

Discovery Engagement Patterns MOOCs Through Cluster Analysis

R. L. Rodrigues, J. L. C. Ramos, J. C. S. Silva and A. S. Gomes

Abstract— Cluster analysis can be used to help researchers identify behavioral patterns of students with regard to engaging in interactions via the forum and during activities during a course in MOOC mode (English, Massive Open Online Course). This article aims to analyze the effectiveness of educational data mining techniques, specifically the cluster analysis to identify students engagement patterns in MOOC courses in mode. The analyzes in this article demonstrate the use of hierarchical clustering method (Ward clustering) and the non-hierarchical clustering method (k-means) to analyze the engagement behavior characteristics, involving carrying out activities and interactions via the forum. For the analysis were taken into account the interaction patterns made in discussion murals in Openredu platform, as well as data access and activities of completeness. The insights found in this study can serve as indications for use by MOOCs designers to meet the diversity of engagement patterns and design interfaces that guide the design of adaptive strategies that allow increasing engagement and fostering a better learning experience.

Keywords— MOOC, Engagement, Clustering Hierarchical method, Clustering Non-hierarchical method.

I. INTRODUÇÃO

AS RECENTES tecnologias web tem ajudado instituições de ensino superior a oferecer oportunidades de aprendizagem on-line que atendem às necessidades de estudantes ao redor do mundo [21]. Com a ajuda destas tecnologias, o número de cursos on-line aumentou significativamente [1].

Os Cursos Online Abertos e Massivos, conhecidos por MOOC (do inglês, *Massive Open Online Course*) representam uma variante de cursos on-line para grande públicos na modalidade de ensino a distância. O termo foi criado em 2008 pela Universidade de Manitoba, no Canadá e atualmente são observadas taxas crescentes de popularidade dessa abordagem.

Esta modalidade de curso corresponde a um modelo escalável e flexível para o aprendizado on-line. Preenchem lacunas da educação formal, apoiando as oportunidades de requalificação e atraindo novos aprendizes independentes, que por algum motivo, não podem frequentar o ensino superior ou formal.

Recentemente várias instituições de ensino superior e professores aderiram a essa modalidade educacional, lançando cursos em diferentes áreas. Plataformas como o Coursera, EdX, MiriadaX ou FutureLearn estão fornecendo apoio aos

professores que desejem implantar cursos abertos que podem permitir a participação de interessados [18].

Apesar do grande interesse, a modalidade é questionada. Grande parte da divulgação dessa modalidade parece estar voltada a realizar uma ação “promocional influenciada por interesses comerciais”. Pouco se sabe sobre a efetividade da experiência de aprendizagem de alunos com diferentes habilidades, motivações e disposições. Uma das principais evidências é a baixa taxa de conclusão dos cursos por parte dos alunos. Pesquisas apontam que apenas 10% dos alunos matriculados, conseguem finalizar o curso nesta modalidade [3]. Ainda são escassas as soluções que refletem um ganho significativo no aprendizado em relação às estratégias convencionais.

Kizilcec [10] afirma que a crítica acima dos MOOCs se fundamenta no fato de a taxa de conclusão ser o único indicador de avaliação desta modalidade. Isso cria uma percepção dicotômica, indicando apenas quem finalizou ou quem não finalizou o curso. Essas críticas baseiam-se em uma classificação binária dos alunos que se comportam de acordo com as expectativas do professor e todos os outros envolvidos. Esta visão obscurece as razões pelas quais um aluno pode não estar engajado, como também, não levam em consideração os perfis de alunos que optam por participar apenas em alguns momentos, permanecendo envolvidos, mas não almejando ganhar uma declaração ou certificado de realização. Críticas como estas não permitem, por exemplo, que os projetistas de MOOC desenvolvam soluções que ampliem a efetividade da modalidade.

A pergunta que convém formular neste artigo nos orienta a propor novos indicadores que permitam uma descrição mais ampla de como a experiência de aprendizagem é experimentada por alunos em MOOCs. Sendo esse um fenômeno relativamente novo, ainda são poucos os estudos que têm explorado estas questões [13].

Este artigo tem como objetivo analisar a efetividade do primeiro MOOC de língua portuguesa no Brasil, através de técnicas de mineração de dados educacionais (*Educational Data Mining*), especificamente a análise de agrupamento, para identificar padrões de engajamento de alunos em cursos na modalidade MOOC. O estudo foi realizado no ambiente Openredu, por alunos com diferentes habilidades, motivações e disposições.

Este artigo está dividido em outras quatro seções: seção II contém a fundamentação teórica e trabalhos relacionados sobre MOOCs. Na seção III, é apresentada as características e definições sobre engajamento. Na seção IV, são apresentadas as análises dos agrupamentos e os resultados obtidos a partir do estudo e as respectivas discussões. Por fim, são

R. L. Rodrigues, Universidade Federal Rural de Pernambuco (UFRPE), Departamento de Educação, Recife, Brasil, rodrigo.linsrodrigues@ufrpe.br

J. L. C. Ramos, Universidade Federal do Vale do São Francisco (UNIVASF), Juazeiro, Bahia, Brasil, jorge.cavalcanti@univasf.edu.br

J. C. S. Silva, Universidade Federal do Vale do São Francisco (UNIVASF), Juazeiro, Bahia, Brasil, joao.sedraz@univasf.edu.br

A. S. Gomes, Universidade Federal de Pernambuco (UFPE), Centro de Informática, Recife, Brasil, asg@cin.ufpe.br

apresentadas as considerações finais e sugestões para trabalhos futuros.

II. FUNDAMENTAÇÃO TEÓRICA

Nas seções a seguir são apresentados os argumentos necessários para o entendimento sobre as diferentes modalidades de MOOCs bem como o entendimento sobre a experiência dos alunos nestas modalidades e por fim abordaremos aspectos da sua efetividade.

A. Os diferentes tipos de MOOCs

A compreensão das diferentes definições do termo MOOC é importante para o entendimento de como experiências educacionais são desenvolvidas e entregues por meio das várias plataformas online. O tipo de abordagem mais utilizada atualmente é desenvolvida por organizações como Coursera, EDX e Udacity. São caracterizadas com o termo "xMOOCs" [16], [15]. Estes cursos são construídos por instituições de ensino consolidadas internacionalmente, normalmente são concebidos como um programa semanal de aulas em vídeo, seguido por testes ou outras atribuições que avaliam a compreensão dos conteúdos pelos alunos. Uma das ferramentas mais utilizadas para a comunicação são os fóruns de discussão.

Como alternativa ao modelo xMOOCs surge o modelo baseado em teorias conectivistas. Eles são intitulados "cMOOCs" [16],[15]. Nesse formato, os próprios alunos geram a maior parte das atividades de aprendizagem. Os alunos podem escrever suas próprias reflexões sobre o que estão aprendendo em blogs e outras plataformas de mídia social. Eles podem: compartilhar tudo de volta com seus pares, estabelecer colaboração com outras pessoas, compartilhar informações, e agregar a colaboração com outros às suas ideias para criar uma experiência de aprendizagem emergente.

Esta forma de aprendizagem coletiva difere do modelo xMOOC em vários aspectos [9]. Em primeiro lugar, os modos de produção diferem drasticamente. Em xMOOCs, equipes de instituições como a Universidade de Stanford, Harvard University, Massachusetts Institute of Technology e universidades parceiras desenvolvem e mediam os cursos. Em cMOOCs, parte-se do princípio de que qualquer indivíduo adulto é capaz de regular sua experiência de aprendizagem.

A concepção do curso e o papel dos alunos são conceituados de forma muito diferente nos dois modelos. Em um xMOOC as responsabilidades do aluno concentram-se em manipular de forma ativa o conteúdo do curso e ser avaliado. Em um cMOOC, os alunos podem utilizar o conteúdo fornecido pelo curso, mas como um recurso entre muitos que podem surgir através da própria busca e pesquisa on-line.

B. A experiência dos estudantes em MOOCs

A conectividade nesta modalidade é geralmente fornecida através de redes sociais, e um conjunto de recursos on-line de acesso livre fornecendo o conteúdo ou material de estudo. Geralmente esses cursos não exigem pré-requisitos, taxas ou exigência de uma quantidade fixa de participações [16]. A participação em um MOOC é completamente voluntária e

depende da pessoa interessada. O espaço colaborativo de um MOOC pode ser dividido entre várias plataformas e tecnologias diferentes. Por exemplo, os participantes de um MOOC podem criar seus próprios posts discutindo aspectos do curso em diferentes espaços do ambiente de aprendizagem.

Nessa modalidade de curso utilizam-se de diversos tipos de ambientes para promover a comunicação e a colaboração entre os membros. Uma das ferramentas mais utilizadas são os fóruns de discussão. No entanto existem diversos estudos que mostram que poucos alunos são pouco ativos via fórum de discussão em um MOOC. Poucos alunos possuem habilidades para participar de discussões durante boa parte do tempo. Breslow [3] mostra que 27,7% dos estudantes utilizam os fóruns para formular perguntas. Desses, 40,6% respondem a perguntas e 36% fazem comentários. A pesquisa também mostrou que, de todos os alunos que finalizaram o curso, apenas 52% utilizam o fórum ativamente.

Em um estudo [2] realizado pela Universidade de Stanford foi feita uma análise da utilização de fórum em 23 MOOCs realizados na plataforma Coursera. Observou-se que de todos os estudantes matriculados, apenas 10% dos alunos interagem via fórum.

Um outro estudo, realizado pela Universidade de Edimburgo [6] foi realizada uma análise com 6 cursos na modalidade MOOC e identificou-se que os alunos são menos propensos a se envolver em discussão via fóruns do que interagir e utilizar vídeos ou avaliações. Ainda são poucos os estudos que caracterizam e avaliam a experiência de engajamento e efetividade das interações nesses ambientes de aprendizagem.

III. ENGAJAMENTO EM AMBIENTES DE APRENDIZAGEM

Há um consenso na literatura de que o engajamento é necessário para promover resultados positivos na aprendizagem [7]. Ainda que seja complexo definir os limites conceituais do termo, o engajamento é um fenômeno multifacetado, por isto tem despertado o interesse de pesquisadores de diversas áreas. É verificado na literatura que o engajamento surge da interação do sujeito com o contexto e que ele responde às mudanças no ambiente [4]. A seguir definimos, de acordo com a literatura, as diferentes abordagens de engajamento em plataformas de MOOC.

A. Diferentes abordagens de engajamento

Na tentativa de melhor compreender o conceito de engajamento e de instrumentalizar sua mensuração, surgiram várias propostas de classificação relacionadas a tipos e/ou aspectos do engajamento, resultando em categorizações bastante diversificadas e frequentemente sobrepostas. O engajamento acadêmico, cognitivo, intelectual, institucional, emocional, afetivo, comportamental, social e psicológico, são apenas algumas das tipologias descritas para o engajamento, segundo uma extensa revisão bibliográfica elaborada por Parsons e Taylor [19]. Podem-se identificar na literatura três principais modelos de engajamento [5], a saber: comportamental, cognitivo e emocional.

O engajamento comportamental está relacionado à noção de participação e é associado a comportamentos mais

facilmente observáveis como a participação efetiva do estudante nas atividades curriculares e extracurriculares propostas. O engajamento comportamental costuma ser tradicionalmente mensurado a partir de aspectos quantitativos, como: frequência, eficácia e tempo despendido na realização de atividades e notas obtidas em testes.

O engajamento cognitivo está relacionado à ideia de investimento, fazer esforço para entender ideias complexas e deter habilidades difíceis. Esse revela-se pelo investimento pessoal, deliberado, persistente e autônomo do aluno para regular sua aprendizagem. Esse tipo de engajamento mobiliza habilidades cognitivas mais complexas com o objetivo de aumentar a própria compreensão, resolver problemas complexos e construir novos conhecimentos, desenvolvendo estratégias específicas para os fins que se pretenda atingir [5].

O engajamento emocional relaciona-se às reações positivas e negativas do estudante aos fatores e recursos escolares. O sentimento de pertencer a um grupo, representado pela comunidade de aprendizagem e pela instituição educacional, assim como a aceitação de seus valores são referências para identificação deste tipo de engajamento. Considera-se que existe engajamento emocional quando os estudantes apresentam atitudes e reações positivas em relação à instituição, aos professores, ao aprendizado e aos seus pares [7]. Para um maior engajamento, Rudduck [20] salienta que é importante para o aluno sentir-se parte de uma comunidade de aprendizagem e que sua participação agregue valor ao resultado final a ser obtido.

Diversos trabalhos vêm sendo desenvolvidos com o objetivo de qualificar e avaliar o engajamento de alunos em atividades de aprendizagem. A maioria destes trabalhos foca na identificação de engajamento em ambientes de aprendizagem presencial. No trabalho de Milligan [17], o autor desenvolve uma maneira de identificar padrões de engajamento utilizando uma abordagem qualitativa de pesquisa em um ambiente de MOOC. O experimento foi realizado entre setembro de 2011 a maio de 2012 em um curso com 2.300 estudantes. A coleta foi realizada por meio de entrevistas semi-estruturadas que explorou o tema do engajamento através de perguntas sobre motivação, comportamento, participação e a composição da rede de aprendizagem pessoal de cada entrevistado. Ao analisar as respostas foi possível detectar três padrões de engajamento: os participantes ativos, *lurkers* e passivos.

Considerar os aspectos comportamentais, emocionais e cognitivos do engajamento significa encará-lo como fenômeno multidimensional para o qual concorrem diversos fatores [5].

No entanto, ainda há poucas iniciativas na literatura que permitam criar uma abordagem ampla para qualificar o engajamento em ambientes de aprendizagem do tipo MOOC. Na seção a seguir analisaremos a adequação de técnicas de análise de agrupamento para avaliar o engajamento nessa modalidade de ensino.

IV. DESCRIÇÃO DO EXPERIMENTO

Para entender quais variáveis descrevem padrões de engajamento em MOOCs, acompanhou-se uma formação

nesta modalidade com 5100 participantes inscritos. O tema do curso foi a Nova Gramática da Língua Portuguesa.

A. Procedimento e Participantes

O MOOC foi realizado utilizando a plataforma Openredu (<http://openredu.org>). Este é um ambiente de aprendizagem híbrido que integra funções de sistema de gestão da aprendizagem (*Learning Management System*, ou LMS) e de rede social.

O curso contou com a participação de estudantes de diferentes idades e níveis de escolaridade. As origens dos participantes variavam por todas as regiões do Brasil e de outros países como: Colômbia, Portugal e Argentina. Este foi o primeiro curso de língua portuguesa, na modalidade MOOC, realizado no Brasil [14].

Os participantes do curso receberam um itinerário de como proceder ao longo do curso e foram acompanhados por uma equipe de tutores. O conteúdo foi apresentado em multimídia nos formatos de texto, vídeos, exercícios e apresentações. Cada módulo do curso foi dividido por conteúdo específicos da língua portuguesa. Os participantes podiam acessar o conteúdo do MOOC e interagir através da rede social Openredu e de uma aplicação móvel.

B. Variáveis e coleta de dados

A maioria dos estudos relacionados ao nível de engajamento tem-se pautado pela medição da frequência ou assiduidade dos alunos. A frequência é usada em diversos estudos para estimar o nível de engajamento. Ela é um dos indicadores mais disponíveis em cursos online [17].

O grau de motivação dos alunos é muitas vezes associado ao grau de engajamento e tem sido definida como um estado interno ou condição que ativa o comportamento e dá-lhe sentido [8].

Para os objetivos deste estudo, definimos 15 variáveis que representaram a frequência de diferentes categorias de postagens, assiduidade e notas.

Os dados utilizados na análise são postagens que os participantes do curso emitiam durante as atividades on-line. Todas as postagens foram extraídas do ambiente de aprendizagem Openredu, preservando a identidade dos participantes.

C. Análise de agrupamento e padrões de engajamento

Para a busca por padrões comportamentais de engajamento foi utilizado a técnica de análise de agrupamento. Esta técnica busca por grupos excludentes, ascendentes, reduzindo as informações de um conjunto de n parcelas para um novo conjunto de k grupos, no qual k é significativamente menor que n [12].

Durante a realização do experimento, foram utilizadas duas formas de análises de agrupamento: a hierárquica e a não hierárquica. A análise de agrupamento hierárquico teve como objetivo encontrar, através da geração de um dendrograma, uma quantidade k de grupos, até então desconhecidos.

Posteriormente foi utilizado o tipo não-hierárquico de agrupamento com o objetivo de classificar os estudantes dentro de terminados grupos. O objetivo foi fazer uma classificação mais precisa das observações em uma quantidade

de grupo já pré-definida, encontrada inicialmente com a análise hierárquica.

V. RESULTADOS E DISCUSSÕES

Nesta seção são apresentadas e interpretadas as análises descritivas e de agrupamento das variáveis e suas relações. O critério de inclusão de sujeitos adotado no estudo leva em consideração o fato de os mesmos terem participado do MOOC postando ao menos duas vezes no mural do curso ou ter realizado pelo menos uma das atividades propostas. O total de estudantes incluídos para fins de análise dos dados a partir do universo de inscritos foi 1569 (31,38%) do total dos matriculados. Dentre os matriculados, apenas 695 (13,6%) usuários postaram pelo menos uma mensagem no fórum de discussão. Dentre estas postagens, tivemos um total de 7195 mensagens, onde 1986 foram do tipo comentários.

A. Análise descritiva dos dados

Entre a variabilidade das postagens, tivemos um máximo de 225 postagens por aluno e uma média de 10 postagens por aluno ($SD=28,7$). O professor da disciplina interagiu através de 439 mensagens postadas. Dos alunos inscritos, apenas 3531 (69,2%) apresentaram-se no fórum da disciplina através de uma única postagem. Eles não realizaram outras atividades ou comentários via mural.

A Tabela I mostra uma matriz de autocorrelação com uma correlação positiva entre a quantidade de respostas a dúvidas e a frequência de atividades realizadas pelo grupo. Essa correlação indica que a intensidade da colaboração em um MOOC está positivamente correlacionada com efetividades na realização das atividades propostas pelos professores.

TABELA I
MATRIZ DE CORRELAÇÃO

Variável	Total	Média \pm DP	Matriz de autocorrelação		
			Atividades realizadas	Comentários sobre postagens	Respostas a dúvidas em fórum
Atividades realizadas	7347	4,69 \pm 4,18	1	0,353	0,408*
Quantidade de comentários sobre postagens	15.221	9,71 \pm 20,83	-	1	0,045
Quantidade de respostas a dúvidas em fórum	4.225	2,69 \pm 27,20	-	-	1

A seguir foi realizada uma análise, através da Tabela II de dupla entrada, na qual verificou-se a relação das frequências de classes das médias das atividades realizadas, em relação às médias de postagens em fórum.

TABELA II
RESULTADOS DA ANÁLISE DE MÉDIAS X POSTAGENS

Média de notas	Total	Percentual	Médias de postagens	
			Média de comentários em	Média de respostas a

fórum \pm DP			dúvidas em fórum \pm DP
< 4	81 (5,51%)	4,69 \pm 4,18	3,370 \pm 5,27
> 4 e < 7	383 (26,05%)	9,71 \pm 20,83	8,900 \pm 18,18
> 7	1006 (68,43%)	2,69 \pm 27,20	10,271 \pm 20,09

Observa-se que à medida que a participação em termos de postagens aumenta, observa-se aumento das médias de notas obtidas. Na Tabela II observa-se que os alunos que obtiveram médias de notas abaixo de 4 mantêm um padrão de interação baixa: 3,37 postagens em média. Essa média inclui as interações do tipo comentários e respostas a dúvidas. Para alunos que tiveram média de nota acima de 7, observa-se que a participação (postagens e comentários) foi de 10,271 em média.

Isto nos dá indícios de que a interação via fórum de discussão parece promover um efetivo engajamento na realização de atividades.

B. Análise de agrupamento

Inicialmente realizou-se um pré-processamento, onde foi feita uma padronização das escalas das variáveis associadas através de uma normalização. Ao todo, foram extraídas 14 variáveis, sendo: 1 variável de frequência de comentários em fórum, 1 variável de respostas a dúvidas e 12 variáveis referente a realização de atividades dos módulos.

Foi necessário realizar a criação da matriz de proximidades utilizando a distância euclidiana, verificada como sendo a melhor para o contexto dos dados da pesquisa. Em seguida foi realizada a construção do agrupamento hierárquico com a aplicação de algoritmos de agrupamento.

Os algoritmos geraram como resultado um gráfico do tipo dendograma, onde, a partir de sua interpretação, identificaram-se grupos com características comportamentais distintas. A interpretação do gráfico (Fig. 1) nos permite visualizar a existência de sete grupos de participantes do MOOC. Este valor foi utilizado como valor de entrada para o algoritmo *K-means* na aplicação da clusterização não-hierárquica.

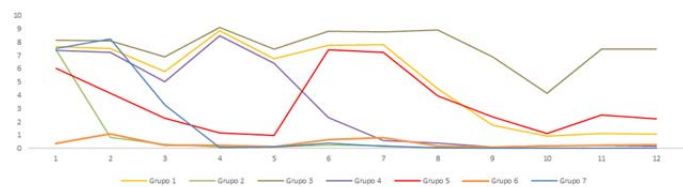


Figura 1. Representação do comportamento dos grupos.

O gráfico ilustra o desempenho dos sete grupos ao longo das doze atividades. Percebe-se claramente que o grupo 6 teve um desempenho insatisfatório em todas as atividades. O grupo 2 apresentou uma queda relevante de desempenho após a primeira atividade e a partir da segunda atividade, também teve desempenho bastante insatisfatório como o grupo 6. O grupo 7 também apresentou essa queda acentuada na média após a segunda atividade, atingindo indicadores similares aos grupos 2 e 6.

Os grupos 1 e 4 tiveram desempenho semelhante até a atividade 5, onde a partir daí verificou-se uma queda acentuada e progressiva na média das notas do grupo 4,

ficando este grupo com baixo nível de desempenho a partir dessa atividade. O grupo 1 permaneceu com média satisfatória até a sétima atividade, a partir da qual o desempenho caiu para níveis bem inferiores, mas ainda assim melhores que os dos grupos 2, 4, 6 e 7. No geral, o grupo 1 obteve a segunda maior média dentre os sete grupos.

O maior grupo de alunos foi o grupo sete, pois agrupou 373 (23,94%) do total de alunos. A maioria dos alunos desse grupo (368 - 96,71%) realizou menos de 50% das atividades. O mesmo grupo obteve uma média de notas de 1,67 ($SD \pm 2,92$). Observou-se uma média de 5,19 ($SD \pm 5,81$) comentários e 0,81 ($SD \pm 3,17$) respostas a dúvidas por aluno.

O grupo que mais se destacou, com relação ao engajamento em atividades, foi o grupo três. Este grupo teve uma quantidade de 249 (15,98%) alunos. A maioria dos alunos (170 - 68,27%) realizaram todas as doze atividades. Este grupo teve uma característica importante no que se refere à interação via fórum de discussão, pois se destacou com uma média de 26,91 ($SD \pm 39,02$) interações do tipo comentários e uma média de 170 ($SD \pm 68,27$) de comentários do tipo esclarecimento de dúvidas de outros alunos (ajuda aos pares). Este grupo também se destacou por obter as maiores médias das notas 7,69 ($SD \pm 1,30$).

Para ter uma ideia do desempenho dos grupos foi necessário observar a variabilidade dos dados em torno da mediana. Utilizamos o Boxplot como medida gráfica para plotar a medida de tendência central, a variabilidade dos dados e avaliar a existência de outliers (valores extremamente altos ou baixos) (Fig. 2).

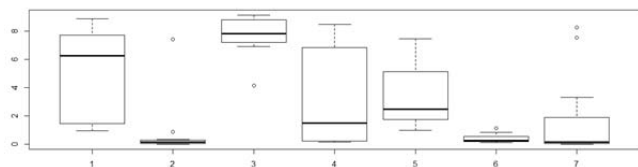


Figura 2. Box plot referente aos grupos.

Podemos observar que no grupo 1 existe uma grande variabilidade dos dados. Trata-se de um grupo heterogêneo. A mediana permanece acima de seis e existem alunos com médias acima de oito.

O grupo 2 foi um dos grupos com menor desempenho dentre os sete grupos, tem mediana perto de zero, pouca variabilidade e existência de *outliers*. O grupo três foi considerado o melhor grupo dentre os sete, teve pouca variabilidade e mediana em torno de oito e valor superior extremo próximo a nove.

O grupo quatro tem mediana próxima a três e variabilidade superior alta com ponto máximo próximo a sete. O grupo seis, graficamente, tem semelhanças ao grupo dois, embora possua um pouco mais de variabilidade. O grupo sete possui mediana inferior a todos os grupos, variabilidade superior alta e presença de *outliers*.

C. Aplicação do teste de comparação de médias

Embora tenhamos encontrado sete grupos com características distintas, foi necessário realizarmos testes inferenciais de comparação de médias e variâncias para testar,

a aceitação ou não, de possíveis igualdades entre os grupos encontrados.

Na Tabela III temos os testes de comparação para cada dois grupos. Como podemos observar, valores de p-valor abaixo de 0,005 rejeitam a hipótese de igualdade entre os grupos.

TABELA III
RESULTADOS DOS TESTES DE MÉDIAS

Grupo	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6
1	-	-	-	-	-	-
2	0.00069	-	-	-	-	-
3	0.01885	0.00000	-	-	-	-
4	0.16390*	0.04888	0.00078	-	-	-
5	0.15070*	0.00715	0.00000	0.83810*	-	-
6	0.00023	0.52730*	0.00000	0.01483	0.00073	-
7	0.01144	0.41640*	0.00000	0.25330*	0.00456	0.17280*

*os grupos são semelhantes

Após a realização do teste de comparação de médias foi possível identificar a presença de grupos estatisticamente iguais. Desta forma, até a análise exploratória (*cluster*) tínhamos a presença de sete grupos distintos. Após o teste inferencial de comparação de médias, verificou-se que alguns grupos tinham características similares e não eram significativamente distintos, fazendo com que estes grupos tenham a mesma característica comportamental dos alunos.

Ao final foi verificado que existia a presença de apenas três categorias distintas, que são: Categoria dos engajados (grupo 3), Categoria dos esporádicos (grupo 1, grupo 4 e grupo 5) e Categoria dos desengajados (grupo 2, grupo 6 e grupo 7), que serão descritos na próxima seção.

D. Descoberta de padrões de engajamento

Os alunos foram distribuídos em sete grupos iniciais, definidos de acordo com as características representativas do engajamento do aluno no MOOC: atividades realizadas, comentários postados e respostas a dúvidas.

A partir da análise resultante da aplicação de teste estatístico das médias, foram encontrados grupos com comportamento similares, restando então três grupos que representaram as categorias de engajamento, onde os alunos puderem ser enquadrados em engajados, esporádicos e não engajados. Este resultado vem a corroborar com o resultado encontrado no trabalho de Milligan [17], onde o autor utilizou uma abordagem totalmente qualitativa. A Tabela IV mostra o resultado das categorias bem como suas principais características comportamentais.

TABELA IV
PADRÕES COMPORTAMENTAIS

Categorias de engajamento	Total de alunos	Características
Engajados (grupo 3)	249 (15,98%)	Esse foi o grupo que mais se destacou dentre todos. Tem uma ótima interação via fórum de discussão, mantém ritmo contínuo na realização das atividades e pouca variabilidade entre as médias de notas. Este grupo tem uma média de 26,91 interações via fórum de discussão e uma média de 7,55 ajudas a outros membros. Em relação a realização de atividades, 68,27% dos alunos realizaram todas as atividades e 31,33% dos alunos realizaram pelo menos nove das doze atividades. Pode ser considerado um grupo de alunos engajados que realizam atividades do início ao fim do curso e com pouca variabilidade comportamental.

<i>Esporádicos (grupos 1, 4 e 5)</i>	<i>406 (26,06%)</i>	<i>Esse grupo tem uma forte característica de mudanças. Consegue atingir picos repentinos durante todo o curso, sendo um grupo de grande variabilidade. Uma das prováveis características é que este grupo passa longos períodos sem entrar no ambiente e perde alguns prazos de realização de atividades, embora tenha um background de conhecimentos que faz com que tenha notas elevadas quando as realizam.</i>
<i>Desengajados (grupos 2, 6 e 7)</i>	<i>903 (57,96%)</i>	<i>Este grupo foi considerado o pior dos sete grupos, com relação ao engajamento nas atividades. Mais de 90% dos alunos só realizou no máximo duas atividades e as notas médias são inferiores aos demais grupos durante todo o curso. Em relação à média de comentários este grupo apresentou os piores resultados. Podemos ter indícios que este grupo é um grupo com características apenas de visualização de material didático e optam por participar apenas em alguns momentos, permanecendo envolvidos, mas não almejando ganhar uma declaração ou certificado de realização, e por este motivo não realizam atividades. Alunos deste grupo deixam de realizar mais de 75% das atividades propostas. É interessante observar que os grupos 6 e 7 iniciam o curso com desempenhos satisfatórios, no entanto no decorrer da segunda semana em diante tendem a não realização das atividades, permanecendo assim o resto do curso.</i>

A opção por três categorias de engajamento permitiu uma caracterização mais objetiva dos grupos formados. A inclusão de mais categorias ou subcategorias de engajamento levaria a um nível de detalhamento mais complexo ou mesmo desnecessário para o que se pretendia obter de conhecimento acerca dos alunos no MOOC realizado.

Evidenciamos que essa classificação é decorrente da técnica e do algoritmo de mineração utilizados neste trabalho. Outros níveis de engajamento ou mesmo outras configurações de agrupamento podem ser obtidas se forem utilizadas outras técnicas/algoritmos de mineração de dados.

VI. CONCLUSÃO

Este artigo buscou identificar padrões de engajamento de alunos participantes de um MOOC em Língua Portuguesa a partir da análise das suas participações online em fóruns de discussão e da realização de atividades.

Neste estudo buscou-se evidências, através da aplicação de técnicas de mineração de dados, especificamente análise de agrupamento, a importância do entendimento dos perfis de engajamento.

A análise de agrupamento foi importante pois serviu para identificar grupos com características comportamentais distintas que muitas vezes são imperceptíveis na visão da gestores e professores de MOOC.

Como trabalhos futuros pretendemos desenvolver soluções de software, acopladas a plataforma Openredu, que promovam um acompanhamento automatizado das ações dos usuários, afim de promover um maior engajamento nas atividades. Outro ponto ser abordado em trabalhos futuros é o estudo mais detalhadamente dos comportamentos de cada grupo na busca de indícios que possam ajudar a explicar a razão dos seus desempenhos e níveis de interação.

AGRADECIMENTOS

Esta pesquisa é apoiada pela FACEPE-APQ No. 0525-1.03/14 e BCT 0016-1.03/15. Alex Sandro Gomes é bolsista DT Nível 2/CNPq, processos nº 310466/2012-1 e nº 475634/2013-6.

REFERÊNCIAS

- [1] Allen, I. E., & Seaman, J. (2007). Online nation: Five years of growth in online learning. Sloan Consortium. PO Box 1238, Newburyport, MA 01950.
- [2] Anderson, A., Huttenlocher, D., Kleinberg, J., & Leskovec, J. (2014, April). Engaging with massive online courses. In Proceedings of the 23rd international conference on World wide web (pp. 687-698). ACM.
- [3] Breslow, L., Pritchard, D. E., DeBoer, J., Stump, G. S., Ho, A. D., & Seaton, D. T. (2013). Studying learning in the worldwide classroom: Research into edX's first MOOC. Research & Practice in Assessment, 8(1), 13-25.
- [4] Finn, J. D., & Rock, D. A. (1997). Academic success among students at risk for school failure. Journal of applied psychology, 82(2), 221.
- [5] Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. Review of educational research, 74(1), 59-109.
- [6] Grainger, B. (2013). Massive open online course (MOOC) report 2013.
- [7] Harris, L. R. (2008). A phenomenographic investigation of teacher conceptions of student engagement in learning. The Australian Educational Researcher, 35(1), 57-79.
- [8] Huitt, W. (1997). The SCANS report revisited. Paper delivered at the Fifth Annual Gulf South Business and Vocational Education Conference, Valdosta State University, Valdosta, GA, April 18. Retrieved December 2001,
- [9] Kafai, Y. B., & Peppler, K. A. (2011). Youth, technology, and DIY developing participatory competencies in creative media production. Review of Research in Education, 35(1), 89-119.
- [10] Kizilcec, R. F., Piech, C., & Schneider, E. (2013, April). Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In Proceedings of the third international conference on learning analytics and knowledge (pp. 170-179). ACM.
- [11] Shen, C. W., & Kuo, C. J. (2015). Learning in massive open online courses: Evidence from social media mining. Computers in Human Behavior (in press).
- [12] Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2009). Análise multivariada de dados. Bookman Editora.
- [13] Mackness, J., Mak, S., & Williams, R. (2010). The ideals and reality of participating in a MOOC. In Networked Learning Conference, Aalborg (pp. 266-274).
- [14] Mattar João. (2013). Fóruns de discussão em educação a distância: Moodle, Facebook e Redu. Tecnologia Educacional, pp. 6-17.
- [15] Siemens, G. (2012). MOOCs are really a platform.
- [16] McAuley, A., Stewart, B., Siemens, G., & Cormier, D. (2010). In the open: The MOOC model for digital practice. Charlottetown, Canada: University of Prince Edward Island.
- [17] Milligan, C., Littlejohn, A., & Margaryan, A. (2013). Patterns of engagement in connectivist MOOCs. MERLOT Journal of Online Learning and Teaching, 9(2).
- [18] Muñoz-Merino, P. J., Ruipérez-Valiente, J. A., Alario-Hoyos, C., Pérez-Sanagustín, M., & Kloos, C. D. (2015). Precise Effectiveness Strategy for analyzing the effectiveness of students with educational resources and activities in MOOCs. Computers in Human Behavior. Volume 47, June 2015, Pages 108-118
- [19] Parsons, J., & Taylor, L. (2011). Student Engagement: What do we know and what should we do?. University of Alberta.
- [20] Rudduck, J. (2007). Student voice, student engagement, and school reform. In International handbook of student experience in elementary and secondary school (pp. 587-610). Springer Netherlands.
- [21] Yukselturk, E. (2009). Do Entry Characteristics of Online Learners Affect Their Satisfaction?. International Journal on E-Learning, 8(2), 263-281.



Rodrigo Lins Rodrigues possui mestrado em Ciência da Computação pelo Cin/UFPE, graduação em Computação e em Estatística. Atualmente é aluno de doutorado em Ciência da Computação. Áreas do conhecimento em Computação: Data Science, Educational Data Mining (EDM) e Technology Enhanced Learning (TEL). Áreas do conhecimento em Estatística: Modelos de Regressão, Testes Paramétricos,

Modelagem estatística computacional.



Jorge Luis Cavalcanti Ramos possui mestrado em Ciência da Computação pelo Cin/UFPE. Atualmente é aluno de doutorado em Ciência da Computação e professor do quadro efetivo da Universidade Federal do Vale do São Francisco. É pesquisador do grupo de pesquisas em Ciências Cognitivas e Tecnologias Educacionais (CCTE) no Cin/UFPE. Áreas do conhecimento: Technology Enhanced Learning (TEL) e Educational Data Mining (EDM).



João Carlos Sedraz Silva, possui graduação em Engenharia Mecânica, especialização em Educação a Distância e mestrado em Ciência da Computação. É docente da Universidade Federal do Vale do São Francisco, onde coordenou a implantação dos cursos de graduação e pós-graduação na modalidade a distância. Atualmente, é aluno de doutorado no Centro de Informática da Universidade Federal de Pernambuco, com pesquisa concentrada na concepção e modelagem de ambientes virtuais de aprendizagem.



Alex Sandro Gomes é Engenheiro Eletrônico (UFPE, 1992), Mestre em Psicologia Cognitiva (UFPE, 1995) e concluiu o doutorado em Ciências da Educação pela Université de Paris v (René Descartes) em 1999. Atualmente é Professor no Centro de Informática da UFPE. É líder do grupo de pesquisa Ciências Cognitivas e Tecnologia Educacional. É Coordenador das comunidades de software livre Amadeus e Openredu.