

# Uso de Técnicas de Mineração de Dados na Extração de Sequências de Navegação de Estudantes de um Massive Open Online Course (MOOC)

## *Use of Data Mining Techniques in Extracting Navigation Sequences from Students in a Massive Open Online Course (MOOC)*

ISSN 2177-8310  
DOI: 10.18264/eadf.v10i2.1070

**Napoliana Silva de Souza<sup>1\*</sup>**  
**Gabriela Trindade Perry<sup>1</sup>**

<sup>1</sup> Universidade Federal do Rio Grande do Sul. Av. Paulo Gama, 110 – Porto Alegre, RS – Brasil.

\*[souzapoliana2@gmail.com](mailto:souzapoliana2@gmail.com).

### Resumo

Estudantes têm objetivos, preferências e ações de interação que se diferenciam e, em MOOCs, o comportamento navegacional pode ser útil para fazer descobertas sobre a aprendizagem. Nesse contexto, este artigo relata uma pesquisa realizada com os propósitos de detectar as sequências de navegação de 906 estudantes de um MOOC com conteúdos ligados à disciplina de Química e identificar os materiais educacionais mais acessados. Técnicas de mineração de dados educacionais foram empregadas, com aplicação do algoritmo Apriori. Os resultados mostraram 18 sequências frequentes e indicaram que os conteúdos do primeiro módulo do curso receberam mais acessos.

**Palavras-chave:** MOOC. Mineração de dados. Padrões sequenciais.



Recebido 11/ 06/ 2020  
Aceito 25/ 08/ 2020  
Publicado 27/ 08/ 20

### COMO CITAR ESTE ARTIGO

**ABNT:** SOUZA, N. S. de; PERRY, G. T. Uso de Técnicas de Mineração de Dados na Extração de Sequências de Navegação de Estudantes de um Massive Open Online Course (MOOC). **EaD em Foco**, v. 10, n. 2, e1070, 2020. DOI: <https://doi.org/10.18264/eadf.v10i2.1070>

## *Use of Data Mining Techniques in Extracting Navigation Sequences from Students in a Massive Open Online Course (MOOC)*

### *Abstract*

*Students have different goals, preferences and interaction actions, and in MOOCs, the navigational behavior can be useful for making discoveries related to learning. In this context, this paper reports a research carried out with the purpose of detecting the navigation sequences of 906 students enrolled in a MOOC with content related to the discipline of Chemistry, and identify the most accessed educational materials. We employ educational data mining techniques, using the Apriori algorithm. The results showed 18 frequent sequences, and indicated that the contents of the first module of the course received more access.*

**Keywords:** MOOC. Data mining. Sequential patterns.

## 1. Introdução

*Massive Open Online Courses (MOOCs)* são ambientes delineados para promover o aprendizado a distância, os quais defendem uma perspectiva social que tem como foco mediar a interação entre os estudantes e incentivar a aprendizagem colaborativa (GREGORI et al., 2018). A proposta pedagógica autoinstrucional incorporada aos MOOCs pretende que os estudantes sejam capazes de administrar sua própria experiência de aprendizagem (FREDERIKS et al., 2018).

A partir das características e comportamentos dos estudantes ao acessar sessões de aprendizagem, tem-se a oportunidade de extrair os seus caminhos de aprendizagem, identificar as suas preferências por determinados conteúdos, contribuindo para reconhecer o perfil de aprendizado do aluno e para recomendar caminhos de aprendizado de forma personalizada e automatizada (CASTRO et al., 2017). Os caminhos de aprendizagem dos estudantes de MOOCs auxiliam na identificação de seus estilos de aprendizagem, que são refletidos nas sequências de interação, comunicando as trilhas de aprendizagem aderidas, bem como os seus padrões comportamentais (BOROUJENI; DILLENBOURG, 2018).

Os MOOCs têm registros diversificados associados aos comportamentos de cliques, ações em fóruns de discussão e em vídeos, bem como dados temporais referentes às datas de acessos aos materiais e a postura do aluno diante dos diversos recursos de aprendizagem, o que inclui a ordem em que os conteúdos são visualizados; o acesso integral ou parcial aos conteúdos mostram as diversas sequências de aprendizagem que os estudantes adotam, sendo úteis para gerar esclarecimentos referentes à aprendizagem (CHEN et al., 2020).

Em termos metodológicos, as sequências de aprendizado de estudantes podem ser extraídas por meio de diferentes técnicas. Uma das possibilidades corresponde ao uso de Mineração de Padrões Sequenciais, sendo uma alternativa para aplicação no âmbito dos MOOCs (DEEVA et al., 2018). Mineração de Padrões Sequenciais é uma técnica de mineração de dados que utiliza algoritmos para analisar um banco de dados contendo as ações de interação de usuários, de modo a extrair um conjunto de sequências de ações frequentes, atuando na busca por padrões de interação (BOSC et al., 2017).

Analisar o comportamento de navegação ou de transição entre os conteúdos por estudantes não se limita à identificação das sequências seguidas - é também uma estratégia para estudar os motivos de os participantes não completarem todas as atividades ou acessarem somente alguns conteúdos, compreender seus progressos e frequência de participação no curso, bem como mapear percursos de aprendizagem que podem variar (RIZVI; RIENTIES; ROGATEN, 2018).

Conforme mencionado por Andrade e Silveira (2016), os MOOCs possuem números altos de inscritos e de interações, e a autonomia é um atributo almejado e importante para que os estudantes sejam bem-sucedidos na aprendizagem. Esses aspectos foram motivadores para a realização da pesquisa descrita nesse artigo, pois o MOOC analisado está hospedado em uma plataforma brasileira que recebe muitos inscritos e registra muitos comportamentos de interação e de aprendizagem dos estudantes, e não há professores ou tutores supervisionando os participantes. Nesse contexto, é fundamental fornecer atenção para os dados comportamentais disponíveis em ambientes on-line de aprendizagem.

Mediante as possibilidades fornecidas pelas técnicas de mineração de sequências, esta pesquisa foi conduzida com o objetivo de detectar os padrões sequenciais frequentes de navegação de 906 estudantes de um MOOC com conteúdos associados à disciplina de Química e identificar quais são os materiais de aprendizagem mais acessados.

A relevância desta pesquisa reside em mostrar uma alternativa automatizada fundamentada em técnicas de mineração de dados para rastrear o comportamento navegacional dos usuários do MOOC, verificar se os materiais de aprendizagem estão sendo acessados e descobrir padrões que contribuirão para implementar futuramente intervenções que permitam recomendar caminhos de aprendizado, bem como identificar e oferecer suporte aos estudantes que estão propensos a abandonar o curso.

Com a temática contextualizada, o artigo foi organizado com a seguinte estrutura: na Seção 2, são descritos alguns trabalhos relacionados, com ênfase na mineração de padrões sequenciais em MOOCs. A Seção 3 descreve a metodologia e caracteriza o MOOC submetido à análise. Na Seção 4, constam os resultados, e a Seção 5 é destinada à apresentação das considerações finais.

## 2. Mineração de Padrões Sequenciais em MOOCs

Técnicas de mineração de padrões sequenciais têm sido empregadas para extrair sequências de navegação de usuários e, em MOOCs, os objetivos das pesquisas que utilizaram esse tipo de técnica são variados. Nesta seção, podem ser conferidas algumas pesquisas correlatas, que tiveram em comum a análise do comportamento navegacional de usuários de MOOCs para encontrar suas sequências de interação, havendo algumas diferenças quanto aos propósitos e aos métodos empregados.

No estudo relatado por Deeva et al. (2018), foram utilizadas técnicas de mineração de padrões sequenciais e mineração de processos em um MOOC denominado “ePSYx Trends in e-Psychology”, disponibilizado na plataforma edX. Os comportamentos sequenciais de 1.394 estudantes foram analisados com o intuito de comparar o potencial das técnicas utilizadas para prever o abandono do MOOC. Para a classificação de sequências foram usados os algoritmos “SCIP” (Sequence Classification based on Interesting Patterns) e “cSPADE” (Sequential PAttern Discovery using Equivalence classes), e no procedimento de mineração de processos foi utilizada uma técnica de classificação denominada “SECPI-technique”. A partir da análise dos dados comportamentais dos estudantes inscritos no MOOC, os autores concluíram que o SECPI utilizado na mineração de processos não foi bom o suficiente para prever a desistência, enquanto os algoritmos cSPADE e SCIP utilizados na mineração de sequências tiveram melhores resultados.

Boroujeni e Dillenbourg (2018) propuseram e empregaram métodos na análise das sequências de interação de 7.527 estudantes de um MOOC intitulado “Functional Programming Principles in Scala”, com

o objetivo de encontrar padrões comportamentais de interação. Os autores esclarecem que os métodos utilizados baseiam-se em tarefas de agrupamentos de sequências, incluindo o uso de clusters que agrupam as trajetórias de ações dos estudantes no decorrer do curso. Para essa finalidade, foi utilizada uma abordagem não supervisionada e temporal, em que as sequências de interação são capturadas com o papel de modelar as diferentes ações de transição entre os conteúdos e encontrar padrões comportamentais. Os principais resultados listados pelos autores indicaram que os padrões mais frequentes consistiram em: (i) visualizar vídeos antes do envio de tarefas; (ii) pular os vídeos e submeter as tarefas; (iii) assistir aos vídeos e não enviar as tarefas. Contudo, os autores mencionam que visualizar vídeos antes do envio de tarefas foi o comportamento predominante.

Maldonado-Mahauad et al. (2018) analisaram as sequências de aprendizagem de 2.035 estudantes de um MOOC sobre Eletrônica disponível na plataforma Coursera. Inicialmente, os autores aplicaram um questionário de autorrelato para identificar o perfil de autorregulação da aprendizagem dos estudantes. Posteriormente, empregaram Mineração de Processos para extrair as sequências frequentes de interação dos estudantes. Os padrões sequenciais e de autorregulação foram usados para prever o sucesso na aprendizagem, que depende da nota alcançada no curso. Os resultados dos modelos preditivos permitiram a confirmação de que os padrões sequenciais juntamente com o comportamento de autorregulação de aprendizado dos estudantes fornecem valor significativo para a predição da nota.

Os exemplos listados nesta seção mostram algumas possibilidades da mineração de padrões sequenciais, tais como encontrar padrões de interação, prever desistência ou o desempenho, bem como pode-se testar diferentes algoritmos para comparar a precisão na detecção de sequências de aprendizado.

### 3. Metodologia

Nessa pesquisa foi empregada uma técnica de mineração de dados para extrair as sequências frequentes de navegação de estudantes inscritos em um MOOC, considerando os seus comportamentos de visualização de materiais de aprendizagem capturados por meio de registros de *logs*.

Para a realização da mineração de sequências de navegação, foram utilizadas as quatro etapas de mineração de dados sugeridas por Márquez-Vera et al. (2013): (i) coleta de dados; (ii) pré-processamento; (iii) mineração de dados; (iv) interpretação. Na etapa de coleta de dados a fonte de dados é consultada com o propósito de identificar os dados necessários e reunir as informações que serão úteis para a análise. A etapa de pré-processamento consiste na preparação dos dados, incluindo tarefas de limpeza e transformação dos atributos, quando necessário. Para a etapa de mineração de dados, quando os dados estiverem adequadamente preparados, ocorre a seleção de um ou mais algoritmos para alcançar um objetivo previamente formulado. A última etapa é direcionada à interpretação dos resultados dos modelos, regras ou padrões encontrados para gerar descobertas (MÁRQUEZ-VERA et al., 2013). Os procedimentos de pré-processamento e mineração de dados foram realizados utilizando a ferramenta RStudio. A análise dos comportamentos de navegação foi efetuada visando responder a duas questões:

- Quais os materiais mais acessados no MOOC?
- Quais são as sequências de navegação mais frequentes?

#### 3.1. Coleta de dados

Para a coleta de dados, foram exportados os *logs* referentes aos comportamentos de interação de estudantes inscritos em um MOOC disponível na plataforma Lúmina<sup>1</sup>. Os dados de interação registram

<sup>1</sup> Plataforma Lúmina: <https://lumina.ufrgs.br/>

informações que incluem os nomes completos dos estudantes, as datas e horários de cliques realizados pelos usuários em cada conteúdo e o *status* de visualização das unidades de aprendizagem que são marcadas automaticamente como “concluídas” quando recebem um clique e quando os materiais ou *links* não são acessados os *status* de visualização ficam vazios.

### 3.2. Pré-processamento

Na etapa de pré-processamento dos dados de *logs*, por questões éticas, os atributos identificadores dos alunos foram excluídos (nome e ID). Também foram removidas entradas duplicadas e, em seguida, a formatação dos dados foi transformada para se adequar ao formato aceito pelo algoritmo utilizado.

### 3.3. Mineração de dados

Para a etapa de mineração de dados foi utilizado o algoritmo APRIORI, que analisa um conjunto de transações e encontra sequências frequentes de itens (BODON, 2005). O algoritmo APRIORI é empregado na descoberta de regras de associação e para investigar o comportamento do usuário ao visitar páginas Web e detectar padrões sequenciais de navegação (KOTIYAL et al., 2013).

Para encontrar os itens frequentes, o algoritmo APRIORI analisa as ocorrências de itens. Um exemplo amplamente utilizado na literatura para compreender o funcionamento do APRIORI diz respeito ao comportamento de clientes em um supermercado. Dado um conjunto de produtos denominados “itens”, disponíveis em um estabelecimento comercial, é possível descobrir itens comprados com frequência, por exemplo: se, no conjunto de dados de registros de compras, é comum que os clientes que compram pão e manteiga também comprem leite, significa que a ocorrência contendo pão e manteiga representa o antecedente, enquanto leite é o elemento consequente (AGRAWAL; IMIELINSKI; SWAMI, 1993).

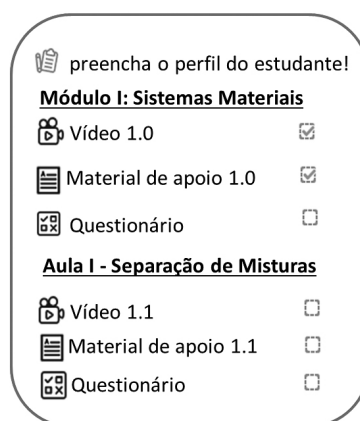
Em páginas Web, o algoritmo APRIORI se comporta de forma semelhante às aplicações comerciais; dada uma sequência de acessos, a partir das visitas dos usuários e cliques em menus de uma página, pode-se descobrir os padrões sequenciais frequentes de navegação (KOTIYAL et al., 2013). As regras determinadas pelo algoritmo APRIORI baseiam-se em cálculos de fatores de suporte e confiança que precisam ser satisfeitos (AGRAWAL; IMIELINSKI; SWAMI, 1993). “O suporte mede o número de transações que contém todos os itens da transação. Já a confiança indica a proporção de vezes que uma transação contendo o elemento A, também contém B. A força da regra é a soma do suporte mais confiança” (AMARAL, 2016, p. 120). Com base nesses esclarecimentos, o APRIORI foi empregado na análise de um *dataset* de um MOOC contendo o comportamento navegacional de 906 estudantes, utilizando o suporte mínimo igual a 0.35 e confiança de 0.6. A justificativa para a especificação desses valores levou em consideração a variabilidade comportamental dos estudantes; ao testar previamente outras configurações, verificou-se que valores de suporte abaixo do especificado geraram números excessivos de sequências, enquanto parâmetros altos resultaram na geração de, no máximo, quatro sequências frequentes. Desta forma, a configuração escolhida representou melhor a heterogeneidade comportamental dos participantes.

### 3.4. Interpretação

Nessa última etapa metodológica da análise, foram gerados gráficos com o uso do pacote “arulesViz”, e realizada a interpretação das sequências de interação mais frequentes encontradas pelo algoritmo, correspondente aos resultados de investigação.

### 3.5. Descrição do MOOC Analisado

O MOOC do qual os dados de interação foram extraídos versa sobre a disciplina de Química; trata-se de um curso de curta duração, vinculado a atividades de ensino, pesquisa e extensão. Os materiais de aprendizagem estão distribuídos em quatro módulos; cada módulo é composto por vídeos de curta duração (hospedados no YouTube), materiais de apoio que possuem um conjunto de arquivos de textos e vídeos educacionais complementares e um questionário de avaliação da aprendizagem. Há também um fórum para compartilhar dúvidas, disponibilizado após o último módulo. O curso é gratuito, não há pré-requisitos para inscrição, disponibilizado para o público em geral; oferece certificado de participação para aqueles que cumprirem os requisitos, não possui tutoria e também não define prazos para finalizar o curso e responder às avaliações, cabendo ao estudante conduzir a própria aprendizagem conforme o seu ritmo. A Figura 1 ilustra como os conteúdos do MOOC são organizados.



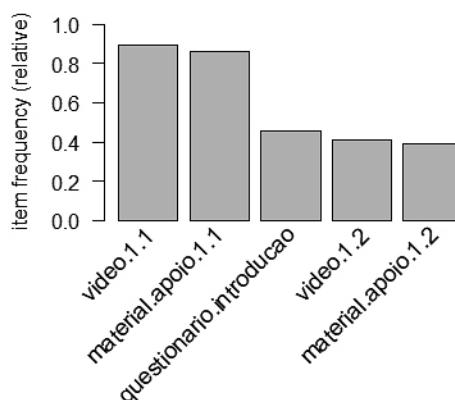
**Figura 1:** Estrutura das Unidades de Aprendizagem

Fonte: Elaborado pelos autores.

Todos os quatro módulos do MOOC possuem o mesmo padrão de organização de conteúdos exibido na Figura 1. Há o total de 20 *links* para acesso aos vídeos, 19 pastas de materiais de apoio e 19 questionários de múltipla escolha para avaliação da aprendizagem. Após a inscrição no curso, o usuário tem acesso a todos os materiais organizados sequencialmente, porém não é imposto ao estudante seguir a ordem pré-definida pelo MOOC.

## 4. Resultados

Os resultados desta pesquisa estão associados aos dois questionamentos formulados. Primeiramente, houve o interesse em descobrir quais os materiais mais acessados no MOOC. Posteriormente, o foco voltou-se para a descoberta das sequências de navegação mais frequentes. Na Figura 2, foi gerado um gráfico com os 5 materiais que receberam mais visualizações.


**Figura 2:** Materiais de aprendizagem mais acessados

**Fonte:** Autores (Gráfico gerado de forma automatizada pela ferramenta Rstudio).

A descoberta que pôde ser extraída da Figura 2 diz respeito à ocorrência de os cinco materiais mais visualizados pertencerem ao primeiro módulo do MOOC. Ao consultar os *logs* para esclarecimento, foi possível constatar que há uma quantidade significativa de estudantes que acessaram os conteúdos do Módulo I, havendo uma redução expressiva no número de visitas aos conteúdos do Módulo II, enquanto nos módulos III e IV, os status de visualização estavam quase todos vazios. Trata-se de uma descoberta relevante, pois os estudantes iniciaram o curso com alta atividade, porém não progrediram no curso. Essa ocorrência pode ser um indício de que, a partir do Módulo II, os estudantes estão sujeitos a reduzirem o ritmo de aprendizagem, ficarem inativos temporariamente ou interromper seu vínculo com o MOOC, de modo a abandonar o curso.

No que se refere às sequências encontradas, ao usar o suporte de 0.35 e confiança igual a 0.6, foram retornadas 18 sequências mais frequentes (Tabela 1) – enumeradas e registradas entre colchetes. As sequências foram ordenadas de acordo com o valor do *Lift* (do maior para o menor). *Lift* é um método que avalia a relevância das regras/sequências encontradas; esse procedimento é feito por meio de cálculos que determinam a existência de correlação entre elementos antecedentes e consequentes (TAIR; EL-HALLEES, 2012). Os cálculos de suporte, confiança e *lift*, bem como as sequências disponíveis na Tabela 1, foram retornados de forma automatizada por meio do uso do algoritmo Apriori.

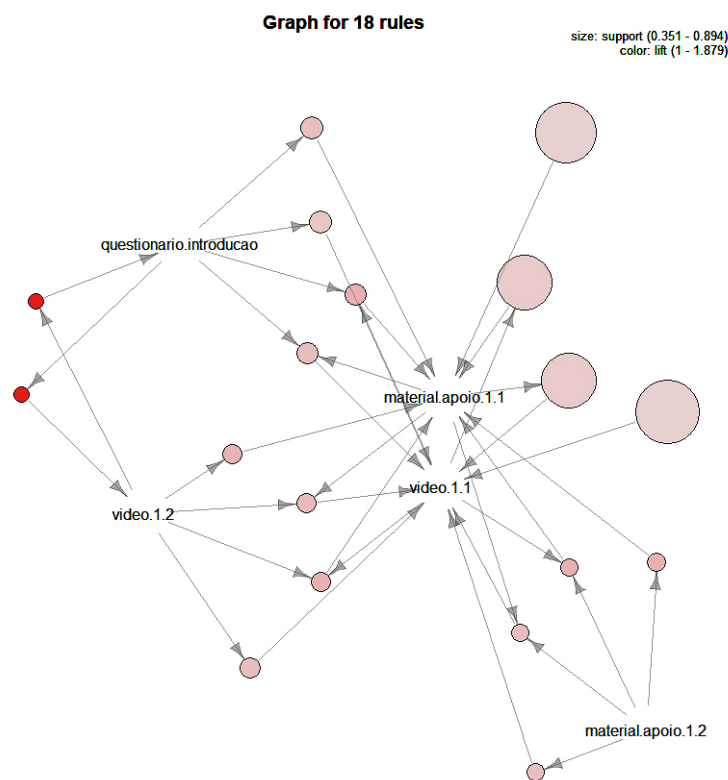
**Tabela 1:** Sequências frequentes de interação

Nº	Sequências	Suporte	Confiança	Lift
[1]	{video.1.2→{questionário.introdução}}	0.3511111	0.8540541	1.879337
[2]	{questionário.introdução→{vídeo.1.2}}	0.3511111	0.7726161	1.879337
[3]	{material.apoio.1.2,vídeo.1.1}→{material.apoio.1.1}	0.3766667	0.9912281	1.155577
[4]	{material.apoio.1.2}→{material.apoio.1.1}	0.3844444	0.9885714	1.152480
[5]	{questionário.introdução,vídeo.1.1}→{material.apoio.1.1}	0.4188889	0.9843342	1.147540
[6]	{vídeo.1.1,vídeo.1.2}→{material.apoio.1.1}	0.3977778	0.9835165	1.146587
[7]	{vídeo.1.2}→{material.apoio.1.1}	0.4011111	0.9756757	1.137446
[8]	{material.apoio.1.1, vídeo.1.2}→{vídeo.1.1}	0.3977778	0.9916898	1.108721
[9]	{vídeo.1.2}→{vídeo.1.1}	0.4044444	0.9837838	1.099882
[10]	{questionário.introdução}→{material.apoio.1.1}	0.4277778	0.9413203	1.097394
[11]	{material.apoio.1.1, material.apoio.1.2}→{vídeo.1.1}	0.3766667	0.9797688	1.095394

[12]	{material.apoio.1.1,questionário.introdução} →{vídeo.1.1}	0.4188889	0.9792208	1.094781
[13]	{material.apoio.1.2}→{vídeo.1.1}	0.3800000	0.9771429	1.092458
[14]	{questionário.introdução}→{vídeo.1.1}	0.4255556	0.9364303	1.046941
[15]	{vídeo.1.1}→{material.apoio.1.1}	0.7933333	0.8869565	1.034017
[16]	{material.apoio.1.1}→ {vídeo.1.1}	0.7933333	0.9248705	1.034017
[17]	{ }→{vídeo.1.1}	0.8944444	0.8944444	1.000000
[18]	{ }→{material.apoio.1.1}	0.8577778	0.8577778	1.000000

A Tabela 1 evidencia que os estudantes exibem maneiras diferentes de navegar pelos conteúdos do Módulo I. Não há um padrão único; a variabilidade comportamental resultou em seis principais padrões sequenciais de transição entre os conteúdos: (i) assistir aos vídeos e, em seguida, responder ao questionário de avaliação; ii) resolver os questionários de avaliação da aprendizagem primeiro e depois visitar os vídeos; ou resolver os questionários e, em seguida, visualizar os materiais de apoio – as sequências de números [2], [5], [10] e [14] da Tabela 1 exibem esses padrões. Esse tipo de comportamento são indicações de usuários que estão em busca de certificado, pois resolver os exercícios primeiro acelera o término do curso. Porém, quando o estudante inicia o aprendizado pelos questionários e não alcança a pontuação mínima, sua estratégia é consultar os materiais disponíveis antes de fazer a próxima tentativa de resolução; (iii) visualizar os materiais de apoio antes dos vídeos; (iv) acessar os vídeos antes do material de apoio; (v) visualizar somente os materiais de apoio; a sequência número [4] expõe esse comportamento; (vi) visitar somente os vídeos, ver, por exemplo, a sequência [9].

Para fins de visualização, foram gerados grafos que exibem o comportamento navegacional dos estudantes. A Figura 3 mostra as 18 sequências frequentes encontradas pelo algoritmo Apriori.

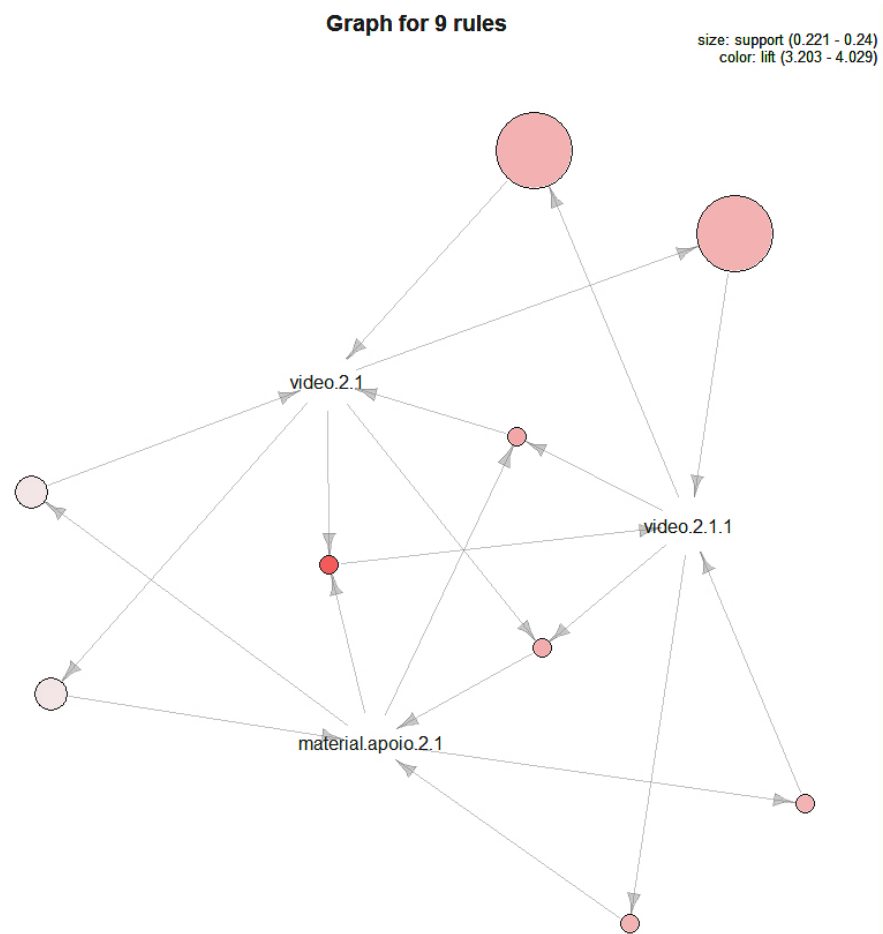


**Figura 3:** Representação visual das sequências de navegação frequentes  
**Fonte:** Autores (Grafo gerado de forma automatizada pela ferramenta RStudio).



A representação visual de sequências em forma de grafo é composta por vértices e arestas, que representam os itens e suas relações, também conhecidas como “regras” (HAHSLER; CHELLUBOINA, 2011). Nas Figuras 3 e 4, os itens são as unidades de aprendizagem mais acessadas, e as setas exibem os trajetos de um item para o outro. O tamanho dos círculos está associado aos valores de suportes, enquanto a cor está ligada aos valores de *lift*. Desta maneira, círculos maiores representam valores de suportes mais altos, enquanto a cor mais intensa relaciona-se com os valores de *lift* mais altos, expressando correlações mais fortes entre elementos.

Mediante a variabilidade comportamental dos estudantes constatada na Tabela 1, e devido à situação de os conteúdos mais visitados pertencerem ao módulo I, foi decidido realizar outra análise excluindo o módulo I e, ao realizar esta ação, o algoritmo identificou apenas os trajetos executados no módulo II, e não encontrou sequências para os módulos III e IV, confirmando que somente os módulos I e II tiveram acessos. As sequências frequentes referentes ao módulo II do MOOC podem ser visualizadas na Figura 4.



**Figura 4:** Comportamento de navegação no módulo II

**Fonte:** Autores (Grafo gerado de forma automatizada pela ferramenta RStudio).

Na Figura 4, é possível constatar uma quantidade menor de sequências no módulo II; essa ocorrência deve-se à queda de visualizações dos materiais, quando comparada ao módulo I. A partir dos trajetos dos estudantes exibidos nas Figuras 3 e 4, fica claro, pelo direcionamento das setas, que há diversos pontos de partida e de chegada, sugerindo que os estudantes percorrem os conteúdos de forma desordenada, sem uma organização prévia ou estratégia de aprendizado formulada. Uma possível justificativa para os padrões comportamentais diversificados pode ser associada ao fato de os participantes se inscreverem com os propósitos de explorar o curso para saber como funciona, avaliar a qualidade ou o nível de dificuldade dos conteúdos, verificar se o curso corresponde às suas necessidades.

Desta forma, as principais descobertas desta pesquisa mostraram que há uma diversidade de padrões navegacionais, e os materiais mais visualizados estão vinculados ao módulo I do MOOC. Embora seja esperado e comum que os estudantes estejam mais engajados no início do curso, o fato de reduzirem os acessos de maneira antecipada, não progredindo no curso, é um aspecto negativo que gerou preocupação, apontando a necessidade de investigar quais fatores provocaram o desengajamento de forma antecipada. Essas constatações revelaram que a desistência do curso ocorre precocemente, e os materiais de módulos mais avançados não chegam a ser acessados. Essas evidências mostram a importância de estudar o comportamento navegacional de estudantes de MOOC, a urgência em estabelecer intervenções para evitar o abandono do curso e definir estratégias que propiciem melhor aproveitamento dos materiais/ sessões de aprendizagem.

Ao consultar os resultados de trabalhos relacionados, algumas semelhanças e diferenças foram encontradas. Em Boroujeni e Dillenbourg (2018), é possível rastrear alguns pontos em comum. Esses autores encontram grupos de estudantes que realizaram os testes antes de visualizarem os outros tipos de materiais de aprendizagem; capturaram estudantes que visualizaram os vídeos, porém não resolveram os testes; também encontraram alunos pouco ativos, e outros que começaram o curso engajados, mas esse envolvimento não permaneceu ao longo do curso. Apesar destes aspectos semelhantes, o estudo de Boroujeni e Dillenbourg (2018) se diferencia da pesquisa descrita neste artigo em razão do uso de *clusters*, e os autores também capturaram um conjunto maior de padrões comportamentais e analisaram a relação dos padrões encontrados com o desempenho/nota.

Embora a pesquisa de Deeva et al. (2018) tenha utilizado métodos diferentes para encontrar as sequências de navegação de estudantes de um MOOC, também notaram-se algumas similaridades. Deeva et al. (2018) constataram que o comportamento navegacional dos estudantes permite prever o abandono. Uma das razões está relacionada aos eventos acumulados. Os estudantes propensos à desistência registram menos ações, quando comparados àqueles que chegaram ao final do curso. De forma semelhante a esses autores, também identificamos que, a partir do momento em que os estudantes reduzem suas ações no curso, não visualizando os materiais, tem-se um forte indicador de insucesso na aprendizagem e maior probabilidade de que estes estudantes não permanecerão no curso.

O trabalho de Maldonado-Mahauad et al. (2018) aponta maiores diferenças, pois além combinar o uso de *clusters* e de regressão em seus métodos, a intenção dos autores foi capturar as sequências de interação para prever o sucesso/aprovação em um MOOC, incluindo o uso de variáveis demográficas, de autorregulação da aprendizagem e notas. Contudo, alguns dos padrões encontrados por esses autores também foram identificados em nossa pesquisa, por exemplo: alunos que se matricularam no curso, mas exibiram poucas ações e não retornaram ao curso; alunos que acessaram vídeos e depois tentaram as avaliações; alunos que preferem fazer as atividades antes de acessar os vídeos.

Mediante os aspectos em comum observados, podemos concluir que há evidências da utilidade dos padrões sequenciais de interação dos estudantes de MOOCs para fins de descobertas de aspectos ocultos que podem explicar fatores de aprendizagem, bem como levantar diferentes atributos que estão associados e que justificam o desempenho, sucesso, e desistências.

## 5. Conclusões

Esta pesquisa foi realizada com os propósitos de estudar o comportamento navegacional de usuários de um MOOC para descobrir como ocorre a transição entre os conteúdos, detectar as sequências mais frequentes e identificar quais os materiais mais acessados. A partir dos procedimentos de mineração de dados, constatou-se que os conteúdos do primeiro módulo do MOOC são os materiais mais visitados pelos estudantes.

No que se refere ao comportamento navegacional, os estudantes percorreram as unidades de aprendizagem a partir de trajetos diferenciados. Os comportamentos navegacionais mais frequentes foram divididos em seis grupos opostos: (i) assistir a vídeos antes de responder ao questionário de avaliação; (ii) resolver os questionários antes de acessar os demais materiais; (iii) visualizar os materiais de apoio antes dos vídeos; (iv) visualizar os vídeos antes dos materiais de apoio; (v) visualizar somente os materiais de apoio; (vi) visitar somente os vídeos. A Tabela 1 e as Figuras 3 e 4 permitiram verificar os itens antecedentes e consequentes, que representam os pontos de partida de uma unidade de aprendizagem para outra, incluindo acessos desordenados e retornos a conteúdos que deveriam ter sido completados primeiro.

Os padrões distintos de navegação sinalizaram que não existe um único trajeto de aprendizagem a ser seguido, porém não se sabe se tais comportamentos correspondem a estratégias individuais de consumo dos materiais educacionais, havendo a necessidade de conduzir investigações complementares na tentativa de compreender os motivos da transição entre os conteúdos serem diferenciadas e como o comportamento navegacional pode influenciar a aprendizagem.

Encontrar as sequências de navegação dos estudantes a partir do uso do algoritmo Apriori foi um passo inicial para conhecer os trajetos percorridos pelos estudantes e verificar se os conteúdos do curso estão sendo acessados. Os resultados divulgados neste artigo contribuirão para nossas pesquisas futuras, principalmente em investigações com ênfase nos efeitos do comportamento navegacional no desempenho em atividades e na conclusão dos cursos, bem como para implantar estratégias futuras para recomendar trajetórias adequadas com o perfil dos estudantes, e agir antecipadamente mediante a detecção de sinais comportamentais indicativos de abandono do curso. A limitação desta pesquisa foi não ter utilizado outros algoritmos de mineração de padrões sequenciais para comparar qual teria o melhor desempenho nos dados do MOOC. Por isso, almejamos que outros algoritmos possam ser estudados e testados na plataforma futuramente.

## Financiamento

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

## Referências

- AGRAWAL, R.; IMIELINSKI, T.; SWAMI, A. Mining Association Rules between Sets of Items in Large Databases. In: Proceedings of the ACM SIGMOD Conference on Management of data. **Anais SIGMOD '93**. Washington, EUA, p. 207-216, 1993.
- AMARAL, F. **Aprenda mineração de dados: teoria e prática**. Rio de Janeiro: Alta Books, 2016.
- ANDRADE, M. V. M.; SILVEIRA, I. F. Panorama da Aplicação de Massive Open Online Course (MOOC) no Ensino Superior: Desafios e Possibilidades. **EaD em Foco**, v. 6, n. 3, 27 dez. 2016.
- BODON, F. A Trie-based APRIORI Implementation for Mining Frequent Item Sequences. In: The Eleventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. **Anais OSDM '05**. Chicago, IL, USA, p. 56 – 65, 2005.
- BOROJENI, M. S.; DILLENBOURG, P. Discovery and Temporal Analysis of Latent Study Patterns in MOOC Interaction Sequences. In: 8th International Conference on Learning Analytics and Knowledge (LAK'18). **Anais LAK '18**. Sydney, Australia, 2018, p. 206-215.

- BOSC., G. et al. Pattern Mining Approach to Study Strategy Balance in RTS Games. **IEEE Transactions on Computational Intelligence and AI in Game**, v. 9, n. 2, Jun. p. 123-132, 2017.
- CASTRO, M. S. O. et al. Seleção de Caminho de Aprendizagem para Grupo de Usuários: uma Abordagem baseada em Perfil. In: XXVIII Simpósio Brasileiro de Informática na Educação (SBIE 2017). **Anais [...]**. Recife, PE, 2017, p. 1167-1176.
- CHEN, Q. et al.. ViSeq: Visual Analytics of Learning Sequence in Massive Open Online Courses. **IEEE Transactions on Visualization and Computer Graphics**, v. 26, n. 3, p. 1622-1636, March. 2020.
- DEEVA G.; et al. Dropout Prediction in MOOCs: A Comparison Between Process and Sequence Mining. In: TENIENTE, E.; WEIDLICH, M. (EDS). **Business Process Management Workshops**. BPM 2017. Lecture Notes in Business Information Processing, v. 308. Cham: Springer, 2018. p. 243-255.
- FREDERIKS, G. et al . Learning by creating a MOOC. In: Learning With MOOCS (LWMOOCS), 2018, Madrid, Spain. **Anais [...]**. Madrid, 2018. p. 127-130.
- GREGORI, E. B. et al. Learner support in moocs: Identifying variables linked to completion. **Computers & Education**, v. 122, p. 153-168, Jul. 2018.
- HAHSLER, M.; CHELLUBOINA, S. **Visualizing association rules**: Introduction to the R-extension package arulesViz. R project module. 2011. Disponível em: <https://cran.r-project.org/web/packages/arulesViz/vignettes/arulesViz.pdf>.
- KOTIYAL, B. et al. User behavior analysis in web log through comparative study of Eclat and Apriori. In 7th International Conference on Intelligent Systems and Control (ISCO). **Anais [...]**. Coimbatore, Tamil Nadu, India, 2013, p. 421-426.
- MALDONADO-MAHAUAD, J. et al. Predicting Learners' Success in a Self-paced MOOC Through Sequence Patterns of Self-regulated Learning. In: Pammer-Schindler, V.; Pérez-Sanagustín, M.; Drachsler, H.; Elferink, R.; Scheffel, M. (eds). **Lifelong Technology-Enhanced Learning**. EC-TEL 2018. Lecture Notes in Computer Science, v. 11082. Cham: Springer, 2018. p. 355-369.
- MÁRQUEZ-VERA, C. et al. Predicting student failure at school using genetic programming and different data mining approaches with high dimensional and imbalanced data. **Applied Intelligence**, v. 38, n. 3, p. 315-330. Abr. 2013.
- RIZVI, S.; RIENTIES, B.; ROGATEN, J. Temporal Dynamics of MOOC Learning Trajectories. In: Proceedings of the First International Conference on Data Science, E-learning and Information Systems. **Anais [...]**. Madrid, Spain, Article No. 39. 2018, p. 1-6.
- TAIR, M. M. A.; EL-HALEES, A. M. Mining Educational Data to Improve Students' Performance: A Case Study. **International Journal of Information and Communication Technology Research**, v. 2, n. 2, p. 140-146, 2012.