

Como Formar Grupos em Ambientes Virtuais de Aprendizagem? Uma Abordagem Híbrida Utilizando Algoritmos Genéticos e Algoritmos de Agrupamento

Milnner Kauan T. Andrade¹, Vinícius A. Silva¹, Hiran Nonato M. Ferreira¹

¹Instituto Federal de Educação, Ciência e Tecnologia do Sul de Minas Gerais
(IFSULDEMINAS) - Campus Passos
Passos – MG – Brasil

milnner.andrade@alunos.ifsuldeminas.edu.br
{vinicius.silva, hiran.ferreira}@ifsuldeminas.edu.br

Abstract. According to the "Censo de Educação Superior de 2022", there was an increase of 189.1% in enrollments in Distance Education between the years 2018 and 2022. In this context, classes can be larger, as there is no physical limitation imposed by a classroom, and the necessary interactions for teaching are facilitated by computing. This work studies another computational facility that can be inserted into the academic context. Therefore, a methodology that combines techniques, WMH K-means and Genetic Algorithms (GA), is proposed to create solutions that seek to optimize relationships among students. The approach was tested with probabilistically generated data and analyzes the hybrid method in relation to an approach solely using GA. The goal is to enable a more efficient and productive learning environment.

Resumo. Segundo o Censo de Educação Superior de 2022, houve uma evolução de 189,1% entre os anos de 2018 e 2022 nas matrículas na Educação a Distância (EaD). No contexto da EaD, as turmas podem ser mais volumosas, não existindo uma limitação física imposta por uma sala e as interações necessárias para o ensino sendo facilitadas pela computação. Este trabalho estuda mais uma facilidade computacional, que pode ser inserida no contexto acadêmico, tanto em ambientes EaD quanto presenciais. Portanto, propõe-se uma metodologia que combina técnicas, o WMH K-means e Algoritmos Genéticos (GA), para criar soluções que buscam otimizar as relações entre os estudantes. A abordagem foi testada com dados gerados probabilisticamente e faz uma análise do método híbrido em relação a uma abordagem apenas por GA. Com o objetivo de possibilitar um ambiente de aprendizado mais eficiente e produtivo.

1. Introdução

A formação de grupos em ambientes escolares é uma prática essencial que influencia significativamente o aprendizado e o desenvolvimento dos estudantes. A composição de equipes de trabalho pode potencializar o engajamento, promover a colaboração e desenvolver habilidades sociais importantes para o contexto educacional dos estudantes [Qureshi et al. 2023]. Ao formar grupos para o trabalho colaborativo, é importante definir quem será responsável por selecionar os membros. No entanto, formar equipes

de maneira eficiente e equitativa é um desafio complexo que exige a consideração de múltiplos fatores, como habilidades individuais, compatibilidade entre membros, e objetivos educacionais [Lappas et al. 2009].

O Problema de Formação de Equipes (do inglês, *Team Formation Problem*, ou TFP) refere-se à tarefa de organizar um conjunto de indivíduos em grupos que atendam a determinados critérios de desempenho e coesão. Este problema tem sido amplamente estudado em diversas áreas, como redes sociais, gestão de projetos, engenharia de software e contextos educacionais [Lappas et al. 2009, Awal and Bharadwaj 2014, Rubin and Bai 2015, Borges et al. 2023]. Em ambientes escolares, a formação de equipes deve considerar não apenas as habilidades técnicas dos alunos, mas também suas características sociais e emocionais, a fim de maximizar o aprendizado colaborativo e minimizar conflitos [Gutiérrez et al. 2016, Singhabghel and Bhavani 2018, Esgario et al. 2019].

Semelhante ao TFP, o Problema de Formação de Múltiplas Equipes (do inglês, *Multiple Team Formation Problem*, ou MTFP) é aplicado para a formação de diversos grupos, os quais devem atender as restrições impostas pela modelagem que descreve como os grupos devem ser. O intuito do MTFP é semelhante ao do TFP, mas aplicado à criação de diversos grupos [Gutiérrez et al. 2016, Esgario et al. 2019]. Ao modelar o MTFP, a solução pode ser multi-objetivo e envolver a otimização de vários critérios simultaneamente, como a maximização da diversidade de habilidades em cada equipe, a minimização dos custos associados à formação dos grupos e a maximização da eficiência geral das equipes formadas. Além disso, o MTFP frequentemente lida com a alocação de recursos escassos entre diferentes grupos, para que cada equipe seja formada de maneira viável e eficaz [Zhang and Zhang 2013, Singh et al. 2023].

É comum, em ambientes educacionais, que a formação de grupos seja realizada pelos próprios estudantes, levando em consideração questões de afinidade entre os integrantes. Quando o contexto educacional ultrapassa questões físicas e acontece em ambientes virtuais essa tarefa se torna ainda mais complexa. Tanto para os estudantes que não conhecem "de perto" seus colegas quanto para o professor/tutor, que não estão próximos da realidade educacional e social desses estudantes [Cunha and Siebra 2017, Miranda et al. 2020]. Diante disso, este trabalho propõe a criação de uma abordagem computacional para auxiliar o processo de formação de grupos em ambientes educacionais. A proposta foi construída a partir de uma abordagem híbrida, na qual são utilizados algoritmos de agrupamento e algoritmos genéticos para agrupar indivíduos caracterizados a partir de um modelo de estudante.

O restante do trabalho está organizado da seguinte maneira: na Seção 2, é apresentada a fundamentação teórica que embasa este estudo; na Seção 3, são apresentados os detalhes da abordagem proposta; na Seção 4, são descritos os detalhes sobre o método de desenvolvimento desta pesquisa; na Seção 5, são apresentados os resultados; na Seção 6, discutem-se os resultados obtidos e, por fim, na Seção 7, são apresentadas as considerações finais.

2. Fundamentação Teórica

Sendo um campo de estudo emergente, a aplicação de técnicas de mineração de dados¹ no contexto educacional visa explorar, identificar e compreender os fenômenos educacionais presentes nos ambientes de ensino, com o objetivo de aprimorar suas dinâmicas e resultados [Ramos et al. 2020]. Neste trabalho, propõe-se a caracterização das condições de aprendizagem através de um método de agrupamento, para produzir grupos mais capazes na solução de problemas em atividades de integração no ambiente acadêmico [Miranda et al. 2020].

Para solucionar o TFP, é preciso garantir que os membros da equipe atendam a certos requisitos estabelecidos, seja para uma tarefa ou qualquer outro motivo. Isso torna o TFP um problema da classe NP-Difícil [Lappas et al. 2009]. O problema se torna ainda mais complexo quando é necessário formar diversas equipes, atendendo ao MTFP. Muitos estudos sugerem usar heurísticas, pois calcular todas as soluções possíveis demandaria um tempo computacional inviável na prática [Daş et al. 2022, Borges et al. 2023, Qureshi et al. 2023, Singh et al. 2023].

A adoção de abordagens utilizando Algoritmos Genéticos (do inglês, “Genetic Algorithm”, ou GA), é frequente em problemas para os quais encontrar soluções ótimas tem alto custo computacional [Júnior and Dorça 2018, Pereira et al. 2019]. Esses algoritmos buscam soluções diferentes iterativamente e podem encontrar diversas respostas “boas”, mas nem sempre a melhor [Awal and Bharadwaj 2014, Esgario et al. 2019, Wang et al. 2022]. Como são métodos estocásticos, executá-los várias vezes pode resultar em diferentes soluções [Awal and Bharadwaj 2014]. Mesmo assim, para problemas em que é inviável encontrar a solução ótima, é aceitável usar essas aproximações.

Para resolver o TFP, pode-se utilizar uma estratégia híbrida. Dois trabalhos trouxeram variações do *K-means* com o objetivo de dividir os alunos em grupos menores e, em seguida, ajustá-los para reduzir o problema [Kalantzi et al. 2022, Singh et al. 2023]. No estudo de [Singh et al. 2023], foi proposto o *K-means* heterogêneo multivariado ponderado (do inglês, *Weighted Multivariate Heterogeneous K-means*, ou *WMH K-means*), que leva em consideração as diferentes características dos alunos para criar grupos balanceados. Esses grupos são então distribuídos pelo algoritmo *round-robin*, que cria novos grupos de tamanho igual. Esse método foi aplicado para reduzir o problema antes de utilizar a *Surrogate Optimization*, o que produziu bons resultados. Já [Kalantzi et al. 2022] sugeriu outra variante do *K-means*, utilizada com o mesmo propósito. Embora semelhante em funcionamento, essa variante se diferencia na forma de calcular as distâncias entre os indivíduos e na produção de grupos balanceados, sendo que, ao invés de aplicar o *round-robin*, seleciona-se os elementos aleatoriamente, grupo a grupo, para formar novos times balanceados. Em ambas as aplicações do método de agrupamento, foi demonstrado que os grupos produzidos pelas variações do *K-means* eram capazes de gerar grupos heterogêneos.

Para o cálculo das distâncias entre os membros dos grupos, é necessário propor um Modelo de Estudante, que será utilizado para relacioná-los entre si [Singh et al. 2023]. Este modelo descreve computacionalmente as características que um estudante possui.

¹Processo de descobrir padrões, correlações e informações úteis a partir de grandes conjuntos de dados utilizando técnicas de estatística.

Após a definição do modelo, o *WMH K-means* é aplicado N vezes para gerar N soluções, as quais serão utilizadas na inicialização do GA abordado neste estudo. Em seguida, é feita uma avaliação da qualidade das soluções finais, que serão comparadas reproduzindo o experimento com a população do GA substituída por indivíduos aleatórios.

3. Abordagem Proposta

Para a abordagem deste trabalho, adotou-se o Modelo de Estudante que pode ser visto na Figura 1. A figura divide-se em dois campos, que descrevem as características dos estudantes e os passos do método aplicado no estudo. As variáveis tentam descrever um modelo hipotético modificável dadas as circunstâncias. Elas se separam entre as que são utilizadas para calcular o conflito e as que calculam a diversidade.

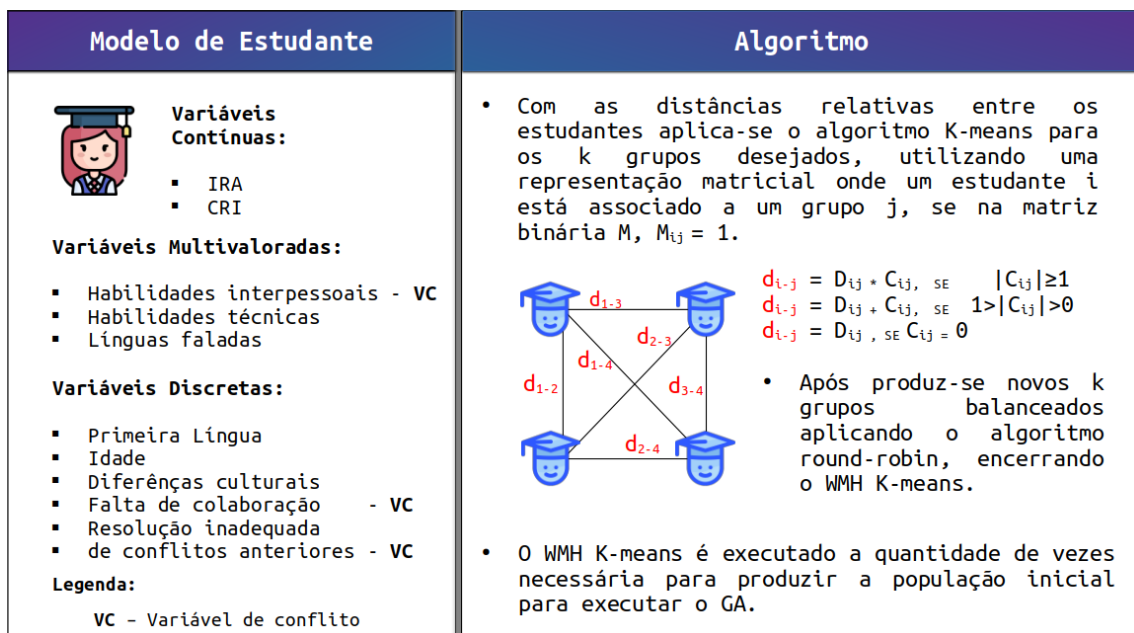


Figura 1. Modelo do Estudante

Outra divisão que se dá nas variáveis é a distinção dos seus tipos, o que também define como elas serão tratadas durante a computação das distâncias. As variáveis contínuas são compostas: (i) pelo índice de rendimento acadêmico (IRA) e (ii) a capacidade de reter informação de um texto escrito (CRI). Já as discretas são formadas: (i) pela primeira língua; (ii) idade; (iii) diferenças culturais; (iv) falta de colaboração; e (v) resolução inadequada de conflitos anteriores. Para tratar as variáveis multi-valoradas, têm-se os valores contidos nos conjuntos: (i) de línguas faladas; (ii) habilidades interpessoais; e (iii) habilidades técnicas. Todas serão analisadas pelo *WMH K-means* que produzirá as respectivas distâncias relativas aos estudantes, que serão sempre as mesmas para um mesmo conjunto.

WMH K-means é aplicado repetidas vezes, tanto quanto necessário, para produzir o conjunto para o GA, sempre alterando as posições iniciais dos centroides, de modo que, ao agrupar, exista sempre a possibilidade de formar novos grupamentos. A razão desta etapa se dá pela expectativa de acelerar a convergência da solução, evitando desperdício de tempo computacional.

3.1. Descrição formal do problema

Dado um grupo de estudantes S deseja-se produzir n sub-grupos disjuntos com os elementos de S . Um estudante $\alpha_i \in S$ pode ser descrito por variáveis que sejam contínuas, discretas e por campos multivalorados [Singh et al. 2023]. Sendo D_{ij} o nível de diversidade entre o par de indivíduos i e j . D_{ij} é computado considerando a distância Euclidiana ponderada $E(X_{c,i}, X_{c,j})$, Equação 1 para variáveis contínuas X_c normalizadas para o intervalo $[0, 1]$, a distância de *Hamming* ponderada $H(X_{d,i}, X_{d,j})$, Equação 2 para variáveis discretas X_d , e a distância de *Jaccard* ponderada $J(X_{m,i}, X_{m,j})$, Equação 3 para variáveis multi-valoradas X_m . As notações $X_{c,i}$, $X_{d,i}$ e $X_{m,i}$ representam os valores dos atributos contínuos, discretos e multi-valorados do indivíduo i , respectivamente.

$$E(X_{c,i}, X_{c,j}) = \sqrt{\sum_{x_c \in X_c} w_{x_c} (x_{c,i} - x_{c,j})^2} \quad (1)$$

$$H(X_{d,i}, X_{d,j}) = \sum_{x_d \in X_d} w_{x_d} 1(x_{d,i} \neq x_{d,j}) \quad (2)$$

$$J(X_{m,i}, X_{m,j}) = \sum_{x_m \in X_m} w_{x_m} \left(1 - \frac{|x_{m,i} \cap x_{m,j}|}{|x_{m,i} \cup x_{m,j}|} \right) \quad (3)$$

Os pesos normalizados w_{x_c} , w_{x_d} e w_{x_m} representam a importância de cada variável nos conjuntos X_c , X_d e X_m , com valores positivos, de modo que $\sum_{x_c \in X_c} w_{x_c} = 1$, $\sum_{x_d \in X_d} w_{x_d} = 1$ e $\sum_{x_m \in X_m} w_{x_m} = 1$, e $1(x_{d,i} \neq x_{d,j})$ é uma função indicadora: que é igual a 1, se os valores discretos dos indivíduos i e j diferem, e 0 caso contrário. Assim, os valores de D_{ij} e C_{ij} são respectivamente definidos pelas Equações 4 e 5.

$$D_{ij} = \frac{1}{3} [E(X_{c,i}, X_{c,j}) + H(X_{d,i}, X_{d,j}) + J(X_{m,i}, X_{m,j})] \quad (4)$$

$$C_{ij} = \frac{1}{3} [E(Y_{c,i}, Y_{c,j}) + H(Y_{d,i}, Y_{d,j}) + J(Y_{m,i}, Y_{m,j})] \quad (5)$$

Os conjuntos Y_c , Y_d e Y_m representam as variáveis de conflito e substituem respectivamente os conjuntos X_c , X_d e X_m aplicados as funções $E(Y_{c,i}, Y_{c,j})$, $H(Y_{d,i}, Y_{d,j})$ e $J(Y_{m,i}, Y_{m,j})$. Ao calcular C_{ij} , os pesos w_{x_c} , w_{x_d} e w_{x_m} são substituídos por w_{y_c} , w_{y_d} e w_{y_m} , analogamente correspondendo a importância de cada variável de conflito, com valores positivos de modo que $\sum_{y_c \in Y_c} w_{y_c} = 1$, $\sum_{y_d \in Y_d} w_{y_d} = 1$ e $\sum_{y_m \in Y_m} w_{y_m} = 1$ seja satisfeito.

4. Materiais e Métodos

Este trabalho tem como objetivo a formação de grupos de estudantes com base em métricas de diversidade e conflito, buscando otimizar esses grupos para: maximizar a diversidade intragrupo, minimizar a diversidade entre estes grupos e minimizar o conflito intragrupo. A metodologia desenvolvida visa fornecer uma ferramenta eficiente para a criação de grupos que atendam a estes critérios, promovendo uma dinâmica mais equilibrada e produtiva entre os membros.

A abordagem adotada tem foco na viabilidade da metodologia proposta, avaliando qual o impacto ao utilizar o algoritmo *WMH K-means* para gerar os indivíduos iniciais de uma população para aplicação de um GA por cruzamento. As etapas do trabalho envolvem: a geração de dados, a aplicação do algoritmo e a visualização dos resultados, que são expostas na Figura 2, a qual descreve suas respectivas importâncias.

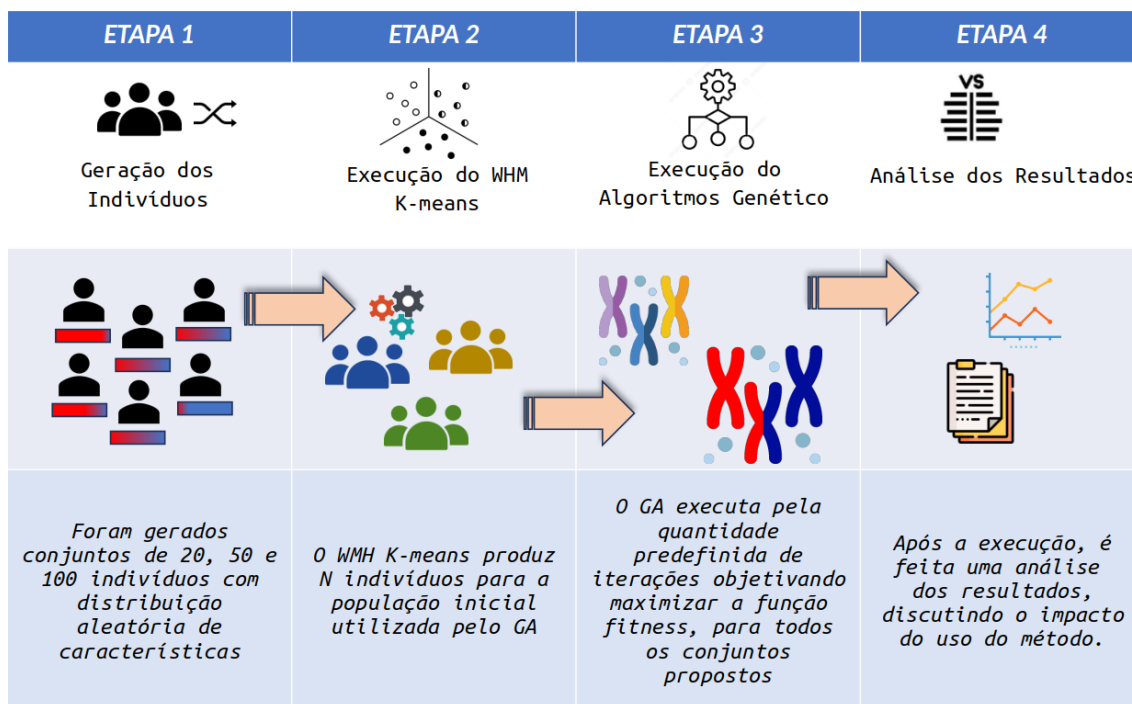


Figura 2. Etapas da metodologia

A geração do conjunto de estudantes é realizada de maneira probabilística, onde as características individuais de cada aluno são determinadas por distribuições de probabilidade específicas para cada variável. Variáveis contínuas, como o IRA e o CRI, seguem uma distribuição uniforme no intervalo de 0 a 1. Variáveis discretas, como Primeira Língua, Idade e características comportamentais, são definidas por distribuições discretas ou binomiais, refletindo a diversidade da população. Já habilidades interpessoais, técnicas e línguas faladas são consideradas variáveis multi-valoradas, com cada aluno possuindo um subconjunto aleatório dessas habilidades, visando representar a diversidade de competências na amostra.

Após, os pesos são atribuídos a cada tipo de atributo e refletem a importância relativa de cada característica no cálculo das métricas de diversidade e conflito. Com os pesos definidos, aplica-se o *WMH K-means*, calculando as distâncias relativas entre os estudantes, executando o *K-means* e, em seguida, balanceando os grupos formados utilizando *round-robin*.

Repete-se o *WMH K-means*, alterando a posição inicial dos centroides para produzir *N* soluções, que serão os *N* indivíduos da população utilizada na aplicação do GA. O GA aplicado é definido em termos de cromossomo, indivíduo, população, seleção, cruzamento, mutação e função de avaliação. O fluxograma do algoritmo pode ser visualizado na Figura 3.

Em termos de GA, um cromossomo é a menor unidade de informação que caracteriza a qual grupo um estudante pertence, fazendo de um indivíduo uma solução candidata no espaço de busca, representando uma possível distribuição de alunos em grupos. Uma população é o conjunto de indivíduos que representam possíveis soluções. A seleção é o processo de escolher indivíduos da população atual, chamados de “pais”, para gerar novos indivíduos, os “filhos”. Neste contexto, utiliza-se a técnica da roleta, na qual a probabilidade de seleção de um indivíduo é proporcional à sua aptidão em relação às outras aptidões presentes na população. Isso significa que indivíduos com maior aptidão têm uma chance maior de serem selecionados, aumentando a probabilidade de que as características desejáveis sejam passadas para a próxima geração. O cruzamento permuta valores dos cromossomos de dois indivíduos “pais” para produzir dois indivíduos “filhos”, que herdaram características de ambos. A mutação é utilizada para introduzir possíveis variações nos genes dos “filhos” para manter a diversidade genética e explorar novas áreas do espaço de busca. Por fim, a função de avaliação é a função que avalia a qualidade de um indivíduo com base em critérios específicos, determinando quão “boa” é a solução que ele representa.

O procedimento de um GA é o seguinte: uma população POP de n cromossomos é inicializada aleatoriamente. A aptidão de cada cromossomo em POP é calculada. Dois cromossomos, digamos C_1 e C_2 , são selecionados da população POP de acordo com o método de seleção adotado. O operador de cruzamento, é aplicado em C_1 e C_2 para produzir um descendente, digamos C_0 . Em seguida, o operador de mutação é aplicado no descendente produzido C_0 com probabilidade de mutação Mp para gerar C'_0 . O novo descendente C'_0 é colocado na nova população [Katoch et al. 2021]. Deste modo, as operações de seleção, cruzamento e mutação serão repetidas na população atual até que a nova população esteja completa.

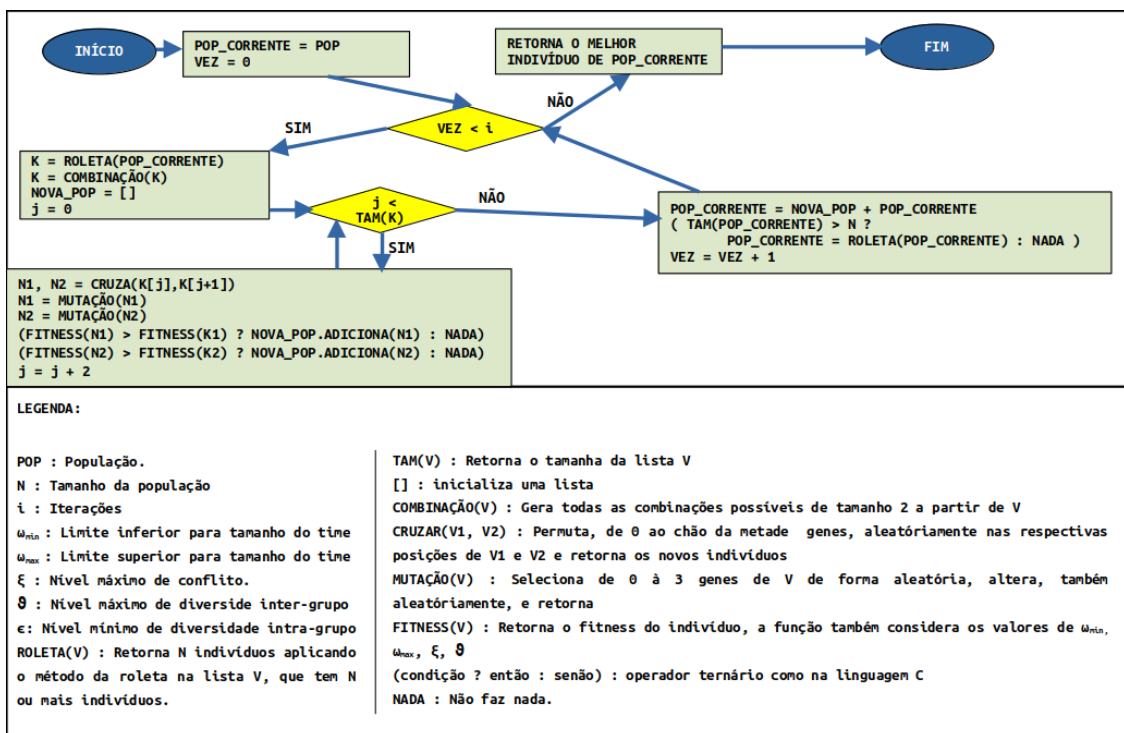


Figura 3. Fluxograma representando o GA

4.1. Definição da Função de Avaliação

Dado dois conjuntos S e G , sendo respectivamente os conjuntos dos alunos e dos grupos, denota-se $S = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$ e $G = \{\beta_1, \beta_2, \dots, \beta_k\}$. Definindo $G_s = \{(\alpha_i, \beta_j) \in S \times G\}$. Com isso pode-se denotar a Equação 6 como a função de avaliação, ou função objetivo:

$$\theta = \max_{g_\pi \in G_s} \left(\frac{1}{|G|} \sum_{x=1}^{|G|} [\lambda_1 DL(\alpha : (\alpha, \beta_x) \in g_\pi) - \lambda_2 CL(\alpha : (\alpha, \beta_x) \in g_\pi)] - \lambda_3 \frac{2}{|G|(|G| - 1)} \sum_{y=1}^{|G|} \sum_{z=y+1}^{|G|} [DL(\alpha : (\alpha, \beta_y) \in g_\pi) - DL(\alpha : (\alpha, \beta_z) \in g_\pi)]^2 \right) \quad (6)$$

Em que g_π deve atender um nível de diversidade mínima intra-grupo ϵ , compreendido como, $DL(\alpha : (\alpha, \beta_i) \in g_\pi) \geq \epsilon, \forall i \in \{1, 2, \dots, |G|\}$, um nível de diversidade máximo inter-grupo ϑ , assim, $[DL(\alpha : (\alpha, \beta_i) \in g_\pi) - DL(\alpha : (\alpha, \beta_j) \in g_\pi)]^2 \leq \vartheta, \forall i, j \in [1, 2, \dots, |G|]$, um intervalo para a quantidade de estudantes de um grupo, sendo um mínimo ω_{min} e um máximo ω_{max} , deste modo, $\omega_{min} \leq |\alpha : (\alpha, \beta_i) \in g_\pi| \leq \omega_{max}, \forall i \in \{1, 2, \dots, |G|\}$, o nível máximo de conflito ξ , portanto, $CL(\alpha | (\alpha, \beta_i) \in g_\pi) \leq \xi, \forall i \in \{1, 2, \dots, |G|\}$, por fim $(\alpha, \beta_i) \in g_\pi$ (α_i está associado a exatamente um β_j).

Sendo $DL(A)$ e $CL(A)$ funções aplicáveis aos conjuntos de alunos α , neste caso associados a um grupo β determinado, elas denotam o nível de diversidade e de conflito de um grupo respectivamente entre estes alunos α . Onde λ_1 é o valor de importância dada a diversidade intragrupo, λ_2 é o valor de importância dada ao nível de conflito e λ_3 é o valor de importância dada a diversidade intergrupo, sendo $\lambda_1 + \lambda_2 + \lambda_3 = 1$. A função $DL(A)$, sendo A sendo um conjunto de pares ordenados. é definida pela Equação 7.

$$DL(A) = \frac{2}{|A|(|A| - 1)} \sum_{\alpha_i, \alpha_j \in A | i < j} D_{ij} \quad (7)$$

A função $CL(A)$, com A sendo um conjunto de pares ordenados, é definida pela Equação 8.

$$CL(A) = \frac{2}{|A|(|A| - 1)} \sum_{\alpha_i, \alpha_j \in A | i < j \in A | i < j} C_{ij} \quad (8)$$

4.2. Visualização dos Indivíduos

Para visualizar os indivíduos em um gráfico, é aplicada a técnica de escalonamento multidimensional (do inglês, *Multidimensional scaling*, ou MDS). Esta é uma ferramenta para quantificar julgamentos de similaridade. Formalmente, o MDS refere-se a um conjunto de procedimentos estatísticos utilizados para análise exploratória de dados e redução de dimensionalidade [Hout et al. 2013]. Assim aplicou-se o MDS as distâncias relativas entre os indivíduos para que fosse possível produzir uma visualização no R^2 .

4.3. Ambiente de Testes

Para executar os testes foi utilizado um Acer Aspire A515-45. O aparelho estava conectado a rede elétrica por seu carregador, com um processador Ryzen 7 5700u e 8GB de memória RAM, rodando o sistema operacional Ubuntu 22.04.4 LTS. E todos os experimentos foram realizados por um núcleo. Todo o desenvolvimento se deu por meio de um Jupyter Notebook, na versão 6.5.4, com um *kernel* executando o Python 3.11.5.

5. Resultados

Para avaliar visualmente o produto do algoritmo, observe a Figura 4, que mostra como os estudantes são distribuídos em cada etapa. Neste caso específico, ele foi aplicado a um conjunto de 100 estudantes distribuídos entre 2 grupos, facilitando a visualização da distribuição graficamente. A parte (a) da Figura 4 mostra o resultado do *WMH K-means*, e a parte (b) exibe o resultado após a execução do GA.

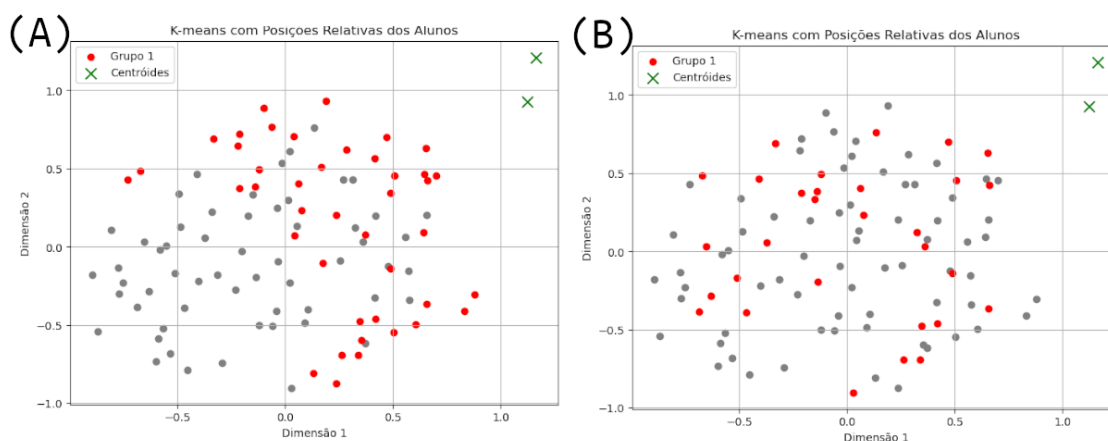


Figura 4. Distribuição de estudantes

É possível observar na Figura 4, parte A, que após aplicar o *K-means* existe uma clara separação dos grupos, com os pontos vermelhos marcando os indivíduos que pertencem ao grupo 1, e os demais pontos, em cinza, marcando os outros indivíduos, que pertencem ao grupo 2. Após realizar o processo de otimização, visualiza-se uma nova construção dos grupos, disponível na parte B da Figura 4. Nesta nova configuração dos grupos, é possível perceber que os indivíduos que compõem um mesmo time estão dispostos de maneira mais esparsa pelo plano.

Para facilitar a leitura e o entendimento das tabelas, os resultados dos testes foram divididos. A Tabela 1 trata de um conjunto de 20 estudantes, distribuídos em 2 grupos. Os algoritmos foram executados para populações de 5, 10, 15, 20 e 25 indivíduos, a fim de compor a população do GA.

Os mesmos testes foram executados para 50 indivíduos, produzindo 5 grupos, que podem ser visualizados na Tabela 2. O mesmo foi feito para 100 estudantes e 10 times, vistos na Tabela 3.

Em todos os experimentos, os valores ω_{min} e ω_{max} , usados para delimitar as quantidades possíveis de estudantes em um grupo, foram calculados da seguinte forma:

$[|S|/|G| - |S| \times 0.2]$ e $[|S|/|G| + |S| \times 0.2]$. Os valores de ξ , ϵ , ϑ e ζ foram mantidos livres, respeitando os intervalos estabelecidos anteriormente neste trabalho. Todos os conjuntos foram submetidos à otimização por 200 iterações, sem um critério de parada além do fim da execução.

Tabela 1. Comparação dos Métodos para (20, 2)

| Método | Tamanho da População | Tempo do Último Ajuste | Tempo Final | Função de Avaliação |
|------------------|----------------------|------------------------|-------------|---------------------|
| GA | 5 | 7.61s | 10.06s | 0.04344159084646459 |
| WMH K-means + GA | | 9,34s | 9,54s | 0.04269570510009474 |
| GA | 10 | 32.76s | 33.07s | 0.04437175447271137 |
| WMH K-means + GA | | 28,24s | 32,08s | 0.0444986868041493 |
| GA | 15 | 49.99s | 58.71s | 0.0448785506163281 |
| WMH K-means + GA | | 28.84s | 56.80s | 0.0448785506163281 |
| GA | 20 | 26.43s | 108.24s | 0.0448785506163281 |
| WMH K-means + GA | | 52.37s | 100.17s | 0.0448785506163281 |
| GA | 25 | 77.39s | 150.72s | 0.0448785506163281 |
| WMH K-means + GA | | 33.89s | 144.35s | 0.0448785506163281 |

Tabela 2. Comparação dos Métodos para (50, 5)

| Método | Tamanho da População | Tempo do Último Ajuste | Tempo Final | Função de Avaliação |
|------------------|----------------------|------------------------|-------------|----------------------|
| GA | 5 | 55.74s | 56.83s | 0.008720289495236995 |
| WMH K-means + GA | | 51.32s | 54.55s | 0.008542559495211675 |
| GA | 10 | 223.06s | 226.92s | 0.011568531346837886 |
| WMH K-means + GA | | 208.44s | 214.70s | 0.011336081849362873 |
| GA | 15 | 396.57s | 398.30s | 0.011504701801990293 |
| WMH K-means + GA | | 368.87s | 380.51s | 0.011579284705771783 |
| GA | 20 | 683.51s | 694.46s | 0.011680888041773387 |
| WMH K-means + GA | | 654.18s | 710.21s | 0.011641799046395106 |
| GA | 25 | 964.30s | 993.09s | 0.011829853914820778 |
| WMH K-means + GA | | 896.32s | 896.32s | 0.011720711114453863 |

Tabela 3. Comparação dos Métodos para (100, 10)

| Método | Tamanho da População | Tempo do Último Ajuste | Tempo Final | Função de Avaliação |
|------------------|----------------------|------------------------|-------------|-----------------------|
| GA | 5 | 194.62s | 202.08s | 0.0021855531885501597 |
| WMH K-means + GA | | 185.33s | 185.33s | 0.0020826737117647745 |
| GA | 10 | 877.31s | 877.31s | 0.0031556031056129688 |
| WMH K-means + GA | | 864.63s | 864.63s | 0.003283455338652493 |
| GA | 15 | 1892.78s | 1925.27s | 0.003943117436833028 |
| WMH K-means + GA | | 1672.11s | 1672.11s | 0.0035135002993322154 |
| GA | 20 | 3741.79s | 3741.79s | 0.0045476420738738005 |
| WMH K-means + GA | | 3421.47s | 3551.48s | 0.0043230319365541265 |
| GA | 25 | 4883.80s | 4945.76s | 0.004504854956634398 |
| WMH K-means + GA | | 4355.49s | 4438.72s | 0.004098088237538268 |

6. Discussão

Analisando as tabelas apresentadas na sessão 5, é possível verificar que os tempos de execução do *WMH K-means + GA* são, geralmente, menores ou comparáveis aos do GA puro, convergindo também mais rapidamente para o melhor valor da função de avaliação. Entretanto, esses valores são, em geral, inferiores aos obtidos executando o GA com uma população inicial gerada aleatoriamente. Ainda assim, podem ser soluções aceitáveis, sendo que, na maioria dos casos, a diferença entre os valores é razoável. Com relação aos tempos de execução, o *WMH K-means + GA* se mostra tão veloz quanto a alternativa, possuindo, no entanto, um tempo elevado para montar as equipes, algo que, para ambas as alternativas, pode ser minimizado com o uso de programação paralela, ainda que não seja o ideal. Isso indica que ambos os métodos alcançam resultados semelhantes em termos de qualidade da solução final.

Em relação ao tamanho das diferentes populações utilizadas para as otimizações, não é possível visualizar uma tendência clara de superioridade de alguma das abordagens com os testes executados. Isso pode sugerir que a escolha desses parâmetros deva ser feita e ajustada de acordo com as necessidades específicas de cada problema a ser abordado.

Com base nos resultados apresentados, é evidente que tanto o GA quanto o *WMH K-means + GA* são eficazes na formação de grupos de estudantes, mas cada um possui suas particularidades em termos de tempo de execução e qualidade das soluções finais. A análise detalhada dos tempos de execução e dos valores da função de avaliação sugere que a combinação dos métodos pode oferecer uma alternativa viável para situações em que a convergência rápida é desejada, embora não seja necessariamente a mais eficiente em termos de qualidade absoluta. A escolha dos parâmetros e a adaptação da metodologia devem ser feitas cuidadosamente, levando em consideração o problema em questão, para garantir que se obtenha o equilíbrio adequado entre velocidade e qualidade na formação dos grupos.

7. Considerações Finais

Neste trabalho, foi apresentada uma abordagem híbrida para solucionar o MTFP, utilizando o *WMH K-means* para produzir a população inicial de um GA e comparando-a com uma abordagem tradicional que utiliza apenas GA. A estratégia adotada para o problema produziu resultados que não se mostram significativamente melhores às alternativas já existentes para solucionar o MTFP, devido ao tempo gasto para obter uma solução.

A proposta contribui ao demonstrar que a aplicação do *WMH K-means* + GA é tão eficiente quanto a aplicação do GA e ao aplicar o MDS para permitir uma visualização gráfica dos indivíduos dos conjuntos formados grupo a grupo. Propor soluções mais eficientes para o MTFP, que sejam velozes ao otimizar os grupos, em tempos inferiores aos vistos neste trabalho, se faz necessário. Isso seria útil para dinâmicas pedagógicas, como a proposta no trabalho de [Silveira et al. 2022], para produzir ambientes interessantes, dadas certas alterações no Modelo de Usuário e nos parâmetros do algoritmo.

Outras técnicas para a criação dos grupos podem se mostrar mais eficientes. Uma possível abordagem para trabalhos futuros é a implementação de Redes Neurais Convolucionais, avaliando seu desempenho em relação aos métodos estabelecidos.

8. Agradecimentos

Os autores agradecem o apoio das agências de pesquisa e instituições CNPq, FAPEMIG e Instituto Federal de Educação, Ciência e Tecnologia do Sul de Minas Gerais - IFSUL-DEMINAS pelo apoio concedido a este trabalho.

Referências

- Awal, G. K. and Bharadwaj, K. K. (2014). Team formation in social networks based on collective intelligence—an evolutionary approach. *Applied intelligence*, 41:627–648.
- Borges, R., Sahlgrens, O., Koivunen, S., Stefanidis, K., Olsson, T., and Laitinen, A. (2023). Computational team assembly with fairness constraints. *arXiv preprint arXiv:2306.07023*.
- Cunha, F. O. M. and Siebra, C. D. A. (2017). Mapeamento sistemático na literatura acadêmico-científica sobre abordagens para formação de grupos em e-learning. *Revista Brasileira de Informática na Educação*, 24:16.
- Daş, G. S., Altinkaynak, B., Göçken, T., and Türker, A. K. (2022). A set partitioning based goal programming model for the team formation problem. *International Transactions in Operational Research*, 29:301–322.
- Esgario, J. G. M., da Silva, I. E., and Krohling, R. A. (2019). Application of genetic algorithms to the multiple team formation problem. *arXiv preprint arXiv:1903.03523*.
- Gutiérrez, J. H., Astudillo, C. A., Ballesteros-Pérez, P., Mora-Melià, D., and Candia-Véjar, A. (2016). The multiple team formation problem using sociometry. *Computers & Operations Research*, 75:150–162.
- Hout, M. C., Papesh, M. H., and Goldinger, S. D. (2013). Multidimensional scaling. *WIREs Cognitive Science*, 4:93–103.

- Júnior, C. B. and Dorça, F. (2018). Uma abordagem para a criação e recomendação de objetos de aprendizagem usando um algoritmo genético, tecnologias da web semântica e uma ontologia. In *Brazilian symposium on computers in education (simpósio brasileiro de informática na educação-sbie)*, page 1533.
- Kalantzi, M., Polyzou, A., and Karypis, G. (2022). Fern: Fair team formation for mutually beneficial collaborative learning. *IEEE Transactions on Learning Technologies*, 15:757–770.
- Katoch, S., Chauhan, S. S., and Kumar, V. (2021). A review on genetic algorithm: past, present, and future. *Multimedia Tools and Applications*, 80:8091–8126.
- Lappas, T., Liu, K., and Terzi, E. (2009). Finding a team of experts in social networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 467–476.
- Miranda, P. B., Mello, R. F., and Nascimento, A. C. (2020). A multi-objective optimization approach for the group formation problem. *Expert Systems with Applications*, 162:113828.
- Pereira, F., Oliveira, E., Fernandes, D., de Carvalho, L. S. G., and Junior, H. (2019). Otimização e automação da predição precoce do desempenho de alunos que utilizam juízes online: uma abordagem com algoritmo genético. In *Anais do XXX SBIE 2019*, page 1451. Brazilian Computer Society (SBC).
- Qureshi, M. A., Khaskheli, A., Qureshi, J. A., Raza, S. A., and Yousufi, S. Q. (2023). Factors affecting students' learning performance through collaborative learning and engagement. *Interactive Learning Environments*, 31:2371–2391.
- Ramos, J. L. C., Santos, L. F. L., Silva, J. C. S., and Rodrigues, R. L. (2020). Identificação de perfis de interação de estudantes de educação a distância por meio de técnicas de agrupamentos. In *Anais do XXXI Simpósio Brasileiro de Informática na Educação*, pages 932–941. Sociedade Brasileira de Computação.
- Rubin, P. A. and Bai, L. (2015). Forming competitively balanced teams. *IIE Transactions*, 47(6):620–633.
- Silveira, P. D. N., Carneiro, S., Moreli, J., de Menezes, C. S., and Cury, D. (2022). Smart learning environments em apoio aos ecossistemas de aprendizagem. In *Anais SBIE*, pages 175–185. Sociedade Brasileira de Computação - SBC.
- Singh, P., Huynh, P. K., Nguyen, D., Le, T., Moreno, W., and Le, T. Q. (2023). Leveraging multi-criteria integer programming optimization for effective team formation. *Authorea Preprints*.
- Singhbaghel, V. and Bhavani, S. D. (2018). Multiple team formation using an evolutionary approach. *2018 11th International Conference on Contemporary Computing, IC3 2018*.
- Wang, H., Li, J., Song, Y., Huang, J., Li, J., and Chen, Y. (2022). An improved genetic algorithm for team formation problem. *Proceedings of the 2022 IEEE Symposium Series on Computational Intelligence, SSCI 2022*, pages 774–781.
- Zhang, L. and Zhang, X. (2013). Multi-objective team formation optimization for new product development. *Computers & Industrial Engineering*, 64(3):804–811.