

Uma Abordagem Baseada em Grafos e Aprendizado de Máquina para Recomendação Curricular em Engenharia

Sarah A. da C. Medeiros¹, Abílio N. Barros¹, Alexandre Magno A. Maciel¹

¹ Universidade de Pernambuco (UPE) Recife – PE – Brazil

sacm@poli.br, abilionbarros@gmail.com, alexandre.maciel@upe.br

Abstract. *Engineering programs present complex prerequisite structures, retention phenomena and heterogeneous enrollment patterns, making academic planning a challenging task. This work investigates the integration of curriculum graphs, student flow graphs and Long Short-Term Memory (LSTM) networks to support personalized academic recommendation. Using real academic records from a Brazilian public university, the proposed approach combines structural, behavioral and temporal information to model student progression and generate personalized curriculum recommendations based on real learning trajectories. The results demonstrate that the proposed hybrid model outperforms traditional recommendation approaches, highlighting its effectiveness in supporting academic progression.*

Resumo. *Cursos de Engenharia apresentam estruturas complexas de pré-requisitos, fenômenos de retenção e padrões heterogêneos de matrícula, tornando o planejamento acadêmico uma tarefa desafiadora. Este trabalho investiga a integração entre grafos curriculares, grafos de fluxo discente e redes Long Short-Term Memory (LSTM) para apoiar recomendação acadêmica personalizada. Utilizando históricos acadêmicos reais de uma universidade pública brasileira, a abordagem proposta combina informações estruturais, comportamentais e temporais para modelar a progressão estudantil e gerar recomendações curriculares personalizadas baseadas em trajetórias reais de aprendizagem. Os resultados demonstram que o modelo híbrido proposto supera abordagens tradicionais de recomendação, evidenciando sua eficácia para apoiar a progressão acadêmica.*

1. Introdução

No contexto do ensino superior brasileiro, assegurar a conclusão dos cursos e que ela ocorra dentro do tempo previsto de integralização constitui um desafio persistente para as instituições de ensino superior. Entre os principais fatores associados a esse cenário destacam-se os fenômenos de evasão e retenção estudantil, sendo a evasão reconhecida como um problema multifatorial que afeta diretamente a permanência e o sucesso acadêmico dos estudantes [Lorenzo-Quiles et al. 2023]. De forma complementar, pesquisas recentes apontam que a retenção estudantil produz impactos relevantes sobre o tempo de formação e os indicadores institucionais de desempenho [Gonçalves et al. 2024].

De acordo com os indicadores de fluxo divulgados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) no âmbito do Censo da Educação

Superior 2024, a taxa de evasão anual no ensino superior brasileiro alcançou aproximadamente 9,5% nos cursos presenciais e 24,1% na modalidade a distância. Além disso, indicadores de trajetória apontam que apenas cerca de 39% dos estudantes que ingressaram na graduação concluíram seus cursos dentro do período analisado, evidenciando desafios persistentes relacionados à permanência e à progressão acadêmica [Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira 2026]. Estudos complementares indicam que, nesses cursos, a evasão acumulada ao longo dos cinco primeiros anos pode superar 50%, evidenciando um cenário crítico de permanência estudantil e progressão acadêmica [Ministério da Educação 2020].

Nesse cenário, o planejamento acadêmico configura-se como uma tarefa complexa tanto para estudantes quanto para gestores educacionais. Escolhas inadequadas podem comprometer a progressão curricular, ampliar o tempo de integralização e aumentar o risco de retenção [McCoy et al. 2023]. Apesar dos avanços na gestão acadêmica, muitas dessas decisões ainda são fundamentadas em orientações generalizadas e na experiência individual de docentes e coordenadores. Além disso, observa-se uma limitada utilização de análises sistemáticas das trajetórias acadêmicas individuais. A literatura de *Learning Analytics* (LA) destaca o potencial da análise de dados educacionais para apoiar decisões acadêmicas e institucionais [Lang et al. 2017]. Já no contexto da análise curricular, estruturas de dependência entre disciplinas têm sido utilizadas para compreender gargalos e atrasos na formação [Heileman et al. 2017].

Apesar dos avanços observados na literatura, as abordagens existentes tendem a enfatizar aspectos distintos da progressão acadêmica. Enquanto representações curriculares baseadas em grafos modelam explicitamente relações de pré-requisitos e dependências entre disciplinas [Heileman et al. 2017], abordagens fundamentadas em dados históricos de estudantes permitem identificar padrões de progressão observados empiricamente nas trajetórias acadêmicas [Salehi and Haghighi 2013]. Nesse contexto, este trabalho investiga a integração entre grafos curriculares, grafos de fluxo discente e redes Long Short-Term Memory (LSTM) para apoiar a recomendação curricular personalizada em cursos de Engenharia a partir de registros reais de histórico acadêmico.

A escolha por modelos sequenciais decorre do fato de que a trajetória acadêmica possui natureza temporal e cumulativa. As disciplinas cursadas, aprovadas ou reprovadas em semestres anteriores influenciam diretamente as possibilidades de progressão futura do estudante. Nesse ponto, redes LSTM têm sido amplamente empregadas em problemas educacionais por sua capacidade de capturar dependências de longo prazo em sequências [Piech et al. 2015a]. Assim sendo, esse trabalho busca responder a seguinte questão: "Como o desalinhamento entre o currículo formal e as trajetórias acadêmicas reais pode ser modelado e explorado para apoiar a recomendação curricular personalizada por semestre em cursos de Engenharia?".

2. Trabalhos Relacionados

A recomendação acadêmica tem sido amplamente investigada no contexto de *Learning Analytics* (LA) e Mineração de Dados Educacionais (EDM), especialmente em aplicações voltadas à retenção estudantil, previsão de desempenho e apoio ao planejamento curricular [Romero and Ventura 2010, Siemens 2013, Viberg et al. 2018]. As abordagens tradicionais de recomendação frequentemente utilizam técnicas de filtragem colaborativa,

métodos baseados em conteúdo ou modelos híbridos [Burke 2002, Ricci et al. 2015]. Embora apresentem resultados satisfatórios em diferentes domínios, essas abordagens possuem limitações quando aplicadas ao contexto acadêmico, principalmente por não incorporarem explicitamente restrições curriculares, relações de pré-requisitos e características estruturais dos cursos.

Com o avanço das pesquisas em modelagem educacional, abordagens baseadas em grafos passaram a ser utilizadas para representar estruturas curriculares e relações entre disciplinas [Montagud et al. 2020, Wu et al. 2021, Hamilton 2020]. Essas representações permitem modelar formalmente dependências curriculares, identificar disciplinas críticas e analisar propriedades associadas à retenção e à progressão estudantil. Além disso, estudos em Redes Neurais em Grafos (*Graph Neural Networks*) e aprendizado relacional demonstram que estruturas baseadas em grafos são capazes de capturar padrões complexos de interação em sistemas de recomendação [Ying et al. 2018], reforçando o potencial da modelagem relacional para análise de trajetórias acadêmicas e fluxos curriculares.

Paralelamente, modelos sequenciais baseados em aprendizado profundo têm sido empregados para modelar a evolução temporal do processo de aprendizagem. Trabalhos como o *Deep Knowledge Tracing* (DKT) [Piech et al. 2015b] e suas extensões baseadas em mecanismos de atenção [Pandey and Karypis 2019] demonstram que modelos recorrentes conseguem capturar dependências temporais complexas relacionadas ao comportamento dos estudantes. De forma semelhante, pesquisas em recomendação sequencial evidenciam que a ordem temporal das interações influencia diretamente o comportamento futuro dos usuários, tornando as dependências sequenciais um elemento fundamental para sistemas de recomendação mais precisos [Quadrana et al. 2018, Tang and Wang 2018].

No contexto educacional, trabalhos como os de Elbadrawy e Karypis [Elbadrawy and Karypis 2016] e Morsy e Karypis [Morsy and Karypis 2019] exploram técnicas de aprendizado profundo para predição de desempenho e recomendação de disciplinas, incorporando informações contextuais dos estudantes.

Além da modelagem preditiva, estudos de análise curricular indicam que estruturas curriculares excessivamente encadeadas podem aumentar a retenção e ampliar o tempo de integralização dos cursos [Heileman et al. 2017, Slim et al. 2014].

Esses trabalhos evidenciam a relevância da análise estrutural do currículo como instrumento de apoio à gestão acadêmica e ao planejamento institucional. Apesar dos avanços observados, permanece uma lacuna na literatura relacionada à ausência de modelos capazes de integrar simultaneamente as estruturas curriculares formais, os padrões reais de fluxo acadêmico e as dependências temporais das trajetórias estudantis.

Este trabalho busca preencher essa lacuna ao desenvolver um modelo híbrido que integre esses três módulos e validando a abordagem proposta por meio de métricas e procedimentos experimentais adequados, avaliando sua capacidade de apoiar a recomendação curricular personalizada nos cursos de Engenharia.

3. Método

O desenvolvimento deste estudo seguiu um *pipeline* estruturado composto por quatro etapas principais: (i) aquisição e pré-processamento dos dados acadêmicos, (ii) definição formal do problema de recomendação curricular, (iii) modelagem estrutural e comporta-

mental por meio de grafos, e (iv) modelagem sequencial das trajetórias estudantis utilizando redes neurais recorrentes. Por fim, essas representações foram integradas em um sistema híbrido de recomendação.

3.1. Aquisição e Processamento dos Dados

Os experimentos foram conduzidos utilizando dados acadêmicos do perfil curricular CP011 do curso de Engenharia da Computação de uma instituição de ensino superior pública. A escolha deste perfil curricular foi motivada pela disponibilidade de histórico acadêmico consolidado, volume suficiente de estudantes para treinamento dos modelos e estrutura curricular fortemente dependente de pré-requisitos, características adequadas para a modelagem baseada em grafos e aprendizado sequencial. A Tabela de caracterização do conjunto de dados resume as principais características da base.

Tabela 1. Caracterização do conjunto de dados

Característica	Valor
Curso	Engenharia da Computação (CP011)
Estudantes	728
Registros acadêmicos	18.924
Disciplinas obrigatórias	51
Ingressos registrados	De 2002.2 a 2020.2
Aprovações e reprovações	64,23% e 31,09% dos registros
Concluintes dentro do tempo previsto	124 estudantes

O conjunto de dados utilizado contém a matriz curricular formal, as relações de pré-requisito entre disciplinas e os históricos acadêmicos completos incluindo registros de aprovação e reprovação dos estudantes. Mas as disciplinas optativas, atividades complementares, equivalências curriculares ou outros cenários não foram considerados na construção dos grafos. Além disso, eventos de trancamento foram mantidos apenas como informação temporal da trajetória do estudante, não sendo tratados como disciplinas elegíveis para recomendação.

O *pipeline* de pré-processamento foi projetado para garantir consistência sem comprometer a dinâmica real das trajetórias acadêmicas. As etapas incluíram:

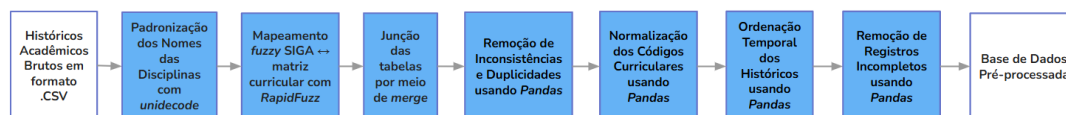


Figura 1. Visualização do *pipeline* de pré-processamento

Ao contrário de outras abordagens, disciplinas repetidas foram mantidas, permitindo preservar padrões reais de retenção e progressão acadêmica. Essa decisão metodológica foi importante, pois reprovações sucessivas representam sinais relevantes de desalinhamento curricular e frequentemente relacionadas a retenção prolongada.

Após as etapas de pré-processamento e ordenação temporal dos históricos, os dados foram transformados em sequências acadêmicas estruturadas. Nessa representação,

cada estudante é modelado como uma sequência cronológica de semestres letivos, preservando a ordem em que as disciplinas foram cursadas ao longo da graduação. Assim, em vez de analisar disciplinas de forma isolada, o modelo passa a considerar a trajetória acadêmica completa do estudante, permitindo capturar padrões de progressão, retenção e reorganização curricular ao longo do tempo. Essa estrutura sequencial constitui a entrada utilizada pela rede LSTM para aprender dependências temporais presentes nos históricos.

3.2. Cuidados Éticos

Os dados utilizados foram anonimizados antes do processamento, com a remoção de matrículas, nomes e quaisquer identificadores diretos dos estudantes. Dessa forma, nenhuma informação pessoal identificável foi utilizada durante as etapas de processamento, modelagem ou avaliação dos modelos propostos. Os dados foram disponibilizados pela coordenação da instituição de ensino superior pública mediante assinatura de um *Non-Disclosure Agreement* (NDA), sendo o acesso restrito exclusivamente aos pesquisadores envolvidos no projeto.

A utilização dos dados ocorreu exclusivamente para fins acadêmicos e científicos, em conformidade com as normas institucionais vigentes e com os princípios estabelecidos pela Lei Geral de Proteção de Dados Pessoais (LGPD – Lei nº 13.709/2018). Por fim, esse estudo não realiza distinções entre estudantes com base em características individuais nem produz decisões automatizadas que possam resultar em benefícios ou prejuízos acadêmicos. Os resultados produzidos podem vir a ser interpretados como mecanismos auxiliares de apoio à decisão pedagógica.

3.3. Definição Formal do Problema

A formulação do problema segue a perspectiva de sistemas de recomendação educacional, nos quais o objetivo consiste em utilizar informações históricas e contextuais para apoiar decisões acadêmicas futuras [Ricci et al. 2015, Nguyen et al. 2020].

Considere um conjunto de estudantes:

$$\mathcal{S} = \{s_1, s_2, \dots, s_n\}$$

e um conjunto de disciplinas:

$$\mathcal{D} = \{d_1, d_2, \dots, d_m\}$$

Cada estudante s_i possui uma trajetória acadêmica representada como uma sequência temporal:

$$T_i = (S_1, S_2, \dots, S_t)$$

onde S_t corresponde ao conjunto de disciplinas cursadas no semestre t .

O problema de recomendação consiste em estimar o conjunto de disciplinas mais adequadas para matrícula no semestre seguinte, considerando simultaneamente as restrições curriculares formais definidas pela matriz, os padrões históricos de progressão acadêmica dos estudantes e as dependências temporais nas trajetórias.

Formalmente, deseja-se aprender uma função:

$$f(T_i, G_c, G_f) \rightarrow R_k$$

onde:

1. T_i representa a trajetória do estudante;
2. G_c representa o grafo curricular formal;
3. G_f representa o grafo de fluxo acadêmico real;
4. R_k representa o conjunto das k disciplinas recomendadas.

3.4. Modelagem por Grafos

Para capturar simultaneamente regras curriculares e padrões reais de progressão, foram construídos dois grafos complementares:

1. **Grafo Curricular Formal (G_c):** modela relações oficiais de pré-requisito entre disciplinas;
2. **Grafo de Fluxo Acadêmico (G_f):** representa transições efetivamente observadas nas trajetórias dos estudantes.

A utilização de grafos para modelagem curricular está alinhada a abordagens recentes que exploram estruturas educacionais como redes complexas [Shen et al. 2018, Montagud et al. 2020, Wu et al. 2021]. Enquanto o grafo curricular garante a consistência estrutural das recomendações, o grafo de fluxo permite capturar padrões emergentes não explicitamente definidos na matriz curricular.

A Figura 2 apresenta a representação do grafo curricular construído a partir da matriz de pré-requisitos do curso. Nesse grafo, os vértices correspondem às disciplinas e as arestas representam relações formais de dependência entre os componentes curriculares. A orientação das arestas indica o fluxo de precedência acadêmica, evidenciando, período a período, quais disciplinas são exigidas para a cursagem de outras. Essa representação permite visualizar de forma clara o encadeamento acadêmico previsto pelo currículo oficial e identificar a estrutura de dependências que organiza a progressão dos estudantes ao longo do curso.

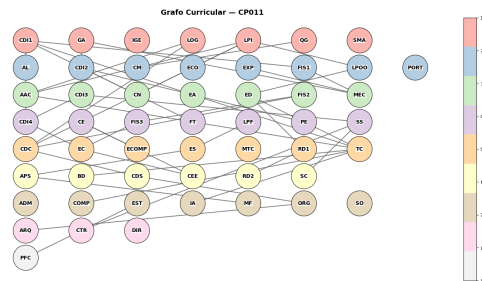


Figura 2. Visualização do grafo curricular ideal

Além da estrutura curricular formal, foi construído um grafo de fluxo acadêmico para representar as trajetórias reais dos estudantes ao longo do curso. Nesse modelo, os vértices correspondem às disciplinas cursadas e as arestas representam a ordem de progressão observada nos históricos acadêmicos. Como a visualização completa do grafo

agregado é excessivamente complexa, a Figura 3 apresenta um recorte simplificado referente a um único estudante, ilustrando como o modelo incorpora informações temporais, aprovações e reprovações não representadas no grafo curricular formal. Na figura, nós em verde claro indicam aprovação por média, nós em verde escuro representam aprovação após avaliação final e nós em vermelho correspondem a disciplinas previamente reprovadas.

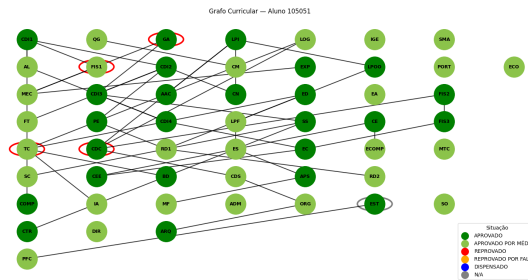


Figura 3. Visualização do grafo de fluxo real de um aluno único

O peso das transições no grafo de fluxo foi definido como:

$$w(u, v) = \sum_s \exp[-\lambda(t' - t)]$$

onde $\lambda = 0.5$ controla a penalização temporal, atribuindo maior relevância a transições mais recentes.

Essa escolha busca equilibrar o fator de preservação dos padrões históricos recorrentes, a priorizar dinâmicas curriculares mais recentes e representativas do contexto institucional atual. Permitindo a capturar não apenas a frequência de transições entre disciplinas, mas também sua relevância temporal no contexto das trajetórias acadêmicas.

3.5. Modelagem Sequencial

Para modelar dependências temporais nas trajetórias estudantis, foi utilizada uma rede Long Short-Term Memory (LSTM), considerando que modelos LSTM são amplamente utilizados para modelagem de dependências temporais em sequências, sendo particularmente eficazes em contextos educacionais para capturar padrões de aprendizagem ao longo do tempo [Hochreiter and Schmidhuber 1997, Piech et al. 2015b].

A modelagem sequencial de interações tem sido amplamente explorada em sistemas de recomendação e em aplicações educacionais. Trabalhos clássicos de recomendação sequencial demonstram que a ordem temporal dos eventos influencia significativamente o comportamento futuro dos usuários [Quadrana et al. 2018, Tang and Wang 2018]. No contexto educacional, abordagens como *Deep Knowledge Tracing* evidenciam o potencial de redes recorrentes para modelar trajetórias de aprendizagem ao longo do tempo [Pandey and Karypis 2019]. O uso de LSTM foi selecionado pelo motivo de trajetórias acadêmicas apresentarem forte dependência histórica, na qual reprovações, atrasos e decisões curriculares anteriores influenciam significativamente escolhas futuras, preservando assim dependências de longo prazo e capturando efeitos acumulativos de retenção acadêmica.

Cada histórico acadêmico foi representado como uma sequência:

$$\mathbf{X}_{student} = (S_1, S_2, \dots, S_T)$$

As disciplinas foram codificadas como índices inteiros e mapeadas para *embeddings* densos treináveis, permitindo a aprendizagem de representações vetoriais dos componentes curriculares. A arquitetura do modelo é composta por uma camada de *Embedding* com 64 dimensões, seguida por uma camada LSTM com 128 unidades para modelagem das dependências temporais presentes nas trajetórias acadêmicas. Em seguida, foi aplicada uma camada de *Dropout* com taxa de 0,2 para fins de regularização, seguida por uma camada densa com função de ativação ReLU e uma camada de saída Softmax responsável pela predição da próxima disciplina.

Optou-se pela utilização de redes LSTM devido à sua capacidade de modelar dependências temporais de longo prazo e ao seu amplo emprego na literatura de modelagem sequencial educacional. A escolha também foi motivada pela adequação dessa arquitetura às características da base de dados utilizada, composta por sequências acadêmicas de tamanho relativamente reduzido.

O treinamento foi realizado utilizando divisão temporal de 70% para treinamento, 15% para validação e 15% para teste, preservando a ordem cronológica. Foram utilizados o otimizador Adam, a função de perda *Sparse Categorical Crossentropy* e *early stopping* para evitar *overfitting*.

3.6. Sistema Híbrido de Recomendação

O sistema proposto integra três componentes complementares. O componente estrutural, representado pelo grafo curricular, garante a viabilidade acadêmica das recomendações ao considerar as relações formais de pré-requisitos. O componente comportamental, representado pelo grafo de fluxo acadêmico, captura padrões reais de progressão observados nos históricos estudantis. Por sua vez, o componente sequencial, implementado por meio de uma rede LSTM, modela as dependências temporais das trajetórias acadêmicas, incorporando informações relacionadas à ordem de cursagem das disciplinas.

A pontuação final atribuída a cada disciplina candidata é definida por:

$$\text{Score}(d) = \alpha P_{LSTM}(d) + \beta P_{fluxo}(d) + \gamma P_{curr}(d)$$

com:

$$(\alpha, \beta, \gamma) = (0.4, 0.35, 0.25)$$

onde:

1. $P_{LSTM}(d)$ representa a probabilidade predita pelo modelo sequencial;
2. $P_{fluxo}(d)$ representa a probabilidade baseada no grafo de fluxo;
3. $P_{curr}(d)$ representa a pontuação de elegibilidade curricular.

Os pesos utilizados na combinação dos componentes do modelo híbrido foram ajustados por meio de uma busca em grade (*grid search*) utilizando validação temporal. Os dados foram particionados em 70% para treinamento, 15% para validação e 15% para

teste. Diferentes combinações de pesos foram avaliadas e a configuração final foi selecionada com base no melhor desempenho obtido no conjunto de validação.

Adicionalmente, os pesos refletem uma interpretação conceitual do problema:

1. o maior peso atribuído ao componente LSTM reflete a importância predominante das dependências temporais;
2. o componente de fluxo recebe peso intermediário por representar comportamento coletivo real dos estudantes;
3. o componente curricular recebe menor peso relativo por atuar principalmente como mecanismo de consistência estrutural.

Após o cálculo da pontuação híbrida, é aplicada uma etapa de filtragem baseada em restrições curriculares formais, garantindo que apenas disciplinas elegíveis sejam recomendadas. A integração é complementar: o componente curricular atua como filtro de elegibilidade, o grafo de fluxo modela padrões históricos coletivos e a LSTM captura características temporais individuais dos estudantes.

4. Resultados e Discussões

De posse dos resultados, podemos responder sobre a pergunta de pesquisa norteadora desse trabalho **“Como o desalinhamento entre o currículo formal e as trajetórias acadêmicas reais pode ser modelado e explorado para apoiar a recomendação curricular personalizada por semestre em cursos de Engenharia?”**.

4.1. Comparação entre Modelos

Os resultados apresentados na Tabela 2 evidenciam diferenças substanciais entre as abordagens avaliadas. Observa-se que o método baseado exclusivamente em regras estruturais (Currículo Formal) apresenta desempenho limitado, com Top-1 de 47.44%. Esse comportamento sugere que abordagens puramente estruturais não são suficientes para representar a complexidade das trajetórias acadêmicas reais, especialmente em cenários caracterizados por complexidade curricular e trajetórias heterogêneas [Heileman et al. 2017].

Modelos baseados em padrões de transição, como Cadeias de Markov [Norris 1998] e o modelo de Fluxo Real construído a partir das trajetórias observadas nos históricos acadêmicos, apresentam melhorias consistentes, sugerindo que a modelagem de dependências locais entre disciplinas contribui significativamente para o desempenho. Esse comportamento é consistente com pesquisas recentes de recomendação sequencial, nas quais modelos recorrentes apresentam vantagens sobre abordagens estáticas por capturarem dependências temporais entre interações sucessivas [Hidasi et al. 2016, Quadrana et al. 2018, Tang and Wang 2018].

O modelo LSTM alcança Top-1 de 61.231%, reforçando evidências recentes de que dependências temporais de longo prazo possuem papel relevante na modelagem de trajetórias educacionais [Piech et al. 2015b, Pandey and Karypis 2019, Tang and Wang 2018]. Além disso, o ganho observado é consistente com trabalhos recentes de recomendação sequencial educacional, que indicam que dependências sequenciais possuem papel central na modelagem da progressão acadêmica [Hidasi et al. 2016, Pandey and Karypis 2019, Quadrana et al. 2018].

Tabela 2. Comparação entre modelos baseline

Modelo	Top-1	Top-5	MRR	NDCG
Markov	59.474%	89.350%	70.264%	76.313%
Fluxo Real	59.889%	87.414%	69.810%	75.587%
Curriculo Formal	47.441%	59.474%	53.432%	58.992%
Collaborative Filtering	23.510%	41.521%	24.217%	48.996%
SVD	50.069%	55.187%	52.673%	55.788%
LSTM	61.231%	84.467%	70.355%	74.811%
Modelo Híbrido	69.710%	90.041%	78.829%	83.457%

O modelo híbrido proposto supera todos os baselines, atingindo 69.710% de Top-1. Esse ganho não ocorre apenas pela combinação de modelos, mas pela capacidade de integrar diferentes dimensões do fenômeno de progressão acadêmica. Enquanto modelos tradicionais capturam padrões isolados, como frequência de disciplinas ou dependências locais, o modelo híbrido consegue representar simultaneamente restrições estruturais, comportamento coletivo e dependências temporais individuais.

4.2. Exemplo de Recomendação e Aplicabilidade Institucional

Tabela 3. Exemplo da recomendação Top-5 para um estudante

Posição	Disciplina recomendada	Score
1	Português Instrumental	0.957
2	Mecânica	0.713
3	Redes I	0.664
4	Engenharia de Software	0.646
5	Teoria da Computação	0.631

Como ilustração do funcionamento do método proposto, foi analisado um estudante cujo histórico acadêmico foi observado até o final do quarto semestre. A partir dessa trajetória parcial, o modelo híbrido gerou um *ranking* das disciplinas consideradas mais adequadas para o período subsequente, conforme apresentado na Tabela 3.

Observa-se que disciplinas como Português Instrumental e Mecânica receberam pontuações mais elevadas por apresentarem elevada compatibilidade com o estado acadêmico atual do estudante, atendendo simultaneamente aos critérios de elegibilidade curricular e aos padrões de progressão identificados nos dados históricos. De forma semelhante, Redes I, Engenharia de Software e Teoria da Computação aparecem entre as recomendações por serem componentes curriculares frequentemente associados às trajetórias observadas para estudantes com perfil acadêmico semelhante.

É importante destacar que embora os resultados indiquem potencial para aplicação em sistemas de apoio ao planejamento acadêmico, a utilização institucional da abordagem ainda deve ser considerada uma perspectiva futura, não tendo sido realizada validação com coordenadores de curso, docentes ou estudantes.

Embora o foco principal deste trabalho esteja relacionado à recomendação curricular, os resultados obtidos possuem implicações relevantes para gestão acadêmica e

planejamento institucional. Por exemplo, a identificação automática de gargalos curriculares pode auxiliar coordenadores de curso na revisão de estruturas de pré-requisito e na identificação de disciplinas associadas à retenção prolongada.

Adicionalmente, os grafos de fluxo permitem compreender trajetórias reais frequentemente adotadas pelos estudantes, fornecendo evidências empíricas que podem apoiar decisões relacionadas à revisão curricular, à flexibilização de encadeamentos de disciplinas, ao planejamento da oferta de componentes curriculares e ao desenvolvimento de políticas institucionais de permanência estudantil.

Esses resultados aproximam-se da literatura de análise curricular. Heileman et al. [Heileman et al. 2017] mostram que a análise estrutural dos currículos pode auxiliar na identificação de gargalos associados ao atraso na formação. Viberg et al. [Viberg et al. 2018] destacam o uso de dados educacionais para apoiar a tomada de decisão institucional, enquanto Matcha et al. [Matcha et al. 2020] destacam o potencial da análise de sequências e trajetórias de aprendizagem para compreender padrões comportamentais em ambientes educacionais. De forma semelhante, os resultados deste trabalho indicam que os grafos de fluxo podem apoiar a identificação de trajetórias críticas e subsidiar ações de gestão acadêmica.

O modelo também apresenta potencial para aplicações de acompanhamento acadêmico individualizado, permitindo identificar estudantes cujas trajetórias acadêmicas se aproximam de padrões historicamente associados à retenção. Em um cenário de aplicação institucional, as recomendações geradas poderiam auxiliar na identificação precoce de estudantes com progressão acadêmica atípica, elevado acúmulo de disciplinas pendentes ou sucessivas reprovações em componentes curriculares críticos, subsidiando ações de orientação acadêmica e acompanhamento pedagógico. Assim, os sistemas de recomendação curricular podem atuar não apenas como ferramentas preditivas, mas também como instrumentos de apoio à tomada de decisão pedagógica e institucional.

4.3. Discussão

Os grafos de fluxo revelaram que estudantes frequentemente adotam trajetórias distintas do percurso curricular ideal, especialmente após retenções em disciplinas fundamentais. Esse comportamento produz padrões recorrentes de reorganização curricular que não são representados explicitamente pela matriz formal de pré-requisitos. Já o grafo curricular representa restrições institucionais e relações oficiais de dependência, o grafo de fluxo acadêmico captura estratégias reais de progressão observadas empiricamente. A modelagem sequencial com LSTM complementa essa representação ao incorporar dependências temporais de longo prazo relacionadas ao histórico individual de cada estudante.

Os resultados também reforçam discussões recentes sobre modelagem relacional e sequencial em sistemas de recomendação educacional como visto nos trabalhos de [Ying et al. 2018, Hamilton 2020, Quadrana et al. 2018]. E os ganhos observados pelo modelo híbrido sugerem que abordagens de recomendação curricular tornam-se mais eficazes quando incorporam simultaneamente dimensões estruturais, temporais e comportamentais da progressão acadêmica.

4.4. Limitações

Apesar dos resultados promissores, algumas limitações devem ser consideradas. O estudo utiliza dados de uma única instituição e de um único curso, podendo incorporar vieses históricos específicos do contexto analisado. Mudanças curriculares ao longo do tempo, qualidade dos registros acadêmicos, ausência de fatores socioeconômicos e a possibilidade de reforçar trajetórias historicamente frequentes, mas não necessariamente ideais do ponto de vista pedagógico, também representam ameaças à validade dos resultados.

Além disso, o modelo ainda não foi avaliado junto a coordenadores, docentes ou estudantes em ambiente real de utilização. Os experimentos foram realizados com dados de um único perfil curricular de um curso de Engenharia, o que restringe a generalização imediata dos resultados para outros contextos acadêmicos. Em adição, fatores socioeconômicos, institucionais e comportamentais que podem influenciar a progressão estudantil não foram modelados explicitamente. Dessa forma, os resultados devem ser interpretados considerando as características específicas do contexto analisado.

5. Considerações Finais e Trabalhos Futuros

Este trabalho investigou como o desalinhamento entre currículo formal e trajetórias acadêmicas reais pode ser explorado para recomendação curricular personalizada em cursos de Engenharia. Os resultados obtidos evidenciaram que a integração entre grafos curriculares, grafos de fluxo acadêmico e modelagem sequencial com LSTM permite representar simultaneamente restrições estruturais, padrões coletivos de progressão e dependências temporais individuais.

Os experimentos demonstraram que o modelo híbrido proposto supera abordagens tradicionais de recomendação curricular, indicando que trajetórias acadêmicas não podem ser interpretadas apenas como sequências isoladas ou estruturas curriculares estáticas. Os grafos de fluxo revelaram padrões recorrentes de retenção, reorganização curricular e trajetórias alternativas frequentemente adotadas por estudantes após sucessivas reprovações em disciplinas críticas.

Além da capacidade preditiva, a abordagem proposta mostrou potencial como ferramenta analítica para identificação de gargalos curriculares, padrões recorrentes de retenção e trajetórias alternativas de progressão. Sob perspectiva institucional, tais informações podem apoiar processos de revisão curricular, planejamento de oferta de disciplinas, acompanhamento acadêmico e políticas de permanência estudantil. Do ponto de vista científico, os resultados sugerem que abordagens capazes de integrar múltiplas perspectivas da trajetória acadêmica podem representar uma alternativa promissora para a modelagem da progressão estudantil, especialmente em cenários caracterizados por forte dependência entre disciplinas e elevada heterogeneidade de trajetórias.

Como trabalhos futuros, pretende-se incorporar variáveis adicionais relacionadas a desempenho acadêmico, engajamento estudantil e fatores socioeconômicos, além de investigar técnicas de explicabilidade para tornar as recomendações mais interpretáveis. Também se pretende investigar sua aplicação em cursos de outras áreas do conhecimento, além das Engenharias, buscando analisar a capacidade de generalização do modelo diante de diferentes estruturas curriculares, padrões de progressão acadêmica e perfis estudantis. Dessa forma, espera-se ampliar a compreensão do impacto pedagógico e da aplicabilidade prática da abordagem proposta.

6. Uso de IA

Ferramentas de Inteligência Artificial Generativa foram utilizadas como apoio durante a elaboração deste trabalho. O ChatGPT (OpenAI) foi empregado para auxílio na revisão textual, aprimoramento da redação acadêmica, reorganização estrutural de seções e adequação do manuscrito às normas de submissão. O Gemini (Google) foi utilizado como ferramenta de apoio à revisão e verificação de trechos de código durante o desenvolvimento da implementação computacional. As ferramentas de IA não foram utilizadas para geração dos dados, condução dos experimentos, análise dos resultados ou definição das conclusões científicas apresentadas.

Referências

- Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*, 12(4):331–370.
- Elbadrawy, A. and Karypis, G. (2016). Domain-aware grade prediction and top-n course recommendation. In *RecSys*.
- Gonçalves, G. S., Serra, F. A. R., Storopoli, J. E., Scafuto, I. C., and Rafael, D. N. (2024). Undergraduate student retention activities: Challenges and research agenda. *SAGE Open*, 14(3).
- Hamilton, W. L. (2020). *Graph Representation Learning*. Morgan & Claypool.
- Heileman, G. L. et al. (2017). Curricular analytics: A framework for reducing time to degree. *Computers & Education*, 115:18–29.
- Hidasi, B. et al. (2016). Session-based recommendations with recurrent neural networks. In *ICLR*.
- Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8):1735–1780.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (2026). Resumo técnico do censo da educação superior 2024. Technical report, INEP, Brasília, DF.
- Lang, C., Siemens, G., Wise, A., and Gašević, D. (2017). Handbook of learning analytics. *Society for Learning Analytics Research*.
- Lorenzo-Quiles, O., Galdón-López, S., and Lendínez-Turón, A. (2023). Factors contributing to university dropout: A review. *Frontiers in Education*, 8:1159864.
- Matcha, W., Gašević, D., Jovanović, J., Uzir, N. A., Oliver, C. W., Murray, A., and Gasevic, D. (2020). Analytics of learning strategies: The association with the personality traits. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, pages 151–160. ACM.
- McCoy, L., Shapiro, C., and Heileman, G. (2023). Curriculum complexity and student success in higher education: A systematic review. *Educational Research Review*, 39:100521.
- Ministério da Educação (2020). Indicadores da educação superior. <https://www.gov.br/mec>.

- Montagud, M. et al. (2020). Graph-based analysis of curriculum structures. *Education Sciences*, 10(8):205.
- Morsy, S. and Karypis, G. (2019). Cnr: Course recommendation using deep learning. *IEEE Access*, 7:2019.
- Nguyen, A., Gardner, L., and Sheridan, D. (2020). A review of student recommendation systems in higher education. *Artificial Intelligence Review*, 54(2):1327–1364.
- Norris, J. R. (1998). *Markov Chains*. Cambridge University Press.
- Pandey, S. and Karypis, G. (2019). A self-attentive model for knowledge tracing. In *IEEE International Conference on Data Mining (ICDM)*.
- Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L., and Sohl-Dickstein, J. (2015a). Deep knowledge tracing. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 28.
- Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L., and Sohl-Dickstein, J. (2015b). Deep knowledge tracing. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 505–513.
- Quadrana, M. et al. (2018). Sequence-aware recommender systems. *ACM Computing Surveys*, 51(4):1–36.
- Ricci, F., Rokach, L., and Shapira, B. (2015). *Recommender Systems Handbook*. Springer.
- Romero, C. and Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics*, 40(6):601–618.
- Salehi, M. and Haghighi, I. N. K. (2013). A course recommender system based on graduated students’ experiences. In *Proceedings of the 5th Conference on Information and Knowledge Technology (IKT)*, pages 139–144. IEEE.
- Shen, J. et al. (2018). Inferring prerequisite structures from student data. *IEEE Transactions on Learning Technologies*, 11(2):195–208.
- Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, 57(10):1380–1400.
- Slim, A. et al. (2014). Curriculum bottlenecks and student progression. In *EDM*.
- Tang, J. and Wang, K. (2018). Personalized top-n sequential recommendation via convolutional sequence embedding. In *WSDM*.
- Viberg, O. et al. (2018). Current trends and issues in learning analytics. *Computers in Human Behavior*, 89:98–110.
- Wu, Q. et al. (2021). Graph neural networks in recommender systems: A survey. *ACM Computing Surveys*, 55(5):1–37.
- Ying, R. et al. (2018). Graph convolutional neural networks for web-scale recommender systems. In *KDD*.