



Pamella Letícia Silva de Oliveira

Vinculando Perfis de Engajamento ao Desempenho Acadêmico por Meio de Análise de Redes Sociais e Análise de Agrupamento nos Dados de Fóruns de Discussão

Recife

2021

Pamella Letícia Silva de Oliveira

Vinculando Perfis de Engajamento ao Desempenho Acadêmico por Meio de Análise de Redes Sociais e Análise de Agrupamento nos Dados de Fóruns de Discussão

Artigo apresentado ao Curso de Bacharelado em Sistemas de Informação da Universidade Federal Rural de Pernambuco, como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação.

Universidade Federal Rural de Pernambuco – UFRPE

Departamento de Estatística e Informática

Curso de Bacharelado em Sistemas de Informação

Orientador: Rodrigo Lins Rodrigues

Recife

2021

Dados Internacionais de Catalogação na Publicação
Universidade Federal Rural de Pernambuco
Sistema Integrado de Bibliotecas
Gerada automaticamente, mediante os dados fornecidos pelo(a) autor(a)

048v

Oliveira, Pamella Letícia Silva de

Vinculando perfis de engajamento ao desempenho acadêmico por meio de análise de redes sociais e análise de agrupamento nos dados de fóruns de discussão / Pamella Letícia Silva de Oliveira. - 2021.
18 f. : il.

Orientador: Rodrigo Lins Rodrigues.
Inclui referências e apêndice(s).

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal Rural de Pernambuco, Bacharelado em Sistemas da Informação, Recife, 2021.

1. Educação a distância. 2. engajamento. 3. desempenho acadêmico. I. Rodrigues, Rodrigo Lins, orient. II. Título

CDD 004

Pamella Letícia Silva de Oliveira

VINCULANDO PERFIS DE ENGAJAMENTO AO
DESEMPENHO ACADÊMICO POR MEIO DE ANÁLISE DE
REDES SOCIAIS E ANÁLISE DE AGRUPAMENTO NOS
DADOS DE FÓRUMS DE DISCUSSÃO

Artigo apresentado ao Curso de Bacharelado em Sistemas de Informação da Universidade Federal Rural de Pernambuco, como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação.

Aprovada em: 18 de Novembro de 2020.

BANCA EXAMINADORA

Rodrigo Lins Rodrigues (Orientador)
Departamento de Educação
Universidade Federal Rural de Pernambuco

Roberta Macedo Marques Gouveia
Departamento Estatística e Informática
Universidade Federal Rural de Pernambuco

Raphael Augusto de Souza
Dourado Centro de Informática
Universidade Federal de Pernambuco

Vinculando Perfis de Engajamento ao Desempenho Acadêmico por Meio de Análise de Redes Sociais e Análise de Agrupamento nos Dados de Fóruns de Discussão

Pamella Letícia Silva de Oliveira¹, Rodrigo Lins Rodrigues²

¹ Departamento de Estatística e Informática, Universidade Federal Rural de Pernambuco, Brazil

² Departamento de Educação, Universidade Federal Rural de Pernambuco, Brazil
{pamella.lesilvao, rodrigomuribec}@gmail.com

Resumo Dada a popularidade crescente da educação a distância, é importante entender como a interação social ocorre nos fóruns de discussão comumente usados nas plataformas que dão suporte a esses cursos — os Ambientes Virtuais de Aprendizagem (AVAs). Uma vez que interação e engajamento são dois fenômenos entrelaçados na aprendizagem online, e ambos influenciam o desempenho acadêmico. Neste trabalho, aplicamos a Análise de Redes Sociais (SNA) e a análise de agrupamento aos dados de fóruns de discussão para identificar grupos de alunos com diferentes perfis de engajamento. Identificamos quatro perfis e, ao analisarmos as particularidades de cada um e relacioná-los às notas dos alunos, encontramos uma ligação entre os perfis de engajamento e o desempenho acadêmico. Portanto, as características de cada perfil de engajamento identificado em nosso trabalho podem auxiliar professores e gestores na definição de estratégias para aumentar a interação dos alunos no ambiente virtual e, assim, prevenir o abandono.

Palavras-chave: educação a distância. engajamento. desempenho acadêmico.

1 Introdução

De acordo com o decreto Nº 9.057, de 25/05/2017, que regulamenta o Art. 80 da Lei 9.394/96 – Diretrizes e Bases da Educação Nacional (LDB), a educação a distância é definida como sendo: “a modalidade educacional na qual a mediação didático-pedagógica nos processos de ensino e aprendizagem ocorra com a utilização de meios e tecnologias de informação e comunicação, com pessoal qualificado, com políticas de acesso, com acompanhamento e avaliação compatíveis, entre outros, e desenvolva atividades educativas por estudantes e profissionais da educação que estejam em lugares e tempos diversos”. A educação a distância também pode ser definida como a troca interativa de dados com base na web, ferramentas e equipamentos tecnológicos avançados entre alunos e docentes distantes uns dos outros [26]. Esse conceito, em vários países, não é mais uma novidade [1]. A crescente aceitação nesta modalidade de ensino pode ser explicada devido à procura e o interesse nos benefícios que a educação a distância possui, tais como: maior alcance geográfico, flexibilidade de horário e custos acessíveis. No entanto, um dos problemas em aberto na literatura sobre esta modalidade é a baixa

interação social aluno-aluno e aluno-professor. No estudo de [12] os autores mencionam o isolamento do aluno e a falta de comunicação como limitações da educação a distância.

O engajamento é um fenômeno multifacetado, e não existe uma definição unificada na literatura para este fenômeno. Desse modo, diversos autores encara ele como um processo complexo de se identificar e mensurar [8]. Além disso, a interação desenvolve o engajamento dos alunos [17]. Visto que, a interação, o diálogo e a colaboração são fatores que determinam a natureza da aprendizagem, uma vez que a qualidade dos cursos a distância está diretamente relacionada a esses fatores [10].

Uma técnica que busca entender e mensurar os aspectos de interação é a Análise de Redes Sociais, conhecida na língua inglesa como *Social Network Analysis* ou (SNA). Este é um campo de estudo que investiga a interação e os padrões de interação entre os sujeitos. Esses padrões, quando mapeados, permitem uma análise detalhada e uma visão aprofundada de grupos de indivíduos [16], e têm como foco identificar as relações entre os atores dos grupos [22]. Segundo [27] a SNA oferece um potente conjunto de conceitos, procedimentos e medidas baseadas em algoritmos relativamente padronizados para a descrição e a explicação das estruturas de participação e de interação. Corroborando com essas afirmações, [9] afirmam que a partir desse tipo de análise é possível compreender e acompanhar de forma mais eficaz a disseminação de informações na rede, a interação entre as pessoas e a identificação de relacionamentos.

Outra técnica muito utilizada é a análise de agrupamento que tem como objetivo principal, encontrar padrões que se agrupam e permitem classificar os dados em diferentes grupos ou categorias inicialmente desconhecidas. Visto que, os padrões são automaticamente identificados através da manipulação das características dos dados [14].

Embora estudos utilizem essas técnicas para entender a interação entre os pares, ainda é perceptível na literatura, a carência de trabalhos que apresentem como a interação entre os alunos pode impactar no engajamento e, como consequência, influenciar o desempenho acadêmico dos alunos. Diante disso, o objetivo desta pesquisa é responder às seguintes questões: 1) É possível encontrar diferentes perfis de engajamento nos fóruns de discussão, por meio das métricas de SNA?; 2) Quais são as relações entre os perfis de engajamento e o desempenho acadêmico dos estudantes?. Para isso, este artigo está dividido em outras cinco seções: entendimento do problema, trabalhos relacionados, metodologia, resultados e discussões.

2 Engajamento e Análise de Redes Sociais

A literatura apresenta três categorias de definições para o engajamento educacional: (1) o cognitivo, que está relacionado à ideia de investimento, ou seja, fazer esforço para entender ideias complexas e desenvolver habilidades difíceis; (2) o comportamental, costuma ser mensurado a partir de aspectos quantitativos em relação às ações dos estudantes; e (3) o emocional, o qual está associado ao sentimento de pertencer a um grupo [20].

Outro ponto a destacar é que na literatura muitas vezes o engajamento está associado com o desempenho do aluno. Segundo [7], pesquisadores estão focando sua atenção no engajamento do aluno porque há uma correlação com os resultados positivos do sucesso dos alunos. Um exemplo disso, é o estudo de [18] onde utilizou o algoritmo Apriori

em seu experimento, e constatou que os alunos com níveis mais altos de engajamento tendem a ter um melhor desempenho no curso.

Além disso, a colaboração é um dos aspectos sociais considerados para identificar o engajamento dos estudantes em atividades educacionais. Visto que a colaboração pode ser medida pela melhora na satisfação, no engajamento, no aprendizado e na qualidade do envolvimento dos alunos [17]. A colaboração possui um papel fundamental no desenvolvimento e na aprendizagem do grupo, pois por meio da reflexão e da discussão em grupo os alunos constroem seus conhecimentos [6]. Ademais, o nível de colaboração tem sido associado à demonstração de habilidades do pensamento crítico [3]. Em [13], por exemplo, foram observadas correlações positivas entre o número de perguntas dos alunos enviadas para o instrutor e suas notas finais. Desse modo, pode-se notar a relação do engajamento na aprendizagem, no que se refere ao desempenho acadêmico, aquisição de conhecimento e motivação.

Para mapear as relações e as interações entre os agentes dentro de grupos em ambientes de participação muitos estudos utilizam a análise de redes sociais [25]. O uso desta técnica em contextos educacionais têm sido evidenciada na literatura em razão da necessidade dos pesquisadores em descobrir o nível de interação entre alunos e aluno-professor. Alguns estudos que utilizaram essa técnica foram: Em [10] para encontrar os padrões de interação em fóruns de discussão. Os autores argumentaram que essas informações, podem servir de apoio aos gestores ou coordenadores de cursos para analisar aspectos comportamentais dos educadores quanto ao envolvimento e interação dos alunos no desenvolvimento do curso; Em [24], onde os autores buscaram verificar se as relações que ocorrem no Ambiente Virtual de Aprendizagem (AVA) têm o professor como principal mediador. Para isso, a metodologia empregada no estudo consistiu na obtenção de métricas de centralidade aplicadas à análise de rede social, registrada por meio das atividades e interações dos usuários no AVA. Além disso, o grau de densidade evoluiu, evidenciando que a motivação para a discussão e troca de mensagens nos fóruns se manteve ao longo da disciplina. Levando em conta que a turma estudada utilizou o AVA como um apoio às aulas presenciais. Logo, percebeu-se que tal peculiaridade contribuiu para a descentralização do professor pois mesmo com a presença (apenas virtual) do professor houve interação entre pares; [2] utilizou para elaborar um sistema capaz de acompanhar as interações nos fóruns; no estudo de [4] foi utilizado técnicas de SNA em conjunto com o algoritmo de predição *Random Forest* para prever e melhorar o desempenho dos alunos, a fim de trabalhar na motivação e acompanhamento individualizado, seja por um professor-tutor em um pequeno curso, como por um sistema automatizado para um curso massivo.

A análise de agrupamento também é mencionada na literatura para auxiliar os professores na melhoria da metodologia de ensino. Por exemplo, em [21], por meio da análise de agrupamento foi encontrado diferentes perfis de engajamento que ofereceu informações particularidade de cada grupo. Essas informações foram analisadas para oferecer a coordenação ou aos professores como ajustar o curso evitando reprovações e desistências. Em [23] os autores em seus experimentos utilizaram a análise de agrupamento para analisar a participação dos alunos em fóruns de discussão, e concluíram que essa análise é importante para identificar grupos com características comportamentais distintas que muitas vezes são imperceptíveis na visão da gestores e professores.

3 Metodologia

A metodologia utilizada seguiu um processo iterativo composto por oito fases. Na Figura 1 são apresentadas as fases e suas relações durante o fluxo de execução.

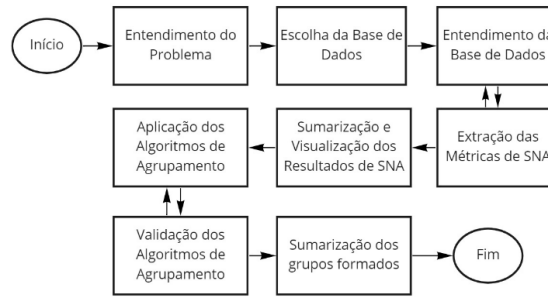


Figura 1: Fluxograma da Metodologia

As seções seguintes descrevem em detalhes cada uma, das fases da pesquisa.

3.1 Entendimento do Problema

Para a compreensão das duas questões desta pesquisa, foi necessário buscar artigos na literatura que tratassem de análise de rede social e de engajamento. Diante disso, foi feita uma leitura na revisão sistemática desenvolvida por [9]. Nessa revisão, os autores investigaram e reuniram de forma sistemática os principais trabalhos científicos já publicados que versam sobre SNA em sistemas *E-learning*. Com isso, os autores conseguiram alcançar contribuições em relação: as interações que se formam em um ambiente *E-Learning*, os estudos publicados acerca desta temática e as possibilidades de pesquisa na área. Em seguida, foi feita a leitura de uma revisão sistemática sobre engajamento, o artigo escolhido foi de [20], onde os autores extraíram o total de onze artigos na revisão, que foram lidos para obter as respostas das cinco questões de pesquisa, que foram: “Quais são os tipos de engajamento em plataformas LMS?”; “Quais são as variáveis que representam o fenômeno de engajamento em plataformas LMS?”; “Quais são as formas de mensurar engajamento em plataforma LMS?”; “Quais as técnicas para visualizar o engajamento?”; e “Por quais motivos o estudo de performance é relacionado à engajamento?”

3.2 Origem dos dados

A coleta de dados foi realizada com os dados do curso de biologia na modalidade a distância, ofertada pela Universidade de Pernambuco. As disciplinas foram disponibilizadas por meio do ambiente Moodle do inglês (*Modular Object-Oriented Dynamic Learning Environment*). Os dados coletados são: as notas das atividades *webquest*, dos fóruns de discussão, das provas presenciais e a interação que ocorreu no chat e nos

fóruns de discussão. O critério de inclusão dos sujeitos adotado no estudo leva em consideração o fato dos mesmos terem postado nos quatro fóruns da disciplina. O total de estudantes incluídos foi de seiscentos e dezesseis, os quais estavam distribuídos em quinze disciplinas, sendo: seis disciplinas do primeiro período, quatro do quinto período e cinco do oitavo período do semestre letivo 2011.1 até 2016.1.

3.3 Extração e entendimento dos dados

Esta seção é composta por cinco etapas: importação da base de dados, entendimento da estrutura relacional, verificação das variáveis relacionadas a interação; escolha do curso e das disciplinas e montar a base de interação.

Na primeira etapa foi utilizado a plataforma *MYSQL Workbench* (ferramenta visual de design de banco de dados que integra desenvolvimento, administração, design, criação e manutenção de SQL em um único ambiente de desenvolvimento integrado para o sistema de banco de dados *MySQL*) para importar a base de dados, ao todo foram importados quatrocentos e sessenta e uma tabelas e quatro mil cento e quarenta e sete colunas. A segunda etapa, se deu pelo entendimento da estrutura relacional de todas as tabelas da base. Em seguida, na terceira etapa, foi possível encontrar as variáveis que estavam associadas com a interação dos alunos e o relacionamento entre elas, como mostra a Figura 2.

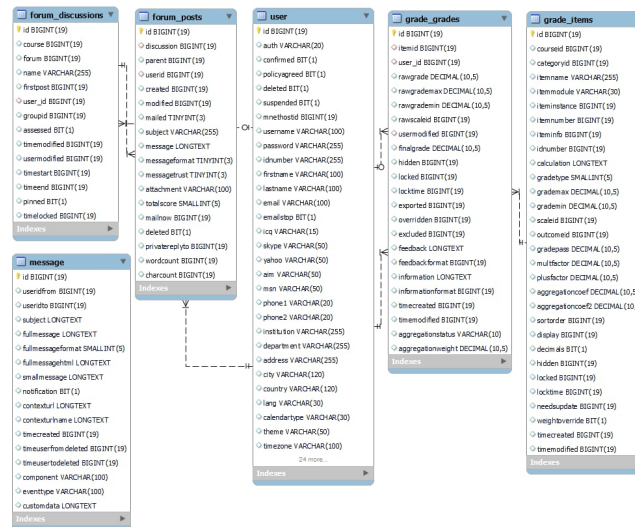


Figura 2: Modelo relacional das tabelas de sumarização

Ao todo seis tabelas foram utilizadas, são elas: *forum.discussion*, que contém as informações sobre os fóruns; *forum_posts*, onde são armazenadas todas as postagens; *grade_grades*, esta tabela mantém notas individuais para cada usuário e cada item, exatamente como importado ou enviado pelos módulos; *grade_items*, esta tabela mantém

Tabela 1: Métricas de *SNA* no contexto educacional.

Métricas	Importância no Contexto Educacional
<i>Indegree</i>	Mede quantas conexões o nó (estudante) recebeu, desta forma é possível perceber a sua popularidade. Além disso, os estudantes que recebem mais conexões, são vistos como líderes.
<i>Outdegree</i>	Mede quantas conexões o nó (estudante) fez, mostrando a contribuição de cada estudante para com os demais.
<i>Degree</i>	Por meio desta métrica é possível ver se a interação está centralizada em um pequeno grupo de estudantes, ou distribuída uniformemente na rede.
<i>Closeness</i>	Mede a distância entre os nós (estudantes) da rede. Logo, os que possuem valores mais altos, são considerados isolados e que receberam pouca informação e influência da rede.
<i>Betweenness</i>	Mostra os estudantes que são responsáveis por distribuir informações entre os nós (estudantes) adjacentes, e com isso, conectar vários grupos.

3.5 Aplicação e Validação dos Algoritmos de Agrupamento

Após a extração das métricas de *SNA* foi utilizada a técnica de análise de agrupamento para identificar os perfis de engajamento. Esta técnica tem como objetivo principal encontrar padrões que se agrupam e permitir classificar os dados em diferentes grupos ou categorias inicialmente desconhecidas. Esses padrões são automaticamente identificados através da manipulação das características dos dados. [14].

Inicialmente, para a aplicação dessa técnica foi utilizada a estatística *Hopkins* para encontrar o melhor conjunto de dados. Em seguida, foram utilizadas duas abordagens para encontrar o número ideal de *clusters*: o método *Elbow* e o hierárquico. O método *Elbow* indica o número ideal dos grupos que é representado por um “cotovelo” mostrado no gráfico. Já, a análise de agrupamento hierárquico teve como objetivo encontrar, através da geração de um dendrograma, uma quantidade *k* de grupos, até então desconhecidos. Por fim, foram usados validações de agrupamento interna e externa, os quais são importante para encontrar padrões em dados aleatórios, bem como comparar agrupamentos de dois algoritmos [15].

4 Resultados e Discussões

Nesta seção, apresentamos os resultados da *SNA* e da análise de cluster. As subseções 4.1 a 4.4 detalham os resultados de cada etapa em nosso processo de análise e como cada uma delas se baseia nos resultados da anterior. A subseção 4.5 discute as relações que encontramos entre os perfis de engajamento e o desempenho acadêmico. Finalmente, a subseção 4.6 caracteriza os quatro perfis de engajamento identificados em nosso trabalho.

4.1 Extração das Métricas de Análise de Redes Sociais

Após a extração dos dados e a identificação dos valores referentes as métricas de *SNA* para cada aluno. Também foi possível encontrar os valores estatísticos das médias das

métricas, e entender quais entre elas obteve melhores resultados, como mostrado na Tabela 2.

Tabela 2: Valores médios para as métricas de SNA

Estatística	<i>Indegree</i>	<i>Outdegree</i>	<i>Degree</i>	<i>Closeness</i>	<i>Betweenness</i>
Média	1.062	1.000	3.012	0.471	0.677
Mediana	0.750	1.000	2.000	0.500	0.000
Desvio Padrão	1.501	2.757	2.867	0.399	3.417
Mínimo	0.000	0.000	0.250	0.000	0.000
Máximo	10.000	19.250	19.250	1.100	41.750

A Tabela 2 mostra a média, mediana, desvio padrão, estatísticas mínimas e máximas para as métricas SNA calculadas para cada aluno em nosso conjunto de dados. Comparando os valores obtidos para média e máximo pode-se considerar que os valores de média são baixo e os de máximo são alto. Logo, os baixos valores médios para as métricas *indegree*, *outdegree*, *degree* e *betweenness* indicam que a maioria dos alunos não interagem muito com o ambiente. Por outro lado, os valores máximos elevados para essas mesmas métricas revelam que alguns alunos apresentam altos níveis de interação com outros alunos. Por isso, existe poucos alunos isolados, como é refletido nos valores baixos para a métrica *closeness* e muitos alunos que se comunicam entre diferentes grupos como mostra o valor mais alto de máximo para *betweenness*.

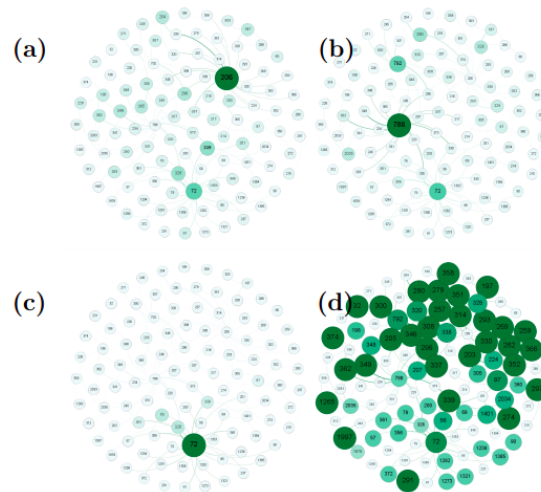


Figura 4: Representação em grafos das métricas de SNA

A Figura 4 mostra, a título de exemplo, a representação gráfica da interação que ocorreu em uma disciplina para as quatro métricas de SNA: (a) *indegree*, (b) *outdegree*,

(c) *closeness* e (d) *betweenness*, gerada com *Gephi*. Cada nó representa um aluno único, cada aresta representa a interação entre dois alunos, o tamanho e a cor dos nós representam o nível de interação do aluno, ou seja, quanto mais verde e maior o nó, maior a interação do aluno. A figura 4 (a) e (b) reforçam a não uniformidade nos níveis de interação dos alunos mostrados na Tabela 2: enquanto alguns alunos interagem muito (os nós mais escuros), a maioria dos outros mostram baixos níveis de interação (os nós mais claros). O grafo (c) devido a pouca quantidade de nós com cores mais escura mostra que existem poucos alunos isolados. Já o grafo (d) apresenta grande quantidade de nós escuros mostrando que os alunos estão se comunicando entre diferentes grupos.

4.2 Clusterização das Métricas de Análise de Redes Sociais

Após a extração e visualização dos grafos formados pelas métricas de *SNA*, buscamos encontrar quais diferentes agrupamentos (*clusters*) de perfis de engajamento podem ser formados com as interações dos fóruns. Para isso, foram utilizados os processos de agrupamentos que serão apresentados nesta seção, que inclui: a validação da tendência de agrupamento; a determinação da quantidade ideal de grupos; a aplicação das medidas internas e externas para validação dos agrupamentos; e a geração da representação visual da dispersão dos grupos.

Primeiro, usamos a estatística de *Hopkins* para encontrar o melhor conjunto de métricas para formar os agrupamentos. Essa estatística avalia a tendência de agrupamento de um conjunto de dados. Logo, cinco combinações considerando os quatro fóruns das quinze disciplinas foram testadas: I. foram utilizadas as métricas: “*indegree*”, “*outdegree*”, “*degree*”, “*closeness*” e “*betweenness*”; II. foram utilizadas as métricas: “*indegree*”, “*outdegree*”, “*degree*”, “*closeness*”; III. foram utilizadas a média das métricas: “*indegree*”, “*outdegree*”, “*degree*”, “*closeness*” e “*betweenness*”; IV. foram utilizadas a média das métricas: “*indegree*”, “*outdegree*”, “*degree*”, “*closeness*”; V foram utilizada a combinação das cinco médias das métricas mais a média das notas nas provas, nas atividades webquest e nos fóruns. Os resultados de cada combinação para a estatística *Hopkins* estão apresentadas na Tabela 3.

Tabela 3: Estatística *Hopkins* para às cinco combinações das métricas.

Métricas	Estatística <i>Hopkins</i>
I	0,07
II	0,08
III	0,10
IV	0,05
V	0,11

Segundo [15], se o valor do *Hopkins* for próximo a zero, significa que o conjunto de dados possui *clusters* significativos. Diante dos resultados da Tabela 3 optou-se por seguir a análise de cluster utilizando a combinação IV, que é o valor mais próximo a zero.

Em seguida, foram aplicados dois métodos para encontrar o número ideal de *clusters*, são eles: o *Elbow* e o Hierárquico. O método *Elbow* testa a variância dos dados em relação ao número de clusters, até o momento em que um aumento do número de *clusters* não representa um valor significativo de ganho [21]. O resultado deste método mostrou para o conjunto de dados desta pesquisa, o número ideal de *clusters* é quatro, como mostra a Figura 5.

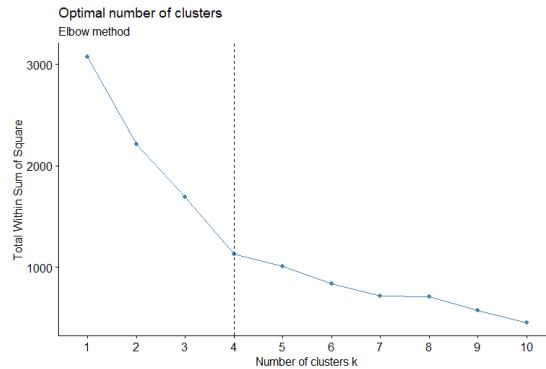


Figura 5: Gráfico gerado pelo método *Elbow*, sendo $K=4$ o número ideal de *clusters*.

Já o método hierárquico consiste em uma série de sucessivos agrupamentos ou sucessivas divisões de elementos, nas quais os elementos são agregados ou desagregados visando construir uma hierarquia de *clusters*. O resultado do agrupamento hierárquico é representado através de uma árvore de *clusters*, também chamada de Dendograma [14], que mostra a formação dos *clusters*. Como resultados desta pesquisa foram encontrado quatro *clusters* principais, como mostra o dendograma da Figura 6.

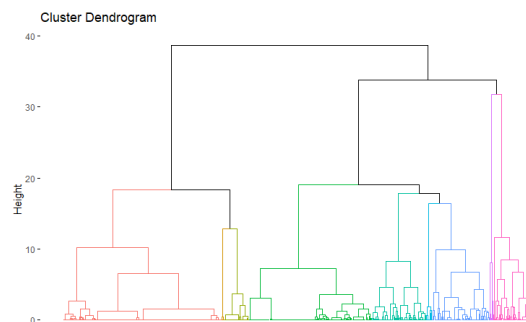


Figura 6: Gráfico gerado pelo método hierárquico, sendo $K=4$ o número ideal de *clusters*.

4.3 Medidas de Validação de Agrupamento

Para avaliar a qualidade de nossos clusters, executamos dois tipos de validação: interna e externa. A interna, que é utilizada para estimar o número de *clusters*; e a externa que é utilizada principalmente para selecionar o algoritmo de agrupamento correto para um conjunto de dados específico.

O índice *Dunn* é uma validação interna, usada para identificar *clusters* compactos e bem separados. Para isso, a distância entre os aglomerados deve ser grande e o diâmetro dos aglomerados deve ser pequeno. Assim, grandes valores do índice indica a presença de aglomerados compactos e bem separados [11]. O coeficiente de silhueta é uma validação de *clusters* onde quando o valor está próximo a 1 indica um bom agrupamento. No entanto, um valor próximo a -1 indica que o objeto está mal agrupado.

Logo, foram analisadas essas validações para determinar o número ideal de *clusters*. Os testes foram feitos utilizando o algoritmo K-Means e com valores entre 2 e 6 para K, como apresentado na Tabela 4.

Tabela 4: Índices Dunn e os coeficientes de silhueta.

K=	Índice Dunn	Coefficiente de Silhueta
2	1,19	0,49
3	0,63	0,45
4	0,56	0,49
5	0,52	0,50
6	0,64	0,52

O valor seis para K apresentou o melhor resultado para o coeficiente de silhueta e o segundo melhor para o índice de *Dunn*. No entanto, os resultados do método hierárquico e do *Elbow* indicaram um número ótimo de quatro *clusters*. Portanto, foi necessário executar uma validação externa para decidir entre usar 4 ou 6 *clusters*. Logo, foi necessário utilizar a validação externa para decidir se o melhor agrupamento é com quatro ou seis grupos. O índice *Rand* é uma validação externa que fornece uma medida para avaliar a semelhança entre duas partições ajustadas ao acaso. Seu intervalo é -1 (sem concordância) a 1 (perfeito acordo) [15]. assim, ele foi utilizado para verificar a diferença do agrupamento com valores de quatro e seis para K. O teste retornou o valor de 0.94 indicando ter um perfeito acordo entre esses agrupamentos, o que possibilita a escolha de qualquer um dos dois.

4.4 Representação Gráfica dos Agrupamentos

Por fim, como o valor de índice *Rand* indicou um acordo entre o agrupamento com quatro e seis grupos, então também foi analisado o balanceamento dos agrupamentos. A divisão dos agrupamentos composto por quatro grupos de alunos ficou dividido em: primeiro grupo, com duzentos e oitenta e dois; segundo grupo, com oitenta e nove; terceiro grupo, com treze; e o quarto grupo, com duzentos e trinta e dois. Já para o

agrupamento composto por seis grupos ficou dividido em: primeiro grupo, com duzentos e vinte e cinco; segundo grupo, com oito; terceiro grupo com duzentos e setenta e um; quarto grupo, com vinte e dois; quinto grupo, com oitenta e cinco; e o sexto grupo, com cinco. Logo, o que mostrou um melhor balanceamento foi o agrupamento com quatro grupos, então ele foi escolhido para seguir com o estudo.

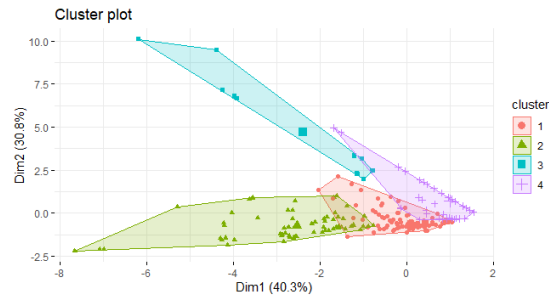


Figura 7: Gráfico gerado pelo algoritmo *K-Means*, sendo $K=4$.

A escolha do algoritmo *K-Means* se deu por ser considerado o algoritmo não hierárquico mais utilizado e quando comparado ao método hierárquico, este método é mais rápido [2]. O gráfico da Figura 7 mostra o agrupamento formado pelo algoritmo *K-Means*. Nele é possível perceber que os quatro grupos possuem pequenas interseções entre si. No entanto, a maioria das observações estão distantes de outros grupos, indicando realmente pertencer ao grupo em que se encontram.

4.5 Relação das Métricas de SNA com o Desempenho Acadêmico

Nesta seção, analisamos o comportamento dos clusters em relação às métricas *SNA* e as notas médias dos alunos nas atividades *webquest*, provas e nas postagem dos fóruns. Esses resultados são apresentados nas Figuras 8 e 9 por meio de *boxplots*, que permitem a identificação de tendências centrais, variabilidade e *outliers*.

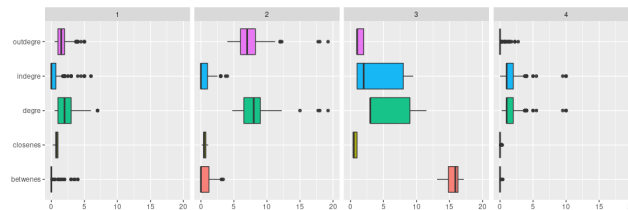


Figura 8: Comportamento dos grupos para as métricas.

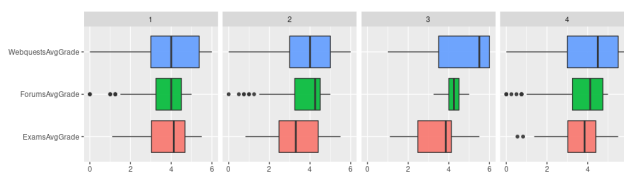


Figura 9: Comportamento dos grupos para as média das notas.

Observando as métricas de cada cluster nas Figuras 8 e 9, podemos identificar várias características distintas. No cluster 1, tanto as métricas *SNA* quanto os valores médios das notas coincidem com a mediana, indicando que os alunos pertencentes a este grupo têm um padrão de interação irregular, o que leva a notas também irregulares. O cluster 2 mostra bons resultados medianos para as métricas *outdegree* e *degree*, bem como para as notas médias; no entanto, as medianas de *indegree* e *ExamsAvgGrade* não são tão boas, o que sugere que os alunos neste grupo interagem de forma consistente nos fóruns, mas outros alunos não interagem de volta com eles, explicando os baixos valores da métrica *ExamsAvgGrade*. O cluster 3 se destacou entre os quatro grupos, por ser o único que não possui *outlier* e apresenta o maior valor mediano para a *betweenness* e *WebquestsAvgGrade*; isso significa que os alunos deste grupo interagem com grupos diferentes, construindo pontes entre eles, o que resultou nas boas notas mostradas para este grupo. Por fim, o cluster 4 apresentou os piores resultados, com valores medianos de zero para três das cinco métricas de interação e muitos outliers nas métricas de notas (Figura 9); portanto, os alunos deste grupo apresentam os níveis de interação mais baixos e alta variação em suas notas.

Analisar como acontece a interação dentro dos fóruns de discussão e identificar os diferentes perfis de grupos é essencial para que os professores possa adotar metodologias baseadas no perfil de cada grupo. Corroborando, com essa afirmação em [28], os autores afirmam que é importante estabelecer um processo relacional em que o professor e o tutor desenvolvam estratégias que garantam a interação dos alunos de forma ativa e autônoma. Além disso, os autores também mencionam, a importância da interação social uma vez que parece ser esse um dos fatores mais importantes nos cursos on-line.

Além disso, nesta pesquisa pode-se constatar que existe uma relação entre o engajamento, a interação e o desempenho acadêmico. Essa relação é confirmada em outros estudos, por exemplo, em [5], os autores afirmam que pesquisas descobriram que os comportamentos de certos alunos em um LMS podem ser associado ao desempenho acadêmico e engajamento dos alunos. Já em [19] também confirma que os estudos sugerem que fatores que afetam os resultados educacionais são: a qualidade do ensino, o engajamento e a motivação do aluno. Sendo assim, a identificação dos diferentes perfis de engajamento é essencial para analisar as características particulares de cada grupo, e fornecer importantes informações que possam prevenir possíveis evasões. Corroborando com essa afirmação, no estudo de [23], foram utilizados testes estatísticos para identificar grupos de engajamento, que muitas vezes são imperceptíveis na visão de gestores e professores.

4.6 Identificação dos Perfis de Engajamento

A Tabela 5 descreve os quatro perfis de engajamento identificados neste trabalho por meio da interação dos alunos em fóruns, além das características particulares de cada grupo. Em geral, pode-se concluir que o grupo 1, os “esporádicos”, são alunos que interagem irregularmente nos fóruns. O grupo 2, “socializador”, representa os alunos mais ativos nos fóruns, aqueles que interagem regularmente. O grupo 3, “integrador”, são os alunos que interagem com diferentes perfis de alunos, construindo pontes na turma. Por fim, o grupo 4, “isolado”, é o que apresenta menor nível de participação.

Tabela 5: Perfis de engajamento em fóruns de discussão.

Grupo	Nome	Características Educacionais
1	Esporádicos	É um grupo composto por alunos que utilizam o recurso do fórum com uma periodicidade irregular. Os alunos deste grupo provavelmente procuram interagir apenas quando têm dúvidas. Então, eles não são isolados porque existe interação em algum momento, no entanto, não buscam interagir muito. Desse modo, eles possuem notas médias que é reflexo da irregularidade na interação.
2	Socializador	É um grupo bastante ativo nos fóruns. Os alunos pertencentes a este grupo, são muito comunicativo e os que mais procuram interagir com os demais alunos, mesmo quando não há uma troca por parte do outro aluno. Então, por mais que eles busquem interagir os outros não correspondem da mesma forma. É possível que estes alunos interajam em busca de sanar dúvidas com os colegas de turma, mas como não há essa troca mútua, os alunos deste grupo apresenta grande variação nas notas.
3	Integrador	Os alunos pertencentes a este grupo têm a capacidade de interagir em grupos com perfis distintos de interação, formando uma ponte entre os grupos. Com isso, esses alunos conseguem obter conhecimento de diferentes formas e contribuem com a colaboração nas discussões geradas nos fóruns. Esta maior aquisição de aprendizagem, acarreta nas notas mais altas em relação aos demais grupos, pois mesmo os gráficos de desempenho sendo parecidos, o desse grupo parece ser o melhor, especialmente, para as atividades webquest.
4	Isolado	Este é o grupo com menor nível de engajamento nas discussões nos fóruns. Os alunos deste grupo são os que menos buscam interagir, no entanto, alguns são requisitados por outros alunos em busca de interação. Logo, é o grupo que apresenta muita variação para notas baixas.

Em suma, esses resultados podem mudar a percepção dos educadores sobre como ocorre a interação nos fóruns, o que pode influenciar a forma como planejam e monitoram essas atividades de aprendizagem em busca de melhorar a interação e o engajamento dos alunos. Além disso, analisar as particularidades dos diferentes perfis de engajamento pode ajudar na adoção de intervenções pedagógicas personalizadas para cada grupo.

5 Conclusões

Este estudo teve como objetivo examinar a interação em fóruns de discussão utilizando a análise de redes sociais para encontrar os valores das métricas de *indegree*, *outdegree*, *degree*, *closeness* e *betweenness*. Isso nos permitiu constatar que a interação nos fóruns não ocorre de forma uniforme. Em seguida, usamos a análise de *cluster* nas métricas SNA para identificar os diferentes perfis dos alunos, que revelaram quatro grupos: esporádicos, solicializador, integrador e isolado.

Além disso, buscamos encontrar a relação entre perfis de engajamento e desempenho acadêmico. Para isso, analisamos o comportamento de cada grupo e as notas que obtiveram nas atividades do curso, o que revelou uma relação entre as métricas e as notas de SNA. Os alunos do grupo 'Integrador' são vistos como pontes entre os grupos e obtêm as melhores notas. O grupo 'Socializador' representa os alunos mais comunicativos, embora haja uma grande variação nas notas. O grupo 'Esporádicos' mostra uma participação irregular nos fóruns e notas médias. Finalmente, o grupo de alunos 'Isolado' é aquele com os níveis de interação mais baixos e também com notas baixas.

De maneira geral, este estudo foi capaz de mostrar que 1) é possível encontrar diferentes perfis de engajamento por meio de métricas SNA, e 2) existe uma relação entre engajamento e desempenho acadêmico. Portanto, o conhecimento desses diferentes perfis de envolvimento pode ajudar discentes e docentes. Onde os alunos podem ser beneficiados pois irão identificar em qual grupo pertencer e monitorar o seu engajamento. Para professores e a instituição será benéfico para aplicar metodologia a fim de melhorar o curso e agir de acordo com o perfil de cada aluno.

Como trabalho futuro, pretendemos aplicar as mesmas técnicas de análise utilizadas neste trabalho para outros conjuntos de dados (como outros cursos ou programas) e também testar outras técnicas de EDM para descobrir novos comportamentos ou prever o comportamento dos alunos.

Referências

1. Alaneme, G.C., Olayiwola, P.O., Reju, C.O.: Combining traditional learning and the e-learning methods in higher distance education: Assessing learners' preference. In: 2010 4th International Conference on Distance Learning and Education. pp. 187–190 (2010)
2. André, S., Álvaro, F.: Depicting online interactions in learning communities. In: Proceedings of the 2012 IEEE Global Engineering Education Conference (EDUCON). pp. 1–8 (2012)
3. Blanchette, J.: Participant interaction in asynchronous learning environments: Evaluating interaction analysis methods. *Linguistics and Education* **23** (2012)
4. Cabeza, S.G., Arredondo, A.L., Massafiero, P., Rubido, N., Hirt, A.M.: Redes profesionales en procesos educativos en línea. In: Proceedings of the 2nd Latin American Conference on Learning Analytics (2019)
5. Celis, S., Lopez, D., Silva, J.: Analyzing the influence of online behaviors and learning approaches on academic performance in first year engineering (03 2019)
6. da Rocha Seixas, L., Gomes, A.S., de Melo Filho, I.J.: Effectiveness of gamification in the engagement of students. *Computers in Human Behavior* **58**, 48–63 (2016)
7. e Jorge Figueiredo e Dalton Serey, M.M.: Online student engagement: A case study in teaching of programming. *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação - SBIE)* **30**(1) (2019)

8. Ding, L., Er, E., Orey, M.: An exploratory study of student engagement in gamified online discussions. *Computers and Education* **120**, 213–226 (2018)
9. Engel, A., Coll, C., Bustos, A.: Distributed teaching presence and communicative patterns in asynchronous learning name versus reply networks. *Computers and Education* pp. 184–196 (2007)
10. Gottardo, E., Noronha, R.V.: Social networks applied to distance education courses. Proceedings of the 18th Brazilian symposium on Multimedia and the web - WebMedia '12 (2012)
11. Halkidi, M., Batistakis, Y., Vazirgiannis, M.: On clustering validation techniques. *Journal of Intelligent Information Systems* **17** (2001)
12. Hasan, K., Ayça, Ç., Memnune, P.: Student opinions about the period of measurement and evaluation in distance education: the difficulties. *Procedia - Social and Behavioral Sciences* **9**, 1597–1601 (2010)
13. He, P.: Evaluating students online discussion performance by using social network analysis. 2012 Ninth International Conference on Information Technology - New Generations pp. 854–855 (2012)
14. Jorge, R., Rodrigo, R., João, S., Alex, G., Ricardo, S.: A comparative study between clustering methods in educational data mining. *IEEE Latin America Transactions* **14**, 3755 (2016)
15. Kassambara, A.: Practical Guide To Principal Component Methods in R. *STHDA* (2017)
16. de Lima, D.P.R., Netto, J.F.d.M., Bremgartner, V.: Applying social network analysis in a course supported by a lms: Report of a case study. In: 2017 IEEE Frontiers in Education Conference (FIE). pp. 1–9 (2017)
17. Messias, I., Morgado, L., Barbas, M.: Students' engagement in distance learning: Creating a scenario with lms and social network aggregation pp. 44–49 (2015)
18. Moubayed, A., Injadat, M., Shami, A., Lutfiyya, H.: Relationship between student engagement and performance in e-learning environment using association rules. pp. 1–6 (2018)
19. Namoun, A., Alshaqiti, A.: Predicting student performance using data mining and learning analytics techniques: A systematic literature review. *Applied Sciences* **11** (2021)
20. Oliveira, P.L.S.D., Souza, A.J.D., Rodrigues, R.L.: Identificação de pesquisas referentes ao engajamento de alunos em plataformas de lms e suas relações com o desempenho acadêmico, anais do xxx simpósio brasileiro de informática na educação (sbie 2019) (2019)
21. Oliveira, P., Rodrigues, R., Ramos, J., Silva, J.: Uma análise de algoritmos de clusterização para descoberta de perfis de engajamento. In: Anais do XXXI Simpósio Brasileiro de Informática na Educação. pp. 1012–1021 (2020)
22. Recuero, R., Bastos, M., Zago, G.: Análise de Redes para Mídia Social. Editora Sulina (2018)
23. Rodrigues, R., Ramos, J., Silva, J., Gomes, A.: Discovery engagement patterns moocs through cluster analysis. *IEEE Latin America Transactions* **14**(9), 4129–4135 (2016)
24. Sacerdote, H.C.D.S., Fernandes, J.H.C.: Investigando as interações em um ambiente virtual de aprendizagem por meio da análise de redes sociais. *InCID: Revista de Ciência da Informação e Documentação* **4**, 129–146 (2013)
25. Saqr, M., Fors, U., Tedre, M.: Como o estudo da aprendizagem colaborativa online pode orientar professores e prever o desempenho dos alunos em um curso de medicina. *BMC Med Educ* **18** **24** (2018)
26. Semih, C., Sibel, S., Deniz, O.: Determining student satisfaction in distance education courses. *Procedia Computer Science* **120**, 529–538 (2017)
27. Silva, J.C., Brito, A.V., Medeiros, F.: Mapeamento sistemático da literatura acadêmico-científica sobre análise de redes sociais aplicada em e-learning. *Revista Brasileira de Informática na Educação* (2015)
28. Wander, B., Gomes, M.Q., Pinto, M.E.B.: Avaliação da interação em fóruns de discussão na especialização de preceptoria em medicina de família e comunidade a distância. *Interface - Comunicação, Saúde, Educação [online]* **24** (2020)