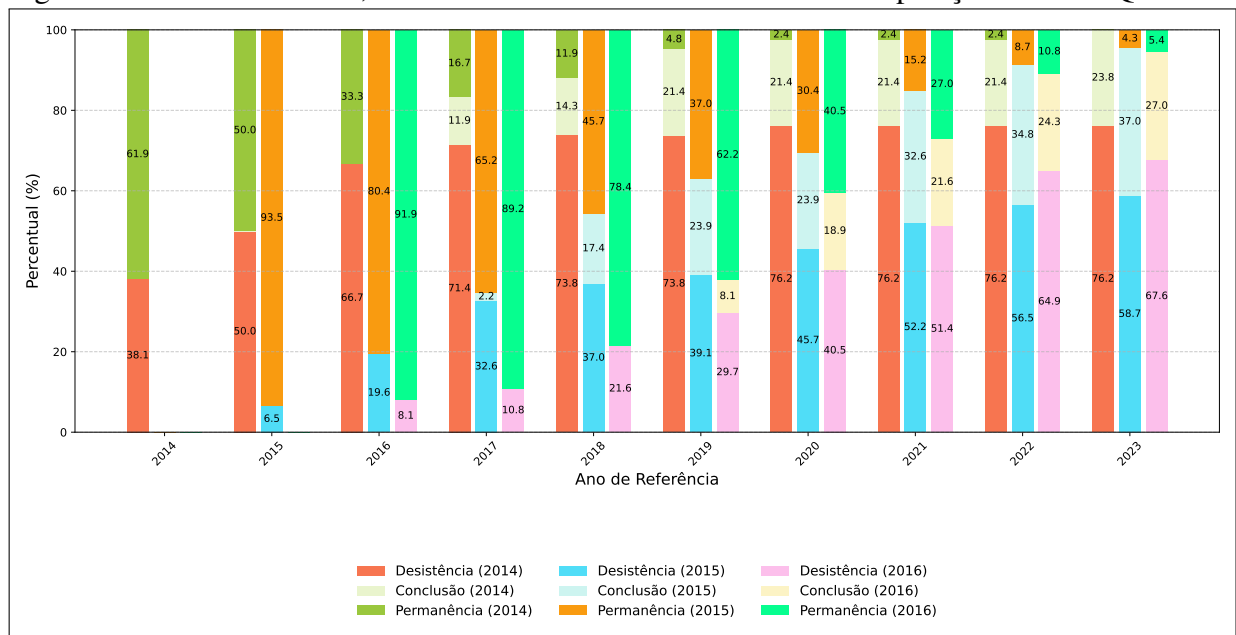
Esse padrão evidencia um problema estrutural no sistema de ensino superior. Conforme destaca Oliveira (2023), a evasão compromete não apenas o investimento público em educação, mas também o desenvolvimento social e econômico do país, ao reduzir a formação de profissionais qualificados.

Outro fator preocupante é a quantidade de discentes que necessitam de mais tempo do que o previsto para concluir seus cursos. A análise da taxa de permanência revela que cerca de 15% dos estudantes demandam mais de cinco anos para finalizar sua graduação, cujo tempo ideal de integralização é, justamente, de cinco anos.

Ao analisar o fluxo acadêmico do curso de Ciência da Computação da UFC, *campus* de Quixadá, no mesmo período, constata-se um cenário semelhante ao observado em nível nacional. A *coorte* de 2014 apresenta, ao final do ciclo (2023), uma taxa de desistência acumulada de aproximadamente 76,2%, enquanto a taxa de conclusão é de 23,8%, não havendo mais discentes vinculados ao curso. Isso indica que, para cada estudante que concluiu a graduação, mais de três a abandonaram ao longo da trajetória.

As *coortes* de 2015 e 2016, ainda em acompanhamento, também revelam padrões

Figura 2 – *Coortes* de 2014, 2015 e 2016 do curso de Ciência da Computação na UFC-Quixadá



Fonte: elaborada pelo autor.

preocupantes: a *coorte* de 2015 acumula 58,7% de desistência, 37,0% de conclusão e 4,3% de permanência; enquanto a *coorte* de 2016 registra 67,6% de evasão, 27,0% de conclusão e 5,4% de estudantes ainda vinculados.

Esses dados reforçam a evasão como um desafio estrutural que persiste mesmo em instituições públicas e bem avaliadas, como a UFC. A elevada taxa de retenção nos anos finais do ciclo — período em que os discentes deveriam estar concluindo sua formação — evidencia entraves pedagógicos, curriculares e institucionais. O fenômeno impacta não apenas o percurso individual dos estudantes, mas também compromete o aproveitamento de vagas públicas, o planejamento institucional e a formação de profissionais qualificados. Nesse sentido, torna-se urgente o fortalecimento de políticas de permanência que articulem dimensões acadêmicas, financeiras, pedagógicas e psicossociais.

2.1.1 Causas da evasão

A evasão no ensino superior é um fenômeno multifatorial e complexo, que abrange desde aspectos individuais até fatores institucionais e pedagógicos. A literatura recente sobre o tema aponta que as causas da evasão não se restringem a uma única dimensão, mas resultam da interação entre dificuldades acadêmicas, barreiras socioeconômicas, questões pessoais e limitações institucionais Souza *et al.* (2015), Saraiva *et al.* (2021), Kantorski *et al.* (2023).

Diversos estudos baseados em mineração de dados educacionais e modelagens predi-

tivas identificam padrões recorrentes nos perfis de estudantes que evadem. Entre os principais fatores associados à evasão, destacam-se:

- **Baixo desempenho acadêmico**, especialmente nos primeiros semestres, com destaque para reprovações em disciplinas como Cálculo, Álgebra e Algoritmos Carmo *et al.* (2022), Viana *et al.* (2022);
- **Trancamentos e reprovações por nota ou falta**, que dificultam o progresso curricular e desestimulam a permanência Unger *et al.* (2022), Kantorski *et al.* (2023);
- **Falta de engajamento com o ambiente virtual de aprendizagem**, no caso dos cursos EaD, associada à baixa interação com conteúdos e colegas Andrade *et al.* (2023);
- **Problemas socioemocionais e pessoais**, como estresse, ansiedade, conflitos familiares e sensação de isolamento Saraiva *et al.* (2021);
- **Mudança de interesses ou insatisfação com o curso**, motivando reopções ou transferências para outras áreas Moraes *et al.* (2021);
- **Condições socioeconômicas adversas**, como necessidade de trabalhar, baixa renda familiar e falta de apoio financeiro, dificultando a dedicação integral aos estudos Colpo *et al.* (2021);
- **Distância geográfica da instituição**, que compromete a frequência e aumenta o custo da permanência, especialmente em cursos presenciais no turno noturno Colpo *et al.* (2021);
- **Fatores institucionais**, como currículos inflexíveis, ausência de políticas efetivas de acolhimento e escassez de tutorias, monitorias ou acompanhamento individualizado Carvalho *et al.* (2023).

Além desses fatores, a literatura destaca que a evasão tende a ocorrer com maior frequência nos três primeiros semestres do curso, o que reforça a importância de estratégias de detecção e intervenção precoce. O uso de abordagens baseadas em mineração de dados, aliado ao acompanhamento pedagógico ativo, surge como alternativa promissora para identificar padrões de risco e orientar ações preventivas. Nesse contexto, é fundamental compreender que a evasão não se configura como um evento isolado, mas como um processo cumulativo que exige atenção contínua das instituições de ensino.

Diante do persistente cenário de alta evasão no ensino superior, diversas pesquisas e iniciativas institucionais têm se concentrado na identificação de estratégias que possibilitem não apenas a compreensão do fenômeno, mas também o desenvolvimento de ações eficazes de prevenção. Nesse cenário, destacam-se as abordagens baseadas em Mineração de Dados

Educacionais (MDE) e Mineração de Processos, que se mostram promissoras na análise de grandes volumes de dados e na extração de padrões relevantes sobre o comportamento acadêmico discente.

A MDE permite a construção de modelos preditivos capazes de antecipar situações de risco, a partir da análise do histórico acadêmico, características demográficas, socioeconômicas e comportamentais dos estudantes Oliveira (2023), Santos *et al.* (2021). Já a mineração de processos contribui para a compreensão de como os discentes percorrem, de fato, os componentes curriculares, revelando desvios, gargalos, ciclos de repetência e padrões de progressão ou abandono Unger *et al.* (2022), Carmo *et al.* (2022).

Este trabalho adota uma abordagem fundamentada na mineração de processos para investigar os fatores que contribuem para a evasão nos cursos de graduação da Universidade Federal do Ceará – *campus* de Quixadá. Por meio da análise de históricos escolares, busca-se mapear as trajetórias dos discentes ao longo do curso, identificar comportamentos recorrentes entre aqueles que evadem e contrastá-los com os percursos de concluintes. Essa análise permite compreender, por exemplo, a frequência e o momento em que ocorrem trancamentos, reprovações ou interrupções no fluxo curricular, além de evidenciar padrões de retenção prolongada.

Os achados obtidos por meio dessa abordagem podem subsidiar a tomada de decisões no âmbito da gestão educacional, orientando a formulação de políticas de permanência mais eficazes, com base em evidências. Além disso, possibilitam o desenvolvimento de sistemas de alerta precoce, tutorias direcionadas, intervenções pedagógicas personalizadas e ações de acolhimento mais alinhadas às reais necessidades do corpo discente.

2.2 Modelagem de processos

Segundo Harte *et al.* (1994), um modelo de processo descreve como se passa de um estado inicial de conhecimento, por meio de uma sequência de estados intermediários, até um determinado estado de meta. Já Cbok (2009) define processo como uma sequência lógica de atividades interligadas. Para Medeiros (2011), um processo é uma atividade realizada por uma organização, composta por uma série de etapas e controles que possibilitam o fluxo de informações.

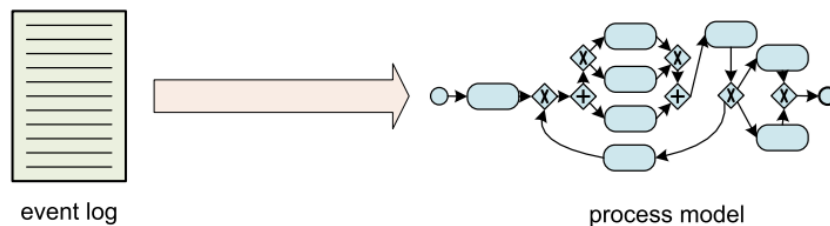
Dessa forma, é possível compreender a estrutura curricular de um curso de graduação como um processo: o discente parte de disciplinas iniciais, que estão interligadas ao longo do curso, e, ao concluir essa sequência, adquire as competências e habilidades necessárias ao

exercício profissional da área escolhida.

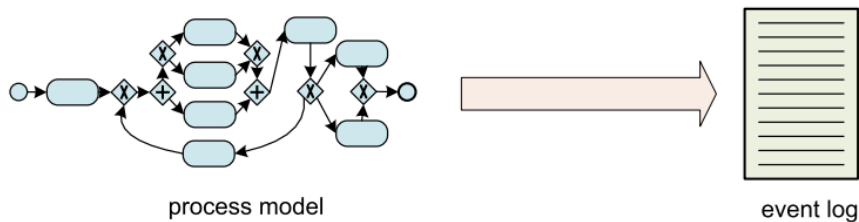
Para alcançar a melhoria dos processos é necessário mapear as atividades que os compõem Jesus *et al.* (2014). Assim, torna-se viável modelar a estrutura curricular e analisá-la com o objetivo de identificar oportunidades de aprimoramento. Para avaliar a representatividade da realidade por meio do modelo, três abordagens fundamentais são utilizadas na análise e modelagem de processos: *Play-In*, *Play-Out* e *Replay* Aalst (2016), conforme ilustrado na Figura 3.

Figura 3 – *Play-In*, *Play-Out* e *Replay*

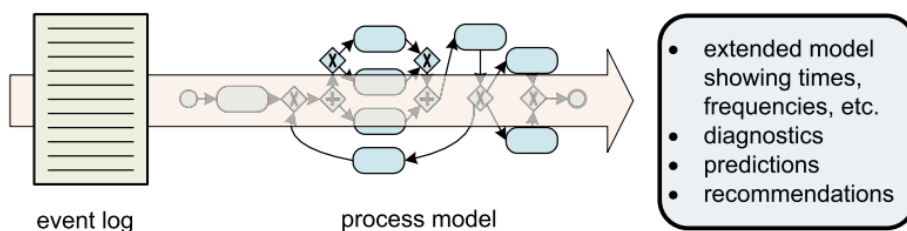
Play-In



Play-Out



Replay



Fonte: Aalst (2016).

O *Play-Out* utiliza um modelo de processo para simular e produzir *traces* possíveis. Essa abordagem permite compreender, na prática, como o processo pode ocorrer, evidenciando a diversidade de caminhos que as atividades podem seguir durante sua execução. Por sua vez, o *Play-In* parte de um *log* de eventos e gera, a partir dos dados observados, o modelo de processo correspondente. Já o *Replay* combina ambos: utiliza um *log* de eventos e um modelo previamente definido para comparar os *traces* registrados com aqueles previstos pelo modelo, avaliando o

grau de conformidade entre eles.

Para representar o currículo como um processo, diversas metodologias de modelagem poderiam ser adotadas. No entanto, este trabalho utiliza redes de Petri como abordagem principal, aplicando especificamente a técnica de *Token Replay* e Alinhamento (*Alignments*) para analisar a conformidade entre o comportamento real dos discentes e o modelo ideal de trajetória curricular.

2.2.1 Redes de Petri

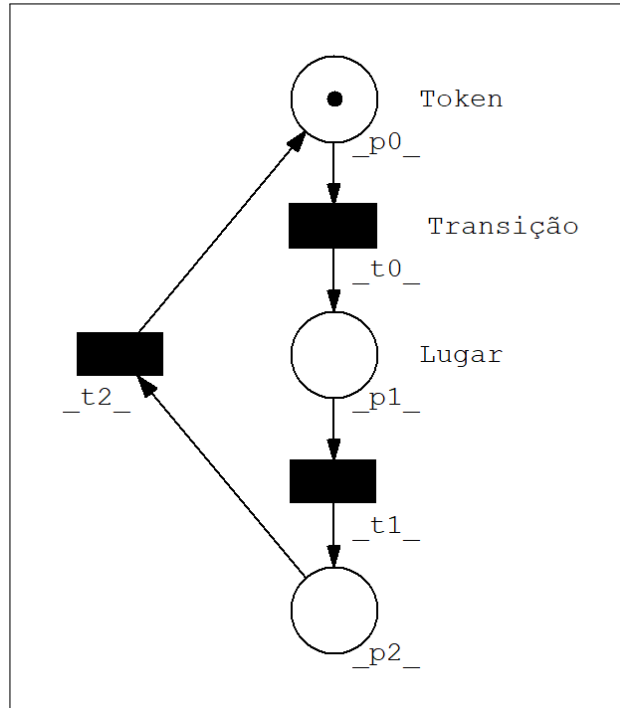
A concepção teórica das redes de Petri teve origem na tese de doutorado de Carl Adam Petri, intitulada “Comunicação com Autômatos”, defendida em 1962, na qual foram introduzidos os princípios fundamentais desse modelo matemático de representação de sistemas concorrentes. Segundo Peterson (1981), a aplicação das redes de Petri ocorre por meio da modelagem, que consiste em uma representação — frequentemente em termos matemáticos — das características consideradas relevantes do objeto ou sistema em estudo. De acordo com Maciel *et al.* (1996), as redes de Petri constituem uma técnica de especificação de sistemas que, além de permitir uma representação matemática formal, oferece mecanismos robustos de análise, possibilitando tanto a verificação de propriedades quanto a avaliação da corretude do sistema especificado.

Segundo Aalst (2016), uma rede de Petri é um grafo bipartido, com uma estrutura composta por dois conjuntos distintos de elementos: lugares e transições. As conexões entre esses elementos são feitas por arcos direcionados, que sempre ligam um lugar a uma transição ou uma transição a um lugar, garantindo que não existam conexões diretas entre dois lugares ou entre duas transições. De acordo com Cardoso e Valette (1997), a rede de Petri é composta por três elementos básicos:

- **Lugar:** pode ser interpretado como uma condição, um estado parcial, uma espera, um procedimento ou um conjunto de recursos. É representado graficamente por um círculo;
- **Transição:** representada graficamente por uma barra ou retângulo, está associada a um evento que ocorre no sistema;
- **Token:** é um marcador que indica que a condição associada ao lugar é verdadeira. Representado graficamente por um ponto preto no interior de um lugar.

A Figura 4 ilustra uma rede de Petri com seus elementos básicos. Os lugares são representados por $_p0_$, $_p1_$, $_p2_$, enquanto as transições são indicadas por $_t0_$, $_t1_$, $_t2_$.

Figura 4 – Elementos básicos de uma rede de Petri



Fonte: elaborada pelo autor.

O *token* é representado pelo ponto preto em $_p0_$. Aalst (2016) define formalmente uma rede de Petri da seguinte forma:

Definição 2.2.1 Aalst (2016) Uma rede de Petri é um tripla $N = (P, T, F)$

onde:

- $\mathbf{P}$ é um conjunto finito de lugares de N ,
- $\mathbf{T}$ é um conjunto finito de transições de N , de modo que $P \cap T = \emptyset$,
- $\mathbf{F}$ é um conjunto de arcos direcionados, chamado de relação de fluxo, de modo que $F \subseteq (P \times T) \cup (T \times P)$.

Conforme Aalst (2016) os conceitos de lugares de entrada e saída são definidos com base na relação de fluxo F , ou seja, um lugar P é um lugar de entrada de uma transição T se, e somente se, existir um arco direcionado de P para T , isto é, $(p, t) \in F$. Da mesma forma, um lugar P é um lugar de saída de uma transição T se, e somente se, $(t, p) \in F$. Para qualquer $p \in P$ e $t \in T$:

- $\bullet t = \{p \mid (p, t) \in F\}$ representa o conjunto de lugares de entrada de T .
- $t \bullet = \{p \mid (t, p) \in F\}$ representa o conjunto de lugares de saída de T .

A Figura 4 ilustra a relação de fluxo entre *lugares de entrada* e *lugares de saída* em uma *rede de Petri*. O lugar $_p1_$ é um lugar de entrada para a transição $_t1_$, pois há um arco

direcionado conectando $_p1_$ a $_t1_$. Da mesma forma, o lugar $_p2_$ é um lugar de saída da transição $_t1_$, pois existe um arco direcionado que liga $_t1_$ a $_p2_$.

Formalmente, aplicando a Definição 2.2.1 à rede de Petri ilustrada na Figura 4 temos:

- $P = \{_p0_, _p1_, _p2_\}$;
- $T = \{_t0_, _t1_, _t2_\}$;
- $F = \{(_p0_, _t0_), (_t0_, _p1_), (_p1_, _t1_), (_t1_, _p2_), (_p2_, _t2_), (_t2_, _p0_\)};$

De acordo com Aalst (2016), a estrutura de uma rede de Petri é *estática*, ou seja, sua topologia não se altera ao longo do tempo. No entanto, os *tokens* podem se mover pela rede, representando a dinâmica do sistema modelado.

O estado de uma rede de Petri é definido pela distribuição dos *tokens* nos lugares da rede, sendo essa distribuição conhecida como *marcação* e representado por M . A marcação determina a configuração atual da rede e influencia a habilitação e o disparo das transições, permitindo a evolução do sistema ao longo do tempo. Cardoso e Valette (1997) define uma rede de Petri marcada como:

Definição 2.2.2 *Cardoso e Valette (1997) Uma rede marcada PN é uma dupla $PN = (N, M)$, onde N é uma rede de Petri, M a marcação inicial dada pela aplicação de $M = P \rightarrow \mathbb{N}$. $M(p)$ é o número de tokens contidas no lugar p . A marcação M é a distribuição dos tokens nos lugares, sendo representada por um vetor coluna cuja dimensão é o número de lugares e elementos $M(p)$.*

A Figura 4 ilustra uma rede de Petri marcada, onde o lugar $_p0_$ contém um *token*, enquanto os lugares $_p1_$ e $_p2_$ estão vazios. Dessa forma, o vetor de marcação M correspondente a essa rede é dado por $M^T = [1 \ 0 \ 0]$ (T é o transposto do vetor). Esse vetor representa a distribuição de *tokens* na rede, indicando que apenas o lugar $_p0_$ está marcado no estado inicial.

A marcação da rede pode ser modificada por meio da *regra de disparo* de uma transição. Esse mecanismo consiste no consumo de *tokens* dos lugares de entrada da transição e na adição de *tokens* aos lugares de saída. Aalst (2016) define formalmente a regra de disparo da seguinte forma:

O estado da rede é alterado de acordo com as seguintes regras:

1. **Condição para uma transição estar habilitada:** Uma transição $t \in T$ está habilitada na marcação M se, e somente se, todos os seus lugares de entrada possuem *tokens* suficientes para satisfazer a condição:

$$\bullet t \leq M$$

setas indicam os arcos (F), conectando lugares e transições, quadrados rotulados correspondem às transições (T), indicando atividades no processo, o ponto preto representam o *tokens*, que marcam a presença de um estado ativo no processo e balões com texto identificam tipos de junções ou divisões no fluxo do processo.

A rede inicia com um *token* no lugar *start*, de onde o *token* segue para a transição *a* (*register request*). A partir desse ponto, o processo se ramifica em múltiplos caminhos, conforme representado pelos arcos em F , que conectam os lugares às transições e vice-versa, criando o fluxo do processo. Cada transição só pode ser disparada quando há um *token* disponível no lugar imediatamente anterior, permitindo o avanço do processo.

A explicação dos elementos visuais da Figura 5 complementa essa descrição, demonstrando como as atividades do processo podem ser executadas simultaneamente ou em sequências alternativas. As junções *AND-split*, *AND-join*, *XOR-split* e *XOR-join* controlam os fluxos paralelos e exclusivos, garantindo a correta modelagem do comportamento do processo. Esse mapeamento possibilita uma análise detalhada, permitindo identificar os diferentes caminhos que podem ser seguidos durante a execução real.

Para que ocorra o disparo de uma transição em uma rede de Petri, é necessário que ela esteja habilitada. Segundo Aalst (2019), uma transição está habilitada quando todos os seus lugares de entrada possuem pelo menos um *token*.

Ao ser disparada, a transição consome os *tokens* de seus lugares de entrada e gera novos *tokens* nos lugares de saída, seguindo a lógica do fluxo definido na rede. Na rede de Petri ilustrada na Figura 5, apenas a transição *a* encontra-se habilitada.

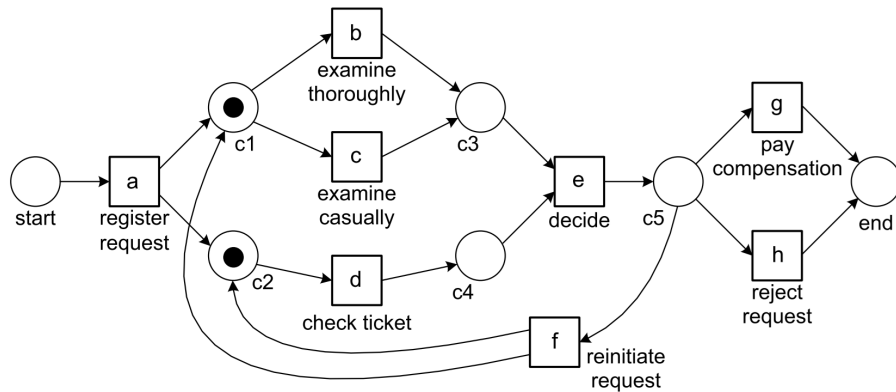
Suponha agora que a transição *a* seja disparada. Nesse caso, o *token* presente no lugar de entrada é consumido, e dois novos *tokens* são gerados: um no lugar *c1* e outro no lugar *c2*, conforme representado na Figura 6.

Neste momento, a marcação da rede encontra-se nos lugares *c1* e *c2*. Consequentemente, as transições *b*, *c* e *d* estão habilitadas.

No lugar *c2*, há apenas um caminho disponível para o *token*, levando à transição *d*. Já no lugar *c1*, existem dois caminhos possíveis para o *token*, um levando à transição *b* e outro à transição *c*.

Ao disparar a transição *d*, o *token* presente em *c2* é consumido e um novo *token* é gerado em *c4*. No caso de *c1*, a transição é disparada de forma que apenas uma das opções ocorra, ou seja, *b* ou *c* pode ser ativada. Como resultado, um novo *token* é produzido em *c3*,

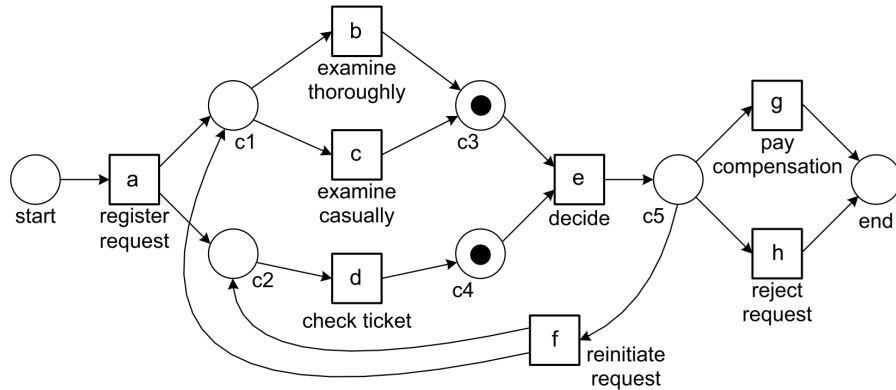
Figura 6 – Exemplo de uma rede de Petri



Fonte: Aalst (2016).

conforme ilustrado na Figura 7.

Figura 7 – Exemplo de uma rede de Petri



Fonte: Aalst (2016).

Neste momento, a rede encontra-se marcada nos lugares *c3* e *c4*, o que habilita a transição *e*.

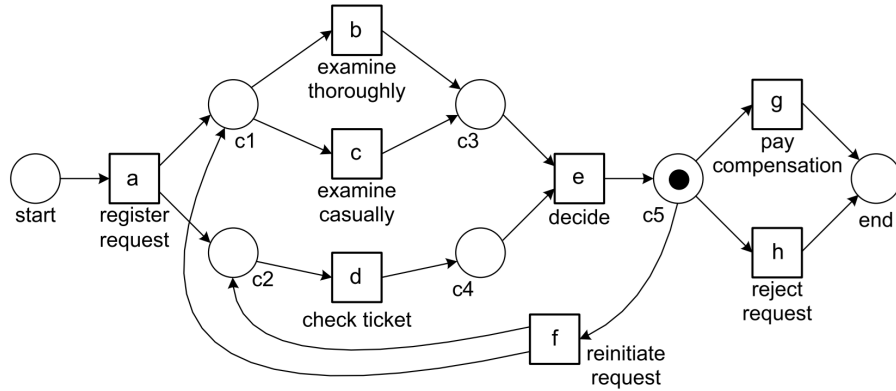
Ao disparar a transição *e*, os *tokens* presentes em *c3* e *c4* são consumidos, e um novo *token* é gerado em *c5*, conforme ilustrado na Figura 8.

Após o disparo da transição *e*, a rede fica marcada no lugar *c5*, habilitando as transições *f*, *g* e *h*. No entanto, apenas uma dessas transições será disparada.

Se a transição *f* for disparada, o *token* presente em *c5* será consumido, e novos *tokens* serão gerados em *c1* e *c2*, reiniciando o processo. Por outro lado, caso a transição *g* ou *h* seja disparada, o *token* de *c5* será consumido, e um novo será gerado no lugar *end*, indicando a finalização do processo.

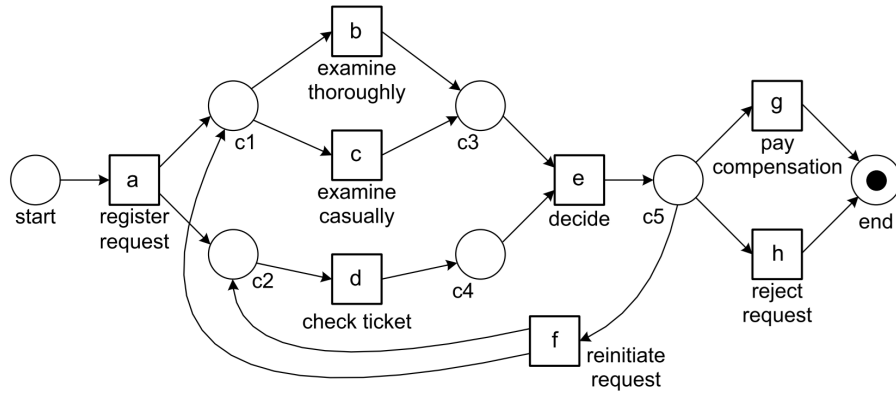
A Figura 9 ilustra o disparo da transição *g*, resultando na marcação do lugar *end*, o que representa o término do fluxo.

Figura 8 – Exemplo de uma rede de Petri



Fonte: Aalst (2016).

Figura 9 – Exemplo de uma rede de Petri



Fonte: Aalst (2016).

2.2.2 Workflow nets

As redes de Petri apresentadas anteriormente podem ser especializadas em uma classe denominada *Workflow nets* (WF-nets), amplamente utilizadas na modelagem de processos organizacionais e administrativos. Segundo Aalst (2016), uma WF-net é uma rede de Petri que garante a existência de um ponto único de início e de fim do processo, além de assegurar que todos os seus elementos estejam conectados dentro de um fluxo completo.

Definição 2.2.3 Aalst (2016) Seja $N = (P, T, F, A, l)$ uma rede de Petri rotulada e seja $\bar{i}$ um identificador tal que $\bar{i} \notin P \cup T$. A rede WN é dita uma *Workflow net* (WF-net) se, e somente se:

1. Existe um lugar de entrada $i \in P$ (também chamado de *source place*) tal que $\bullet i = \emptyset$;
2. Existe um lugar de saída $o \in P$ (também chamado de *sink place*) tal que $o \bullet = \emptyset$;
3. Seja $\bar{N} = (P, T \cup \{\bar{i}\}, F \cup \{(o, \bar{i}), (\bar{i}, i)\}, A \cup \{\tau\}, l \cup \{(\bar{i}, \tau)\})$, então $\bar{N}$ deve ser fortemente conexa, ou seja, deve existir um caminho direcionado entre qualquer par de nós em $\bar{N}$.

Nessa definição, o lugar i representa o ponto inicial do processo e o lugar o , o ponto

final. A condição de forte conectividade garante que todos os elementos da rede estão integrados a um fluxo completo, do início ao fim.

As WF-nets são especialmente relevantes na modelagem de processos acadêmicos, pois representam de forma natural o ciclo de vida de uma instância de processo — como a trajetória de um estudante desde o ingresso até a conclusão do curso. Cada aluno pode ser interpretado como um *token* navegando pela rede, e múltiplas execuções simultâneas são modeladas independentemente.

Dessa forma, a utilização de WF-nets neste trabalho permite representar e analisar formalmente o fluxo acadêmico, possibilitando a identificação de gargalos, desvios, ciclos e padrões de evasão ou conclusão.

2.3 Mineração de processos

A mineração de processos (*process mining*) é um elo entre a ciência de dados e da ciência de processos, tendo como objetivo usar os dados de eventos para extrair informações relacionadas ao processo Aalst (2016). Logo, é possível descobrir, monitorar e aprimorar processos reais.

Segundo Aalst (2016), os *logs* de eventos podem ser usados para conduzir três tipos de mineração de processos: descoberta de processos, verificação de conformidade e aprimoramento de processos.

- Descoberta de processos consistem em utilizar algoritmos para produzir um modelo representativo com as informações dos registros de eventos.
- Verificação de conformidade utiliza-se de um processo existente para compará os registro de eventos e assim, identificar o quanto o modelo representa os dados e vice-versa.
- Aprimoramento de processos visa melhorar um modelo de processo existente usando informações sobre o processo real no registrado de eventos.

Os dados brutos são o ponto de partida para a mineração de processos. Estes dados são coletados, de uma ou mais fontes, tratados e gerado um *log* de eventos. Segundo Aalst (2016), o *log* concentra os eventos de um processo que estão associados a uma atividade e se refere a uma única instância do processo, chamada de *case*. O *log* não se limita apenas ao registro de atividades, podendo incluir informações adicionais como data, hora, custos e recursos relacionados ao evento. A seguir, apresenta-se a definição formal de um *log*.

Definição 2.3.1 Aalst (2016) *Seja A um conjunto de finito de atividades. Um trace $\sigma \in A^*$ é uma sequência de atividades $\sigma = \langle a_1, a_2, \dots, a_n \in A^* \rangle$. Um log de eventos $L = \{\sigma^1, \sigma^2, \dots, \sigma^m\}$ é um conjunto de múltiplos traces sobre A .*

Aalst (2016) exemplifica um *log* de eventos apresentado na Tabela 2. A primeira coluna encontra-se o ID do *trace*, na segunda coluna ID do evento. Na terceira coluna o carimbo de data e hora em que ocorre o evento. Na quarta coluna a atividade realizada. Na quinta e sexta coluna encontro-se o responsável pela execução da atividade e o custo da atividade respectivamente.

Tabela 2 – Exemplo de um *log* de eventos.

<i>Case Id</i>	<i>Event Id</i>	<i>Timestamp</i>	<i>Activity</i>	<i>Resource</i>	<i>cost</i>
1	35654423	30-12-2010:11.02	register request	Pete	50
	35654424	31-12-2010:10.06	examine thoroughly	Sue	400
	35654425	05-01-2011:15.12	check ticket	Mike	100
	35654426	06-01-2011:11.18	decide	Sara	200
	35654427	07-01-2011:14.24	reject request	Pete	200
2	35654483	30-12-2010:11.32	register request	Mike	50
	35654485	30-12-2010:12.12	check ticket	Mike	100
	35654487	30-12-2010:14.16	Examine casually	Pete	400
	35654488	05-01-2011:11.22	decide	Sara	200
	35654489	08-01-2011:12.05	pay compensation	Ellen	200
3	35654521	30-12-2010:14.32	register request	Pete	50
	35654522	30-12-2010:15.06	examine thoroughly	Mike	400
	35654524	30-12-2010:16.34	check ticket	Ellen	100
	35654525	06-01-2011:09.18	decide	Sara	200
	35654526	06-01-2011:12.18	reinitiate request	Sara	200
	35654527	06-01-2011:13.06	examine thoroughly	Sean	400
	35654530	08-01-2011:11.43	check ticket	Pete	100
	35654531	09-01-2011:09.55	decide	Sara	200
	35654533	15-01-2011:10.45	pay compensation	Ellen	200
...					

Fonte: Aalst (2016)

Para simplificar a representação das atividades descritas na Tabela 2, foram adotadas as seguintes abreviações para cada atividade: *register request* (a), *examine thoroughly* (b), *examine casually* (c), *check ticket* (d), *decide* (e), *reinitiate request* (f), *pay compensation* (g), *reject request* (h) e construído o *trace* seguindo a ordem de ocorrência apresentado na coluna *Timestamp*, conforme apresentado na Tabela 3.

Cada *trace* contém um conjunto ordenado de atividades realizadas em um caso

Tabela 3 – Simplificação do *log* de eventos da Tabela 2.

<i>Case Id</i>	<i>Trace</i>
1	$\langle a, c, d, e, h \rangle$
2	$\langle a, d, c, e, g \rangle$
3	$\langle a, c, d, e, f, b, d, e, g \rangle$
4	$\langle a, d, b, e, h \rangle$
5	$\langle a, c, d, e, f, c, d, e, h \rangle$
6	$\langle a, c, d, e, g \rangle$
...	

Fonte: Aalst (2016)

ou instância do processo, representando o caminho efetivamente seguido desde o início até a conclusão da execução. A análise de *trace* permite identificar variações no fluxo do processo, mapear comportamentos comuns ou excepcionais e compreender como as instâncias realmente se desenrolam na prática, contribuindo para a detecção de padrões e potenciais pontos de melhoria no modelo de processo.

A partir dos *traces* que registram a sequência real de eventos do processo, diferentes abordagens podem ser aplicadas para compreender e aprimorar a execução dos fluxos processuais. Entre essas técnicas, destacam-se o *Play-In*, *Play-Out* e *Replay*, cada uma utilizando o *log* de eventos de maneira distinta.

O *Play-In* tem como objetivo a descoberta automática de um modelo de processo a partir do *log* de eventos, permitindo a geração de um fluxo baseado nas execuções reais observadas. O *Play-Out*, por sua vez, utiliza um modelo já estabelecido para simular a execução do processo, garantindo que ele siga a estrutura definida e verificando sua consistência. Já o *Replay* compara a execução real capturada nos *traces* com o modelo teórico, avaliando a conformidade do processo e identificando desvios, gargalos ou oportunidades de otimização. Essas técnicas são fundamentais para a análise e aprimoramento contínuo dos processos, fornecendo uma base para ajustes estratégicos e melhorias operacionais.

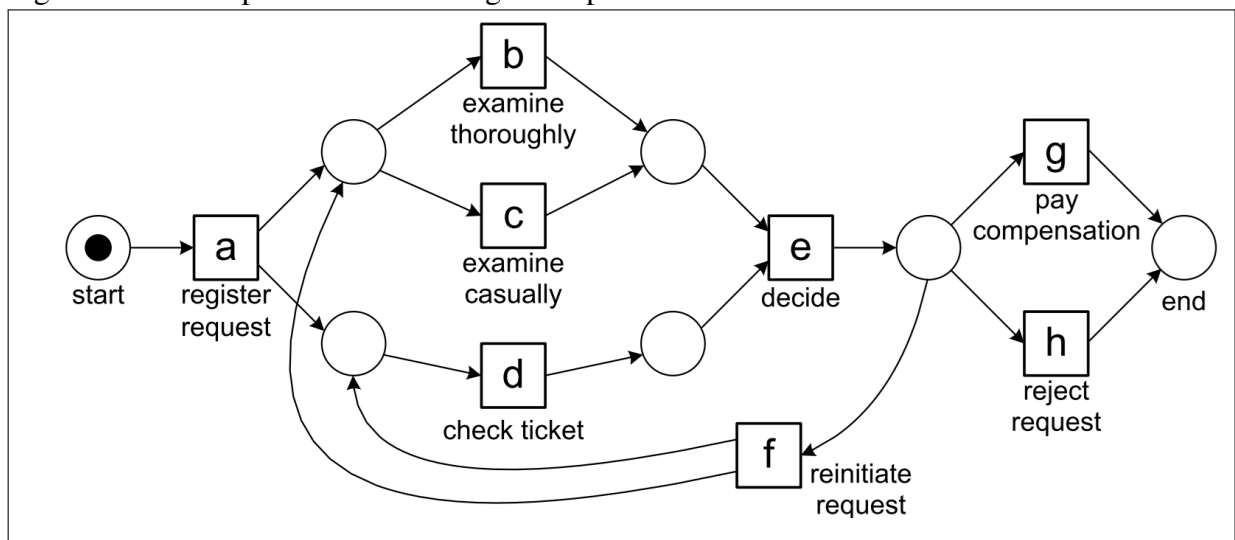
2.4 Critérios de avaliação de processos

Considerar que um modelo de processo seja capaz de capturar integralmente todo o comportamento registrado em um *log* de eventos seria uma suposição ingênua. Para avaliar o quanto o modelo representa adequadamente a realidade expressa no *log*, a área de mineração de processos utiliza quatro métricas principais de qualidade de modelos: adequação, precisão, generalização e simplicidade Munoz-Gama *et al.* (2014).

- Adequação (*Fitness*): esta métrica representa o quanto o modelo incorporou as atividades do *log* de eventos, ou seja, todos os *traces* são reproduzidos no modelo do início ao fim.
- Precisão (*Precision*): esta métrica avalia e quantifica os comportamentos inexistentes previsto pelo modelo.
- Generalização (*Generalization*): esta métrica avalia o comportamento presente no *log* de eventos e a capacidade reproduzir *traces* futuros.
- Simplicidade (*Simplicity*): esta métrica avalia a objetividade do modelo, ou seja, a utilização do menor número de componentes para modelar o comportamento do *log* de eventos.

A Figura 10 apresenta o modelo já descrito em 2.2.1. Do ponto de vista da métrica *Fitness*, o modelo reflete com precisão os *traces* da Tabela 3. Ao analisar o *trace* $\langle a, b, d, e, h \rangle$, pode-se observar sua execução completa, iniciando pela atividade *register request*. Em seguida, ocorre a execução paralela *examine thoroughly* e *check ticket*. Após essas etapas, a atividade *decide* é realizada, finalizando com *pay compensation*, concluindo assim o fluxo do processo. Utilizando o mesmo *trace*, o modelo apresentaria uma baixa *precision* caso permitisse a execução de um *trace* que não foi registrado no *log* de eventos. Por exemplo, se o modelo aceitasse o *trace* $\langle a, b, d, h \rangle$, no qual a atividade *decide* é pulada, isso indicaria um comportamento incorreto, já que, de acordo com o *log*, essa atividade é essencial antes da conclusão do processo. Permitir essa variação sem a devida execução de *decide* refletiria uma baixa precisão do modelo, ao permitir um fluxo não observado na realidade.

Figura 10 – Exemplo de uma modelagem de processo



Fonte: Aalst (2016).

A generalização permite o modelo executar *traces* que não estão presentes no *log* de eventos mas que podem ocorrer na realidade. Por exemplo, o *trace* $\langle a, c, d, e, g \rangle$ está registrado no *log* e pode ser executado pelo modelo. No entanto, se em uma variação ocorrer no processo, e a atividade *examine thoroughly* não for executada e a sequência seguir diretamente da atividade *check ticket* para *decide*, e mantiver as regras do processo, o modelo demonstraria uma boa capacidade de generalização. Isso indica que o modelo é flexível o suficiente para lidar com cenários imprevistos.

A simplicidade refere-se ao quão objetivo e compreensível o modelo de processo é, sem ser complexo ou detalhado. Por exemplo, imagine um processo simples no qual um pedido é registrado, revisado e aprovado ou rejeitado. Se o modelo incluir diversas exceções ou ramificações desnecessárias para cada pequena variação observada no *log* de eventos, ele se tornaria muito complexo e difícil de interpretar.

Para avaliar a conformidade entre o comportamento observado no *log* de eventos e o modelo de processo formalizado, são utilizadas técnicas de verificação de conformidade (*conformance checking*) como *token replay* e *alignments*. Essas técnicas permitem identificar divergências, validar hipóteses e fornecer métricas quantitativas que expressam a aderência entre o modelo teórico e a execução real do processo.

2.4.1 Verificação de conformidade por token replay

A verificação de conformidade por meio do *token replay* requer que cada *trace* no *log* de eventos esteja alinhado com uma sequência de execução válida do modelo de processo, conforme descrito por Aalst (2016). Esse método permite avaliar o quão bem o processo segue o fluxo esperado, verificando a aderência de cada sequência de eventos ao modelo teórico do processo. Durante a execução do *token replay*, o *trace* é percorrido passo a passo, respeitando a ordem das atividades registradas no *log* de eventos e sua correspondência com ao modelo de processo utilizado.

Ao longo dessa execução, dois cenários podem ser observados, indicando possíveis desalinhamentos entre o comportamento real presente no *log* de eventos e o modelo de referência: (i) ausência de *token* no lugar de entrada no momento da execução de uma atividade, sugerindo que a transição foi ativada sem que os requisitos prévios fossem cumpridos, o que pode refletir inconsistências ou desvios no fluxo do processo; (ii) produção de *token* no lugar de saída após a execução de uma atividade sem que estes sejam consumidos posteriormente, evidenciando a

criação de estados no modelo que não são devidamente utilizados na sequência lógica esperada.

A principal característica desse método é a noção de *fitness*, que avalia o grau de conformidade de cada *trace* com o modelo de fluxo de processo estabelecido. Essa métrica é calculada no nível de cada instância individual do processo, permitindo medir o quão bem uma instância específica segue a sequência de atividades definidas no modelo.

O *fitness* de um *trace* avalia o alinhamento de uma instância do processo ao modelo de referência. Ele quantifica a proporção de eventos registrados no *log* que podem ser reproduzidos corretamente no modelo. Quanto maior o *trace fitness*, maior a aderência da instância do processo ao fluxo esperado. Essa métrica é formalizada pela Equação 2.1, aplicada individualmente a cada instância do processo.

$$fitness(\sigma, N) = 0,5 \left(1 - \frac{m}{c} \right) + 0,5 \left(1 - \frac{r}{p} \right) \quad (2.1)$$

Onde:

- σ : representa cada *trace* do *log* de eventos.
- N : é o modelo de processo representado pela rede de Petri.
- m : quantidade de *tokens* inseridos a força;
- c : quantidade de *tokens* consumidos;
- r : quantidade de *tokens* ao final do *Replay*;
- p : quantidade de *tokens* produzidos;

Já o *fitness* do *log* completo avalia a conformidade do *log* de eventos como um todo, agregando os valores de *trace fitness* de todas as instâncias do processo para fornecer uma visão geral da aderência do fluxo de execução ao modelo de referência. Um alto valor de *fitness* indica que a maioria das instâncias do processo seguiu a sequência de atividades conforme especificado no modelo, enquanto valores baixos sugerem a ocorrência frequente de desvios, reiteraões ou interrupções no fluxo processual. O cálculo dessa métrica é formalizado pela Equação 2.2, permitindo quantificar a adequação do *log* de eventos ao modelo do processo.

$$fitness(L, N) = 0,5 \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times m_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times c_{N,\sigma}} \right) + 0,5 \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N,\sigma}}{\sum_{\sigma \in L} L(\sigma) \times p_{N,\sigma}} \right) \quad (2.2)$$

Onde:

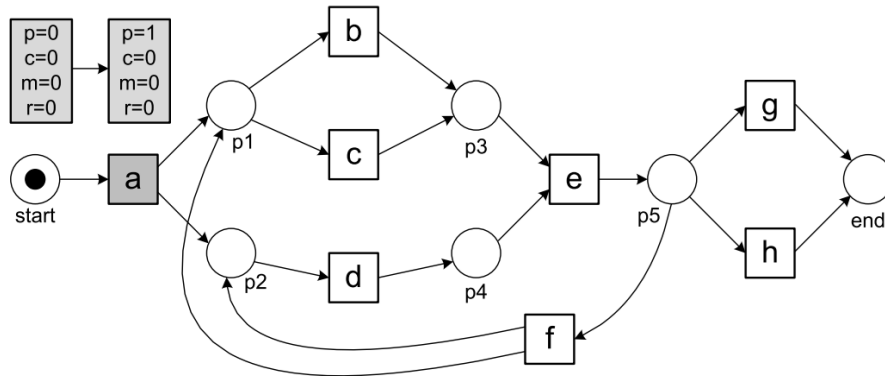
- $L(\sigma)$: representa a frequência do *trace* σ no *log* de eventos L .
- N : é o modelo de processo representado pela rede de Petri.

- $\sigma \in L$: representa cada *trace* do *log* de eventos.
- $m_{N,\sigma}$ número total de *tokens* ausentes durante a execução do *trace* σ no modelo N .
- $c_{N,\sigma}$ número total de *tokens* consumidos durante a execução do *trace* σ .
- $r_{N,\sigma}$ número total de *tokens* remanescentes no modelo ao final da execução do *trace* σ .
- $p_{N,\sigma}$ número total de *tokens* produzidos durante a execução do *trace* σ .

Aalst (2016) apresenta um exemplo de cálculo do *trace fitness* aplicado ao modelo representado na Figura 10, observando a produção, consumo, ausências de *tokens* e *tokens* remanescente ao final do *token replay*. Considere a ocorrência do *trace* $\sigma_1 = \langle a, c, d, e, h \rangle$ no *log* de eventos, conforme descrito na Tabela 3.

No estado inicial, o modelo insere um *token* no lugar *start*, modificando a marcação da rede com $p = 0 + 1 = 1$ e demais zero, conforme ilustrado na Figura 11. Esse *token* indica que a transição *a* está habilitada para ser disparada, desde que esteja presente no *trace*.

Figura 11 – Passo um: *Replay* do *trace* $\sigma_1 = \langle a, c, d, e, h \rangle$ em modelo PN_1



Fonte: Aalst (2016).

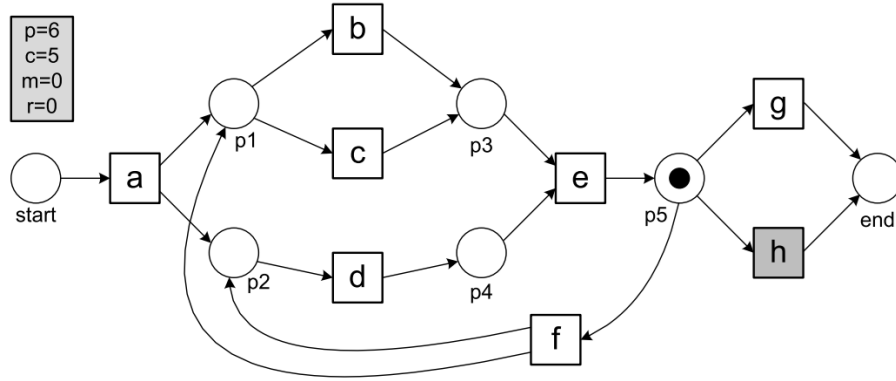
Assim que a transição *a* ocorre no *trace*, o *token* presente em *start* é consumido, e novos *token* são produzidos nos lugares de saída de *a*, que, neste caso, são *p1* e *p2*, conforme ilustrado na Figura 12. Ao mudar a marcação da rede, os valores de $p = 1 + 2 = 3$ e $c = 0 + 1 = 1$ são atualizados.

Dando continuidade à execução do *trace*, a próxima atividade a ser disparada é *c*. Quando essa transição ocorre, o *token* presente no lugar de entrada (*p1*) é consumido, e um novo *token* é produzido no lugar de saída, *p3*, conforme ilustrado na Figura 13, sendo atualizado os valores de $p = 3 + 1 = 4$ e $c = 1 + 1 = 2$.

O lugar de entrada (*p2*) da atividade *d* contém um *token*, o que a torna habilitada para disparo. Como a próxima atividade na sequência do *trace* é *d*, essa transição é disparada, consumindo o *token* de *p2* e produzindo um novo *token* no lugar de saída *p4*, conforme ilustrado

Ao disparar e , os *tokens* de $p3$ e $p4$ são consumidos, e um novo *token* é produzido no lugar de saída $p5$. Como consequência, os valores são atualizados para $p = 5 + 1 = 6$ e $c = 3 + 2 = 5$, refletindo a evolução do processo conforme ilustrado na Figura 15.

Figura 15 – Passo cinco: *Replay* do trace $\sigma_1 = \langle a, c, d, e, h \rangle$ em modelo PN_1



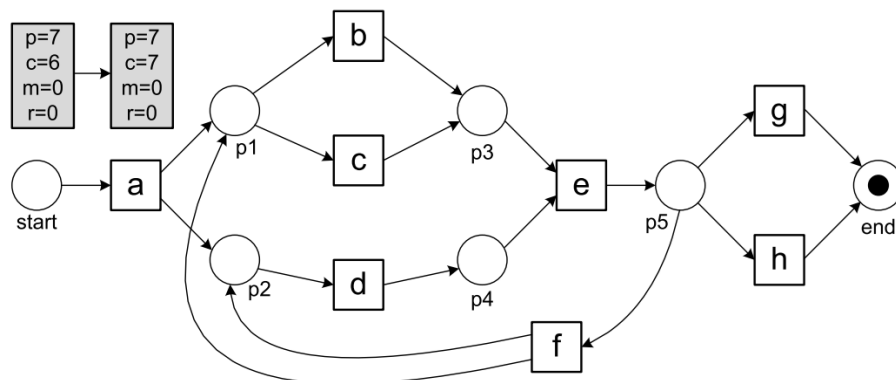
Fonte: Aalst (2016).

Com um *token* em $p5$, que é o lugar de entrada para três transições, inicia-se uma disputa pelo consumo desse *token*, habilitando simultaneamente as transições f , g e h . Caso a transição f seja disparada, o *token* de $p5$ será consumido, e dois novos *tokens* serão produzidos nos lugares $p1$ e $p2$, reiniciando a execução do processo a partir desses pontos.

Por outro lado, caso g ou h sejam disparadas, o *token* consumido de $p5$ resultará na produção de um novo *token* no lugar end , indicando a finalização do processo.

Como a última atividade no trace é h , a transição correspondente é disparada, consumindo o *token* de $p5$ e produzindo um novo *token* no lugar end . Esse evento provoca a atualização dos valores para $p = 6 + 1 = 7$ e $c = 5 + 1 = 6$, conforme ilustrado na Figura 16.

Figura 16 – Passo seis: *Replay* do trace $\sigma_1 = \langle a, c, d, e, h \rangle$ em modelo PN_1



Fonte: Aalst (2016).

Ao final do *token replay*, o *token* presente no lugar end é consumido, resultando

na última atualização de valores para $p = 7$ e $c = 7$. Com isso, a rede de Petri é reiniciada, permitindo que um novo *trace* seja simulado.

Em resumo, ao final da execução do *trace* $\sigma_1 = \langle a, c, d, e, h \rangle$, utilizando a técnica de *token replay*, foram produzidos 7 *tokens* e consumidos 7 *tokens*. Não foram identificados *tokens* ausentes, e nenhum *token* residual permaneceu ao término do *token replay*, indicando que a execução seguiu exatamente o fluxo modelado na rede de Petri. Dessa forma, os valores de p , c , m e r podem ser aplicados na Equação 2.1 para calcular o *fitness* do *trace* σ_1 :

$$0,5 \left(1 - \frac{0}{7} \right) + 0,5 \left(1 - \frac{0}{7} \right) = 1$$

O valor resultante da Equação 2.1 varia entre 0 e 1, onde:

- 0 indica que o *trace* não é reconhecido pelo modelo, ou seja, nenhuma de suas transições é suportada pela estrutura da rede de Petri.
- 1 significa que o *trace* é completamente compatível com o modelo, sem qualquer desvio ou necessidade de ajustes.

Ao aplicar os valores na Equação 2.1, verificamos que o resultado obtido é 1, indicando que todas as transições do *trace* σ_1 foram corretamente representadas e reconhecidas pelo modelo. Isso demonstra que a execução seguiu exatamente o fluxo esperado, sem a necessidade de inserção ou remoção de *tokens* para compensar divergências.

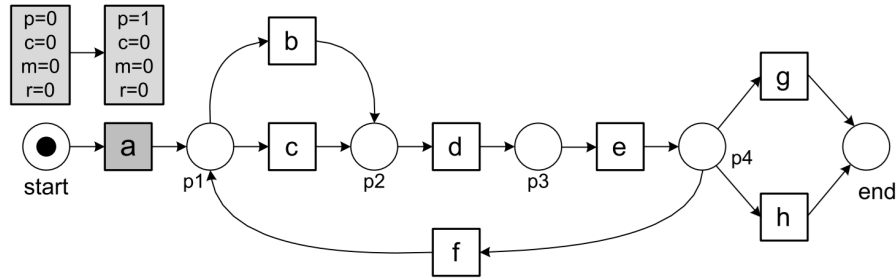
Aalst (2016) também apresenta um cenário em que o cálculo do *trace fitness* resulta em um valor inferior a 1, indicando a existência de desvios entre o *trace* e o modelo do processo. Para ilustrar essa situação, consideremos o *trace* $\sigma_2 = \langle a, d, c, e, h \rangle$ e o modelo de processo PN_2 , representado na Figura 17.

No estado inicial, um *token* é produzido no lugar *start*, modificando a marcação da rede. Como consequência, o valor de p é atualizado para 1, enquanto os demais parâmetros permanecem 0. Esse evento representa o primeiro passo da execução dentro da rede de Petri, preparando o sistema para a sequência de transições subsequentes.

Na sequência do *trace*, a atividade a é executada. Como resultado, o *token* presente no lugar de entrada (*start*) é consumido, e um novo *token* é produzido no lugar de saída $p1$, conforme ilustrado na Figura 18. Com essa transição, os valores da marcação da rede são atualizados para $p = 2$ e $c = 1$, refletindo a evolução do processo dentro da rede de Petri.

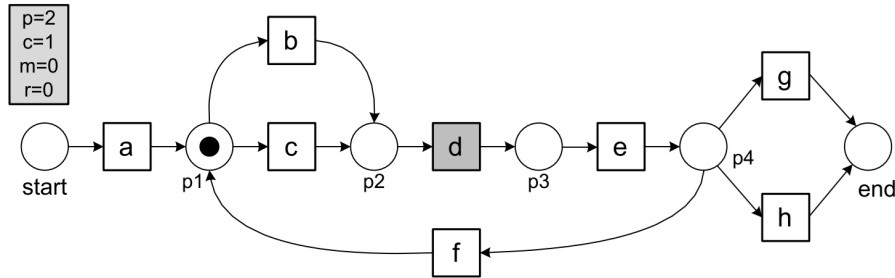
Após a execução da transição a , as transições c e b passam a disputar o *token* presente no lugar de entrada $p1$, uma vez que ambas se tornam habilitadas. No entanto, surge

Figura 17 – Passo um: *Replay* do trace $\sigma_2 = \langle a, d, c, e, h \rangle$ em modelo PN_2



Fonte: Aalst (2016).

Figura 18 – Passo um: *Replay* do trace $\sigma_2 = \langle d, d, c, e, h \rangle$ em modelo PN_2



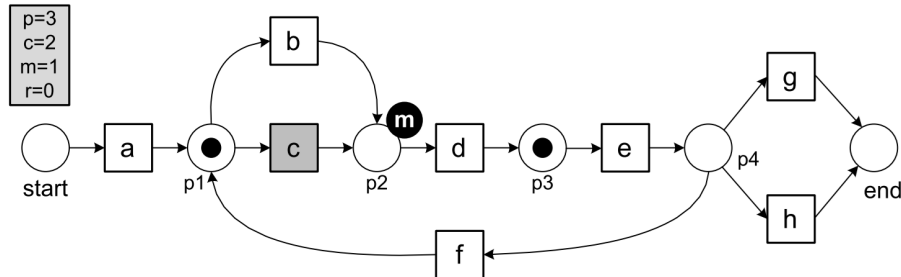
Fonte: Aalst (2016).

um problema: a sequência de atividades registrada no *trace* não segue com *c* ou *b*, mas sim com a transição *d*.

Para que *d* possa ser disparada, seria necessário que houvesse um *token* no seu lugar de entrada, *p2*. No entanto, como esse *token* não está presente, *d* não poderia ser ativada naturalmente. Para permitir que a execução do *trace* continue, um *token* é adicionado forçadamente no lugar de entrada *p2*, permitindo que *d* seja disparada e produza um *token* em *p3*.

Essa ação caracteriza um desvio na adequação entre o *trace* e o modelo, pois indica que, para seguir a sequência do *trace*, foi necessário inserir um *token* ausente. Como consequência, o valor da métrica *m* (*missing tokens*) é atualizado para 1, refletindo a necessidade de inserção desse *token* artificial para corrigir o fluxo da execução. Além disso, os valores são ajustados para $p = 3$ e $c = 2$, conforme ilustrado na Figura 19.

Figura 19 – Passo um: *Replay* do trace $\sigma_2 = \langle d, d, c, e, h \rangle$ em modelo PN_2



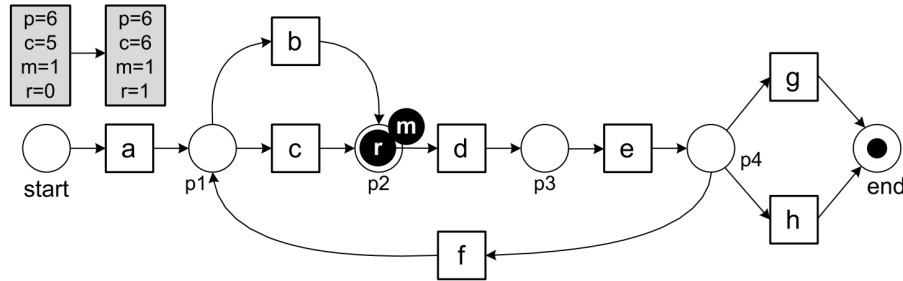
Fonte: Aalst (2016).

por nenhuma transição. A única transição que poderia consumi-lo seria d , mas ela não voltou a ser executada ao longo do processo, ou seja, não reapareceu na sequência do trace σ_2 . Isso caracteriza a presença de um *token* remanescente, indicando que houve um desvio na execução em relação ao modelo.

Como consequência, o valor da métrica r (remaining tokens) é atualizado para 1, e os valores da rede são ajustados para $p = 6$, $c = 5$ e $m = 1$.

Por fim, o *token* do lugar end é consumido, resultando na atualização para $c = 6$, e a rede de Petri é reiniciada para a simulação do próximo *trace*, conforme ilustrado na Figura 22.

Figura 22 – Passo um: *Replay* do trace $\sigma_2 = \langle a, d, c, e, h \rangle$ em modelo PN_2



Fonte: Aalst (2016).

Ao término da simulação do *token replay* para o trace σ_2 utilizando o modelo PN_2 , os valores finais dos parâmetros são: $p = 6$, $c = 6$, $m = 1$ e $r = 1$. Com esses valores, podemos aplicar a Equação 2.1 para calcular o *fitness* do trace σ_2 , que resulta em 0,833.

$$0,5 \left(1 - \frac{1}{6} \right) + 0,5 \left(1 - \frac{1}{6} \right) = 0,833$$

Esse valor reflete um desvio parcial na execução do *trace* em relação ao modelo, indicando a necessidade de inserção de *tokens* ausentes (m) e a presença de *tokens* remanescentes (r), o que impacta a conformidade do *log* de eventos com a rede de Petri.

2.4.2 Verificação de conformidade por alinhamento

Outra técnica utilizada na análise de conformidade é o alinhamento (*alignment*). De acordo com Carmona *et al.* (2018), o alinhamento consiste em estabelecer uma correspondência entre um *trace* do *log* de eventos e uma sequência de execução do modelo. Esse alinhamento é representado por uma matriz de duas linhas, onde: (i) a primeira linha corresponde ao *trace* registrado no *log* de eventos e (ii) a segunda linha representa a sequência de execução esperada

pelo modelo. A Figura 23 representa um alinhamento ótimo, nele é possível verificar que a cada movimento (transição) presente no *trace* existe uma correspondência no modelo.

Figura 23 – Alinhamento do *trace* $\sigma = \langle a, c, d, e, h \rangle$

$$\gamma_1 = \begin{array}{|c|c|c|c|c|} \hline a & d & b & e & h \\ \hline a & d & b & e & h \\ \hline \end{array}$$

Fonte: Aalst (2016).

Por meio dessa estrutura, é possível verificar se as atividades registradas no *log* coincidem com as definidas no modelo, identificando correspondências e desvios. Isso permite analisar se uma determinada atividade ocorreu conforme o previsto, se houve a ausência de uma atividade esperada ou, ainda, se ocorreu uma atividade não prevista no modelo.

Para cada movimento realizado no alinhamento, utiliza-se a notação $\gg$ para indicar quando há uma não correspondência entre as atividades. Caso esse desalinhamento ocorra no modelo — isto é, existe um movimento presente no *log* de eventos, mas não há, naquele momento, uma atividade prevista no modelo — a marcação $\gg$ aparece na linha superior. Por outro lado, se o movimento registrado no *log* não puder ser imitado por uma transição correspondente no modelo, a marcação $\gg$ é inserida na linha inferior, conforme ilustrado na Figura 24.

Figura 24 – Alinhamento do *trace* $\sigma = \langle a, c, d, e, h \rangle$

$$\gamma_{2a} = \begin{array}{|c|c|c|c|c|} \hline a & \gg & d & b & e & h \\ \hline a & b & d & \gg & e & h \\ \hline \end{array} \quad \gamma_{2b} = \begin{array}{|c|c|c|c|c|} \hline a & \gg & d & b & e & h \\ \hline a & c & d & \gg & e & h \\ \hline \end{array} \quad \gamma_{2c} = \begin{array}{|c|c|c|c|c|} \hline a & d & b & \gg & e & h \\ \hline a & \gg & b & d & e & h \\ \hline \end{array}$$

Fonte: Aalst (2016).

Portanto, um movimento é representado como um par $(x, (y, t))$, em que x se refere ao movimento do *log* e (y, t) ao movimento do modelo. Para exemplificar, consideremos a presença de uma atividade a tanto no *log* quanto no modelo. Esse alinhamento é representado por $(a, (a, t_1))$, ou seja, o *log* realiza um movimento a e o modelo também o realiza por meio da transição t_1 . Caso a atividade a esteja prevista apenas no modelo, e não ocorra no *log*, teremos a representação $(\gg, (a, t_1))$, indicando que a é prevista pelo modelo, mas não está presente no *log*. Por fim, se a estiver presente apenas no *log*, mas não puder ser simulada por nenhuma transição no modelo, a representação será $(a, \gg)$, indicando a ausência de correspondência no modelo para essa atividade.

Segundo Aalst (2016), um movimento é considerado legal se atender a um dos quatro casos abaixo:

1. $x = y$ e y é uma atividade representada por uma transição no modelo (movimento síncrono);
2. $x = \gg$ e y é uma atividade prevista pelo modelo (movimento do modelo);
3. $x = \gg$, $y = \tau$ e a transição t é silenciosa (movimento invisível do modelo);
4. $x \neq \gg$ e $(y, t) = \gg$ (movimento do *log*).

A partir desses movimentos legais, é possível construir uma sequência estruturada chamada alinhamento (*alignment*), que estabelece uma correspondência passo a passo entre os eventos observados no *log* e os eventos esperados pelo modelo. Essa estrutura é essencial para avaliar a conformidade entre o comportamento real e o comportamento modelado. Aalst (2016) define um alinhamento como:

Definição 2.4.1 *Aalst (2016) Um alinhamento (*alignment*) é uma sequência de movimentos legais tal que, ao remover todos os símbolos $\gg$ (que indicam desalinhamentos), a linha superior corresponde à sequência de eventos do *log* (ou seja, o *trace* real executado), e a linha inferior corresponde a uma sequência de disparos válida no modelo de processo, levando de um estado inicial a um estado final.*

Para selecionar o alinhamento mais adequado, associamos os custos ao efeito indesejados do movimento e seleciona um alinhamento com os menores custo (*costs*), isto é, o caminho mais eficiente que partir do início chega ao fim do processo. Segundo Aalst (2016) a função de custo δ atribui valores aos movimentos legais, que são:

1. Movimentos nos quais o *log* e o modelo concordam não possuem custo, ou seja, $\delta(x, (y, t)) = 0$ para movimentos síncronos (com $x = y$).
2. Movimentos realizados apenas pelo modelo não possuem custo se a transição for invisível, isto é, $\delta(\gg, (\tau, t)) = 0$ para movimentos invisíveis do modelo.
3. Já $\delta(\gg, (y, t)) > 0$ representa o custo quando o modelo realiza um movimento “ y ” sem uma correspondência no *log* (movimento visível do modelo).
4. $\delta(x, \gg) > 0$ é o custo de um movimento “ x ” presente apenas no *log* (movimento do *log*).

O valor do custo de desalinhamento depende da natureza da atividade. A função de custo padrão, denotada por δ , atribui custo 1 a todos os movimentos visíveis do modelo e do *log*, ou seja, $\delta(\gg, (y, t)) = 1$ e $\delta(x, \gg) = 1$. Para cada símbolo $\gg$ presente na linha superior (relativa ao *trace*) ou inferior (relativa ao modelo), contabiliza-se um custo. Um alinhamento é considerado ótimo quando não há nenhuma alternativa com custo total menor.

Esses custos podem ser convertidos em um valor de *fitness* entre 0 e 1, onde 0 representa um alinhamento completamente desalinhado (custo máximo) e 1 representa um

alinhamento perfeito (sem custo). Aalst (2016) define o *fitness* de um único *trace* conforme a Equação 2.3:

$$fitness(\sigma, N) = 1 - \frac{\delta(\lambda_{opt}^N(\sigma))}{\delta(\lambda_{worst}^N(\sigma))} \quad (2.3)$$

Onde:

- σ : Um *trace* do *log* de eventos.
- N : A rede de Petri com a qual o *trace* será alinhado.
- λ_{opt}^N : O alinhamento ótimo entre o *trace* σ e o modelo N , ou seja, a melhor forma de ajustar o *trace* ao modelo com menor custo.
- λ_{worst}^N : O pior alinhamento possível, que representa o maior custo que poderia ser gerado para o *trace* σ .
- $\delta(x)$: Função de custo do alinhamento, que soma os custos dos movimentos.
- $fitness(\sigma, N)$: Grau de conformidade entre o *trace* σ e o modelo N .

A definição de *fitness* pode ser estendida para todo o *log* de eventos, conforme a Equação 2.4:

$$fitness(L, N) = 1 - \frac{\sum_{\sigma \in L} L(\sigma) \times \delta(\lambda_{opt}^N(\sigma))}{\sum_{\sigma \in L} L(\sigma) \times \delta(\lambda_{worst}^N(\sigma))} \quad (2.4)$$

onde:

- L : o *log* de eventos, conjunto de todos os *trace*.
- $\sigma \in L$: representa cada *trace* do *log* de eventos.
- $L(\sigma)$: representa a frequência do *trace* σ no *log* de eventos L .
- λ_{opt}^N : O alinhamento ótimo entre o *trace* σ e o modelo N , ou seja, a melhor forma de ajustar o *trace* ao modelo com menor custo.
- λ_{worst}^N : O pior alinhamento possível, que representa o maior custo que poderia ser gerado para o *trace* σ .
- $\delta(x)$: Função de custo do alinhamento, que soma os custos dos movimentos.
- $fitness(L, N)$: Grau de conformidade entre o *trace* σ e o modelo N .

Note que $\sum_{\sigma \in L} L(\sigma) \times \delta(\lambda_{opt}^N(\sigma))$ representa a soma de todos os custos ao reproduzir todo o *log* de eventos utilizando os alinhamentos ótimos. Esse valor é então dividido pelo pior cenário possível para se obter um valor de *fitness* global normalizado entre 0 e 1.

3 TRABALHOS RELACIONADOS

Neste capítulo será abordado trabalhos que utilizam técnicas diversas para prever, analisar, estudar a evasão escolar no âmbito universitário.

3.1 Análise da trajetória de aprendizagem do aluno no ensino a distância por meio da mineração de processos

O trabalho de Unger *et al.* (2022) utiliza mineração de processos no contexto da educação a distância com intuito de analisar a trajetória de aprendizagem executado pelos estudantes da Universidade Virtual do Estado de São Paulo (UNESP) dentro de um ambiente virtual de aprendizagem (AVA). Para isto, os autores utilizaram *Process Mining Project Methodology (PM²)* que consiste em seis estágios: planejamento, extração, processamento de dados, mineração e análise, avaliação e refinamento.

Os autores adotaram o método indutivo para analisar os *logs* de atividades dos alunos no AVA em 4 disciplinas e em 7 cursos, ao longo de 4 bimestres, classificando-as em grupos de humanas e exatas. As disciplinas escolhidas foram *Didática*, *Algoritmos* e *Programação de Computadores I*, *Sistemas de Informação* e *Administração I*. Essas disciplinas foram selecionadas por estarem no início do curso, uma fase em que os alunos geralmente apresentam maior motivação, e devido ao alcance abrangente que elas possuem.

Os *logs* de eventos, após tratados, foram aplicados na ferramenta Celonis e foi gerados modelos representativos das trajetórias dos estudantes. Assim, foi possível inferir que as cinco atividades com maior execução foram: *wiki*, *quizzes*, *external_urls*, *topics* e *assgnmentes*. Outro conhecimento adquirido foi identificado no acesso a conteúdos das disciplinas (*wiki*), onde os alunos dos cursos de exatas apresentam 30% mais acesso em comparação ao grupo das humanas. Por outro lado, os alunos das humanas acessam duas vezes mais os fóruns de discussão (*topics*) do que os alunos de exatas.

Outra análise consiste no comportamento dos alunos aprovados e reprovados, pois o baixo engajamento nas atividades propostas, menor duração do processo, abandono na disciplina, maior presença de gargalos e grandes períodos sem nenhum acesso ao AVA é característica dos alunos reprovados.

3.2 Identificação de trajetórias de aprendizagem em um curso de graduação e sua relação com a evasão escolar

O trabalho de Carmo *et al.* (2022) utiliza técnicas de mineração de dados e grafos para investigar a relação entre a evasão e as trajetórias de aprendizagem dos estudantes do curso de Ciências da Computação da Universidade do Estado de Santa Catarina (UDESC). Para isto, é adotada duas abordagens distintas: estruturada e visual. Ao utilizar a abordagem estruturada, os autores são capazes de analisar a trajetória individual do estudante, isto é, o estudante será representado por uma sequência $s = A, B, C, D, \dots$, de modo que A, B, C e D representam o caminho/disciplinas e que esteja ordenado por semestre letivo e alfabeticamente. Enquanto que utilizar a abordagem visual, é possível visualizar um aglomerado de trajetórias estudantis, sendo possível identificar a relação entre concluintes e evadidos.

A base de dados utilizada pelos autores, consiste nas informações demográficas e de ingresso, créditos cumpridos e histórico das disciplinas contento o *status* final do estudante, ao final, o conjunto de dados era formado por 218 estudantes matriculados entre 2009 e 2020, sendo 50 com status de concluintes e 168 evadidos. Para a abordagem estruturada, os autores utilizaram o algoritmo *GSP (Generalized Sequential Pattern)* disponível no software *Weka*.

A análise consistiu em duas partes: a primeira, analisando três dimensões (disciplina, semestre e status), sendo possível inferir que menos de 70% dos estudantes concluintes cursam e são aprovados em 5 das 6 disciplinais do primeiro semestre. Nesta etapa, não foi possível identificar padrões no grupos de alunos evadidos, por isso foi analisado em segunda etapa, onde a análise consiste em duas dimensões (disciplina e semestre, disciplina e status), sendo possível detectar que os alunos evadidos reprovam nas disciplinas *Álgebra Linear* e *Geometria Analítica* e *Cálculo Diferencial I* ao verificar as dimensões disciplina e status.

Para a abordagem visual, os autores utilizaram o *Gephi* e, nesta análise, é possível observar o comportamento dos alunos evadidos, sendo estudantes que desiste antes do início do segundo semestre, cursam apenas as seis primeiras disciplinas, cursam as disciplinas fora do proposto pelo currículo acadêmico.

Os resultados do estudo destacam que os estudantes, tanto concluintes quanto evadidos, tendem a seguir trajetórias diferentes da grade curricular proposta. Além disso, observa-se que a maioria dos alunos que evadem o fazem ao final do primeiro semestre, após reprovações nas disciplinas iniciais. Aqueles que permanecem enfrentam dificuldades para avançar no curso, já que as disciplinas iniciais são pré-requisito para as demais.

3.3 Mineração de dados educacionais para predição da evasão em cursos de graduação presenciais no ensino superior

O trabalho de Kantorski *et al.* (2023) realiza uma análise de métodos de mineração de dados educacionais (MDE) e algoritmos de aprendizado de máquina para classificar o risco de evasão dos cursos de graduação presencial. Para isto, os autores realizam uma adaptação da metodologia *Cross Industry Standard Process for Data Mining* (CRISP-DM) ao trabalho.

O conjunto de dados utilizados como entrada nos algoritmos era composto por 106 cursos de graduação de três semestres letivos entre 2022 a 2023.1, totalizando 61.644 instâncias de treino e 15.367 de teste. Os algoritmos treinados foram: *C4.5*, *CART*, *RIPPER*, *Naïve Bayes*, *Multilayer perceptron* e *KNN*. Foi utilizado como métricas de avaliação a acurácia, precisão, *recall* e *f-measure*, buscando o melhor resultado de *recall* em compensação à precisão, isto é, o número de acertos de evasão seja o maior possível e o maior risco de evasão seja o menor possível.

O algoritmo com maior *recall* foi *C4.5* com 97% e menor *Naïve Bayes* com 87% e a precisão de manteve em praticamente 2% em todos os algoritmos. Ao analisar a interseção entre as previsões dos algoritmos foi inferido uma acurácia de 53%, precisão de 2% e *recall* de 88%.

Isso demonstra que a identificação de potenciais evasões é viável e relevante em uma análise abrangente dos cursos da instituição. O objetivo foi não apenas detectar a evasão, mas também identificar alunos com perfil semelhante aos que abandonaram o curso, permitindo identificar futuras evasões.

3.4 Análise de dados pré-universitários para predizer a evasão de alunos ingressantes em uma instituição de ensino superior

O trabalho de Falcão *et al.* (2023) utiliza técnica de mineração de dados para classificar o risco de evasão utilizando dados pré-universitários dos estudantes ingressantes em cursos de graduação de universidades públicas. Semelhante ao trabalho Kantorski *et al.* (2023), os autores utilizaram a metodologia *Cross Industry Standard Process for Data Mining* (CRISP-DM).

O conjunto de dados era composto por informações de matrículas entre 2014 e 2018 e notas do ensino médio nas disciplinas de matemática, português, química, física, biologia, história e geografia, totalizando 1086 registros com 566 com status de evadidos e 520 com status

não evadidos. Para a construção do modelo foi adotado o algoritmo J48 disponível no *Weka*.

Os resultados obtidos apresentaram-se promissores, embora não sendo o resultado ideal esperado pelos autores. O modelo conseguiu indicar alguns fatores que contribuem para a evasão dos estudantes, como: turno, disciplina de química, região de moradia.

3.5 Modelo de predição de evasão escolar com base em dados de autoavaliação de cursos de graduação

O trabalho de Oliveira e Medeiros (2024) tem como objetivo utilizar as autoavaliações realizada pelos estudantes de ensino superior para a predição do risco de evasão, adotando técnicas de mineração de dados educacionais e aprendizagem de máquina. Para atingir a finalidade proposta, os autores utilizaram a metodologia *CRoss-Industry Standard Process for Educational Data Mining* (CRISP-EDM). Realizada semestralmente e de forma compulsória, a autoavaliação coleta informações dos estudantes durante o período letivo concluído.

Os autores utilizaram os algoritmos *Support Vector Machine* (SVM), *Random Forest* (RF) e *Decision Tree* (DT). Para análise do desempenho dos algoritmos, os autores levaram em consideração o desbalanceamento do conjunto de dados. Logo, foi adotado técnicas adicionais que levassem ao balanceamento dos dados e verificar o impacto no comportamento dos algoritmos. Em ambas as análises, o algoritmo RF obteve o melhor desempenho entre os demais.

Chegando à conclusão de que algumas variáveis favorecem a evasão, como a área de residência em diversos cursos, o desempenho nas disciplinas de Química, Português e Biologia em cursos específicos, e o turno noturno, que possui grande impacto no indicador. Além disso, o ano de conclusão do ensino médio também é relevante, pois quanto maior o intervalo entre a conclusão do ensino médio e o ingresso no ensino superior, maior a possibilidade de o aluno evadir.

3.6 Quadro comparativo

O Quadro 1 apresenta uma comparação entre as abordagens deste estudo e as dos trabalhos relacionados, levando em conta as técnicas aplicadas, a fonte e o tipo dos dados coletados, além da modalidade de ensino analisada.

Quadro 1 – Quadro comparativo das abordagens dos trabalhos relacionados e deste trabalho.

Trabalho	Objetivo do Estudo	Método/Técnica	Origem dos Dados
<i>Presente trabalho</i>	Analisar a conformidade entre a estrutura curricular formal e as trajetórias acadêmicas reais dos estudantes para identificar gargalos e pontos críticos associados à evasão	Mineração de Processos, Redes de Petri, <i>Token Replay</i> , <i>Alignments</i>	Histórico acadêmico (event logs com disciplinas, status, semestre e ID anonimizado)
<i>Unger et al. (2022)</i>	Compreender padrões de comportamento dos discentes em um AVA para traçar sua trajetória de aprendizagem em cursos EaD	Mineração de Processos, BPMN, Heuristics Miner	<i>Logs</i> de interação com ambiente virtual de aprendizagem (AVA)
<i>Carmo et al. (2022)</i>	Identificar correlações entre padrões de trajetórias de aprendizagem e evasão no ensino superior	Mineração de Dados Educacionais, Análise de Grafos, Algoritmos de Agrupamento (Clustering)	Histórico acadêmico, dados socioeconômicos e socioculturais dos discentes
<i>Kantorski et al. (2023)</i>	Construir modelos preditivos de risco de evasão em cursos presenciais a partir de dados históricos	Aprendizado de Máquina (SVM, Árvore de Decisão, Random Forest)	Histórico acadêmico (matrículas, reprovações, coeficiente de rendimento e frequência)
<i>Falcão et al. (2023)</i>	Predizer a evasão acadêmica com base em características pré-universitárias dos estudantes ingressantes	Aprendizado de Máquina (Árvores de Decisão, Regressão Logística)	Dados do ENEM, dados demográficos e socioeconômicos obtidos no ingresso dos estudantes
<i>Oliveira e Medeiros (2024)</i>	Avaliar o risco de evasão com base em autoavaliações periódicas feitas pelos próprios estudantes	Aprendizado de Máquina (XGBoost, KNN, Decision Tree)	Autoavaliações aplicadas ao longo do curso com percepção de dificuldades e desempenho

Fonte: Elaborado pelo autor.

4 REPRESENTAÇÃO DA ESTRUTURA CURRICULAR COM REDE DE PETRI

Neste capítulo, é apresentada a modelagem formal da estrutura curricular do curso de Ciência da Computação da Universidade Federal do Ceará (UFC), Campus de Quixadá, por meio de uma rede de Petri. Essa modelagem permite representar de forma estruturada a progressão acadêmica dos discentes, contemplando os componentes curriculares, os semestres letivos, os estados de aprovação e não aprovação, bem como os pré-requisitos entre disciplinas.

Além da modelagem, este capítulo também explora a simulação preliminar do fluxo acadêmico utilizando a técnica de *token replay*. Essa abordagem permite visualizar diferentes trajetórias estudantis, como o percurso ideal, com aprovações contínuas; trajetórias com reprovações parciais; e trajetórias interrompidas, permitindo compreender os efeitos dessas variações sobre o tempo de formação dos discentes.

A estrutura do capítulo está organizada da seguinte forma: a Seção 4.1 descreve o processo de preparação dos dados e construção do *log* de eventos utilizado nas análises. Em seguida, a Seção 4.2 apresenta a modelagem da estrutura curricular em rede de Petri e ilustra diferentes trajetórias acadêmicas simuladas a partir do modelo, evidenciando as implicações de cada cenário sobre a progressão dos discentes.

4.1 Preparação dos dados e criação do *log* de eventos

Este estudo utiliza um conjunto de dados acadêmicos dos discentes do curso de Bacharelado em Ciência da Computação da Universidade Federal do Ceará (UFC), *Campus* de Quixadá. Os dados foram obtidos a partir do Sistema Integrado de Gestão de Atividades Acadêmicas (SIGAA), disponibilizado pela Pró-Reitoria de Graduação (PROGRAD), e abrangem o período de 2013 a 2024.

O conjunto de dados é composto por 550 discentes, abrangendo estudantes com matrículas ativas e inativas, incluindo aqueles que concluíram o curso, abandonaram ou ainda estão matriculados. Para garantir a privacidade dos discentes, todos os identificadores pessoais foram removidos e as matrículas originais foram substituídas por *IDs* genéricos, assegurando a conformidade com as diretrizes de proteção de dados, Lei nº 13.709/2018. A Tabela 4 apresenta o conjunto de dados brutos.

A Tabela 5 apresenta os dados acadêmicos brutos dos discentes, iniciando pelo *ano*, que indica o ano letivo em que a turma e a matrícula estão registradas (por exemplo,

Tabela 4 – Conjunto de dados brutos (as seções exibem diferentes colunas de uma única tabela, cujas linhas compartilham a mesma posição e referência de matrícula)

Parte 1				
ano	periodo	codigo_turma	codigo	nome_disciplina
2014	1	1	QXD0005	ARQUITETURA DE COMPUTADORES
2014	1	1	QXD0005	ARQUITETURA DE COMPUTADORES
2014	1	1	QXD0005	ARQUITETURA DE COMPUTADORES
2014	1	1	QXD0005	ARQUITETURA DE COMPUTADORES
...	...	...	...	...

Parte 2				
id_tipo_componente	descricao_ds	id_situacao_turma	descricao_dp	id_unidade
2	DISCIPLINA	3	CONSOLIDADA	1020
2	DISCIPLINA	3	CONSOLIDADA	1020
2	DISCIPLINA	3	CONSOLIDADA	1020
2	DISCIPLINA	3	CONSOLIDADA	1020
...	...	...	...	...

Parte 3				
capacidade_aluno	id_discente	id_situacao_matricula	descricao	id_curso
55	999003	7	REP. FALTA	1371512
55	999448	6	REPROVADO	1371512
55	999236	7	APROVADO	1371512
55	999086	6	REPROVADO	1371512
...	...	...	...	...

Parte 4				
curso	id_curriculo	ano_ingresso	periodo_ingresso	id_forma_ingresso
CIÊNCIA DA COMPUTAÇÃO	13730020	2014	1	45
CIÊNCIA DA COMPUTAÇÃO	13730020	2014	1	45
CIÊNCIA DA COMPUTAÇÃO	13730020	2014	1	45
CIÊNCIA DA COMPUTAÇÃO	13730020	2014	1	45
...	...	...	...	...

Fonte: elaborada pelo autor.

2023). O *periodo* refere-se ao período letivo, dividido em semestres (1º, 2º), durante o qual o componente curricular é ofertado. O *codigo_turma* é um identificador único atribuído a cada turma, enquanto o *codigo* do componente curricular serve para a categorização acadêmica do componente curricular. O *nome_disciplina* especifica o título do componente curricular ofertada na turma. O *id_tipo_componente* e a *descricao_ds* classificam o componente curricular como disciplina. O *id_situacao_turma* e a *descricao_dp* identificam o status da turma, como

"Aberta" ou "Consolidada". A *id_unidade* e o *nome_unidade* da unidade referem-se à unidade acadêmica responsável pelo componente curricular. A *ch_total* total indica a carga horária total do componente curricular, enquanto a *ch_dedicada_periodo* diz respeito a carga horária do docente. O *docente* identifica qual o docente responsável por lecionar aquele componente curricular. A *capacidade_aluno* define o número máximo de alunos por turma. O *id_discente* é um identificador único do aluno matriculado, e o *id_situacao_matricula* junto com a *descricao* detalham o *status* da matrícula do aluno (por exemplo, "MATRICULADO", "APROVADO" ou "REPROVADO"). Finalmente, o *id_curso* e *curso* especificam a que curso a disciplina pertence, e o *id_curriculo*, *ano_ingresso*, *periodo_ingresso* e *id_forma_ingresso* fornecem informações sobre o currículo acadêmico do curso e o ingresso do aluno, como vestibular ou transferência.

Tabela 5 – Dados brutos acadêmicos dos discentes do curso de Ciência da Computação

Variáveis	Descrição
ano	ano letivo de oferta
periodo	semestre letivo de oferta
codigo_turma	código identificador da turma
codigo	código do componente curricular
nome_disciplina	nome da disciplina
id_tipo_componente	identificador do tipo de componente curricular
descricao_ds	descrição do tipo de componente curricular
id_situacao_turma	identificador da situação da turma
descricao	descrição da situação da turma
id_unidade	identificador da unidade acadêmica
nome_unidade	nome da unidade acadêmica
ch_total	carga horária total da disciplina
docente	nome do docente responsável pela disciplina
ch_dedicada_periodo	carga horária do docente
capacidade_aluno	capacidade máxima de alunos na turma
id_discente	identificador do aluno
id_situacao_matricula	identificador da situação da matrícula do aluno
descricao	descrição da situação da matrícula
id_curso	identificador do curso
curso	nome do curso
id_curriculo	identificador do currículo
ano_ingresso	ano de ingresso do aluno no curso
periodo_ingresso	período de ingresso do aluno no curso
id_forma_ingresso	identificador da forma de ingresso

Fonte: elaborada pelo autor.

Dentre as variáveis disponíveis no conjunto de dados, foram selecionadas cinco principais colunas para compor o *log* de eventos: *ano*, *periodo*, *id_discente*, *codigo* e *descricao*. A escolha dessas variáveis se deve à sua relevância na modelagem do trajetória acadêmica dos discentes.

4.1.1 *Limpeza dos dados*

Para acompanhar o fluxo de ingressantes no curso, foi selecionado os discentes da *coorte* de 2014, 2015 e 2016, de modo que seja viável acompanhar a progressão longitudinal dos ingressantes até a conclusão ou evasão do curso presente na base de dados. Essa delimitação permite analisar os estudantes que tiveram o tempo mínimo e máximo para conclusão do curso, ou seja, de oito a doze semestres consecutivos, conforme estabelecido pelo Projeto Pedagógico do Curso (PPC) como o prazo mínimo e máximo para finalização do curso e, neste período, o currículo permaneceu inalterado. Após aplicação deste filtro, o conjunto de dados utilizado neste estudo contabiliza 114 discentes.

Uma prática comum nas instituições de ensino superior é a alocação de dois ou mais docentes para ministrar o mesmo componente curricular em um único período letivo. Como consequência, os dados referentes a um mesmo discente foram duplicados, gerando a necessidade de remover 141 registros que apresentavam essa condição. Além disso, para garantir uma análise mais precisa e alinhada à trajetória acadêmica obrigatória dos discentes, foram selecionados apenas os componentes curriculares de caráter obrigatório, excluindo-se as disciplinas optativas e eletivas. Essa decisão se justifica pela ausência de participação significativa dos discentes nessas disciplinas, uma vez que a alta taxa de evasão no curso impede muitos alunos de alcançarem essa fase da formação. Portanto, a inclusão desses componentes comprometeria a análise, além de aumentar desnecessariamente a demanda computacional sem trazer ganhos analíticos relevantes.

As colunas que registram o ano e período letivo de oferta do componente curricular foram combinadas para gerar uma nova coluna chama *Semestre_letivo*, que registra a data do semestre letivo de oferta do componente curricular. Essas datas foram fixadas no mês de fevereiro para o primeiro período letivo de oferta e agosto para o segundo período letivo de oferta do ano, em seguida as colunas ano e período foram excluídas dos dados. Vale destacar, no entanto, que os semestres nem sempre se iniciaram exatamente nesses meses, devido a eventos como greves, pandemias e outras situações que alteraram o calendário acadêmico.

A variável *codigo* da disciplina identifica os componentes curriculares nos quais o discente esteve matriculado ao longo de sua trajetória acadêmica. Já a variável *descricao* reflete o *status* do aluno ao final de cada semestre em relação a cada disciplina cursada. Os possíveis valores dessa variável incluem: "MATRICULADO", "APROVADO", "REPROVADO", "REPROVADO POR FALTA", "TRANCADO", "SUPRIMIDO" e "TRANCADO TOTAL".

Para a modelagem do processo, foram considerados apenas dois *status* principais:

APROVADO e **NAOAPROVADO**. Dessa forma, todas as demais classificações, com exceção de "MATRICULADO" e "APROVADO", foram convertidas para **NAOAPROVADO**, uma vez que cada uma dessas situações implica na necessidade de o discente cursar novamente o componente curricular em um semestre posterior. A condição matriculado no componente curricular não indica uma condição terminativa naquele componente curricular, logo este *status* foi mantido.

Os componentes curriculares Análise e Projeto de Sistemas, Engenharia de Software, Interação Humano-Computador, Introdução à Ciência da Computação e Ética, Direito e Legislação apresentavam, no conjunto de dados, códigos diferentes daqueles registrados na estrutura curricular oficial. Por esse motivo, foi necessário renomeá-los de acordo com os códigos utilizados no currículo adotado neste estudo.

Portanto, ao final do processo de limpeza, o conjunto de dados passou a conter 114 discentes, cada um registrando o código do componente curricular, a identificação do discente, a descrição e o semestre em que o componente foi cursado. Essas informações estão representadas na Tabela 6.

Tabela 6 – Conjunto de dados após limpeza

codigo	id_discente	descricao	Semestre_letivo
QXD0005	999003	NAOAPROVADO	2014-02-01
QXD0103	999141	NAOAPROVADO	2014-02-01
QXD0001	999003	NAOAPROVADO	2014-02-01
QXD0108	999448	APROVADO	2014-02-01
QXD0056	999208	NAOAPROVADO	2014-02-01
		⋮	

É importante destacar que o status "suprimido" foi utilizado durante a pandemia de COVID-19 como um mecanismo para atender às necessidades dos estudantes que enfrentavam dificuldades em acompanhar as atividades letivas. Essa medida visou garantir que esses alunos não fossem prejudicados no Índice de Rendimento Acadêmico (IRA).

4.1.2 Construção do log de eventos

O requisito essencial para a aplicação da mineração de processos é a disponibilidade ou a construção de um *log* de eventos. Este *log* de eventos serve como base para a análise do fluxo de atividades, permitindo que se extraia informações valiosas sobre o comportamento do processo. Um *log* de eventos adequado deve atender a três condições fundamentais:

1. **Case Id:** O identificador da instância do processo (neste caso, o *id_discente*) associa todas

as atividades realizadas por ele ao longo de sua trajetória acadêmica, como as disciplinas cursadas.

2. **Activity:** As atividades consideradas neste contexto correspondem a quatro tipos de eventos registrados para cada discente: código dos componentes curriculares, combinação do código e *status*, semestres e a condição de formados para cada discente.
3. **Timestamp:** O carimbo temporal é uma dimensão essencial para a mineração de processos, neste contexto, os semestres letivos.

Tendo em vista esta necessidade de um *log* de evento bem definido e estruturado, foi necessário realizar alguns procedimentos adicionais no conjunto de dados apresentado na Tabela 6.

A partir das colunas *descrição* e *Semestre_letivo*, foi criado um novo conjunto de dados na qual essas colunas foram concatenadas para gerar uma nova representação a condição acadêmica no componente curricular ao final do semestre letivo. Esse procedimento possibilita uma rastreabilidade mais precisa dos registros ao longo do tempo, garantindo que cada evento tenha uma identificação única.

Por exemplo, para o componente curricular *QXD0001*, se a descrição original for *APROVADO*, será adicionado ao final da coluna código a atividade representada como *QXD0001_APROVADO*. Esse padrão se repete para todas as combinações existentes no conjunto de dados, permitindo diferenciar a condição final para cada componente curricular. Por fim, a coluna *descricao* foi removida do conjunto de dados. A Tabela 7 apresenta os dados após essas transformações.

Tabela 7 – Conjunto de dados após concatenação das colunas código e descrição

codigo	id_discente	Semestre_letivo
QXD0001	999003	2014-02-01
QXD0001	999448	2014-02-01
QXD0001	999236	2014-02-01
QXD0001_APROVADO	999003	2014-02-01
QXD0001_NAOAPROVADO	999448	2014-02-01
QXD0001_APROVADO	999236	2014-02-01
	⋮	

Para cada discente, foram inseridas duas atividades complementares ao seu *log* de eventos, proporcionando um rastreamento mais estruturado do fluxo acadêmico. As atividades adicionadas foram:

1. **semestre:** Esta atividade foi implementada para acompanhar a progressão dos discentes

ao longo dos semestres letivos, associando as disciplinas cursadas ao respectivo período acadêmico do discente. Para padronização, as datas dessas atividades segue a seguinte regra:

- janeiro para o 1º período letivo do ano.
- julho para o 2º período letivo do ano.

2. **formado:** Esta atividade foi incorporada para indicar os discentes que concluíram o curso com êxito, tendo como data a adição de um dia após realizar o último componente curricular.

Após a limpeza e adição das atividades descritas anteriormente, o conjunto de dados está preparado para ser convertido em um *log* de eventos, conforme ilustrado na Tabela 8. Esse *log* é composto por todos os componentes curriculares nos quais o discente se matriculou ao longo de seu vínculo com a instituição, registrando sua trajetória acadêmica. Além das atividades que representam a matrícula nas disciplinas, o *log* inclui os estados finais de cada componente curricular, que podem indicar aprovação, reprovação, suprimido, trancamento parcial ou trancamento total, e ao final, a atividade “formado” para indicar aqueles discentes que conseguiram concluir o curso.

Tabela 8 – Conjunto de dados após inserção das atividades de semestre e formado

Semestre letivo	id_discente	codigo
2016-01-01	999382	semestre01
2016-02-01	999382	QXD0005
2016-02-01	999382	QXD0001
2016-02-01	999382	QXD0108
2016-02-01	999382	QXD0056
...	...	...
2019-08-01	999382	QXD0038
2019-08-01	999382	QXD0043
2019-11-01	999382	QXD0038_APROVADO
2019-11-01	999382	QXD0043_APROVADO
2019-11-02	999382	formado

A Tabela 9 exemplifica o *log* de eventos desenvolvido para o curso de Ciência da Computação. Nele, a coluna *id_discente* é utilizada como *case_id*, a coluna *codigo* representa a *activity* e a coluna *Semestre_letivo* é empregada como *timestamp*. O *log* é composto por um total de 114 *case_id*, cada um correspondendo à trajetória acadêmica de um discente, permitindo a análise da progressão curricular e dos eventos associados à sua jornada acadêmica.

A Tabela 10 apresenta o *trace* de evento com o conjunto de atividades realizadas pelo

Tabela 9 – Log de eventos do curso de Ciência da Computação

Semestre_letivo	id_discente	codigo	case:concept:name	concept:name	time:timestamp	@_index	@_case_index
2015-01-01 00:00:00+00:00	999001	semestre01	999001	semestre01	2015-01-01 00:00:00+00:00	0	0
2015-02-01 00:00:00+00:00	999001	QXD0005	999001	QXD0005	2015-02-01 00:00:00+00:00	1	0
2015-02-01 00:00:00+00:00	999001	QXD0103	999001	QXD0103	2015-02-01 00:00:00+00:00	2	0
2015-02-01 00:00:00+00:00	999001	QXD0001	999001	QXD0001	2015-02-01 00:00:00+00:00	3	0
2015-02-01 00:00:00+00:00	999001	QXD0108	999001	QXD0108	2015-02-01 00:00:00+00:00	4	0
...	...	...	...	...	...	...	...
2017-06-01 00:00:00+00:00	999545	QXD0110_APROVADO	999545	QXD0110_APROVADO	2017-06-01 00:00:00+00:00	7417	110
2017-07-01 00:00:00+00:00	999545	semestre08	999545	semestre08	2017-07-01 00:00:00+00:00	7418	110
2017-08-01 00:00:00+00:00	999545	QXD0038	999545	QXD0038	2017-08-01 00:00:00+00:00	7419	110
2017-11-01 00:00:00+00:00	999545	QXD0038_APROVADO	999545	QXD0038_APROVADO	2017-11-01 00:00:00+00:00	7420	110
2017-11-02 00:00:00+00:00	999545	formado	999545	formado	2017-11-02 00:00:00+00:00	7421	110

discente identificado pelo *id* genérico 999001 ao longo de sua trajetória acadêmica. Observa-se que o estudante levou nove semestres para concluir o curso, ultrapassando o período mínimo ideal. Esse semestre adicional decorre de cinco reprovações em componentes curriculares que desempenham o papel de pré-requisitos para outras disciplinas, impactando diretamente a progressão acadêmica e exigindo a reprogramação do percurso formativo.

4.1.3 Ferramentas utilizadas

Para a preparação e transformação dos dados em um *log* de eventos compatível com as técnicas de mineração de processos, foi utilizada a linguagem de programação *Python*, com o apoio de bibliotecas como *pandas* e *datetime*, para manipulação, limpeza e agregação dos dados acadêmicos. As atividades foram organizadas cronologicamente e agrupadas por discentes (*case_id*), seguindo a estrutura exigida por abordagens de Mineração de Processos.

A modelagem do currículo do curso foi realizada por meio da ferramenta *Snoopy*, que oferece um ambiente gráfico para a construção de redes de Petri. O modelo construído representa os semestres letivos, componentes curriculares, estados de aprovação ou não aprovação e condições de pré-requisitos. A utilização do *Snoopy* permitiu a criação de um modelo formal verificável, exportável em formato PNML, utilizado nas etapas subsequentes de simulação e análise com a biblioteca *PM4PY*.

4.2 Modelo formal da estrutura curricular do Curso de Ciência da Computação em rede de Petri

A matriz curricular do curso de Ciência da Computação na UFC-Quixadá, Figura 25, está organizada em sete eixos de formação, conforme estabelecido no Projeto Pedagógico do Curso (PPC). Esses eixos estruturam o conhecimento em diferentes áreas da computação, proporcionando uma formação ampla e interdisciplinar. Os eixos são:

1. Formação básica em Matemática

Tabela 10 – Log de eventos do discente 999001

<i>timestamp</i>	<i>id_discente</i>	<i>activity</i>
2015-01-01	999001	semestre01
2015-02-01	999001	QXD0005
2015-02-01	999001	QXD0103
2015-02-01	999001	QXD0001
2015-02-01	999001	QXD0108
2015-02-01	999001	QXD0056
2015-02-01	999001	QXD0109
2015-06-01	999001	QXD0005_NAOAPROVADO
2015-06-01	999001	QXD0103_APROVADO
2015-06-01	999001	QXD0001_APROVADO
2015-06-01	999001	QXD0108_APROVADO
2015-06-01	999001	QXD0056_APROVADO
2015-06-01	999001	QXD0109_NAOAPROVADO
2015-07-01	999001	semestre02
2015-08-01	999001	QXD0005
2015-08-01	999001	QXD0010
2015-08-01	999001	QXD0008
2015-08-01	999001	QXD0109
2015-08-01	999001	QXD0007
2015-11-01	999001	QXD0005_APROVADO
2015-11-01	999001	QXD0010_APROVADO
2015-11-01	999001	QXD0008_NAOAPROVADO
2015-11-01	999001	QXD0109_APROVADO
2015-11-01	999001	QXD0007_APROVADO
...	...	...
2019-01-01	999001	semestre09
2019-02-01	999001	QXD0029
2019-06-01	999001	QXD0029_APROVADO
2019-06-03	999001	formado

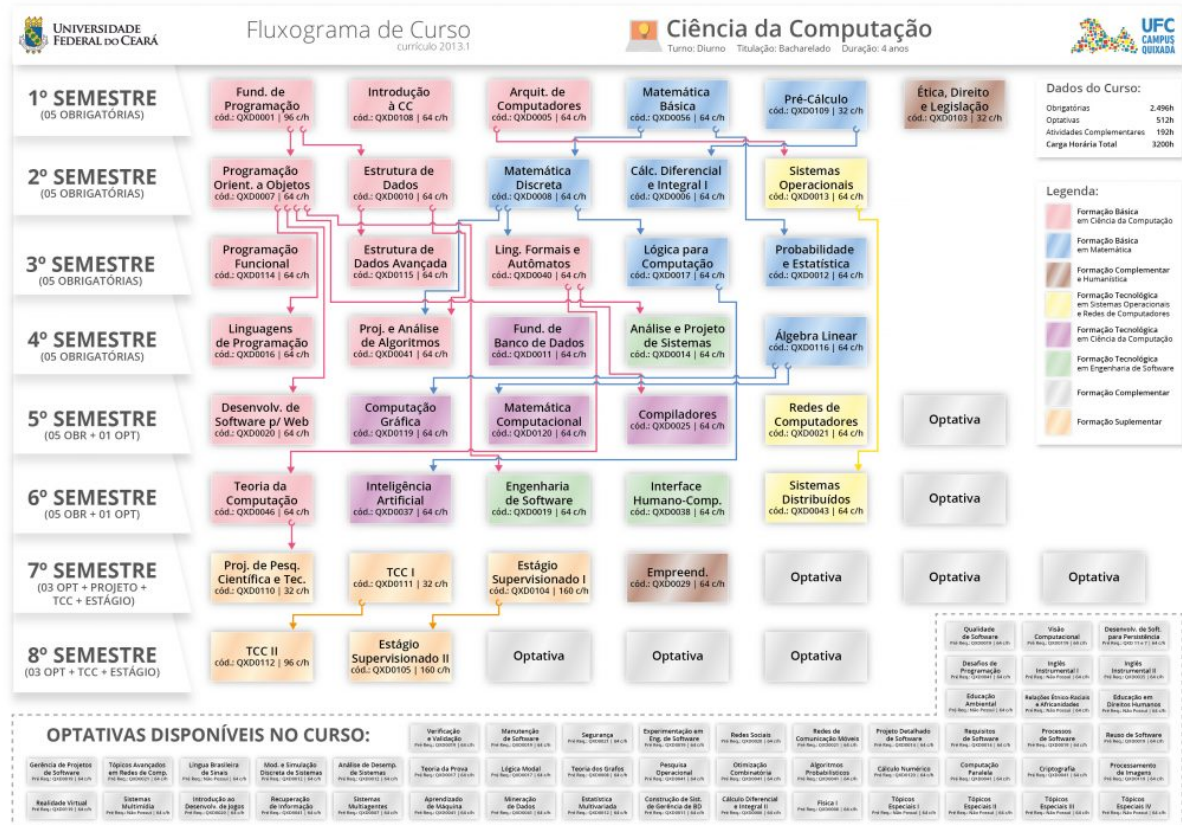
Fonte: elaborada pelo autor.

2. Formação básica em Ciência da Computação
3. Formação tecnológica em Computação
4. Formação tecnológica em Engenharia de Software
5. Formação tecnológica em Sistemas Operacionais e Redes de Computadores
6. Formação Complementar e Humanística
7. Formação Suplementar

Esses eixos, em conjunto, fornecem a base teórica e prática necessária para que os estudantes desenvolvam competências essenciais à atuação profissional em áreas diversas da Computação. Ao longo de sua trajetória acadêmica, o discente deve transitar pelos diferentes componentes curriculares distribuídos ao longo de oito semestres letivos, cursando disciplinas

conforme os requisitos e pré-requisitos do curso. A Figura 26 apresenta a modelagem das transições entre os semestres letivos, representando o fluxo acadêmico necessário para a progressão dos estudantes.

Figura 25 – Estrutura curricular do curso de Ciência da Computação da UFC do Campus de Quixadá.

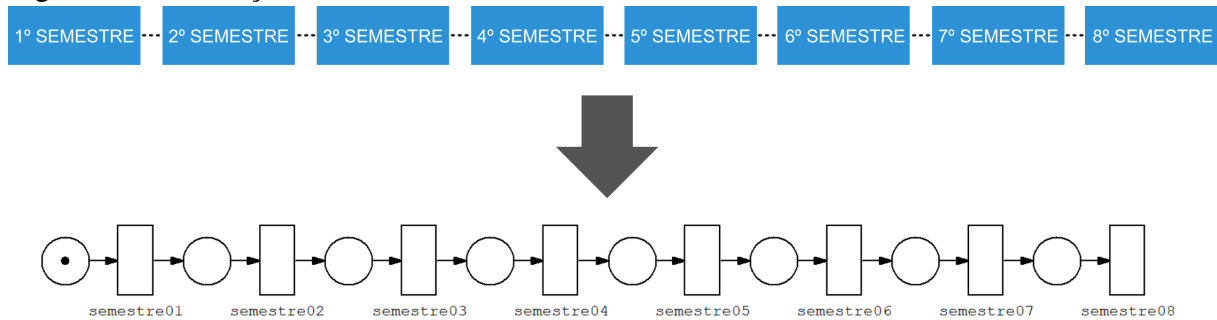


Fonte: Disponível em <https://cc.quixada.ufc.br/estrutura-curricular/fluxograma-de-disciplinas/>

Na modelagem em rede de Petri de cada semestre letivo é representado por uma transição (retângulo), enquanto as entradas e saídas de cada semestre são modeladas como lugares (círculos). Esses lugares indicam os estados acadêmicos intermediários, nos quais os discentes aguardam a conclusão do semestre atual antes de prosseguir para o próximo. O fluxo de progressão é determinado pelos arcs direcionais, que conectam os lugares às transições, garantindo que um discente só possa avançar para o próximo semestre após a conclusão do período letivo vigente.

Ao ingressar em cada semestre letivo, o discente tem à sua disposição um conjunto de componentes curriculares a serem cursados. A Figura 27 apresenta a modelagem desse processo, onde a progressão do estudante é representada por meio da transição de semestres e da matrícula em cada componente curricular.

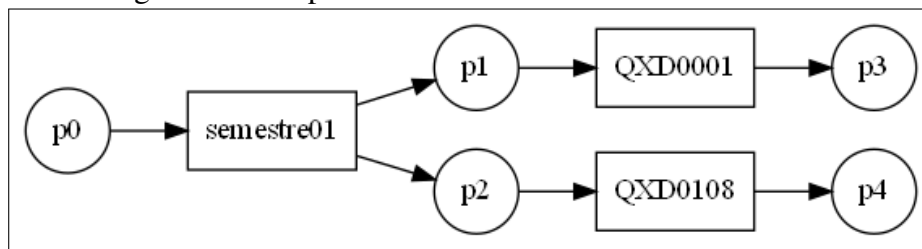
Figura 26 – Transição dos semestres.



Fonte: elaborada pelo autor.

Após o disparo da transição *semestre01*, correspondente ao início do 1º semestre letivo, os lugares *p1* e *p2* recebem um *token* cada, representando que o discente está apto a cursar os componentes curriculares *QXD0001* e *QXD0108*, respectivamente.

Figura 27 – Modelagem das disciplinas



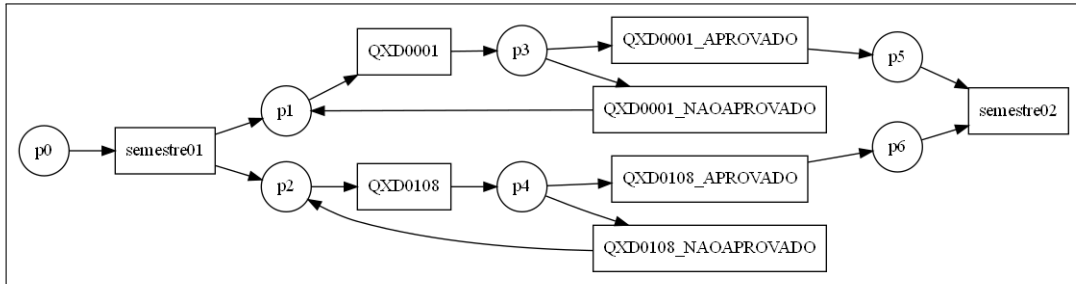
Fonte: elaborada pelo autor.

Cada componente curricular é modelada por uma transição (*QXD0001* e *QXD0108*) simbolizando a matrícula no componente ao longo do semestre. Quando um estudante executa atividades acadêmicas associadas ao componente curricular e essa aparece no seu *trace* de eventos, a transição correspondente é disparada. Como resultado, os *tokens* presentes em *p1* e *p2* são consumidos e um novo *token* é produzido em *p3* e *p4*, representando a condição de cursando a disciplina.

Na modelagem, a matrícula em um componente curricular gera um *token* nos lugares *p3* e *p4*, que representam o estado intermediário onde o estudante está cursando a disciplina. Esse *token* será então consumido por uma das transições possíveis ao final do semestre, levando a um dos dois estados finais: aprovação ou não aprovação no componente curricular, apresentado na Figura 28.

Caso o discente seja aprovado, a transição *QXD0001_APROVADO* é disparada, permitindo que o estudante avance no fluxo acadêmico e prossiga para os componentes curriculares subsequentes que dependem da sua aprovação. O *token* então migra para o lugar correspondente ao avanço no curso (*p5*).

Figura 28 – Modelagem dos estado de aprovação ou não aprovação nos componentes curriculares.



Fonte: elaborada pelo autor.

Se o estudante for não aprovado, a transição *QXD0001_NAOAPROVADO* será disparada, consumindo o *token* em *p3* e redirecionando o fluxo para um estado onde o discente precisará refazer a disciplina em um semestre futuro. Esse cenário de reprovação cria um impacto direto na trajetória acadêmica do estudante, gerando atrasos na progressão e aumentando a probabilidade de evasão, especialmente se múltiplas reprovações ocorrerem ao longo do curso.

A distribuição curricular do curso é estruturada de forma a garantir que os discentes adquiram gradualmente os conhecimentos necessários, desde os fundamentos iniciais até os conceitos mais avançados. Esse encadeamento de aprendizado estabelece uma dependência sequencial entre as disciplinas, resultando na necessidade de pré-requisitos a partir do 2º semestre letivo.

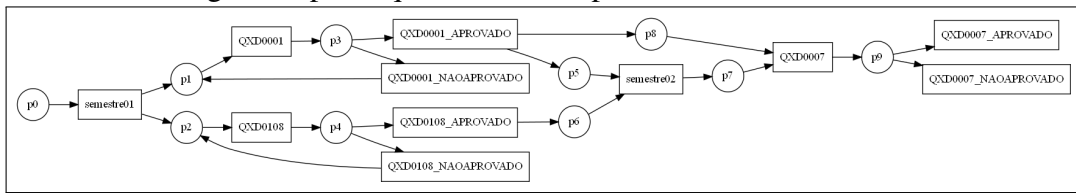
Para capturar essa restrição no modelo formal, a rede de Petri inclui uma condição de controle que regula o avanço do estudante para disciplinas que possuem pré-requisitos. A Figura 29 ilustra essa lógica, onde o lugar *p8* atua como um mecanismo de verificação para garantir que apenas discentes que tenham sido aprovados na disciplina *QXD0001* possam prosseguir para a disciplina *QXD0007*.

O funcionamento desse controle ocorre da seguinte forma:

1. Quando a transição *QXD0001_APROVADO* é disparada, o *token* presente no lugar *p3* (estado de cursando a disciplina) é consumido e novos *tokens* são produzidos em *p5* e *p8*.
2. O *token* gerado em *p5* permite que o estudante avance normalmente para as próximas disciplinas do fluxo acadêmico.
3. O *token* gerado em *p8* representa a autorização para cursar a disciplina dependente *QXD0007*, funcionando como um mecanismo de controle. A transição *QXD0007* só poderá ser disparada se houver um *token* presente em *p8*, ou seja, apenas se o estudante tiver sido aprovado em *QXD0001*.

Por fim, para a conclusão do curso, o discente deve cumprir os critérios estabelecidos

Figura 29 – Modelagem do pré-requisito das disciplinas.



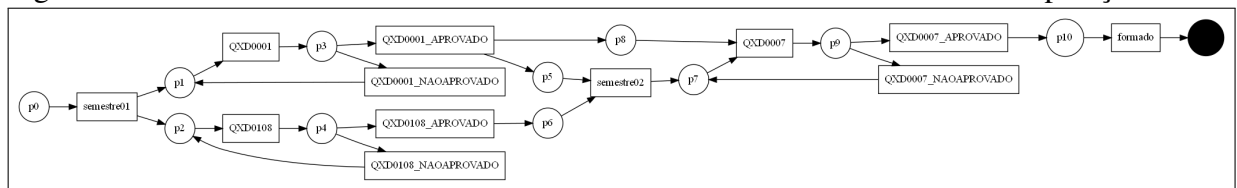
Fonte: elaborada pelo autor.

no Projeto Pedagógico do Curso.

A Figura 30 ilustra a modelagem do processo de conclusão com êxito do discente. No modelo, a transição *formado* representa a finalização bem-sucedida do curso, funcionando como um mecanismo de controle para identificar os discentes que atenderam a todas as exigências mínimas para a conclusão.

Essa transição permite distinguir os estudantes que completaram integralmente sua trajetória acadêmica daqueles que, por diferentes motivos, não atingiram os critérios necessários para a formação, seja por evasão ou retenção prolongada.

Figura 30 – Modelo reduzido da estrutura curricular do curso de Ciência da Computação

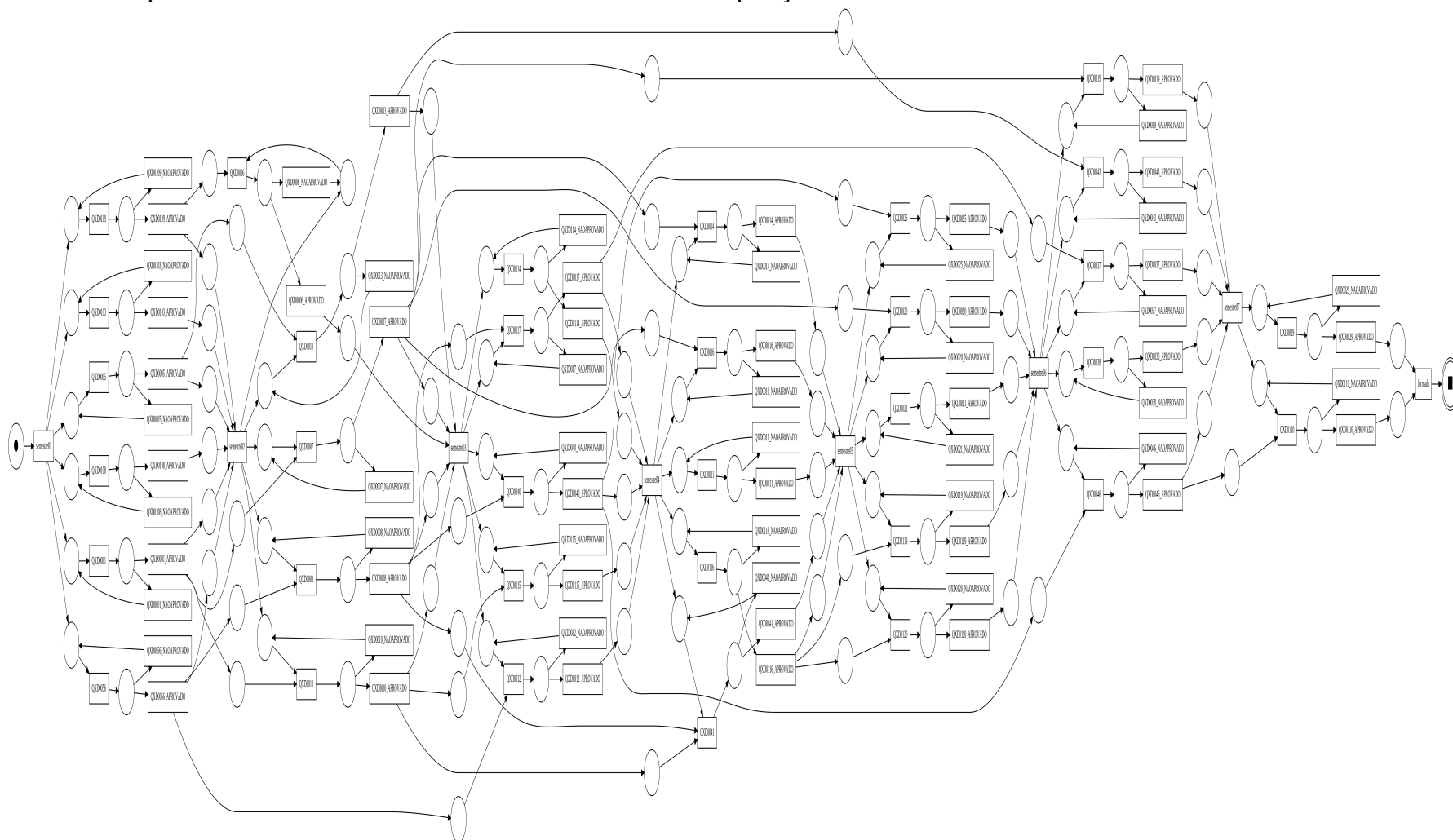


Fonte: elaborada pelo autor.

A Figura 31 apresenta o modelo completo da trajetória acadêmica dos discentes do curso de Ciência da Computação. A modelagem é composta por 123 lugares, 107 transições e 310 arcos, representando detalhadamente o fluxo acadêmico dos estudantes, desde o ingresso no curso até a conclusão ou interrupção da trajetória acadêmica.

A adoção de uma rede de Petri do tipo *WF-nets*, conforme descrito na Subseção 2.2.2, como ferramenta de modelagem do percurso acadêmico permite uma análise detalhada do comportamento dos discentes ao longo do curso. A partir dessa abordagem, é possível identificar padrões de progressão, gargalos acadêmicos e fatores associados à retenção e evasão. Além disso, a simulação do fluxo estudantil por meio do *token replay* possibilita a validação de diferentes trajetórias e o estudo de cenários hipotéticos, contribuindo para a melhoria contínua da estrutura curricular e para a implementação de políticas acadêmicas mais eficazes.

Figura 31 – Modelo completo da estrutura curricular do curso de Ciência da Computação

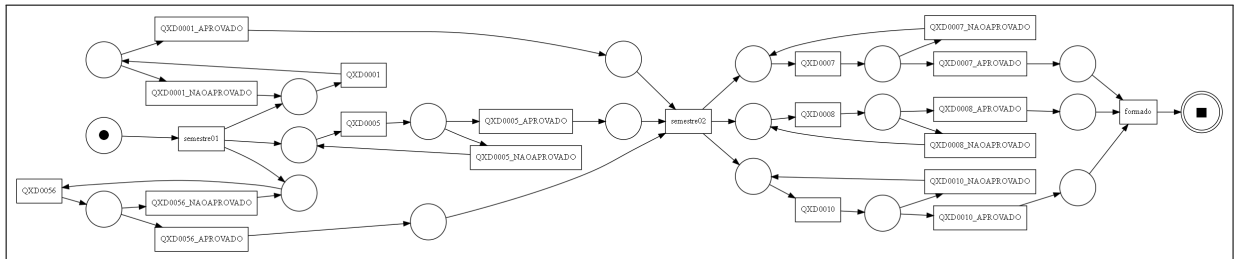


Fonte: Elaborada pelo autor.

4.2.1 Simulação preliminar com token replay

Nesta seção, serão simulados três cenários possíveis presentes no *log* de eventos analisado neste estudo. Para essa simulação, será utilizado um *log* de eventos reduzido, permitindo uma representação mais clara dos diferentes padrões de trajetória acadêmica. Além disso, será empregada uma modelagem simplificada em rede de Petri, composta por um subconjunto de disciplinas e abrangendo apenas dois semestres do curso, conforme ilustrado na Figura 32.

Figura 32 – Modelo reduzido da estrutura curricular para simulação



Fonte: elaborada pelo autor.

Os cenários simulados são os seguintes:

1. **Cenário (i):** O discente segue rigorosamente a estrutura curricular do curso, cursando todos os componentes curriculares no semestre previsto e sendo aprovado em cada um deles. Esse cenário representa o percurso acadêmico ideal, com total conformidade ao fluxo acadêmico planejado.
2. **Cenário (ii):** O discente enfrenta reprovação em um ou mais componentes curriculares e, conseqüentemente, precisa se matricular novamente na disciplina no semestre seguinte. Esse cenário reflete o impacto das reprovações na trajetória acadêmica do discente.
3. **Cenário (iii):** O discente não obtém aprovação em todos os componentes curriculares do 1º semestre, o que o impede de progredir para cursar plenamente os componentes do semestre seguinte. Dessa forma, sua trajetória acadêmica é interrompida temporariamente, exigindo um semestre adicional para que possa recuperar os componentes reprovados e, posteriormente, concluir o curso.

4.2.1.1 Cenário discente ideal

Para simular o cenário (i), é utilizado o *trace* de eventos descrito na Tabela 11, que registra os componentes curriculares cursados e aprovados pelo discente, o *timestamp* correspondente a cada atividade e, ao final, a transição que indica sua conclusão no curso.

A estrutura do *trace* evidencia um percurso acadêmico ideal, no qual o discente segue rigorosamente a estrutura curricular, cursando as disciplinas dentro do semestre previsto e sendo aprovado em todas as atividades acadêmicas sem interrupções ou reprovações. O evento final, representado pela transição *formado*, indica que o discente atingiu todos os requisitos necessários para a conclusão do curso dentro do tempo ideal.

Esse *trace* serve como referência para comparação com os demais cenários, permitindo avaliar o impacto de reprovações, desvios na cronologia prevista e demais variáveis que podem influenciar a progressão acadêmica dos discentes.

Tabela 11 – *Trace* de eventos do discente 1001

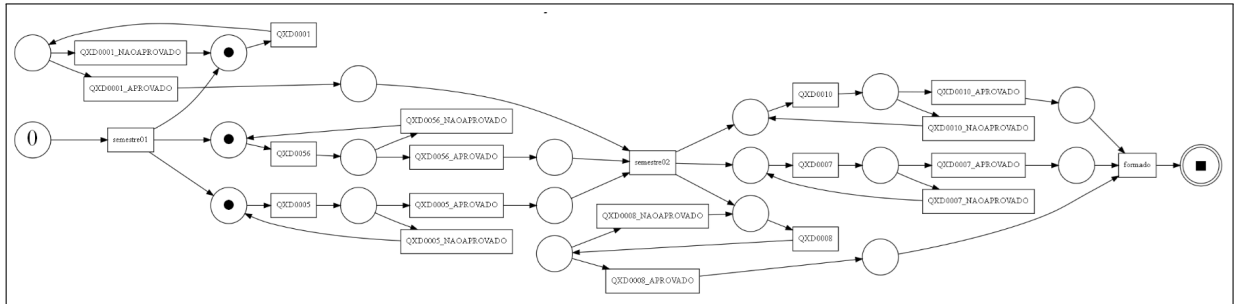
timestamp	id_discente	activity
01/01/2017	1001	semestre01
01/02/2017	1001	QXD0001
01/02/2017	1001	QXD0005
01/02/2017	1001	QXD0056
01/06/2017	1001	QXD0001_APROVADO
01/06/2017	1001	QXD0005_APROVADO
01/06/2017	1001	QXD0056_APROVADO
01/07/2017	1001	semestre02
01/08/2017	1001	QXD0007
01/08/2017	1001	QXD0010
01/08/2017	1001	QXD0008
01/11/2017	1001	QXD0007_APROVADO
01/11/2017	1001	QXD0008_APROVADO
01/11/2017	1001	QXD0010_APROVADO
02/11/2017	1001	formado

A atividade *semestre01* é a primeira a ser registrada no *trace* de eventos, indicando a matrícula do discente nos componentes curriculares do 1º semestre letivo. Esse evento representa o vínculo oficial do discente com a instituição de ensino, marcando o início de sua trajetória acadêmica.

A Figura 33 ilustra o fluxo da rede de Petri no momento dessa transição. Nela, observa-se o consumo do *token* no lugar de entrada da transição *semestre01*, sinalizando que a matrícula foi efetivada. Como resultado, a transição produz três novos *tokens* nos lugares de saída, representando a distribuição do discente entre os componentes curriculares do semestre. Nesse momento, o lugar de entrada da transição *semestre01* fica vazio (0 *tokens*), indicando que o processo de ingresso foi concluído.

As próximas atividades registradas no *trace* correspondem às matrículas nos compo-

Figura 33 – Modelo reduzido da estrutura curricular para simulação

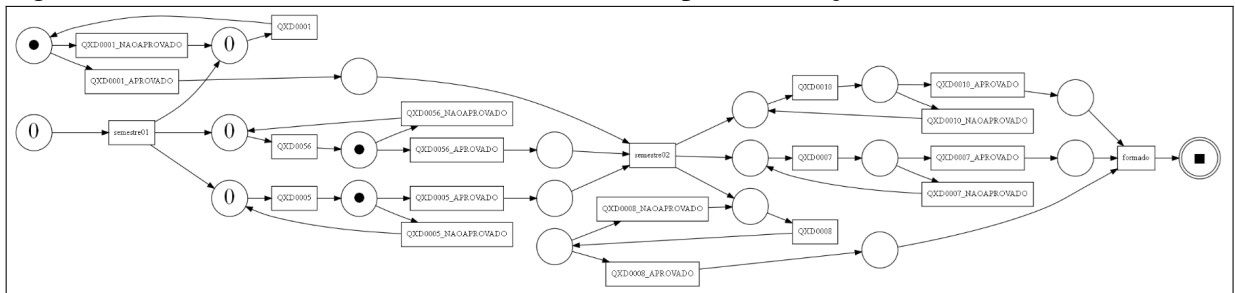


Fonte: elaborada pelo autor.

nentes curriculares *QXD0001*, *QXD0005* e *QXD0056*, todos pertencentes ao 1º semestre letivo. Como essas atividades ocorrem em paralelo, os lugares de entrada correspondentes a cada uma delas são consumidos simultaneamente, e novos *tokens* são produzidos nos lugares de saída. Esse comportamento é ilustrado na Figura 34, que representa a transição para o *status* de cursando cada disciplina.

Neste ponto da rede de Petri, o discente encontra-se formalmente matriculado e cursando os componentes curriculares. Ao final do semestre letivo, ocorrerá a avaliação do desempenho do discente, definindo sua condição acadêmica para cada disciplina, que poderá resultar em aprovação ou reprovação, determinando sua continuidade no fluxo acadêmico.

Figura 34 – Modelo reduzido da estrutura curricular para simulação

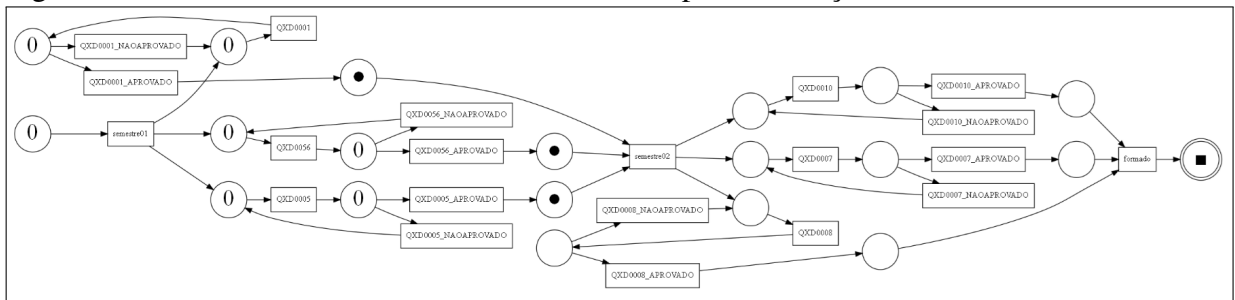


Fonte: elaborada pelo autor.

Ao verificar a situação final em cada componente curricular, são registradas no *trace* de eventos as atividades *QXD0001_APROVADO*, *QXD0005_APROVADO* e *QXD0056_APROVADO*. Essas transições são disparadas, consumindo os *tokens* de seus respectivos lugares de entrada e produzindo novos *tokens* nos lugares de saída, conforme ilustrado na Figura 35. Neste ponto, o discente conclui com êxito todos os componentes curriculares do 1º semestre e encontra-se apto a prosseguir para o 2º semestre e se matricular nos respectivos componentes curriculares.

Seguindo o fluxo do *trace*, a próxima atividade registrada é *semestre02*, que se torna habilitada para disparo devido ao preenchimento de *tokens* em todos os seus lugares de

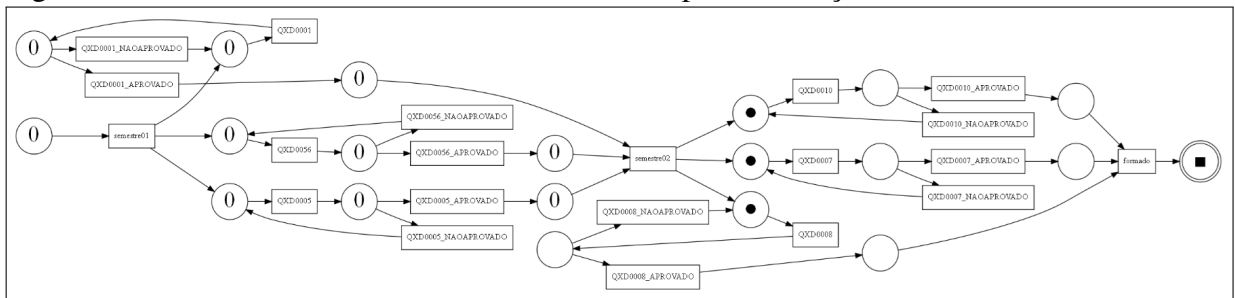
Figura 35 – Modelo reduzido da estrutura curricular para simulação



Fonte: elaborada pelo autor.

entrada. Assim, a transição *semestre02* dispara, consumindo os *tokens* de entrada e produzindo novos *tokens* nos lugares de saída, indicando que o discente está pronto para se matricular nos componentes curriculares do 2º semestre, conforme ilustrado na Figura 36.

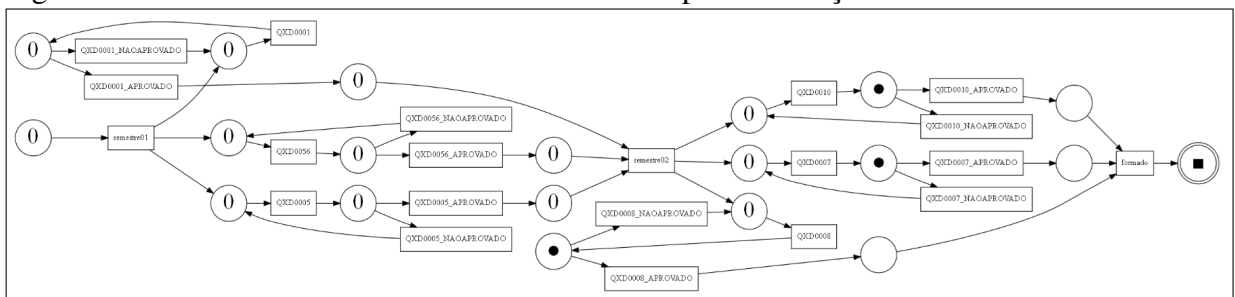
Figura 36 – Modelo reduzido da estrutura curricular para simulação



Fonte: elaborada pelo autor.

Ao realizar a matrícula nos componentes curriculares do 2º semestre, essa ação é convertida em atividades que passam a integrar o *trace* do discente. As próximas atividades disparadas são *QXD0007*, *QXD0008* e *QXD0010*. Como resultado, os *tokens* de seus respectivos lugares de entrada são consumidos e novos *tokens* são produzidos nos lugares de saída, indicando que o discente está cursando esses componentes curriculares, conforme ilustrado na Figura 37.

Figura 37 – Modelo reduzido da estrutura curricular para simulação

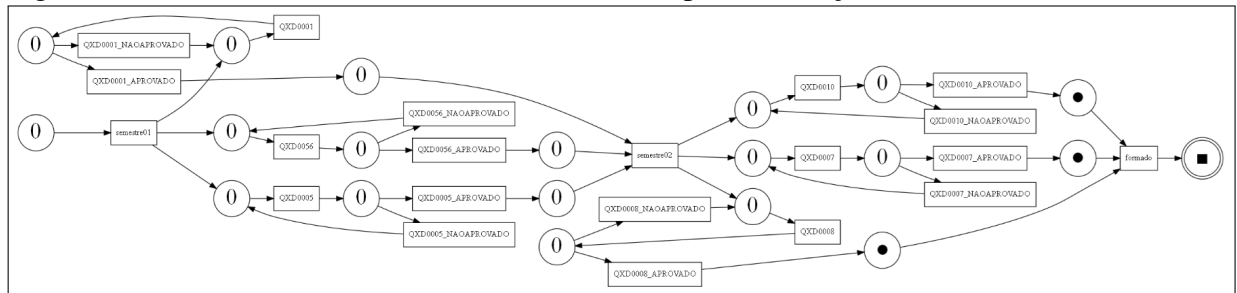


Fonte: elaborada pelo autor.

Neste ponto da trajetória acadêmica, ao final do semestre letivo, os *status* das discipli-

nas são verificados e convertidos em novas atividades: *QXD0007_APROVADO*, *QXD0008_APROVADO* e *QXD0010_APROVADO*. Assim, essas transições são disparadas, consumindo os *tokens* dos lugares de entrada e produzindo novos *tokens* nos lugares de saída, conforme demonstrado na Figura 38..

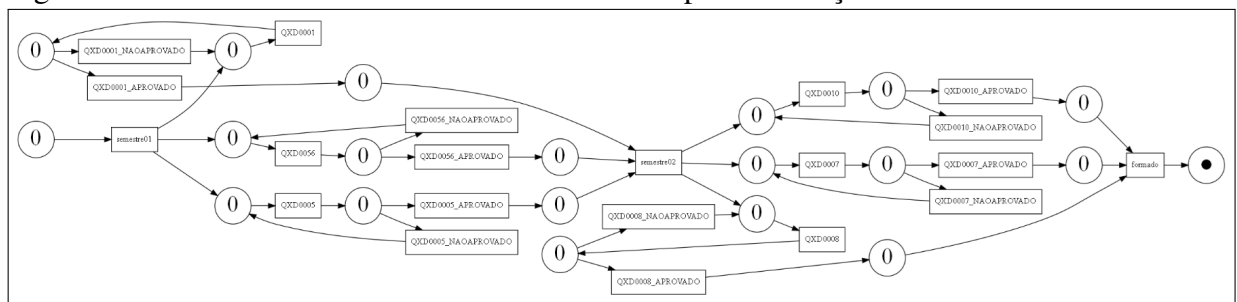
Figura 38 – Modelo reduzido da estrutura curricular para simulação



Fonte: elaborada pelo autor.

Ao atingir o final da trajetória acadêmica, a próxima atividade registrada no *trace* é *formado*, uma transição que verifica se o discente cursou com êxito todos os componentes curriculares necessários para a conclusão do curso. Neste cenário, o discente cumpriu todos os requisitos acadêmicos, finalizando com sucesso o curso. Dessa forma, a transição *formado* dispara, consumindo o *token* do lugar de entrada e produzindo um *token* no lugar de saída, marcando o fim do *trace*, conforme ilustrado na Figura 39.

Figura 39 – Modelo reduzido da estrutura curricular para simulação



Fonte: elaborada pelo autor.

Durante a simulação, os parâmetros de produção e consumo de tokens, ausência de tokens e tokens remanescentes foram monitorados e atualizados ao longo do processo. Neste cenário, foram contabilizados: 20 *tokens* produzidos e consumidos, 0 *tokens* remanescentes e 0 *tokens* ausentes.

Com esses valores, o *fitness* do *trace* é calculado de acordo com a Equação 2.1:

$$0,5 \left(1 - \frac{0}{20} \right) + 0,5 \left(1 - \frac{0}{20} \right) = 1,0$$

O valor do *fitness* igual a 1,0 reflete o comportamento ideal esperado dos discentes, totalmente alinhado com a estrutura curricular do curso. Isso significa que o discente realizou todos os componentes curriculares nos semestres previstos, dentro do tempo estabelecido de 8 semestres, e obteve aprovação em todas as componentes, sem interrupções ou reprovações ao longo de sua trajetória acadêmica.

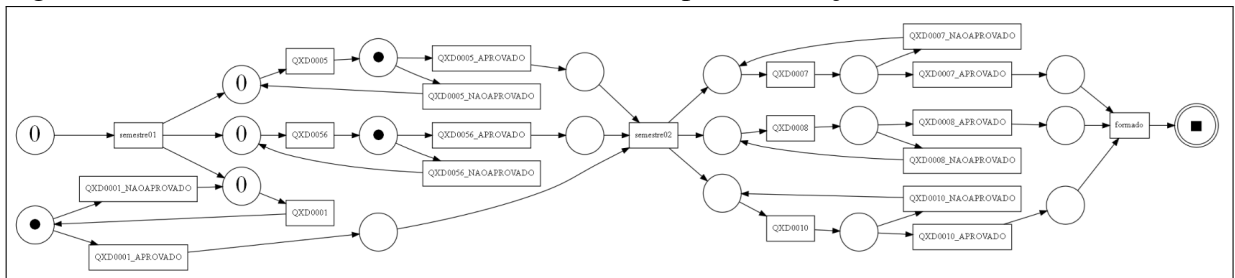
4.2.1.2 Cenário discente com reprovações

O cenário (ii) representa os discentes que não obtêm aprovação em todos os componentes curriculares cursados em um determinado semestre, resultando em reprovações parciais. Esse comportamento é evidenciado na Tabela 12, que detalha o *trace* de eventos do discente identificado com *id* 1002 e mostra que o aluno não conseguiu ser aprovado nas disciplinas *QXD0001* e *QXD0056*, pertencentes ao primeiro semestre do curso.

A execução inicial segue o mesmo padrão observado no cenário (i). O discente inicia sua trajetória acadêmica no 1º semestre, realizando matrícula simultânea nos três componentes curriculares obrigatórios: *QXD0001*, *QXD0005* e *QXD0056*, conforme apresentado na Figura 40.

Após cursar essas disciplinas, ao final do semestre, apenas o componente *QXD0005* é concluído com aprovação, enquanto os demais — *QXD0001* e *QXD0056* — resultam em reprovação. Essa dinâmica é capturada pela rede de Petri através da ativação das transições associadas ao *status* final de cada componente.

Figura 40 – Modelo reduzido da estrutura curricular para simulação



Fonte: elaborada pelo autor.

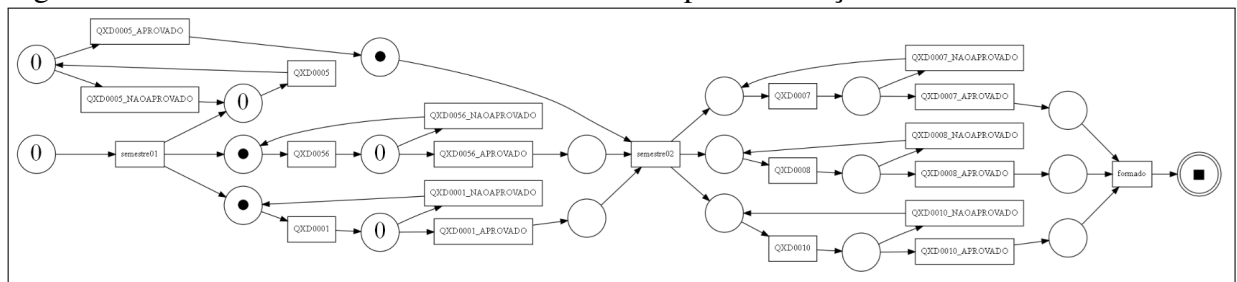
A transição *QXD0005_APROVADO* é disparada normalmente, representando o fluxo esperado em caso de sucesso acadêmico. Por outro lado, as transições *QXD0001_NAOAPROVADO*

Tabela 12 – *Trace* de eventos do discente 1002

timestamp	id_discente	activity
01/01/2017	1002	semestre01
01/02/2017	1002	QXD0001
01/02/2017	1002	QXD0005
01/02/2017	1002	QXD0056
01/06/2017	1002	QXD0001_NAOAPROVADO
01/06/2017	1002	QXD0005_APROVADO
01/06/2017	1002	QXD0056_NAOAPROVADO
01/07/2017	1002	semestre02
01/08/2017	1002	QXD0007
01/08/2017	1002	QXD0010
01/08/2017	1002	QXD0008
01/02/2017	1002	QXD0001
01/02/2017	1002	QXD0056
01/11/2017	1002	QXD0007_APROVADO
01/11/2017	1002	QXD0008_APROVADO
01/11/2017	1002	QXD0010_APROVADO
01/11/2017	1002	QXD0001_APROVADO
01/11/2017	1002	QXD0056_APROVADO
02/11/2017	1002	formado

e *QXD0056_NAOAPROVADO* indicam a reprovação e realimentam o fluxo, retornando o *token* para o lugar de entrada correspondente às disciplinas *QXD0001* e *QXD0056*, ilustrado pela Figura 41. Esse retorno modela o comportamento real do sistema acadêmico, no qual o discente precisa repetir os componentes nos quais não obteve êxito, reafirmando a característica cíclica da trajetória de formação em cursos de graduação.

Figura 41 – Modelo reduzido da estrutura curricular para simulação



Fonte: elaborada pelo autor.

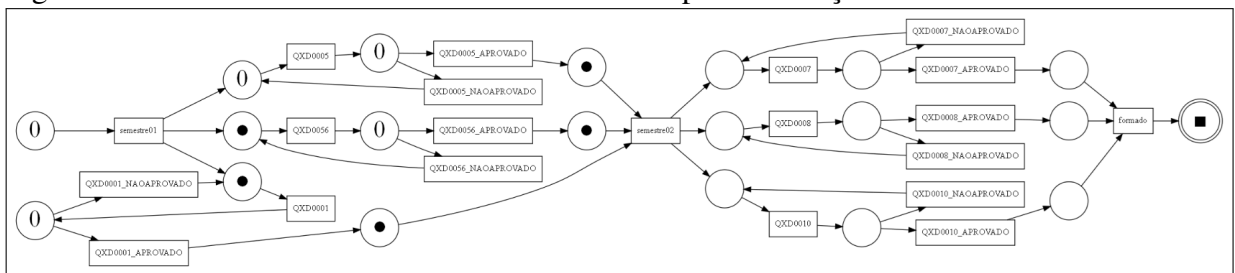
Neste ponto da simulação, a próxima atividade prevista é a transição *semestre02*, que representa a passagem temporal do discente para o segundo semestre letivo, independentemente de ele estar ou não cursando todos os componentes curriculares associados a esse novo período. Contudo, segundo as regras formais de disparo de transições em uma rede de Petri, a transição *semestre02* não pode ser ativada, pois não há *tokens* suficientes em todos os seus lugares

de entrada. Essa situação decorre das reprovações ocorridas nos componentes *QXD0001* e *QXD0056* durante o primeiro semestre, cujos lugares de saída não forneceram *tokens* para *semestre02*.

Como resultado, a ausência de dois *tokens* nas entradas da transição *semestre02* impede a continuidade natural do fluxo da rede. Essa condição é evidenciada visualmente na Figura 41, onde é possível observar a desativação da transição e a lacuna estrutural causada pelas reprovações.

Para dar prosseguimento o fluxo — e representar a lógica do avanço do tempo acadêmico, mesmo diante de pendências —, são inseridos dois *tokens* nos respectivos lugares de entrada, simulando a continuidade da trajetória acadêmica. Essa intervenção permite a habilitação da transição *semestre02*, restaurando o fluxo da rede, conforme ilustra a Figura 42.

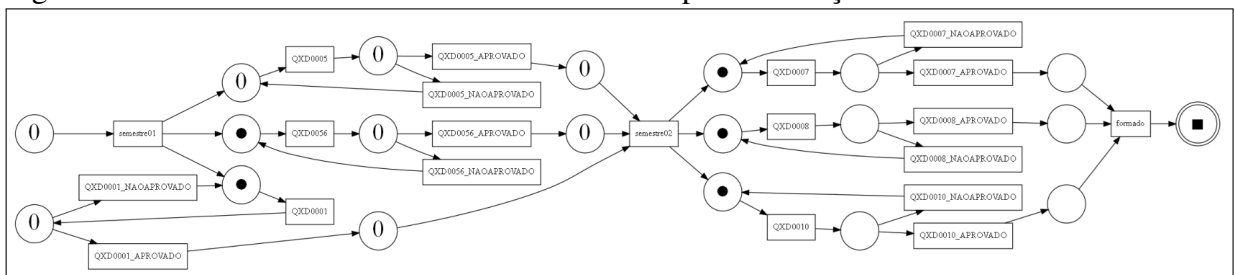
Figura 42 – Modelo reduzido da estrutura curricular para simulação



Fonte: elaborada pelo autor.

É importante destacar que os lugares de entrada dos componentes *QXD0001* e *QXD0056* mantêm *tokens* posicionados, representando o estado de espera para cursar novamente as disciplinas reprovadas. Essa modelagem reforça a capacidade da rede de Petri em representar cenários reais de progressão acadêmica com reprovações parciais, mantendo os ciclos ativos para tentativas futuras. Reestabelecida o fluxo da rede, a transição *semestre02* dispara e produz *tokens* nos lugares de saída, conforme a Figura 43.

Figura 43 – Modelo reduzido da estrutura curricular para simulação

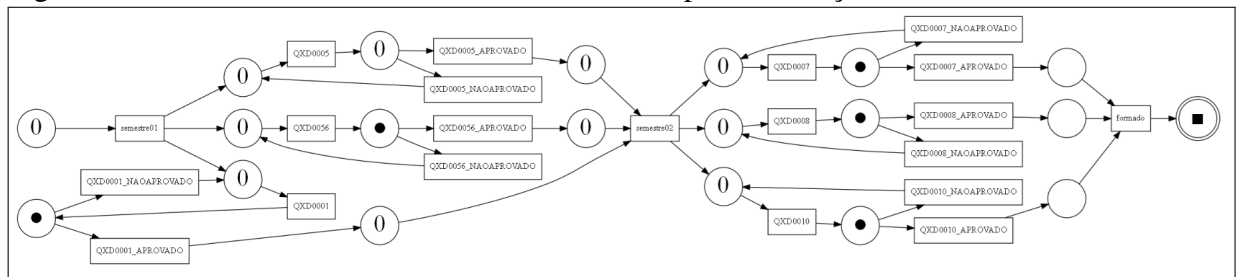


Fonte: elaborada pelo autor.

Dando continuidade à execução do *trace* do discente, as próximas atividades consistem na realização da matrícula nos componentes curriculares do segundo semestre, bem como na rematrícula nas disciplinas que foram reprovadas no semestre anterior. Esse comportamento acadêmico é modelado por meio do disparo das seguintes transições: *QXD0007*, *QXD0008* e *QXD0010*, que correspondem às disciplinas regulares do segundo semestre, e *QXD0001* e *QXD0056*, que representam os componentes reprovados que devem ser recursados.

Cada uma dessas transições, ao ser disparada, consome os *tokens* disponíveis em seus lugares de entrada e produz novos *tokens* nos respectivos lugares de saída, simbolizando o ingresso do discente em novas tentativas de conclusão dessas disciplinas. Esse momento da simulação está representado na Figura 44, que demonstra o estado atualizado da rede após a matrícula simultânea em cinco componentes curriculares. A modelagem reforça a flexibilidade da rede de Petri em lidar com situações complexas de reprovação, rematrícula e avanço parcial no percurso formativo de um discente.

Figura 44 – Modelo reduzido da estrutura curricular para simulação



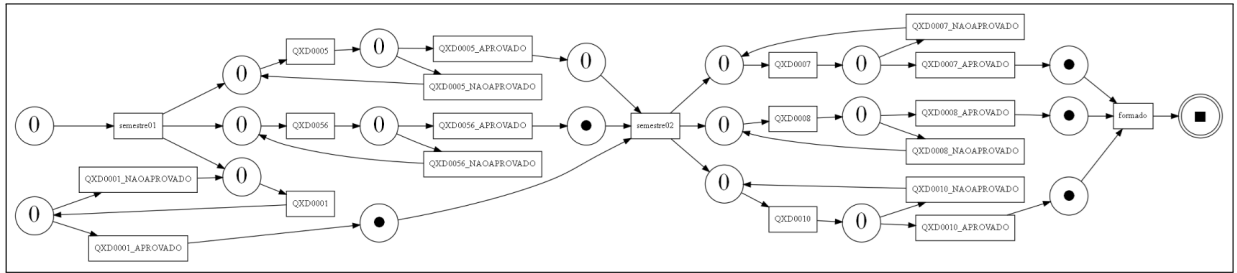
Fonte: elaborada pelo autor.

Seguindo o fluxo estabelecido no *trace*, as próximas transições a serem disparadas são: *QXD0007_APROVADO*, *QXD0008_APROVADO*, *QXD0010_APROVADO*, *QXD0001_APROVADO* e *QXD0056_APROVADO*. Essas transições representam a conclusão bem-sucedida de todas as disciplinas cursadas no segundo semestre, incluindo aquelas que haviam sido reprovadas anteriormente e foram cursadas novamente pelo discente neste ciclo.

A ativação dessas transições sinaliza que o aluno obteve êxito em todos os componentes curriculares em andamento. Ao serem disparadas, as transições consomem os *tokens* posicionados nos seus respectivos lugares de entrada e produzem *tokens* nos lugares de saída, representando a finalização dessas atividades acadêmicas com aprovação.

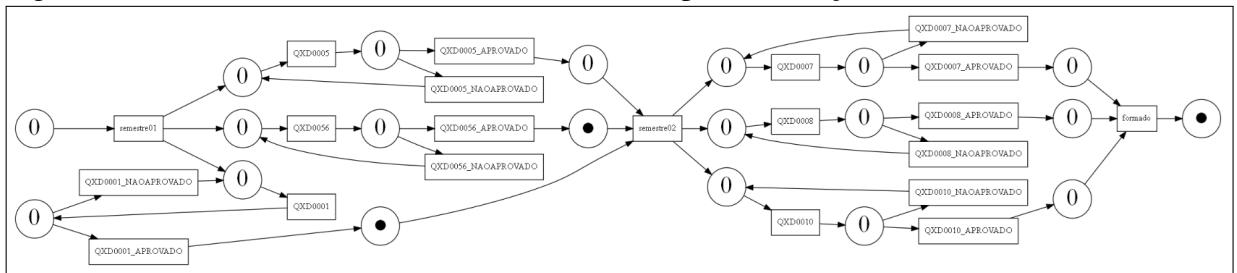
Por fim, a transição *formado* está habilitada pela presença dos *tokens* em todos os seus lugares de entrada, logo essa transição dispara consumindo os *tokens* e concluindo a simulação do *trace*.

Figura 45 – Modelo reduzido da estrutura curricular para simulação



Fonte: elaborada pelo autor.

Figura 46 – Modelo reduzido da estrutura curricular para simulação



Fonte: elaborada pelo autor.

Ao verificar a marcação final da rede de Petri ao término da simulação, observa-se a presença de três *tokens*: um posicionado no lugar final da rede — que representa a conclusão da trajetória do discente — e outros dois localizados nos lugares de entrada da transição *semestre02*. O *token* posicionado no lugar final é interpretado como parte do fluxo esperado e, portanto, é naturalmente consumido pelo sistema, indicando o encerramento do ciclo acadêmico daquele *trace*. Em contrapartida, os dois *tokens* presentes nas entradas da transição *semestre02* representam *tokens* remanescentes, ou seja, *tokens* que não foram devidamente consumidos ao longo da execução da rede e que permanecem no sistema ao final da simulação.

Esses *tokens* remanescentes surgem como consequência direta da intervenção realizada durante a simulação, em que dois *tokens* adicionados para permitir o disparo da transição *semestre02*, mesmo diante da ausência de pré-condições necessárias (devido à reprovação em dois componentes curriculares no primeiro semestre). Embora essa intervenção tenha sido necessária para assegurar a progressão temporal da trajetória do discente, ela introduz um desvio em relação ao comportamento ideal da rede, impactando diretamente nas métricas de avaliação de aderência ao modelo.

Ao final do *token replay*, foi registrado um total de 24 *tokens* produzidos e 24 *tokens* consumidos, além de 2 *tokens* ausentes e 2 *tokens* remanescentes. Com base nesses valores, é possível calcular o *fitness* do *trace* utilizando a Equação 2.1:

$$0,5 \left(1 - \frac{2}{24}\right) + 0,5 \left(1 - \frac{2}{24}\right) = 0,91667$$

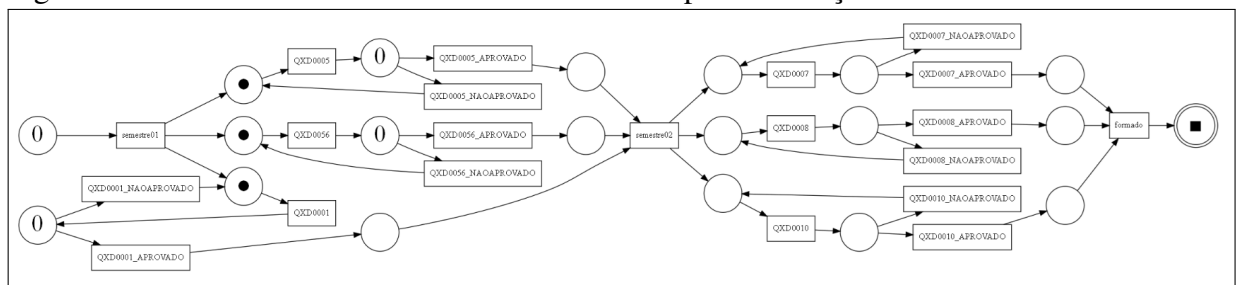
Esse valor de *fitness* evidencia que, embora o fluxo principal tenha sido preservado, a presença de reprovações — e consequentemente, a necessidade de intervir na dinâmica da rede para permitir a continuidade — impôs um impacto negativo na conformidade entre o comportamento observado e o modelo idealizado. Assim, mesmo que o discente tenha concluído o curso, os desvios intermediários causados pelas reprovações influenciam diretamente a métrica de desempenho do processo, refletindo uma trajetória com menor aderência ao caminho esperado.

4.2.1.3 Cenário de reprovações em todos os componentes de um semestre

O cenário (iii) ilustra um comportamento recorrente, especialmente entre os discentes que eventualmente evadem do curso: a reprovação em todos os componentes curriculares do 1º semestre. Nesse caso, o discente não obtém êxito em nenhuma das disciplinas cursadas, comprometendo seu avanço na estrutura curricular. O *trace* correspondente ao discente 1003 é apresentado na Tabela 13.

A simulação se inicia da mesma forma que nos cenários (i) e (ii): o discente ingressa no curso, realiza matrícula nos componentes curriculares do 1º semestre — *QXD0001*, *QXD0005* e *QXD0056* — e os cursa. No entanto, ao final do semestre, sua situação acadêmica é representada pelas atividades *QXD0001_NAOAPROVADO*, *QXD0005_NAOAPROVADO* e *QXD0056_NAOAPROVADO*, indicando reprovação integral. Como resultado, essas transições são disparadas e os *tokens* retornam para os lugares de entrada correspondentes, conforme representado na Figura 47.

Figura 47 – Modelo reduzido da estrutura curricular para simulação



Fonte: elaborada pelo autor.

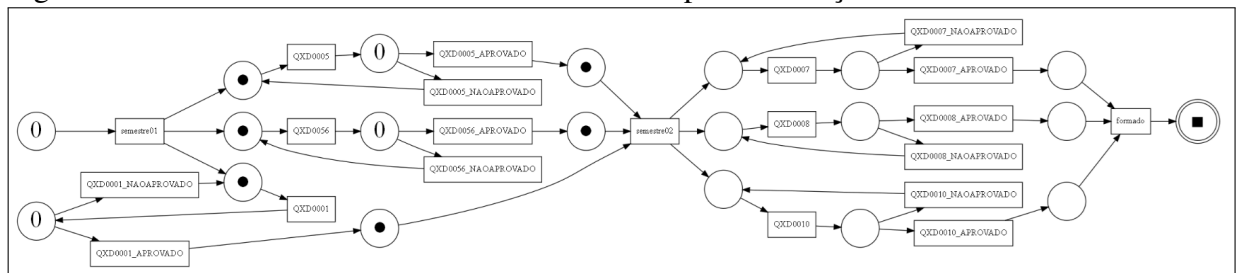
Diante dessa situação, o discente necessita de, pelo menos, um semestre adicional além do tempo idealizado pela estrutura curricular para conseguir recuperar os componentes não

Tabela 13 – *Trace* de eventos do discente 1003

timestamp	id_discente	activity
01/01/2017	1003	semestre01
01/02/2017	1003	QXD0001
01/02/2017	1003	QXD0005
01/02/2017	1003	QXD0056
01/06/2017	1003	QXD0001_NAOAPROVADO
01/06/2017	1003	QXD0005_NAOAPROVADO
01/06/2017	1003	QXD0056_NAOAPROVADO
01/01/2017	1003	semestre02
01/02/2017	1003	QXD0001
01/02/2017	1003	QXD0005
01/02/2017	1003	QXD0056
01/06/2017	1003	QXD0001_APROVADO
01/06/2017	1003	QXD0005_APROVADO
01/06/2017	1003	QXD0056_APROVADO
01/07/2017	1003	semestre03
01/08/2017	1003	QXD0007
01/08/2017	1003	QXD0010
01/08/2017	1003	QXD0008
01/11/2017	1003	QXD0007_APROVADO
01/11/2017	1003	QXD0008_APROVADO
01/11/2017	1003	QXD0010_APROVADO
02/11/2017	1003	formado

concluídos. Na sequência do *trace*, observa-se a tentativa de disparo da transição *semestre02*, que representa o avanço temporal do curso. No entanto, ela não pode ser disparada, pois não está habilitada — ou seja, os lugares de entrada não contêm os *tokens* necessários devido às reprovações anteriores. Assim, registram-se três *tokens* ausentes, que precisam ser adicionados à rede para habilitar a transição, conforme ilustrado na Figura 48. Durante esse período, os *tokens* referentes às disciplinas reprovadas permanecem estacionados, aguardando nova tentativa de matrícula.

Figura 48 – Modelo reduzido da estrutura curricular para simulação

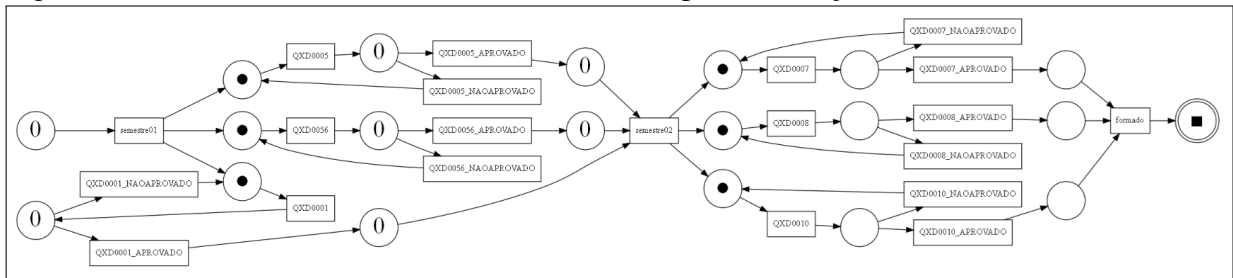


Fonte: elaborada pelo autor.

Após a adição dos *tokens* ausentes, a transição *semestre02* é disparada, consumindo

os *tokens* inseridos e produzindo novos nos seus respectivos lugares de saída, como mostrado na Figura 49. Neste estágio do fluxo, os *tokens* aguardam a matrícula do discente nas disciplinas. Embora o discente esteja, temporalmente, no 2º semestre, ele se matricula novamente nos componentes do 1º semestre devido às reprovações, reiniciando o ciclo para esses componentes.

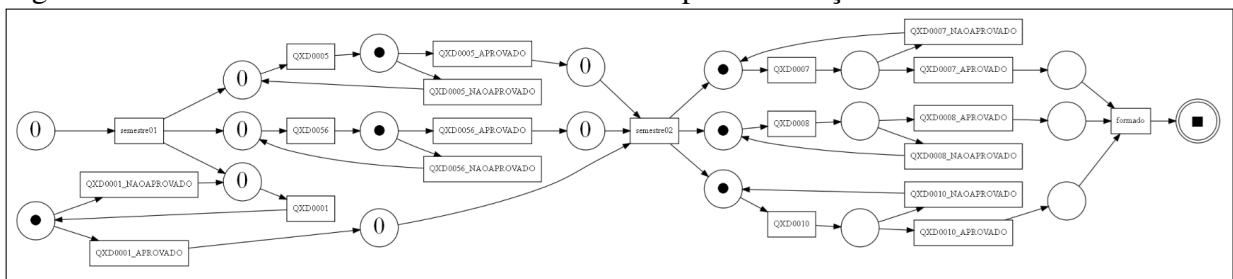
Figura 49 – Modelo reduzido da estrutura curricular para simulação



Fonte: elaborada pelo autor.

Na sequência do *trace*, são registradas as atividades *QXD0001*, *QXD0005* e *QXD0056*, indicando que o discente irá cursar novamente os mesmos componentes do 1º semestre. As respectivas transições são disparadas, consumindo os *tokens* de entrada e produzindo novos nos lugares de saída, como ilustrado na Figura 50.

Figura 50 – Modelo reduzido da estrutura curricular para simulação

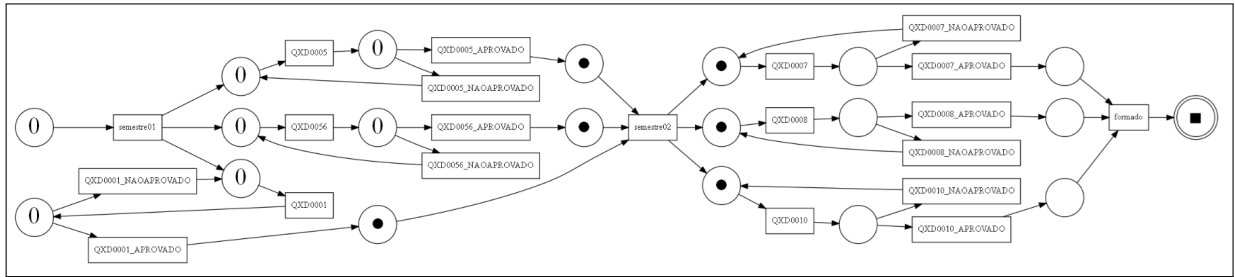


Fonte: elaborada pelo autor.

Dando continuidade ao *trace*, são registradas as atividades que representam o *status* final de cada componente: *QXD0001_APROVADO*, *QXD0005_APROVADO* e *QXD0056_APROVADO*. Como o discente obteve aprovação em todas as disciplinas recursadas, as transições são disparadas, consumindo os *tokens* de entrada e produzindo novos nos lugares de saída, conforme demonstrado na Figura 51.

Em seguida, a próxima atividade registrada é *semestre03*, que não está presente no modelo da rede de Petri. Por esse motivo, essa atividade não resulta na produção nem no consumo de *tokens*. O fluxo prossegue com o registro das matrículas do discente nos componentes do 2º semestre: *QXD0007*, *QXD0008* e *QXD0010*. As transições correspondentes são disparadas,

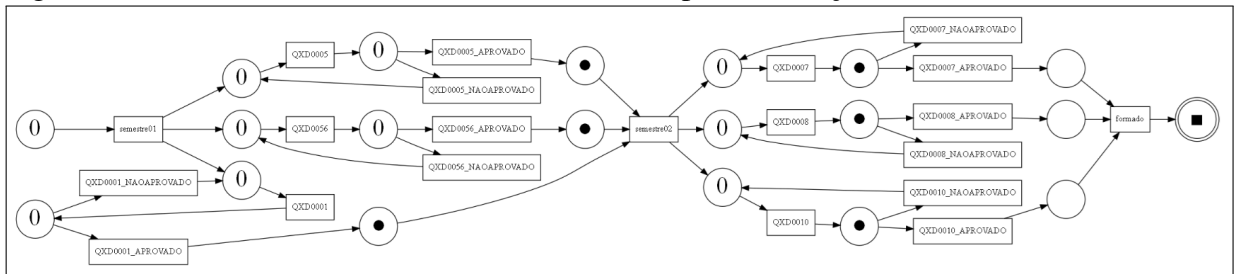
Figura 51 – Modelo reduzido da estrutura curricular para simulação



Fonte: elaborada pelo autor.

consumindo os *tokens* dos lugares de entrada e produzindo nos respectivos lugares de saída, como apresentado na Figura 52.

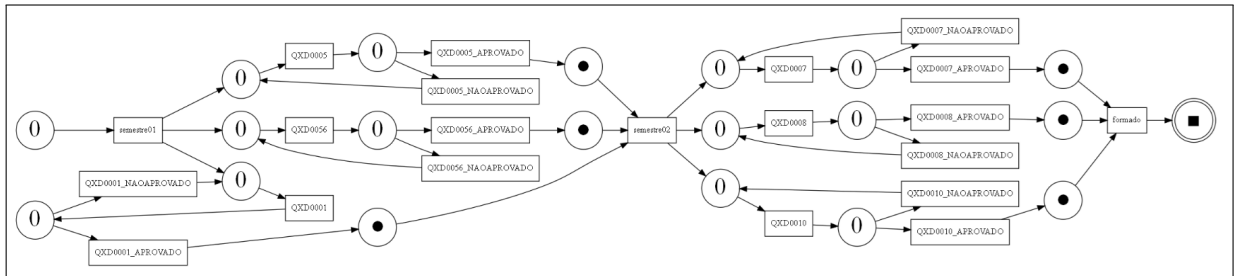
Figura 52 – Modelo reduzido da estrutura curricular para simulação



Fonte: elaborada pelo autor.

Na sequência, as atividades *QXD0007_APROVADO*, *QXD0008_APROVADO* e *QXD0010_APROVADO* são registradas, indicando que o discente foi aprovado em todos os componentes em que estava matriculado. Assim, os *tokens* são consumidos das transições de entrada e novos são produzidos nas saídas, completando o ciclo das disciplinas do 2º semestre, conforme ilustrado na Figura 53.

Figura 53 – Modelo reduzido da estrutura curricular para simulação

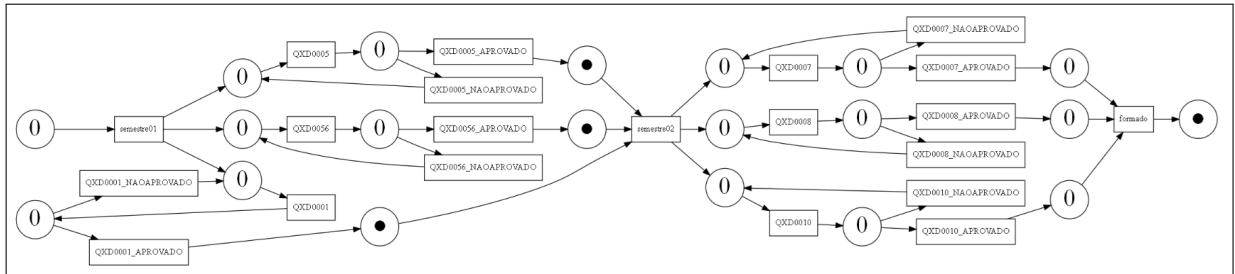


Fonte: elaborada pelo autor.

Por fim, a transição *formado*, que se encontra habilitada, é disparada, consumindo os três *tokens* dos lugares de entrada e produzindo um novo *token* no lugar final da rede, como representado na Figura 54. Neste ponto, a rede contém quatro *tokens* remanescentes: um no

lugar final — que será consumido naturalmente para permitir a execução do próximo *trace* — e três nos lugares de entrada da transição *semestre02*, caracterizando *tokens* não consumidos ao final da simulação.

Figura 54 – Modelo reduzido da estrutura curricular para simulação



Fonte: elaborada pelo autor.

Ao final da simulação, foram registrados 26 *tokens* produzidos, 26 *tokens* consumidos, 3 *tokens* ausentes e 3 *tokens* remanescentes. Esses valores alimentam o cálculo do *fitness* do *trace*, utilizando a Equação 2.1, que expressa o grau de conformidade entre a execução real e o comportamento esperado pelo modelo de rede de Petri.

$$0,5 \left(1 - \frac{3}{26}\right) + 0,5 \left(1 - \frac{3}{26}\right) = 0,8846$$

A presença de *tokens* ausentes e remanescentes evidencia que o fluxo do discente apresentou desvios relevantes em relação ao modelo acadêmico idealizado. Esses desvios ocorreram, principalmente, devido às reprovações em componentes curriculares nos semestres iniciais, que exigiram a inserção de *tokens* para permitir a progressão na rede. Isso compromete o alinhamento entre o comportamento observado e o processo esperado, impactando negativamente o valor do *fitness*.

Assim, o valor de 0,8846 não apenas quantifica a aderência parcial do *trace* ao modelo, mas também reforça a sensibilidade da métrica de *fitness* em capturar os efeitos estruturais das reprovações acadêmicas, evidenciando a capacidade de captura do fenômenos inesperado e identificação de pontos críticos no processo educacional.

5 EXPERIMENTOS

Neste capítulo, são apresentadas as análises realizadas a partir da aplicação de técnicas de verificação de conformidade (*conformance checking*) sobre o modelo formal da estrutura curricular do curso de Ciência da Computação da Universidade Federal do Ceará (UFC), Campus de Quixadá. O objetivo é avaliar o grau de aderência das trajetórias acadêmicas dos discentes ao fluxo ideal definido pela matriz curricular, com base no *log* de eventos construído a partir de dados reais.

Para isso, são exploradas duas técnicas centrais da Mineração de Processos: *token replay* e alinhamentos (*alignments*). A técnica de *token replay* permite simular a execução das trajetórias dos discentes dentro da rede de Petri, identificando padrões de progressão, retenções, reprovações e evasão com base na movimentação de *tokens* no modelo. Já a técnica de alinhamentos possibilita uma comparação detalhada entre os eventos registrados no *log* e a sequência ideal de atividades prevista pelo modelo, revelando desvios estruturais como disciplinas cursadas fora da ordem, atrasos e trancamentos.

A estrutura do capítulo está organizada da seguinte forma: a Seção 5.1 apresenta a aplicação do *token replay*, destacando métricas de conformidade e pontos críticos no fluxo acadêmico. Em seguida, a Seção 5.2 discute os resultados obtidos por meio da técnica de alinhamentos, evidenciando os padrões de desalinhamento entre o modelo e a realidade acadêmica. Por fim, na Seção 5.3 são discutidas implicações dos achados para a gestão curricular e a permanência estudantil.

5.1 Análise de conformidade utilizando a técnica de *token replay*

A técnica de *token replay* fundamenta-se na reprodução das trajetórias acadêmicas dos discentes em uma rede de Petri, na qual os *tokens* representam a evolução dos estudantes ao longo dos componentes curriculares do curso. Essa simulação permite verificar o grau de conformidade entre os caminhos percorridos pelos alunos e o fluxo acadêmico idealizado na estrutura curricular formal. Durante a execução do *replay*, são monitorados quatro indicadores principais: *tokens* produzidos, *tokens* consumidos, *tokens* ausentes e *tokens* restantes, que possibilitam quantificar desvios, gargalos e interrupções nas trajetórias acadêmicas.

Em um cenário ideal — no qual o estudante cursa todas as disciplinas nos semestres previstos e obtém aprovação em todas elas — o *trace fitness* atinge o valor máximo de 1,0. Neste

caso, o número de *tokens* produzidos e consumidos é igual a 123, não havendo *tokens* ausentes ou remanescentes, o que caracteriza uma trajetória perfeitamente aderente ao modelo.

No entanto, desvios em relação ao fluxo ideal resultam em alterações nos indicadores. Os *tokens* ausentes são contabilizados quando uma transição é disparada sem que todos os pré-requisitos estejam satisfeitos, o que pode ocorrer, por exemplo, quando um componente curricular é cursado fora do semestre ideal. Já os *tokens* restantes correspondem a *tokens* que permanecem na rede após a finalização do *trace*, indicando interrupções no percurso, como trancamentos, não aprovações não superadas ou evasão.

5.1.1 Resultados para discentes formados

A Tabela 14 apresenta as estatísticas obtidas ao final do *token replay* para os 38 discentes que concluíram todas as disciplinas obrigatórias. O valor médio de *trace fitness* observado foi 0,9217, indicando alta aderência das trajetórias ao modelo acadêmico ideal. O desvio padrão de 0,0658 revela baixa variabilidade entre os formandos, sugerindo que a maior parte seguiu percursos próximos ao previsto.

Tabela 14 – Estatísticas dos alunos formados

	<i>trace fitness</i>	<i>tokens produzidos</i>	<i>tokens consumidos</i>	<i>tokens ausentes</i>	<i>tokens restantes</i>
<i>count</i>	38.000000	38.000000	38.000000	38.000000	38.000000
<i>mean</i>	0.921701	136.526316	142.184211	14.868421	9.210526
<i>std</i>	0.065766	15.508911	21.778809	14.171099	8.412211
<i>min</i>	0.800262	123.000000	123.000000	0.000000	0.000000
25%	0.880575	123.000000	123.000000	2.250000	2.000000
50%	0.932086	129.000000	132.500000	9.500000	8.000000
75%	0.982829	145.000000	155.250000	25.250000	13.000000
<i>max</i>	1.000000	185.000000	206.000000	50.000000	29.000000

Fonte: elaborada pelo autor.

Os valores mínimos e máximos de *trace fitness* foram 0,8003 e 1,0, respectivamente, refletindo, por um lado, discentes que enfrentaram não aprovações ou alterações no fluxo e, por outro, alunos cuja trajetória foi totalmente compatível com a matriz curricular.

Quanto à movimentação de *tokens*, a média de *tokens* produzidos e consumidos foi de 136,5 e 142,2, respectivamente, números superiores ao ideal, o que indica a ocorrência de disciplinas repetidas ao longo da formação. A média de *tokens* ausentes foi 14,87, enquanto a de remanescentes foi 9,21, com um desvio padrão de 8,41. Esses valores revelam pequenas dissonâncias que, embora não tenham impedido a conclusão do curso, representam obstáculos pontuais no processo formativo.

A análise por *quartis* mostra que:

- No 1º *quartil* (25%), os estudantes produziram e consumiram exatamente 123 *tokens*, com valores medianos de apenas 2 *tokens* ausentes e remanescentes.
- Na mediana, esses números aumentam, apontando trajetórias com maiores ajustes, como repetência e alterações de ordem.
- No 3º *quartil* (75%), os valores saltam para 145 *tokens* produzidos e 155,25 consumidos, com 25,25 *tokens* ausentes e 13 remanescentes, indicando maior número de não aprovações.

5.1.2 Resultados para discentes evadidos

A Tabela 15 apresenta os dados estatísticos para os 76 discentes que não completaram o curso. O valor médio de *trace fitness* foi 0,5815, com desvio padrão de 0,2105, evidenciando trajetórias mais irregulares e heterogêneas. Os valores extremos (0,2169 a 0,9527) indicam desde evasão precoce até casos em que a conclusão esteve próxima, mas não foi atingida.

Em termos de movimentação de *tokens*, os evadidos produziram em média 95,28 *tokens* e consumiram 99,54, o que sugere que a maioria não percorreu integralmente a matriz curricular. A média de *tokens* ausentes foi 33,86, e de remanescentes, 29,59 — números significativamente superiores aos dos discentes formados.

Tabela 15 – Estatísticas dos alunos evadidos

	<i>trace fitness</i>	<i>tokens produzidos</i>	<i>tokens consumidos</i>	<i>tokens ausentes</i>	<i>tokens restantes</i>
<i>count</i>	76.000000	76.000000	76.000000	76.000000	76.000000
<i>mean</i>	0.581503	95.276316	99.539474	33.855263	29.592105
<i>std</i>	0.210471	45.273568	52.728346	8.634356	7.031695
<i>min</i>	0.216899	42.000000	41.000000	9.000000	4.000000
25%	0.389835	54.000000	51.000000	31.000000	28.000000
50%	0.589310	79.000000	83.000000	32.000000	32.000000
75%	0.778881	137.750000	142.500000	38.000000	33.250000
<i>max</i>	0.952701	194.000000	220.000000	55.000000	41.000000

Fonte: elaborada pelo autor.

A análise dos *quartis* reforça esses achados:

- No 1º *quartil*, os discentes apresentaram *fitness* inferior a 0,39 e produziram menos de 55 *tokens*, revelando evasão precoce.
- No 3º *quartil*, há casos de maior envolvimento, com até 137 *tokens* produzidos, mas com 38 *missing tokens* e 33,25 *remaining tokens*, apontando tentativas repetidas e fracassadas de progressão.

5.1.3 Comparação entre grupos

A comparação entre formados e evadidos evidencia contrastes marcantes. Os discentes formados demonstram maior conformidade ao fluxo acadêmico ideal, com valores de *trace fitness* significativamente superiores e menores quantidades de *tokens* ausentes e remanescentes. Já os evadidos apresentam trajetórias fragmentadas, com maior quantidade de desvios, interrupções e não aprovações.

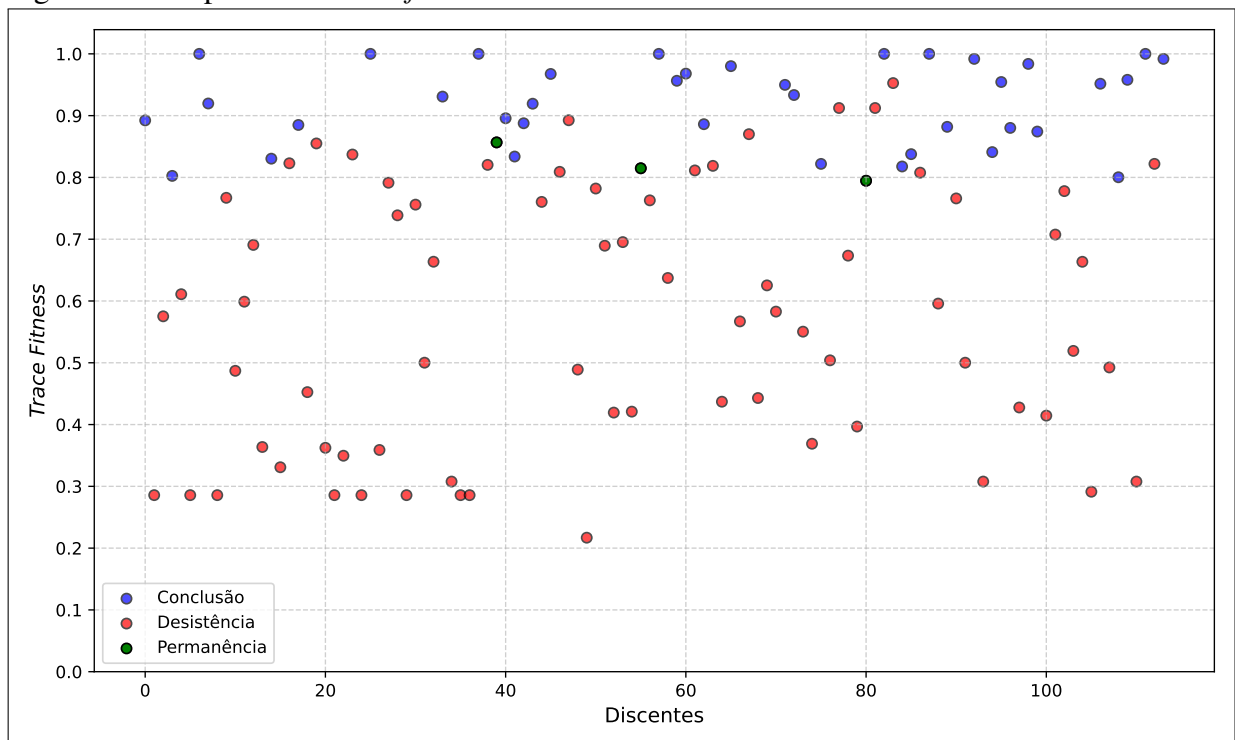
Esses resultados reforçam a importância de ações institucionais de acompanhamento e apoio acadêmico, especialmente nos semestres iniciais, onde ocorrem os primeiros indícios de desalinhamento. A identificação precoce de discentes com alto acúmulo de *tokens* ausentes ou remanescentes pode viabilizar intervenções para mitigar a evasão.

5.1.4 Distribuição do *trace fitness* e padrões de progressão acadêmica

A Figura 55 apresenta a dispersão dos valores de *trace fitness* obtidos para os discentes, separados em três grupos: alunos que concluíram o curso (formados), alunos que evadiram (não formados) e alunos que ainda estão com vínculo ativo (matriculados). Essa visualização permite identificar o grau de conformidade de cada trajetória em relação ao modelo de referência baseado na matriz curricular ideal.

Os discentes que concluíram o curso são representados pelos pontos azuis, esses discentes apresentam valores de *trace fitness* majoritariamente elevados, concentrando-se entre 0,90 e 1,00. Isso indica que suas trajetórias estiveram em alta conformidade com a modelagem curricular formal, com baixo número de reprovações, trancamentos ou alterações na ordem de disciplinas. Alguns alunos, mesmo com desvios leves (*fitness* entre 0,80 e 0,90), conseguiram ajustar sua trajetória e concluir o curso, o que demonstra certa flexibilidade na progressão acadêmica.

Os discentes evadidos representados pelos pontos vermelhos, os discentes evadidos demonstram ampla dispersão nos valores de *fitness*, que variam entre 0,20 e 0,95. Essa heterogeneidade revela diferentes padrões de evasão. Parte deles abandonou precocemente o curso, ainda nos primeiros semestres (*fitness* abaixo de 0,50), enquanto outros avançaram até etapas intermediárias, mas interromperam a trajetória após acumularem reprovações ou atrasos críticos. A presença de alunos com *trace fitness* relativamente altos entre os evadidos evidencia que, em alguns casos, a evasão não se deve apenas ao desempenho acadêmico, mas possivelmente a

Figura 55 – Dispersão do *trace fitness* dos discentes

Fonte: elaborada pelo autor.

fatores externos à estrutura curricular.

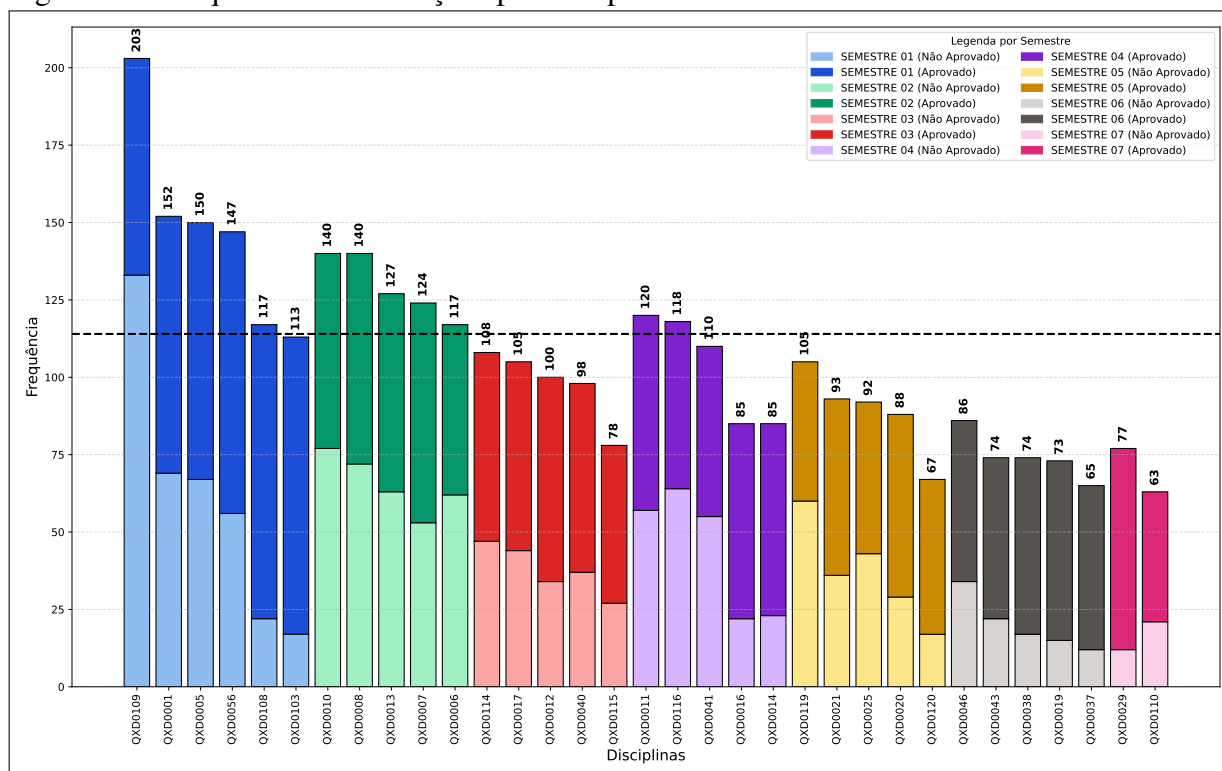
Os discentes em permanência indicados pelos pontos verdes, os estudantes ainda vinculados ao curso, apresentam *trace fitness* intermediário, geralmente entre 0,80 e 0,85. Essa faixa sugere que, até o momento da análise, esses discentes mantêm certa regularidade em sua progressão, embora já apresentem indícios de possíveis desafios, como pequenas reprovações ou alterações de fluxo. A continuidade do acompanhamento desses alunos é essencial para que estratégias de apoio possam ser implementadas de forma preventiva, evitando que migrem para o grupo dos evadidos.

5.1.5 Fatores estruturais que influenciam a evasão

A Figura 56 apresenta a frequência com que as transições associadas a cada componente curricular foram ativadas durante a simulação com a técnica de *token replay*. Cada transição corresponde à execução de uma disciplina, sendo ativada sempre que um *token* percorre aquele ponto da rede durante a trajetória acadêmica de um discente.

No cenário ideal, cada transição seria ativada uma única vez para cada estudante, refletindo o fluxo acadêmico conforme planejado pela matriz curricular. Com um total de 114 discentes no *log* de eventos, a frequência esperada por disciplina seria de 114 ativações

Figura 56 – Frequência das transições por disciplinas



Fonte: elaborada pelo autor.

— assumindo que todos os alunos cursaram a disciplina uma única vez e não houve evasões, reprovações ou trancamentos.

Contudo, os dados da figura revelam dois comportamentos distintos, que podem ser interpretados da seguinte forma:

1. **Frequência superior à esperada:** indica que a disciplina foi cursada mais de uma vez por um número significativo de discentes. Esse comportamento está diretamente relacionado a não aprovações ou trancamentos, que levam à repetição do componente curricular. Disciplinas com frequências elevadas sugerem maior grau de dificuldade, tornando-se potenciais gargalos no fluxo acadêmico.
2. **Frequência inferior à esperada:** reflete evasões totais ou parciais. Neste caso, o discente abandonou o curso ou interrompeu sua trajetória antes de cursar determinada disciplina, impedindo a ativação da transição correspondente. Essas ocorrências revelam pontos de interrupção frequentes ao longo da trajetória.

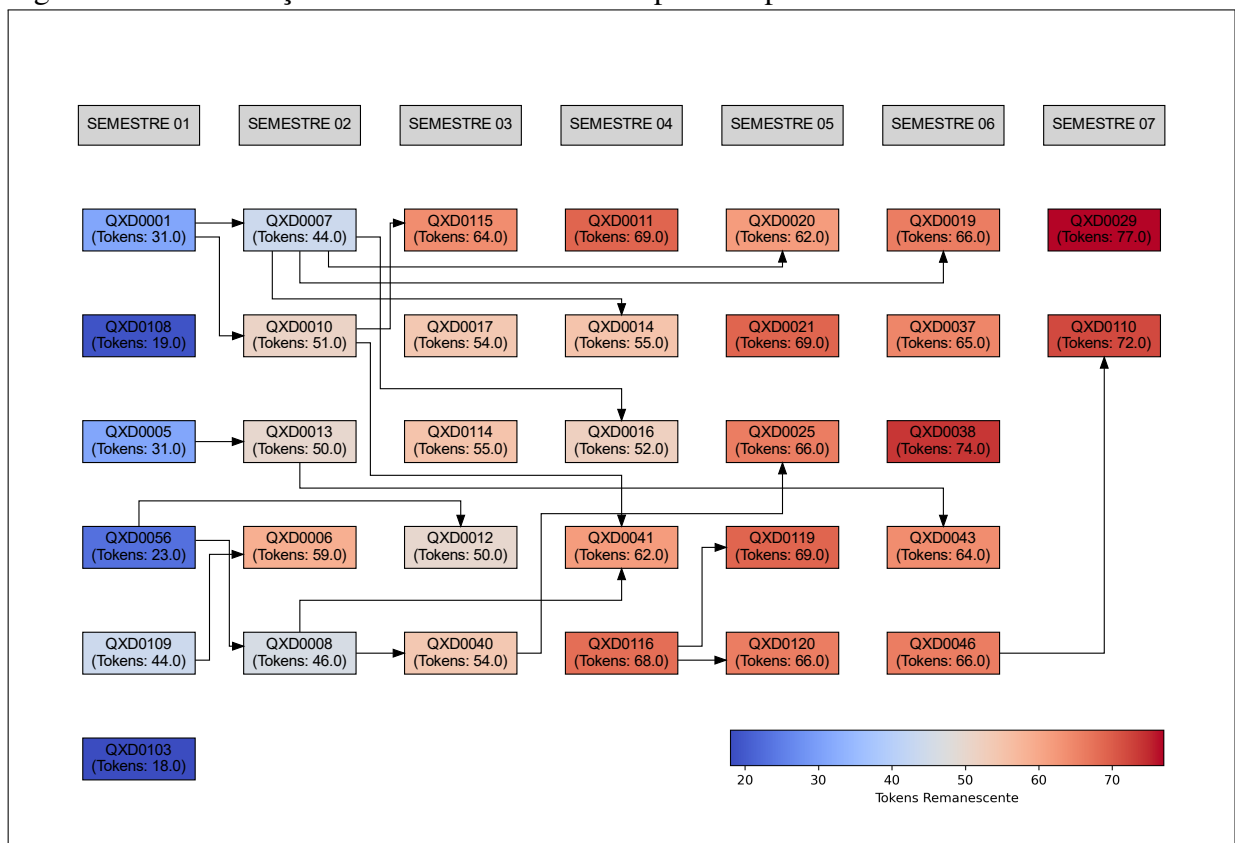
A análise da figura destaca que os maiores desvios de frequência estão concentrados nas disciplinas dos 1º e 2º semestres. Todas as transições correspondentes a esses períodos apresentam ativações significativamente superiores à referência de 114 execuções, indicando que essas disciplinas foram repetidamente cursadas por diversos discentes.

Esses resultados indicam a necessidade de atenção especial às disciplinas introdutórias da matriz curricular, especialmente aquelas que atuam como pré-requisitos para outros componentes. Intervenções pedagógicas direcionadas, revisão de ementas e metodologias, além de apoio acadêmico sistemático nesses primeiros semestres, são medidas fundamentais para mitigar as taxas de reprovação e retenção, melhorando o fluxo de formação e reduzindo a evasão.

5.1.6 Análise dos pontos críticos nas trajetórias acadêmicas

A Figura 57 evidencia os locais da rede de Petri onde os discentes interromperam sua trajetória acadêmica, a partir da distribuição dos *tokens* remanescentes ao final da simulação com a técnica de *token replay*. Esses *tokens* remanescentes indicam os pontos da estrutura curricular onde houve retenção, abandono ou evasão.

Figura 57 – Distribuição de *tokens* remanescentes por disciplinas



Fonte: elaborada pelo autor.

A análise revela uma elevada concentração de *tokens* remanescentes a partir do 3º semestre. Esse padrão sugere que os principais desafios à permanência acadêmica não se limitam aos períodos iniciais do curso, mas tendem a se intensificar nos semestres intermediários. A alta retenção de *tokens* nessas disciplinas indica que muitos discentes enfrentam dificuldades

persistentes ao longo da trajetória formativa, possivelmente relacionadas à dificuldades de aprendizagem, reprovações recorrentes ou falta de adaptação ao ensino superior.

Esses pontos críticos estão relacionados, principalmente, a disciplinas introdutórias que fazem parte dos eixos formativos em *Formação básica em matemática* e *Formação básica em ciência da computação*. São componentes que funcionam como pré-requisitos para diversas outras disciplinas, o que amplia o impacto de uma eventual reprovação. Quando o discente não obtém êxito nessas disciplinas, sua capacidade de avançar na matriz curricular é comprometida, gerando um bloqueio estrutural que favorece a evasão.

Para ilustrar esse fenômeno, destaca-se uma trajetória acadêmica crítica composta pelas disciplinas *QXD0056*, *QXD0008*, *QXD0040* e *QXD0025*, que se distribuem entre o 1º e o 5º semestre do curso. A análise mostra que, já na transição do 1º para o 2º semestre, o número de *tokens* remanescentes dobra, evidenciando um gargalo na progressão. Ao final dessa trajetória, observa-se quase uma triplicação na quantidade de *tokens*, indicando retenções sucessivas ao longo do tempo.

Padrões semelhantes também são observados em outras ramificações da estrutura curricular, como:

- *QXD0001* → *QXD0007* → *QXD0020*, *QXD0019* ou *QXD0014*;
- *QXD0005* → *QXD0013* → *QXD0043*;
- *QXD0056* → *QXD0008* → *QXD0041*;
- *QXD0056* → *QXD0008* → *QXD0040* → *QXD0025*.

Em todas essas ramificações, observa-se um acúmulo progressivo de *tokens* remanescentes, o que reforça a presença de gargalos nos caminhos curriculares dependentes de sequência, limitando a fluidez da trajetória acadêmica.

O acúmulo de *tokens* remanescentes evidencia que a evasão dos discentes ocorre de forma progressiva ao longo de múltiplos períodos. Muitos estudantes permanecem retidos em disciplinas específicas por diversos semestres, sem conseguir superá-las, o que impede o cumprimento de pré-requisitos essenciais para a continuidade no curso. Essa interrupção sequencial compromete o avanço nas etapas posteriores da formação, reduz a ocupação das disciplinas intermediárias e finais, e cria um efeito cascata que afeta diretamente a fluidez do currículo e a eficiência da oferta institucional.

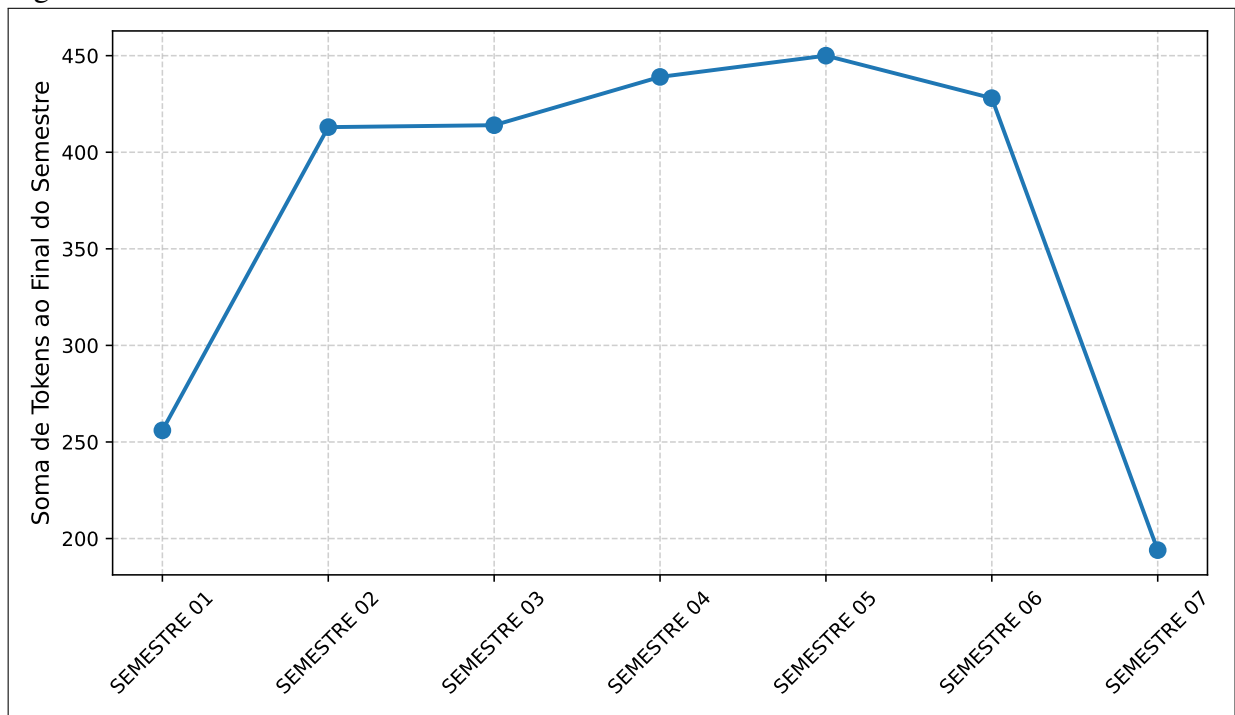
Além dos impactos na trajetória acadêmica, a permanência prolongada em componentes com alta taxa de reprovação acarreta consequências econômicas e estruturais. A repetição

constante de determinados componentes implica na reoferta frequente das mesmas disciplinas, sobrecarregando recursos docentes, ocupando vagas que poderiam ser direcionadas a outros alunos e dificultando o planejamento pedagógico da instituição.

5.1.7 Distribuição de tokens remanescentes por número de semestre letivo até a evasão

A Figura 58 apresenta a distribuição da quantidade de *tokens* remanescentes ao final da execução da técnica de *token replay* por semestre letivo. Essa métrica representa, para cada período da matriz curricular, o volume acumulado de discentes que interromperam sua trajetória naquele estágio específico do curso.

Figura 58 – Tokens remanescentes ao final do Semestre



Fonte: elaborada pelo autor.

Os dados revelam que o 5º semestre concentra o maior número de *tokens* remanescentes, seguido de perto pelo 4º e 6º semestres. O acúmulo de *tokens* nos semestres 4º, 5º e 6º não necessariamente indica que os estudantes estejam cursando integralmente os componentes curriculares previstos para esses períodos. Em muitos casos, esses discentes ainda estão tentando superar reprovações em disciplinas dos semestres iniciais, o que impede a progressão regular dentro da matriz curricular.

Esse fenômeno revela uma sobreposição de desafios: além da complexidade crescente dos conteúdos técnicos exigidos nas fases intermediárias do curso, há um efeito de retenção

acumulada que arrasta reprovações anteriores ao longo da trajetória acadêmica. Assim, a permanência aparente em semestres mais avançados pode mascarar um avanço real limitado, marcado por tentativas repetidas de aprovação em componentes fundamentais não vencidos nos primeiros períodos.

Observa-se também um aumento expressivo no acúmulo de *tokens* entre o 1º e 2º semestres, indicando que o início da trajetória já representa um ponto de inflexão para parte dos estudantes. A partir do 2º semestre, os valores mantêm-se elevados até o 6º, com relativa estabilidade entre o 2º, 3º e 4º períodos. Esse padrão reforça a ideia de que dificuldades acumuladas nos primeiros semestres se propagam, impactando negativamente o fluxo acadêmico nos períodos subsequentes.

O 7º semestre apresenta queda acentuada na quantidade de *tokens* remanescentes, o que pode indicar que poucos alunos atingem esse estágio sem antes terem interrompido o percurso. Isso reforça a hipótese de que grande parte da evasão ocorre antes da consolidação das etapas finais da graduação. Além disso, o acúmulo de *tokens* não se restringe a disciplinas específicas, mas reflete um conjunto de fatores interligados: reprovações em componentes com forte dependência curricular, dificuldade em vencer pré-requisitos e desmotivação frente à repetência sucessiva.

5.2 Análise de conformidade utilizando a técnica de alinhamentos (*Alignments*)

A técnica de alinhamentos (*alignments*) é amplamente utilizada na verificação de conformidade entre um *log* de eventos e um modelo de processo. Seu objetivo é mapear as diferenças entre a sequência de atividades observadas na execução real e o comportamento ideal prescrito pelo modelo. Para isso, são considerados três tipos de movimentos:

- **Movimentos síncronos:** quando o evento observado no *log* coincide com a transição prevista no modelo.
- **Movimentos do modelo:** quando uma transição esperada pelo modelo não é observada no *log*.
- **Movimentos do *log*:** quando uma atividade registrada no *log* não é contemplada no modelo.

Esses movimentos são avaliados em sequência para cada instância do processo e servem de base para o cálculo do *fitness* — métrica que quantifica o grau de aderência da trajetória observada ao modelo ideal. Valores de *fitness* próximos de 1 indicam forte conformidade,

enquanto valores mais baixos sugerem desvios significativos.

A Tabela 16 exemplifica esse processo, apresentando os movimentos identificados no alinhamento do primeiro semestre de um discente. Inicialmente, observam-se movimentos síncronos entre as disciplinas cursadas e aquelas previstas pela matriz curricular. Em seguida, o modelo espera a ocorrência de aprovações nas disciplinas — eventos que não são encontrados no *log*, gerando movimentos do modelo. Por fim, as reprovações registradas no *log*, como *QXD0005_NAOAPROVADO*, não estão contempladas no modelo, resultando em movimentos exclusivos do *log*.

Tabela 16 – Resultado do alinhamento do 1º semestre de um discente.

<i>Trace log</i>	Evento Modelo
semestre01	semestre01
QXD0005	QXD0005
QXD0103	QXD0103
QXD0001	QXD0001
QXD0108	QXD0108
QXD0056	QXD0056
QXD0109	QXD0109
>>	QXD0005_APROVADO
>>	QXD0103_APROVADO
>>	QXD0001_APROVADO
>>	QXD0108_APROVADO
>>	QXD0056_APROVADO
>>	QXD0109_APROVADO
QXD0005_NAOAPROVADO	>>
QXD0103_NAOAPROVADO	>>
QXD0001_NAOAPROVADO	>>
QXD0108_NAOAPROVADO	>>
QXD0056_NAOAPROVADO	>>
QXD0109_NAOAPROVADO	>>

Fonte: elaborada pelo autor.

Embora o modelo utilizado represente formalmente caminhos de não aprovação, o algoritmo de alinhamento busca o percurso ótimo, isto é, o que conduz do estado inicial ao estado final com o menor custo. Como resultado, o alinhamento geralmente privilegia caminhos bem-sucedidos e penaliza desvios como reprovações, trancamentos ou disciplinas fora de ordem.

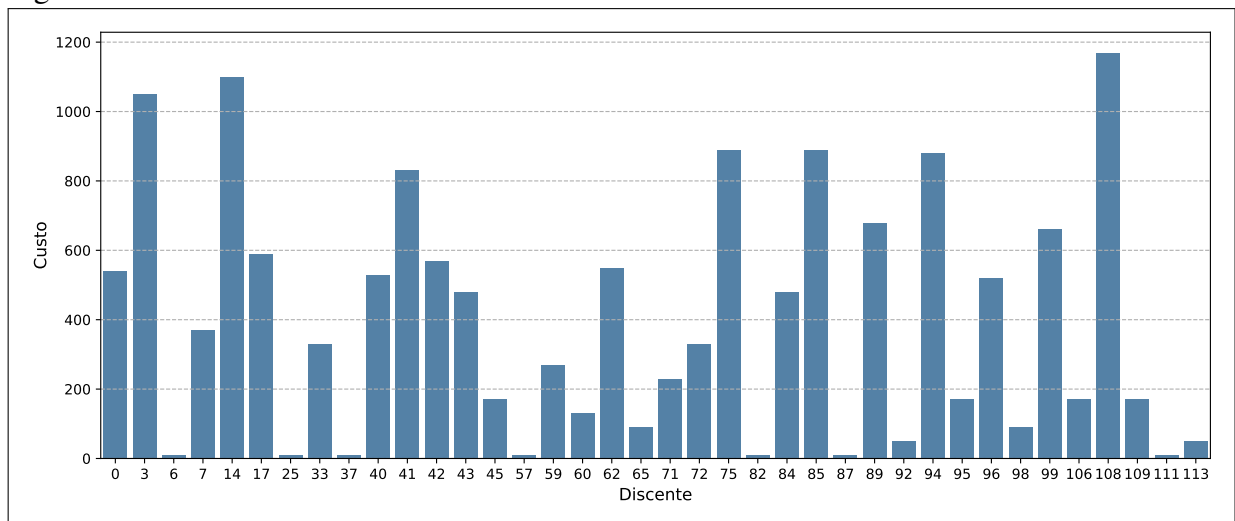
A partir desse procedimento, foi possível gerar métricas quantitativas que distinguem o comportamento dos discentes formados, não formados e em curso. A seguir, essa análise é segmentada por grupo de estudantes para aprofundar os padrões de conformidade e desvios observados.

5.2.1 Alinhamento dos discentes formados

A Figura 59 apresenta os custos de alinhamento individuais dos estudantes que concluíram o curso. Essa métrica representa o esforço necessário para adequar cada *trace* do *log* ao modelo idealizado da trajetória acadêmica, considerando os desvios identificados durante o alinhamento. Custos baixos indicam trajetórias próximas do modelo esperado, enquanto valores elevados sugerem múltiplas inconformidades, como reprovações, trancamentos e mudanças na ordem das disciplinas cursadas.

Ao observar a distribuição dos custos, verifica-se uma significativa heterogeneidade entre os discentes formados. Estudantes como 6, 25, 37, 57, 82, 87 e 111 apresentam custos inferiores a 100, sinalizando uma adesão quase total ao modelo acadêmico, com percursos regulares e aprovação nos componentes curriculares nos semestres esperados. Por outro lado, discentes como 3, 14, 41, 75, 85, 94 e 108 exibem custos superiores a 1.000, evidenciando percursos marcados por diversos desvios e múltiplas tentativas de aprovação em disciplinas, ainda que tenham conseguido finalizar o curso com êxito.

Figura 59 – Custo de alinhamento dos discentes formados

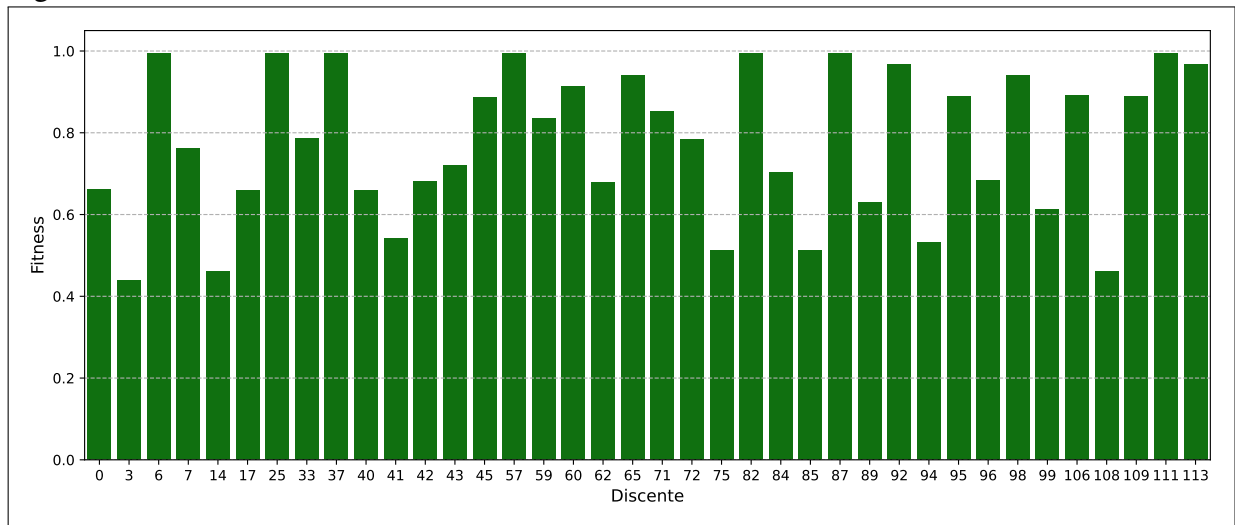


Fonte: elaborada pelo autor.

A Figura 60 complementa essa análise ao apresentar os valores de *trace fitness* para o mesmo grupo de estudantes. Essa métrica, variando entre 0 e 1, indica o grau de conformidade do *trace* com o modelo formal. A maioria dos discentes formados apresenta valores acima de 0,7, sinalizando uma trajetória relativamente aderente, mesmo que não isenta de pequenos desvios.

A comparação entre as Figuras 59 e 60 confirma a relação inversamente proporcional entre as métricas: quanto maior o custo de alinhamento, menor tende a ser o valor de *fitness*.

Figura 60 – *Fitness* de alinhamento dos discentes formados



Fonte: elaborada pelo autor.

Essa relação decorre diretamente do funcionamento da técnica de alinhamento, na qual cada desvio do modelo ideal acarreta penalizações cumulativas que reduzem a aderência do *trace*.

Apesar disso, mesmo entre discentes com *fitness* inferiores a 0,6, foi possível observar a conclusão do curso, o que demonstra a robustez da trajetória desses estudantes frente às dificuldades enfrentadas. Esses casos ilustram a complexidade dos percursos acadêmicos reais, reforçando a importância de considerar tanto o custo total quanto o valor do *fitness* para uma avaliação mais abrangente do desempenho estudantil.

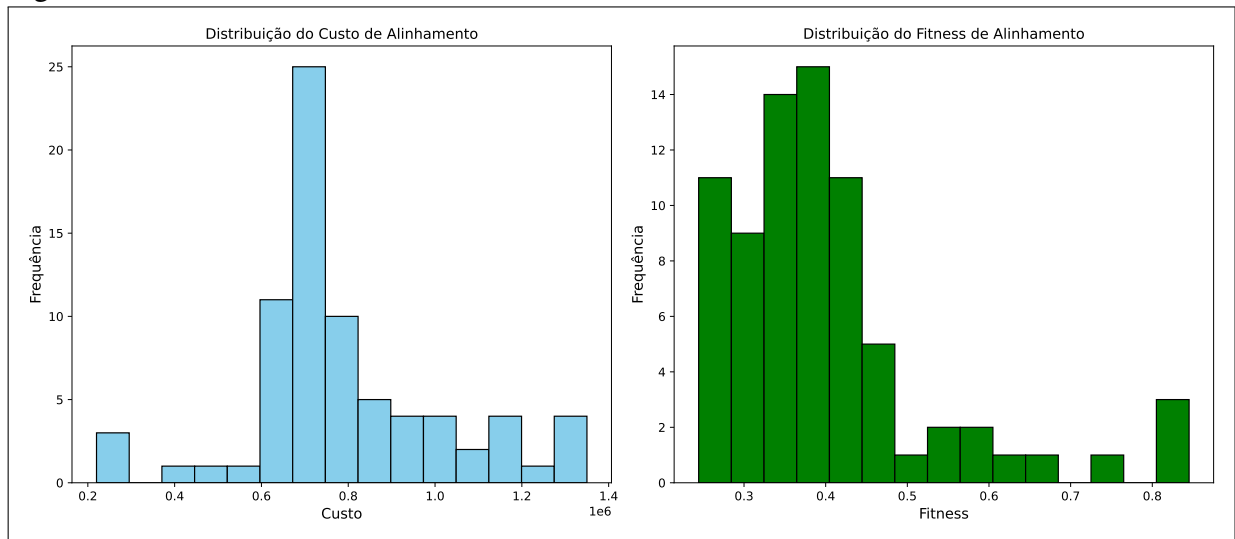
5.2.2 Alinhamento dos discentes não formados

A Figura 61 apresenta os histogramas da distribuição dos custos de alinhamento (à esquerda) e dos valores de *trace fitness* (à direita) para os discentes que não concluíram o curso. Esses estudantes abandonaram a formação em estágios distintos da trajetória acadêmica e, por isso, apresentam diferentes graus de aderência ao modelo curricular.

A análise do histograma de custos revela uma concentração majoritária entre 600.000 e 900.000 unidades, indicando que a maioria dos estudantes não formados apresentou desvios significativos em relação ao modelo de referência. Esses desvios representam eventos como reprovações consecutivas, trancamentos recorrentes, desistências parciais e tentativas frustradas de retomada do curso. Casos extremos, com custos superiores a 1.300.000, refletem trajetórias altamente desalinhadas e fortemente comprometidas desde os semestres iniciais.

O histograma de *fitness* evidencia que os valores mais frequentes entre os discentes evadidos estão entre 0,3 e 0,6, o que denota uma baixa conformidade com a estrutura curricular.

Figura 61 – Custo e *Fitness* de alinhamento dos discentes não formados



Fonte: elaborada pelo autor.

Apenas uma minoria se aproxima de valores superiores a 0,7, sugerindo que mesmo estudantes com trajetórias aparentemente promissoras eventualmente interromperam o curso. Essa interrupção pode decorrer de fatores não diretamente capturados pelo alinhamento, como aspectos pessoais, financeiros ou institucionais.

A relação inversa entre custo e *fitness* — já observada entre os discentes formados — também se confirma neste grupo. Discentes com altos custos de alinhamento geralmente apresentam baixos valores de *fitness*, o que reforça a validade dessa métrica como indicador do grau de aderência ao percurso acadêmico previsto.

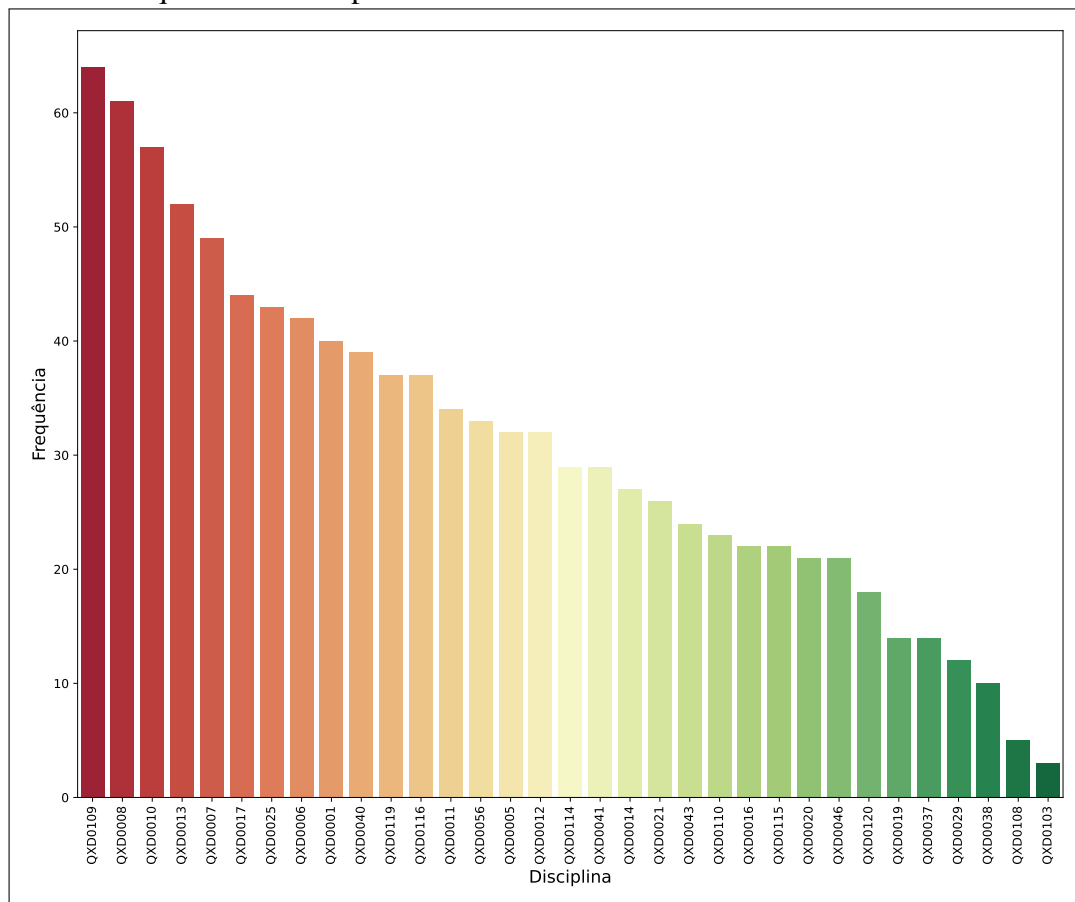
Esses resultados sugerem que trajetórias altamente irregulares, caracterizadas por repetições, reprovações múltiplas e evasões precoces, geram desalinhamentos severos com o modelo curricular. A análise por meio de alinhamentos, portanto, fornece subsídios importantes para identificar perfis de risco e antecipar potenciais evasões com base no histórico de execução do currículo.

5.2.2.1 *Frequência de componentes curriculares cursados fora do semestre ideal*

A Figura 62 apresenta a frequência com que os componentes curriculares foram cursados em semestres posteriores ao ideal previsto na matriz curricular. Esses atrasos refletem tentativas subsequentes de aprovação, motivadas por reprovações por nota ou falta, trancamentos (parciais ou totais), ou ainda por fatores estruturais que dificultaram o cumprimento da trajetória prevista.

A análise dos atrasos permite identificar os componentes que mais contribuíram para

Figura 62 – Frequência de componentes curriculares cursados fora do semestre ideal



Fonte: elaborada pelo autor.

o desalinhamento da trajetória acadêmica, destacando gargalos estruturais do curso. Entre as disciplinas com maior frequência de atraso estão *QXD0008*, *QXD0010*, *QXD0056*, *QXD0001*, *QXD0013* e *QXD0007*, todas pertencentes aos dois primeiros semestres da matriz curricular.

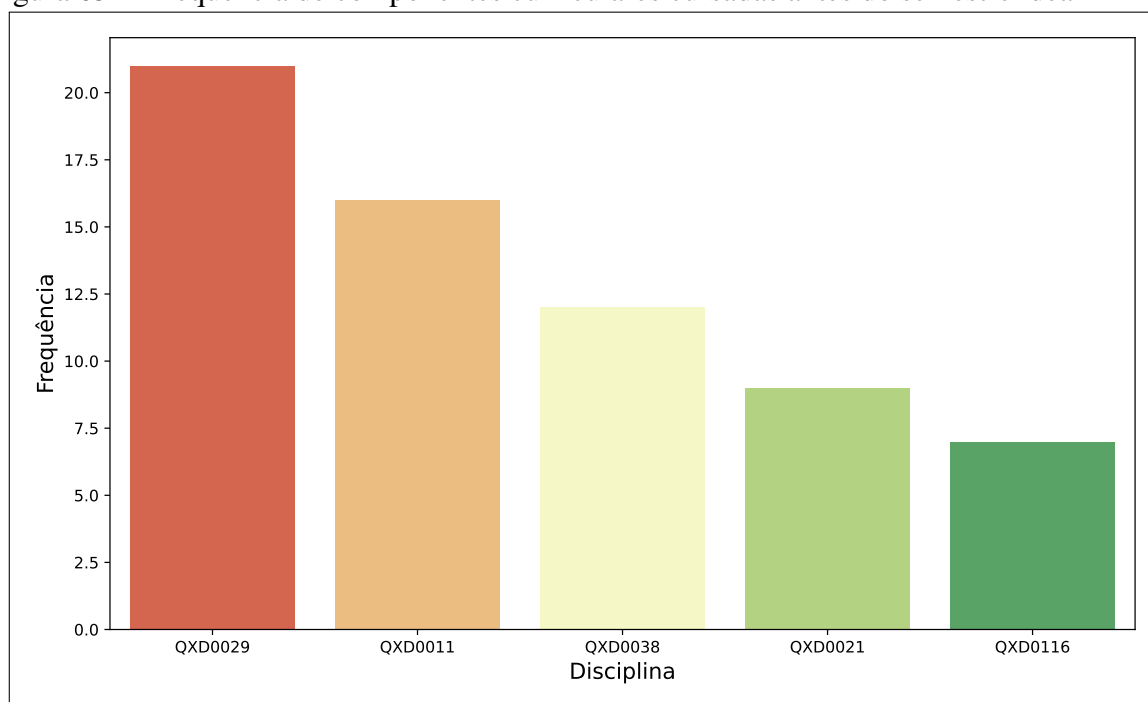
Esse padrão reforça os achados anteriores desta pesquisa — oriundos da análise com *token replay* — que já apontavam os 1º e 2º semestres como períodos críticos para a permanência acadêmica. A recorrência de atrasos nesses semestres iniciais sugere que dificuldades enfrentadas logo após o ingresso no curso têm papel decisivo na formação de trajetórias desviantes.

Além disso, a elevada frequência de atrasos nessas disciplinas pode indicar um duplo efeito: de um lado, sua complexidade técnica ou falta de pré-requisitos consolidados; de outro, um possível descompasso entre a formação prévia dos discentes e as exigências do currículo. Isso evidencia a necessidade de ações pedagógicas específicas para fortalecer o suporte aos alunos nos primeiros semestres, com foco na mitigação dos efeitos acumulativos de reprovações e na promoção de uma progressão mais fluida.

5.2.2.2 *Frequência de componentes curriculares cursados antes do semestre ideal*

A Figura 63 apresenta os componentes curriculares que foram cursados de forma adiantada, ou seja, antes do semestre previsto na matriz curricular. Esse comportamento é comum entre discentes que enfrentam dificuldades no fluxo acadêmico e, como estratégia, antecipam disciplinas que não possuem pré-requisitos formais ou apresentam menor complexidade percebida.

Figura 63 – Frequência de componentes curriculares cursadas antes do semestre ideal



Fonte: elaborada pelo autor.

Essas antecipações geralmente envolvem disciplinas alocadas nos semestres intermediários ou finais do curso, como *QXD0029*, *QXD0011* e *QXD0038*. Tais componentes, por não possuírem dependências explícitas, tornam-se mais acessíveis para matrícula, mesmo quando o discente ainda não cumpriu partes significativas da matriz ideal.

Contudo, embora essa prática possa parecer vantajosa para manter a carga horária ativa ou tentar compensar reprovações anteriores, ela pode acarretar riscos pedagógicos e acadêmicos. Ao cursar disciplinas fora da sequência lógica de aprendizagem, o estudante pode não dispor do conhecimento prévio necessário para um bom desempenho, o que eleva o risco de novas reprovações e contribui para o ciclo de retenção.

Além disso, a antecipação contribui para a desorganização do fluxo acadêmico, com consequências que afetam não apenas o rendimento do estudante, mas também o planejamento

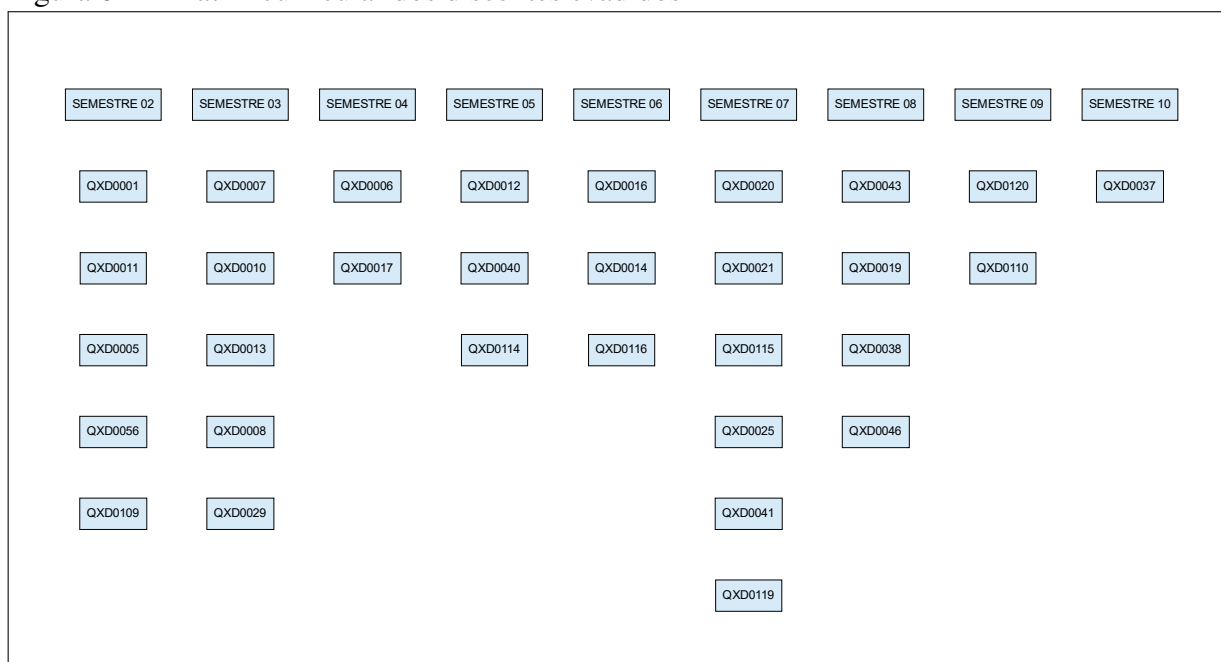
institucional, como distribuição de turmas e dimensionamento de oferta.

Portanto, a frequência elevada de adiantamentos em certos componentes revela a importância de refletir sobre a flexibilidade curricular e suas implicações, assim como a necessidade de políticas institucionais que acompanhem e orientem os discentes em decisões de matrícula, de modo a evitar o agravamento de trajetórias desviantes.

5.2.2.3 *Matriz curricular empírica dos discentes evadidos*

Com base na análise da execução dos *traces* dos discentes evadidos, foi possível construir uma matriz curricular empírica que representa o comportamento real dos estudantes ao longo de sua trajetória no curso. A Figura 64 apresenta essa matriz, na qual cada componente curricular é posicionado no semestre em que houve maior concentração de matrículas por parte dos discentes que não concluíram o curso.

Figura 64 – Matriz curricular dos discentes evadidos



Fonte: elaborada pelo autor.

No 1º semestre, não são observadas variações significativas, já que a matrícula nos componentes curriculares desse período é obrigatória para todos os ingressantes. Assim, os desvios começam a se evidenciar a partir do 2º semestre, onde se observa uma sobreposição de componentes originalmente previstos para o 1º semestre, resultado de reprovações ou trancamentos. Essa repetição precoce evidencia o impacto imediato de insucessos iniciais na organização do percurso formativo.

Do 2º ao 5º semestre, observa-se a persistência de tentativas de aprovação em disciplinas dos semestres anteriores, concomitante a tentativas de antecipação em componentes sem pré-requisitos formais, como *QXD0011* e *QXD0029*. Essa prática fragmenta a trajetória dos discentes e revela esforços isolados de manutenção da carga horária mínima, sem que haja uma progressão lógica no desenvolvimento das competências esperadas.

Nos semestres finais (6º ao 8º), constata-se um acúmulo de componentes curriculares originalmente alocados no meio do curso. Esse deslocamento é consequência direta das dificuldades acumuladas nos períodos iniciais, que impedem o avanço conforme o planejado. A disciplina *QXD0110*, por exemplo, aparece no penúltimo semestre da matriz empírica, mesmo sendo prevista para semestres anteriores. Esse comportamento decorre de reprovações sucessivas nos seus pré-requisitos, como *QXD0056* e *QXD0008*, dificultando sua matrícula no tempo ideal.

A matriz empírica dos evadidos, portanto, revela uma trajetória desalinhada e disfuncional, caracterizada por atrasos, repetições e antecipações mal planejadas. Esse padrão reforça a importância de mecanismos institucionais de monitoramento e apoio, sobretudo nos semestres iniciais, onde os desvios começam a se intensificar. A identificação dessas trajetórias alternativas permite à gestão acadêmica propor estratégias de intervenção para mitigar a evasão, promovendo a permanência e o sucesso dos discentes ao longo do curso.

5.3 Considerações finais

A partir das análises conduzidas neste capítulo, foi possível identificar com precisão diversos padrões de comportamento acadêmico que influenciam diretamente a trajetória dos discentes no curso de Ciência da Computação. As técnicas de *token replay* e alinhamentos permitiram não apenas mensurar a conformidade das trajetórias com o modelo ideal, mas também localizar, com base empírica, os pontos de maior fragilidade na progressão acadêmica.

A técnica de *token replay* revelou que os semestres iniciais, especialmente o 1º e 2º, concentram os maiores gargalos estruturais. Reprovações em componentes curriculares fundamentais comprometem não apenas a progressão imediata, mas desencadeiam um efeito cascata que pode culminar na evasão. A análise dos *tokens* remanescentes e ausentes permitiu mapear esses bloqueios com clareza, reforçando a necessidade de políticas institucionais voltadas à permanência desde os primeiros semestres.

Complementarmente, a técnica de alinhamentos demonstrou que muitos discentes seguem trajetórias desalinhadas em relação ao currículo ideal, cursando disciplinas fora da

ordem, acumulando reprovações e antecipando componentes curriculares sem pré-requisitos formais. A matriz empírica construída com base nos discentes evadidos evidencia esse padrão fragmentado e desestruturado de progressão, resultado direto de dificuldades enfrentadas nos períodos iniciais do curso.

Diante desses achados, algumas direções práticas podem ser consideradas:

- **Apoio acadêmico focalizado:** a oferta de monitorias, reforço e ações de nivelamento em disciplinas críticas dos primeiros semestres pode reduzir as taxas de reprovação e melhorar a fluidez no fluxo acadêmico.
- **Revisão curricular:** os dados sugerem a necessidade de revisão dos pré-requisitos, da distribuição de carga horária e da complexidade das disciplinas dos semestres iniciais, buscando maior equilíbrio e aderência às condições reais dos ingressantes.
- **Monitoramento preditivo:** os padrões identificados podem alimentar sistemas de alerta precoce, permitindo à instituição identificar estudantes em risco de evasão com base em seu histórico de reprovações, atrasos e desvios da matriz ideal.
- **Flexibilização controlada:** a antecipação de componentes curriculares sem a devida base pode ser um sinal de desorganização acadêmica. Modelos de matrícula poderiam incorporar restrições ou recomendações baseadas em pré-requisitos lógicos, e não apenas formais.

Em síntese, as análises realizadas neste capítulo oferecem um diagnóstico das trajetórias acadêmicas dos discentes, evidenciando padrões de sucesso, permanência e evasão. Esses resultados fornecem subsídios concretos para a formulação de políticas institucionais mais eficazes, orientadas por dados, com foco na melhoria da permanência e na ampliação das taxas de conclusão.

6 CONCLUSÕES E TRABALHOS FUTUROS

A evasão estudantil configura-se como um dos maiores desafios enfrentados pelas instituições de ensino superior, exigindo abordagens analíticas que permitam compreender e mitigar seus efeitos. Neste trabalho, foi proposta uma modelagem da estrutura curricular do curso de Ciência da Computação da Universidade Federal do Ceará – *Campus* de Quixadá, utilizando redes de Petri para representar formalmente as trajetórias acadêmicas. Com base nos dados extraídos dos históricos acadêmicos dos discentes, foi construído um *log* de eventos com as atividades realizadas ao longo dos períodos letivos.

A aplicação de técnicas de mineração de processos permitiu simular diferentes trajetórias, como a ideal (sem reprovações), a real (com reprovações, trancamentos e desvios) e a de discentes evadidos. A análise por meio de *token replay* e alinhamento (*alignments*) evidenciou os pontos críticos do processo acadêmico, destacando disciplinas que funcionam como gargalos recorrentes e as variações no fluxo esperado. Além disso, a comparação entre a trajetória ideal e as trajetórias reais revelou importantes desvios que podem estar associados à desistência dos alunos, fornecendo subsídios para ações institucionais mais direcionadas.

Os resultados obtidos demonstram a viabilidade do uso de mineração de processos no contexto educacional, sobretudo na visualização do comportamento dos estudantes ao longo do curso e na identificação de fatores que contribuem para a evasão. A proposta apresentada se mostra promissora como ferramenta de apoio à gestão acadêmica, oferecendo uma base sólida para o desenvolvimento de estratégias preventivas e políticas públicas voltadas à permanência estudantil.

Com base nos resultados obtidos e nas limitações encontradas, apresentam-se as seguintes sugestões para trabalhos futuros:

- Expandir a análise para outros cursos de graduação da Universidade Federal do Ceará (UFC), permitindo identificar padrões comuns ou específicos de evasão entre diferentes áreas de conhecimento.
- Aplicar técnicas de aprendizado de máquina aos *logs* de eventos para desenvolver modelos preditivos capazes de antecipar comportamentos de risco, como prováveis reprovações, trancamentos ou evasão.
- Incluir disciplinas optativas e eletivas na modelagem para oferecer uma visão mais completa da progressão curricular, principalmente em cursos com maior flexibilização na escolha de componentes.

- Realizar estudos longitudinais que avaliem os impactos de intervenções institucionais a partir das informações extraídas da análise de conformidade.
- Implementar uma interface visual interativa que permita aos gestores acadêmicos explorar dinamicamente os dados de conformidade e fluxo estudantil.
- Explorar técnicas de *Object-Centric Process Mining* (OCPM), que permitem modelar múltiplos objetos inter-relacionados (por exemplo, discentes, disciplinas, períodos letivos), oferecendo uma visão mais rica e flexível da trajetória acadêmica em ambientes complexos.

Este trabalho apresenta contribuições significativas para a compreensão da trajetória acadêmica dos discentes de graduação com foco na evasão estudantil, por meio da aplicação de técnicas de Mineração de Processos em dados educacionais. As principais contribuições incluem:

- **Modelagem formal da estrutura curricular:** foi realizada a representação da matriz curricular do curso de Ciência da Computação da UFC *Campus* de Quixadá por meio de redes de Petri, permitindo capturar o fluxo ideal de progressão dos discentes.
- **Construção de logs de eventos educacionais:** o trabalho estruturou dados históricos de eventos acadêmicos em um formato compatível com os padrões da Mineração de Processos, considerando o percurso real dos estudantes em suas interações com os componentes curriculares.
- **Aplicação de técnicas de verificação de conformidade:** foram empregadas abordagens como *token-replay*, *alignments* para avaliar a aderência das trajetórias dos alunos em relação ao modelo acadêmico ideal, permitindo a identificação de gargalos, desvios e padrões associados à evasão.
- **Análise de pontos críticos:** o estudo revelou disciplinas e períodos específicos que concentram taxas elevadas de reprovação ou abandono, destacando potenciais alvos para intervenções pedagógicas e institucionais.
- **Proposta de uso da Mineração de Processos no contexto educacional:** o trabalho reforça a viabilidade e a importância do uso de Process Mining como ferramenta estratégica de apoio à gestão acadêmica, evidenciando o potencial para subsidiar políticas de permanência estudantil.

Essas contribuições demonstram como a integração entre técnicas de modelagem de processos e dados educacionais pode oferecer novos caminhos para o enfrentamento da evasão no ensino superior.

REFERÊNCIAS

- AALST, W.; ADRIANSYAH, A.; MEDEIROS, A.; ARCIERI, F.; BAIER, T.; BLICKLE, T.; R.P., J. C. B.; BRAND, P.; BRANDTJEN, R.; BUIJS, J.; BURATTIN, A.; CARMONA, J.; CASTELLANOS, M.; CLAES, J.; COOK, J.; COSTANTINI, N.; CURBERA, F.; DAMIANI, E.; LEONI, M. de; WYNN, M. Process mining manifesto. In: . [S. l.: s. n.], 2011. v. 99, p. 169–194. ISBN 978-3-642-28107-5.
- AALST, W. M. van der. Everything you always wanted to know about petri nets, but were afraid to ask. In: SPRINGER. **Business Process Management: 17th International Conference, BPM 2019, Vienna, Austria, September 1–6, 2019, Proceedings 17**. [S. l.], 2019.
- AALST, W. V. D. Process mining: Overview and opportunities. **ACM Transactions on Management Information Systems (TMIS)**, ACM New York, NY, USA, v. 3, n. 2, p. 1–17, 2012.
- AALST, W. V. D. Data science in action. In: **Process mining: Data science in action**. [S. l.]: Springer, 2016.
- ANDIFES, A.; ABRUEM, A. *et al.* Diplomação, retenção e evasão nos cursos de graduação em instituições de ensino superior públicas: resumo do relatório apresentado a andifes, abruem e sesu/mec pela comissão especial. **Avaliação: Revista da Avaliação da Educação Superior**, v. 1, n. 2, 1996.
- ANDRADE, T.; ALMEIDA, C.; BARBOSA, J.; RIGO, S. Avaliação de um modelo de sistema de recomendação que integra metodologias ativas e mineração de dados educacionais para mitigar a evasão na educação a distância. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 34., 2023. **Anais [...]**. Porto Alegre, RS, Brasil: SBC, 2023. p. 187–198. Disponível em: <https://sol.sbc.org.br/index.php/sbie/article/view/26661>. Acesso em: 01 ago. 2024.
- CARDOSO, J.; VALETTE, R. **Redes de petri**. [S. l.]: Editora da UFSC Florianópolis, SC., 1997.
- CARMO Êrica; GASPARINI, I.; OLIVEIRA, E. Identificação de trajetórias de aprendizagem em um curso de graduação e sua relação com a evasão escolar. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 33., 2022. **Anais [...]**. Porto Alegre, RS, Brasil: SBC, 2022. p. 323–333. Disponível em: <https://sol.sbc.org.br/index.php/sbie/article/view/22419>. Acesso em: 01 ago. 2024.
- CARMONA, J.; DONGEN, B. van; SOLTI, A.; WEIDLICH, M. Conformance checking. **Switzerland: Springer.[Google Scholar]**, Springer, v. 56, p. 12, 2018.
- CARVALHO, C.; MATTOS, J.; AGUIAR, M. Avaliação da interpretabilidade de modelos por meio da clusterização de explicações no contexto da predição de evasão no ensino superior. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 34., 2023. **Anais [...]**. Porto Alegre, RS, Brasil: SBC, 2023. p. 1191–1201. Disponível em: <https://sol.sbc.org.br/index.php/sbie/article/view/26746>. Acesso em: 01 ago. 2024.
- CARVALHO, L.; SANTOS, A.; NAKAMURA, F.; OLIVEIRA, E. Detecção precoce de evasão em cursos de graduação presencial em computação: um estudo preliminar. In: WORKSHOP SOBRE EDUCAÇÃO EM COMPUTAÇÃO, 27., 2019. **Anais [...]**. Porto Alegre, RS, Brasil:

SBC, 2019. p. 233–243. Disponível em: <https://sol.sbc.org.br/index.php/wei/article/view/6632>. Acesso em: 01 ago. 2024.

CBOK, B. Guide to the business process management common body of knowledge. **Versão**, v. 2, p. 2009, 2009.

COLPO, M.; PRIMO, T.; AGUIAR, M. Predição da evasão estudantil: uma análise comparativa de diferentes representações de treino na aprendizagem de modelos genéricos. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 32., 2021. **Anais [...]**. Porto Alegre, RS, Brasil: SBC, 2021. p. 873–884. Disponível em: <https://sol.sbc.org.br/index.php/sbie/article/view/18114>. Acesso em: 01 ago. 2024.

FALCÃO, A.; VILLWOCK, R.; MILOCA, S. Análise de dados pré-universidade para prever a evasão de alunos ingressantes em uma instituição de ensino superior. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 34., 2023. **Anais [...]**. Porto Alegre, RS, Brasil: SBC, 2023. p. 1293–1304. Disponível em: <https://sol.sbc.org.br/index.php/sbie/article/view/26755>. Acesso em: 01 ago. 2024.

HARTE, J. M.; WESTENBERG, M. R.; SOMEREN, M. van. Process models of decision making. **Acta Psychologica**, Elsevier, v. 87, n. 2-3, p. 95–120, 1994.

INEP, D. d. E. E. **Metodologia de Cálculo dos indicadores de fluxo da educação superior**. [S. l.]: INEP, Brasília, 2017.

INEP, D. d. E. E. **Indicadores de fluxo da educação superior**. INEP, Brasília, 2023. Disponível em: <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/indicadores-educacionais>. Acesso em: 01 ago. 2024.

JESUS, H. O. de; RODRIGUEZ, L. C.; JUNIOR, A. d. O. C. Predição de evasão escolar na licenciatura em computação. **Revista Brasileira de Informática na Educação**, v. 29, p. 255–272, 2021.

JESUS, P. A. S. d.; RODRIGUES, M. d. S. T.; TORRES, J. B. **Proposta de mapeamento de processos de armazenagem em centro de distribuição de supermercados de Fortaleza/CE utilizando BPMN**. 2014. Disponível em: <http://www.repositorio.ufc.br/handle/riufc/61995>. Acesso em: 01 ago. 2024.

KANTORSKI, G.; MARTINS, R.; BALEJO, A.; FRICK, M. Mineração de dados educacionais para predição da evasão em cursos de graduação presenciais no ensino superior. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 34., 2023. **Anais [...]**. Porto Alegre, RS, Brasil: SBC, 2023. p. 1133–1142. Disponível em: <https://sol.sbc.org.br/index.php/sbie/article/view/26741>. Acesso em: 01 ago. 2024.

MACIEL, P. R.; LINS, R. D.; CUNHA, P. R. **Introdução às redes de Petri e aplicações**. [S. l.]: UNICAMP-Instituto de Computacao Sao Paulo, Brazil, 1996.

MEDEIROS, R. S. **Modelagem de processos em projetos de parques eólicos, uma aplicação do BPMN (Business Process Modeling Notation)**. 2011. Disponível em: <http://repositorio.ufc.br/handle/riufc/35694>. Acesso em: 01 ago. 2024.

MORAIS, F.; MELO, A.; MOUTINHO, M.; FAGUNDES, R. Modelos de regressão aplicados na previsão da evasão escolar do ensino básico: uma revisão sistemática da

literatura. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 32., 2021. **Anais [...]**. Porto Alegre, RS, Brasil: SBC, 2021. p. 168–178. Disponível em: <https://sol.sbc.org.br/index.php/sbie/article/view/18053>. Acesso em: 01 ago. 2024.

MUNOZ-GAMA, J. *et al.* **Conformance checking and diagnosis in process mining**. Tese (Doutorado) – Springer, 2014.

MURATA, T. Petri nets: Properties, analysis and applications. **Proceedings of the IEEE**, IEEE, v. 77, n. 4, p. 541–580, 1989.

OLIVEIRA, R. d. S. **Modelo de predição de evasão escolar com base em dados de autoavaliação de cursos de graduação**. 2023. Disponível em: <https://repositorio.ifpb.edu.br/handle/177683/3206>. Acesso em: 01 ago. 2024.

OLIVEIRA, R. dos S.; MEDEIROS, F. P. A. de. Modelo de predição de evasão escolar com base em dados de autoavaliação de cursos de graduação. **Revista Brasileira de Informática na Educação**, v. 32, p. 1–21, 2024.

PETERSON, J. L. **Petri net theory and the modeling of systems**. [S. l.]: Prentice Hall PTR, 1981.

ROMERO, C.; VENTURA, S. Educational data mining: a review of the state of the art. **IEEE Transactions on Systems, Man, and Cybernetics, Part C (applications and reviews)**, Ieee, v. 40, n. 6, p. 601–618, 2010.

SANTOS, C. H.; MARTINS, S.; PLASTINO, A. É possível prever evasão com base apenas no desempenho acadêmico? In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 32., 2021. **Anais [...]**. Porto Alegre, RS, Brasil: SBC, 2021. p. 792–802. Disponível em: <https://sol.sbc.org.br/index.php/sbie/article/view/18107>. Acesso em: 01 ago. 2024.

SARAIVA, D.; PEREIRA, S.; BRAGA, R.; OLIVEIRA, C. Análise de agrupamentos para caracterização de indicadores de evasão. In: WORKSHOP SOBRE EDUCAÇÃO EM COMPUTAÇÃO, 29., 2021. **Anais [...]**. Porto Alegre, RS, Brasil: SBC, 2021. p. 238–247. Disponível em: <https://sol.sbc.org.br/index.php/wei/article/view/15915>. Acesso em: 01 ago. 2024.

SILVA, E.; SILVA, J.; ALBUQUERQUE, C. de. Uma análise da evasão escolar nos cursos de tecnologia da informação: Um estudo de caso em floresta/pe. In: WORKSHOP SOBRE EDUCAÇÃO EM COMPUTAÇÃO, 24., 2016. **Anais [...]**. Porto Alegre, RS, Brasil: SBC, 2016. p. 2413–2421. Disponível em: <https://sol.sbc.org.br/index.php/wei/article/view/9685>. Acesso em: 01 ago. 2024.

SILVA, R.; F., B. A.; FERREIRA, M. de F.; SANTOS, I.; ANDRADE, R. Evasão em computação na ufc sob a perspectiva dos alunos. In: WORKSHOP SOBRE EDUCAÇÃO EM COMPUTAÇÃO, 29., 2021. **Anais [...]**. Porto Alegre, RS, Brasil: SBC, 2021. p. 338–347. Disponível em: <https://sol.sbc.org.br/index.php/wei/article/view/15925>. Acesso em: 01 ago. 2024.

SOUSA, M.; ALBUQUERQUE, D.; LEAL, R.; MAIA, M.; GOMES, A.; PERKUSICH, M. Análise dos fatores de evasão dos alunos dos cursos técnicos da Área de informática no ifpb campus campina grande. In: WORKSHOP SOBRE EDUCAÇÃO EM COMPUTAÇÃO, 30., 2022. **Anais [...]**. Porto Alegre, RS, Brasil: SBC, 2022. p. 369–380. Disponível em: <https://sol.sbc.org.br/index.php/wei/article/view/20845>. Acesso em: 01 ago. 2024.

SOUZA, O.; MORAIS, P.; JÚNIOR, F. S. Um estudo sobre a evasão no curso de licenciatura em informática do ifrn – campus natal – zona norte. In: WORKSHOP SOBRE EDUCAÇÃO EM COMPUTAÇÃO, 23., 2015. **Anais [...]**. Porto Alegre, RS, Brasil: SBC, 2015. p. 216–225. Disponível em: <https://sol.sbc.org.br/index.php/wei/article/view/10238>. Acesso em: 01 ago. 2024.

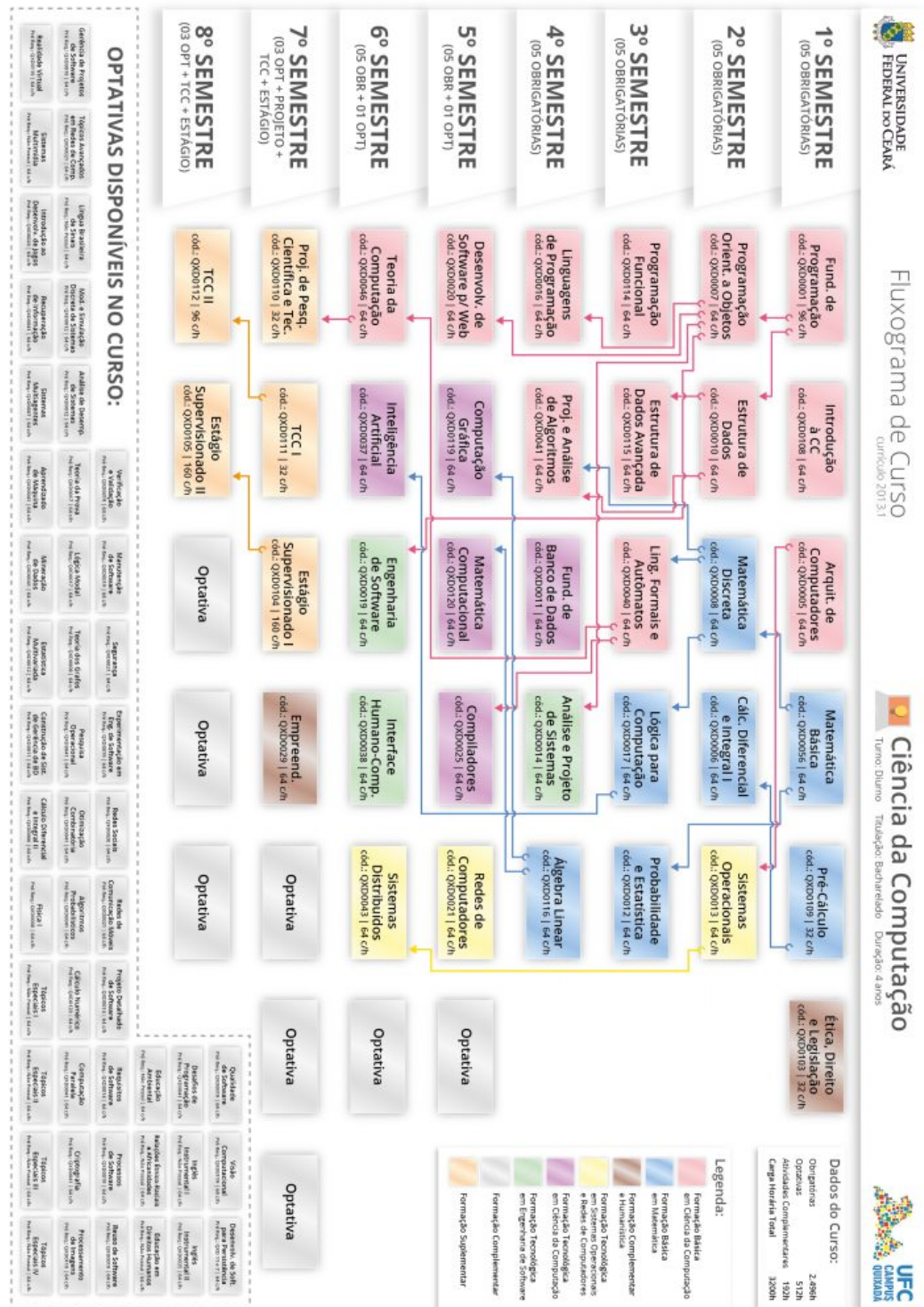
UNGER, A.; JUNIOR, D. L.; LIMA, F.; GERALDO, I.; VENERO, S.; AMBROSIO, R. Análise da trajetória de aprendizagem do aluno no ensino a distância por meio da mineração de processos. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 33., 2022. **Anais [...]**. Porto Alegre, RS, Brasil: SBC, 2022. p. 1163–1172. Disponível em: <https://sol.sbc.org.br/index.php/sbie/article/view/22490>. Acesso em: 01 ago. 2024.

VASCONCELOS, V.; ANDRADE, E. Análise da evasão de alunos na licenciatura em computação. In: WORKSHOP SOBRE EDUCAÇÃO EM COMPUTAÇÃO, 26., 2018. **Anais [...]**. Porto Alegre, RS, Brasil: SBC, 2018. Disponível em: <https://sol.sbc.org.br/index.php/wei/article/view/3482>. Acesso em: 01 ago. 2024.

VIANA, F.; SANTANA, A.; RABÊLO, R. Avaliação de classificadores para predição de evasão no ensino superior utilizando janela semestral. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 33., 2022. **Anais [...]**. Porto Alegre, RS, Brasil: SBC, 2022. p. 908–919. Disponível em: <https://sol.sbc.org.br/index.php/sbie/article/view/22469>. Acesso em: 01 ago. 2024.

7 ESTRUTURA CURRICULAR DO CURSO DE CIÊNCIA DA COMPUTAÇÃO

Figura 65 – Estrutura curricular do curso de Ciência da Computação da UFC do Campus de Quixadá.



Fonte: Disponível em <https://cc.quixada.ufc.br/estrutura-curricular/fluxograma-de-disciplinas/>