

# Uma Abordagem Baseada em Algoritmo Genético para Formação de Grupos de Estudos em Ambientes Virtuais de Aprendizagem

Andson M. Balieiro<sup>1</sup>, Igor Melo<sup>1</sup>, Débora Araújo<sup>3</sup>, Sebastião R Neto<sup>2</sup>, Eraylson Galdino<sup>3</sup>, Anselmo Gomes<sup>1</sup>

<sup>1</sup>Universidade de Pernambuco – Campus Garanhuns (UPE)  
Rua Cap. Pedro Rodrigues, 105 – São José – Garanhuns – PE - Brasil

<sup>2</sup>Instituto de Computação – Universidade Federal de Alagoas (UFAL)  
Caixa Postal 57072-900 – Maceió – Alagoas - Brasil

<sup>3</sup>Centro de Informática – Universidade Federal de Pernambuco (UFPE)  
Caixa Postal 50740-560 – Recife – PE – Brasil

{andson.balieiro@upe.br, igor-sugar@hotmail.com, dca2@cin.ufpe.br, srsn@ic.ufal.br, egs@cin.ufpe.br, anselmodsi@gmail.com}

**Abstract.** *Advances in Communication and Information Technology have driven the growth of distance education mode (DEM). A great ally of DEM is the learning virtual environment (LVE), which aims to represent the classroom in a virtual way. Although the LVE presents many tools (e.g. chats and forum), the study group composition is still neglected. It is a NP-hard problem and presents different characteristics from the physical environment. This paper formulates the problem of study group composition in LVE and proposes an approach based on Genetic Algorithms (GA) to solve it. Results show that the proposed scheme outperforms the random one by up to 25 %.*

**Resumo.** *O avanço da Tecnologia de Informação e Comunicação tem impulsionado o crescimento da modalidade de educação a distância (EAD). Um grande aliado da EAD é o ambiente virtual de aprendizagem (AVA), que busca representar a sala de aula de maneira virtual. Embora apresentem diversas ferramentas (ex. chat e fórum), a formação de grupos de estudos em AVAs ainda é negligenciada. Ela é um problema NP – difícil e apresenta características distintas do ambiente presencial. Este artigo formula o problema de formação de grupos em AVAs e propõe uma abordagem baseada em Algoritmo Genético para resolvê-lo. Resultados mostram que o esquema proposto é superior em até 25% o esquema aleatório.*

## 1. Introdução

Nos últimos anos, o avanço da tecnologia de informação e comunicação (TIC) tem proporcionado mudanças na forma como se pensa e faz educação. Entre elas, está o desenvolvimento da modalidade de educação à distância (EAD), que possibilita novas formas de aprender, modificando os métodos tradicionais de ensino. Nesse novo cenário o acesso a educação não fica limitado aos espaços físicos da sala de aula (Viana e Oliveira, 2015). De acordo com dados do censo da Associação Brasileira de Educação a Distância (ABED 2015/2016), existem 5.048.912 de alunos matriculados na EAD, sendo 1.108.021 em cursos regulamentados totalmente a distância e semipresenciais e 3.940.891 em cursos livres corporativos ou não corporativos. Considerando o contexto de ensino em EAD, a Lei de Diretrizes e Bases da Educação Nacional (LDBN) aponta a

educação à distância como uma modalidade educacional na qual a mediação didático-pedagógica dos processos de ensino e aprendizagem ocorre com a utilização de TIC, proporcionando a estudantes e professores o desenvolvimento de suas atividades educativas em lugares e tempos diversos (BRASIL, 1996).

Os ambientes virtuais de aprendizagem (AVAs) são grandes aliados da EAD, pois buscam representar virtualmente a sala de aula, promovendo a interação entre professor e aluno através de ferramentas como fóruns, chats, mecanismos de avaliação e áreas com os diversos materiais utilizados ao decorrer do curso (Viana e Oliveira, 2015). Embora existam características da sala de aula representadas nos AVAs, algumas ainda são negligenciadas. A formação de grupos de estudo de alunos é uma delas (Silveira, 2006). Uma aproximação dessa característica colaborativa são os fóruns de discussão, entretanto eles são criados para proporcionar interação entre uma turma toda e não entre um grupo particular. Grupos de estudo são configurações de aprendizado autodidatas. Eles têm o objetivo de suprir dificuldades por meio de interações entre duas ou mais pessoas, as quais se diferenciam por vários fatores, como nível de conhecimento do assunto e tipos de tomada de decisão, por exemplo. O estudo em grupo é bastante utilizado para ampliar o conhecimento fora de aula (Thomaz, 2016).

Uma vez que a proximidade física dos alunos não é um aspecto preponderante na EAD, a formação de grupos em AVAs deve considerar outras variáveis, que diferem daquelas dos ambientes presenciais, tendo em vista que a formação de grupos adequados possibilita melhor interação entre os seus membros e impacto positivo na aprendizagem. Neste aspecto, este trabalho propõe um esquema de formação de grupos baseado em algoritmo genético (AG) para AVA. O esquema considera informações estáticas (ex. idade, áreas de interesse), e dinâmicas (ex. tempo de acesso, hora de acesso) dos alunos, e sugere formação de grupos de estudos aos usuários.

De acordo com os aspectos acima delineados, este artigo está assim organizado: a Seção 2 apresenta os trabalhos relacionados; a Seção 3 descreve o problema de formação de grupos de estudos em AVAs; o esquema proposto para resolver este problema é descrito na Seção 4; a avaliação do esquema é apresentada na Seção 5; e, por fim, a Seção 6 conclui este artigo e direciona trabalhos futuros.

## 2. Trabalhos Relacionados

Os algoritmos genéticos vêm sendo bastante utilizados na resolução de diversos problemas no âmbito educacional. Nesse sentido, os autores em Jácome Junior *et al.* (2012), utilizam AGs para seleção de objetos de aprendizagem sensível ao contexto do estudante em ambientes móveis. Para isso, os autores fazem uso de informações como afinidade com o curso em questão e com dispositivos móveis, horários de estudos e localidade na qual o estudante se encontra.

Moreno *et al.* (2011) propõe um algoritmo genético que atua na formação de grupos considerando as características dos estudantes. Para validação do método proposto, os autores utilizaram características de estudantes ingressantes nos cursos de engenharia da University of Colombia, campus Medellin. Dois esquemas para formação de grupos foram avaliados: aleatório e auto-organizado. Não houve diferença significativa de desempenho entre os métodos. Outros trabalhos nesta linha podem ser encontrados em (Lin *et al.*, 2010; e Yannibelli e Amandi, 2012). Uma similaridade entre os trabalhos é que eles são voltados para o ambiente presencial, não considerando AVA.

Em Cunha *et al.* (2015) é discutida a formação de grupos de aprendizagem em EAD. Os autores verificam que as abordagens existentes não consideram as preferências de tutores, o que pode acarretar no fracasso do trabalho em grupo. Com

base em pesquisa realizada com tutores, foi feita a especificação de perfis de grupos considerando critérios e preferências de tutores, porém os resultados dessa abordagem são muito vagos. Ademais, como a preferência dos tutores, que são informações estáticas inseridas no sistema, é um fator predominante na formação de grupo, isso pode acarretar baixa interação dos alunos de um grupo e afetar a aprendizagem, pois as preferências dos alunos e o uso de informações dinâmicas são informações importantes que devem ser levadas em contas durante o processo de formação de grupos.

Em Silveira (2006) é apresentada uma arquitetura modelada com algoritmo genético (AG) para formação de grupos de estudo colaborativos. Silveira faz uso predominantemente de informações estáticas presentes no ambiente TelEduc (ex. sexo, estilo cognitivo e idade), negligenciando informações dinâmicas tais como duração e hora de acesso na formação de grupos. Embora os autores indiquem o uso de AG no trabalho, nenhuma descrição da estrutura do AG em termos de estrutura do cromossomo, função de aptidão, operadores genéticos utilizados e critério de parada, por exemplo, é apresentada. Além disso, diferente da avaliação adotada no artigo aqui proposto, os autores adotam apenas avaliação qualitativa, baseada em questionários respondidos pelos alunos e professores e não apresentam uma formulação para o problema proposto.

Neste sentido, o presente trabalho se difere dos demais, pois formula o problema de formação de grupos de estudos em AVA, levando em consideração tanto informações estáticas quanto dinâmicas, e propõe uma abordagem baseada em AG para resolvê-lo. Além disso, ele realiza a avaliação de desempenho do esquema proposto em comparação a abordagem aleatória.

### 3. Formação de Grupos em AVAs

Os AVAs têm sido uma valiosa ferramenta para o desenvolvimento de cursos na modalidade à distância. Eles possuem diversos mecanismos que buscam virtualizar o ambiente de sala de aula e prover boa interação entre professores e alunos. Entretanto, em geral, mecanismos automáticos de formação de grupos ainda são negligenciados nestes ambientes. A formação de grupos em AVAs difere do ambiente presencial, pois, como não há necessariamente o contato físico entre os alunos, outras características devem ser consideradas para proporcionar uma boa formação de grupos. Neste aspecto, tanto informações estáticas, como nome, idade, área de interesse e cidade, por exemplo, que são definidas no perfil dos alunos, quanto dinâmicas, como tempo de acesso, hora de acesso do usuário, tipo de dispositivo adotado no acesso, por exemplo, que são capturadas automaticamente pelo ambiente utilizado, devem ser considerados no processo de formação de grupos em tais ambientes. Além disso, o número de alunos nos AVAs pode ser muito maior do que em ambientes presenciais, sala de aula, por exemplo.

#### 3.1. Formulação do Problema

O problema de formação de grupos em AVAs pode ser visto como sendo a construção de um dado número de grupos que reúnam o maior número de características (estáticas e dinâmicas) similares de alunos no mesmo grupo. Uma vez que os valores das características podem apresentar magnitudes diferentes (ex. idade, em anos, e altura em metros), faz-se necessário normalizá-los antes de computar o nível de semelhança entre os alunos, garantindo que todas as características tenham o mesmo peso (importância) no agrupamento.

Dado  $K$  alunos,  $A = \{a_1, a_2, a_3, \dots, a_K\}$ , com cada aluno possuindo  $P$

características, isto é,  $a_i = \{a_{i1}, a_{i2}, a_{i3}, \dots, a_{iP}\}$ , com  $i = 1, 2, \dots, K$ , neste trabalho, os valores das características foram normalizados no intervalo [0 1] de acordo com o Eq. 1, onde  $a_{ij}^*$  e  $a_{ij}$  são os valores real e normalizado, respectivamente, da característica  $j$  do aluno  $i$ .  $\min_{t \in \{1, 2, \dots, z\}} \{a_{tj}\}$  e  $\max_{t \in \{1, 2, \dots, z\}} \{a_{tj}\}$  representam o menor e o maior valor da característica  $j$  considerando todos os alunos.

Uma vez efetuada a normalização dos valores e adotando a distância euclidiana como medida de semelhança para quantificar o grau de associatividade entre os alunos, é possível descrever o agrupamento de alunos como o seguinte problema de otimização. Dado  $K$  alunos (conjunto  $A$ ), com  $P$  características e  $G$  grupos para formar, encontrar uma partição  $X$  de  $A$  em  $G$  grupos ( $A_h$ , com  $h = 1, 2, \dots, G$ ) de forma que a soma das distâncias euclidianas entre cada aluno  $a_i$  e o centróide  $\bar{a}_h$  do grupo  $A_h$  seja mínima. Formalmente, tem-se (ver Eq. 2)

$$a_{ij}^* = \frac{a_{ij} - \min_{t \in \{1, 2, \dots, z\}} \{a_{tj}\}}{\max_{t \in \{1, 2, \dots, z\}} \{a_{tj}\} - \min_{t \in \{1, 2, \dots, z\}} \{a_{tj}\}} \quad (1)$$

$$\text{Minimizar } f(X) = \sum_{h=1}^G \sum_{a_i \in A_h} \|a_i - \bar{a}_h\| \quad (2)$$

### 3.2. Características dos Alunos

Em nossa abordagem, 4 características dos alunos inscritos no AVA foram consideradas, as quais são: idade, tópico de interesse (disciplina), tempo de acesso e hora de acesso. A idade é um aspecto importante a ser considerado na formação de grupos, pois em geral, as pessoas tendem a formar grupos com outras de faixas etárias similares. A área de interesse (representado pela “disciplina”) dos alunos é outro ponto relevante, uma vez que pessoas de interesses similares podem auxiliar melhor umas às outras. Neste trabalho, consideraram-se 4 disciplinas, representadas por números inteiros (1 a 4). As duas últimas características, que indicam quando e a duração de acesso dos alunos ao AVA, proporcionam que estudantes que tenham padrões similares de acesso e maior possibilidade de contato online no sistema possam ser inseridos no mesmo grupo de estudos.

As duas primeiras podem ser obtidas através de informação de perfil do aluno, indicadas durante a criação de conta no AVA. As duas últimas podem ser capturadas pelo AVA durante a utilização do ambiente pelo usuário. É importante ressaltar que embora este trabalho adota apenas 4 características, nosso esquema é facilmente extensível para englobar mais aspectos dos alunos.

### 4. Esquema Proposto

Um esquema baseado em Algoritmo Genético (AG) para resolver problema de formação de grupos em AVAs é proposto. AG é um algoritmo de busca baseado nos princípios de seleção natural e genética das espécies. Ele atua evoluindo um conjunto de soluções, representadas pelos chamados cromossomos, durante um período de tempo (número de gerações). Eventualmente, através dos operadores genéticos (seleção, cruzamento e mutação) uma boa solução é encontrada através da combinação de diferentes possíveis soluções (Chen *et al.*, 2010; Balieiro *et al.*, 2014). Embora o AG

seja conhecido por possuir largo tempo na obtenção da solução, mecanismos para reduzir este tempo têm sido propostos na literatura (Chen *et al.*, 2010).

Algumas características do AG tornam a sua adoção interessante para o nosso problema. Diferente de outros métodos, o AG manipula várias soluções simultaneamente a cada iteração e as evolui a fim de alcançar uma resposta para o problema. Além disso, tendo em vista que a formação de grupos é um problema NP-difícil (Yannibelli e Amandi, 2012), o AG tem sido amplamente adotado na literatura para resolver problemas deste tipo, alcançando bons resultados (Balieiro *et al.*, 2014).

#### 4.1. Estrutura do Cromossomo e Função de Fitness

No esquema proposto, o indivíduo ou cromossomo  $X_i$  é representado por vetor de posições (ver Fig. 1), onde  $K$  é o número de alunos inscritos no AVA. Cada gene (posição) do cromossomo se refere a um aluno do AVA. O valor de cada gene indica o grupo ao qual o aluno está inserido. Assim, dado  $G$  grupos a serem formados, tem-se que o valor de cada gene pode ser um número inteiro entre 1 e  $G$ . A Fig. 1, ilustra um exemplo de cromossomo cujo tamanho é 5 (alunos) e a quantidade de grupos ( $G$ ) a serem formados é 2. Assim cada gene pode ter o valor 1 ou 2. Na figura, os alunos 1 e 4 formam o grupo 2, por exemplo. O esquema proposto considera o paralelismo intrínseco do AG para alcançar uma solução ótima ou sub-ótima para o problema de formação de grupos em AVAs.

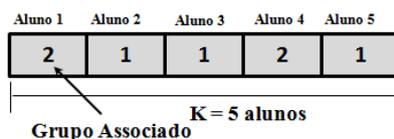


Figura 1. Estrutura do Cromossomo

A função de fitness definida para avaliar os indivíduos (soluções) durante o processo de evolução do AG é expressa na Eq. 3. Uma vez que quanto menor a soma das distâncias euclidiana (definida na Eq. 2) alcançada por um dado indivíduo (solução), maior tem que ser a sua aptidão, adotou-se o inverso da Eq. 2 na formulação da função de fitness juntamente com um fator de escala  $\beta$ . Neste trabalho, adotou-se  $\beta = 100$ .

$$fitness(X_i) = \frac{\beta}{1 + f(X_i)} \quad (2)$$

#### 4.2. Operadores Genéticos

O processo de evolução do AG é conduzido pelas operações de seleção, cruzamento e mutação. Para operação de seleção, que emula o processo de seleção natural que atua nas espécies biológicas (Goldberg, 1989), adotou-se o roulette wheel. Este operador seleciona os indivíduos que participarão do processo de cruzamento baseado no valor de fitness. Assim, indivíduos com maiores valores de fitness serão mais suscetíveis de sobreviverem para as gerações futuras. Na operação de cruzamento, que seleciona genes dos pais para criação de novos filhos (indivíduos), cuja idéia é simular a mistura de material genético que ocorre quando os organismos se reproduzem, adotou-se o operador uniforme. Este operador utiliza uma máscara de cruzamento binária e com tamanho igual ao do cromossomo. Ela indica como os pais serão combinados para gerar os filhos. Para compor cada filho, a seguinte regra é usada. Para o primeiro filho, caso o  $i$ -ésimo bit da máscara é 0, então o valor do gene do primeiro pai é usado na posição  $i$  do filho. Caso contrário, o código genético do segundo pai é usado nesta posição. Para o segundo filho, caso a máscara de cruzamento tenha o bit 0 na posição  $i$ , então o código

do segundo pai é usado na posição  $i$  do segundo filho. Caso contrário, o material genético para esta posição vem do primeiro pai. A Fig. 2 ilustra um exemplo do operador de cruzamento uniforme, considerando cromossomos com tamanho 5 e 2 grupos a serem formados.

A mutação é um processo crítico para o sucesso do AG visto que ele determina a direção de busca e evita convergência a ótimos locais através da inserção de diversidade na população (Goldberg, 1989). Ela ocorre após o cruzamento e altera aleatoriamente os novos filhos. Em nossa abordagem, a mutação de gene foi adotada (Goldberg, 1989). Nela, cada gene dos indivíduos pode sofrer mutação segundo uma dada probabilidade, denotada como probabilidade de mutação ( $pm$ ). Assim, para cada gene do cromossomo, um número aleatório ( $num$ ) entre 0 e 1 é gerado segundo uma distribuição uniforme e comparado a probabilidade de mutação. Caso ele seja menor, então o gene do indivíduo tem seu valor alterado dentro do domínio de valores do gene (1,2,3...,  $G$ ), excetuando seu valor atual. Caso contrário, ele permanece inalterado. A Fig. 3 ilustra um exemplo de mutação com um cromossomo de tamanho igual a 5 e dois grupos a serem formados, onde o primeiro e terceiro genes sofreram mutação.

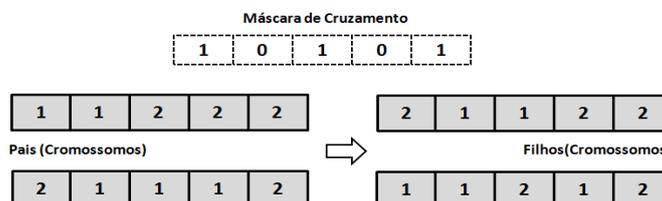


Figura 2. Exemplo de cruzamento uniforme



Figura 3. Exemplo de Mutação de Gene

É importante definir parâmetros corretos na configuração do AG para se obter bons resultados. Dois parâmetros-chave no AG são as probabilidades de cruzamento ( $pc$ ) e mutação ( $pm$ ), visto que elas expressam a frequência com que as operações de cruzamento e mutação são executadas. Uma alta probabilidade de mutação pode resultar a perda de boas soluções durante a evolução do AG, devido a maior possibilidade de alterações nas soluções, especialmente se estratégia elitistas não são empregadas. Por outro lado, baixas taxas de mutação podem ocasionar falta de diversidade na população, bem como convergência prematura do AG (Balieiro *et al.*, 2014).

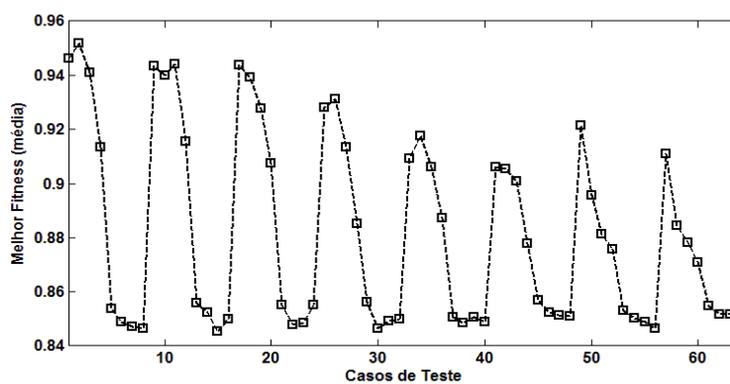
O número de cromossomos que realizam o cruzamento é influenciado pela  $pc$ . Assim, adotar uma  $pc$  pequena pode ocasionar um número insuficiente de novos filhos (soluções). Em contrapartida, quando uma probabilidade de cruzamento alta é adotada, um grande número de novos cromossomos é adicionado na população, o que pode resultar na perda de bons blocos (esquemas) dentro de um cromossomo (Goldberg, 1989). Assim, vários testes foram conduzidos para definir os valores de  $pc$  e  $pm$  que seriam adotados em nosso esquema. Isto envolveu empregar oito valores de teste, dentro do intervalo  $[0,1 \ 0,8]$ , para a probabilidade de crossover e oito valores de teste, definidos em  $[0,01 \ 0,8]$ , para a probabilidade de mutação, conforme apresentado na Tabela 1. A combinação de todos estes casos resulta em 64 casos de teste avaliados. Para cada caso, 10 instâncias de execução do AG foram realizadas em um cenário composto de 100 alunos, cujas características de tempo de acesso (em minutos), hora de acesso, idade (em anos), tópico de interesse (disciplina) foram uniformemente

distribuídas nos intervalos [1 90], [0 23], [10 60] e [1 4], respectivamente. Além disso, considerou-se a formação de 5 grupos de estudos.

A Fig.4 apresenta os resultados dos casos de teste para o nosso AG. O caso de teste selecionado foi aquele que alcançou o maior valor de fitness (em média) para a população da última geração. O caso de teste 2 foi o obteve melhor desempenho. Ele tem as probabilidades de cruzamento e mutação iguais a 0.1 e 0.03, respectivamente.

**Tabela 1. Valores dos Casos de Teste**

Pc	0,1/0,2/0,3/0,4/0,5/0,6/0,7/0,8
Pm	0,01/0,03/0,05/0,1/0,3/0,5/0,7/0,8



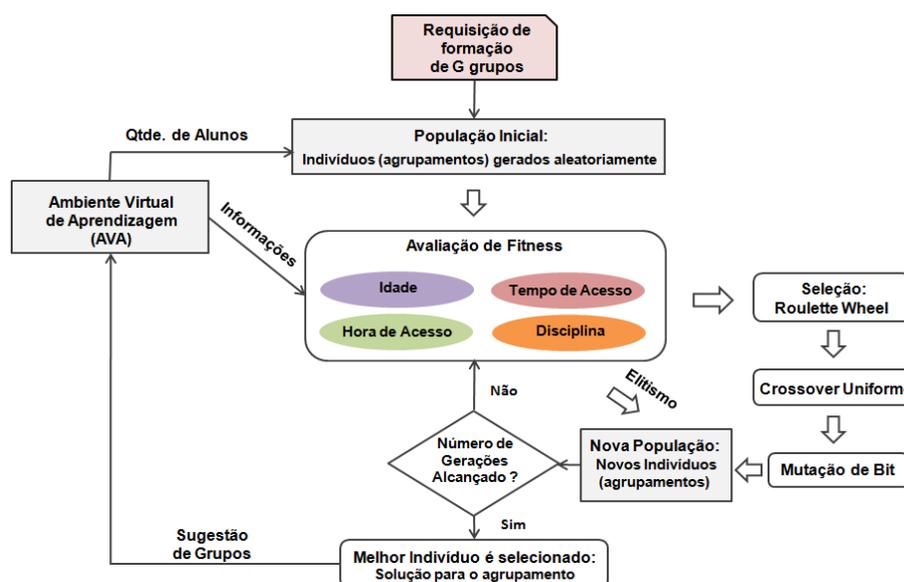
**Figura 4. Resultados dos casos de teste para seleção de pc e pm**

Além das probabilidades de cruzamento e mutação, o tamanho da população (S) e número de gerações(N) foram definidos iguais a 100. Tabela 2 sumariza os parâmetros do AG empregados neste trabalho.

**Tabela 2. Parâmetros do AG adotados**

Parameter	Value
Número de gerações (N)	100
Tamanho da população (L)	100
Probabilidade de cruzamento (pc)	0,1
Probabilidade de mutação (pm)	0,03

### 4.3. Fluxo de Execução do Esquema



**Figura 5. Fluxo de Execução do Esquema Proposto**

O fluxo de execução do nosso esquema é ilustrado na Fig. 5. Inicialmente, dado uma

solicitação de criação de  $G$  grupos, uma população (indivíduos) é gerada aleatoriamente para prover soluções candidatas para o problema. Neste processo, informação sobre a quantidade de alunos inscritos no AVA é considerada. Em seguida, os indivíduos são avaliados de acordo com a função de fitness adotada em nosso AG, a qual considera características dos alunos tais como tempo de acesso, hora de acesso, idade e tópico de interesse (disciplina) que são obtidas através do log do AVA e perfil definido pelo usuário. Após isso, os indivíduos são submetidos aos operadores de seleção, cruzamento e mutação. Além disso, uma estratégia elitista é empregada para garantir que o indivíduo com melhor fitness não será perdido durante o processo de seleção e será conduzido para compor a população da próxima geração. Após estas operações terem sido completadas, uma nova geração de soluções candidatas é formada e o critério de parada, determinado pelo número de gerações ( $N$ ), é avaliado. Caso ele não seja satisfeito, o processo é repetido a partir do estágio de avaliação de fitness. Caso contrário, o melhor indivíduo é escolhido para solução para o problema. Este indivíduo indica quais alunos formarão quais grupos. Assim, a sugestão automática de grupos é sugerida ao usuário do AVA

## 5. Avaliação do Esquema

Três cenários de avaliação foram definidos para analisar o desempenho do esquema proposto. Para todos os cenários, as características dos alunos como disciplina e hora de acesso ( $h$ ) foram geradas segundo uma distribuição uniforme no intervalo  $[1\ 4]$  e  $[0\ 23]$ , respectivamente. No primeiro cenário, a fim de verificar o comportamento do esquema sob diferentes demandas de alunos a serem agrupados, o número de alunos no sistema variou de 100 a 500, com incremento de 100. Neste cenário, o número de grupos foi configurado em 5. No segundo cenário, variou-se a quantidade de grupos a serem formados entre 10 e 50, com passo de 10, para avaliar o esquema sob diferentes demandas de formação de grupos. Neste cenário, consideraram-se 200 alunos no sistema. Assim com no cenário anterior, o tempo de acesso (em min) e a idade dos alunos (em anos) seguiram uma distribuição uniforme no intervalo  $[1\ 90]$  e  $[10\ 60]$  anos, respectivamente.

O terceiro cenário buscou avaliar o esquema quando os alunos apresentaram características mais similares (caso “homogêneo”) ou mais distintas (caso “heterogêneo”). Para isso, as características de tempo de acesso e idade foram definidas considerando intervalos mais estreitos, no caso homogêneo, e mais largos, no heterogêneo. Assim, no primeiro, o tempo de acesso e a idade foram uniformemente distribuídos nos intervalos  $[1\ 60]$  e  $[10\ 30]$ , respectivamente. Já no segundo, adotaram-se os intervalos  $[1\ 120]$  e  $[10\ 60]$ . O número de alunos e grupos foram 100 e 5, respectivamente, em ambos os casos. O esquema foi avaliado em termos de soma das distâncias euclidianas. A estratégia de agrupamento aleatório foi adotada na comparação de desempenho. Para cada ponto avaliado, 30 instâncias de simulação foram realizadas. Os resultados médios são apresentados com um nível de confiança de 95%, obtidos através do método de Bootstrap.

A Fig. 6.a mostra o desempenho dos esquemas quando o número de alunos varia. Nota-se o esquema proposto apresenta melhor desempenho do que o aleatório para todos os diferentes números de alunos considerados, alcançando uma melhoria de 15,95% até 24,43%. Além disso, quando o número de alunos cresce, a diferença entre os esquemas tende a aumentar em termos absolutos, mostrando a capacidade do esquema proposto de alcançar uma reduzida soma de diferenças no agrupamento de um elevado número de alunos. A Fig. 6.b apresenta resultado dos esquemas quando se varia o

número de grupos. O esquema proposto apresenta melhor desempenho que o aleatório. Ele consegue reduzir em até 19,16% a soma das distâncias em relação ao esquema aleatório. Nota-se que ambos os esquemas apresentam um decaimento na soma das distâncias quando o número de grupos aumenta. Isso acontece, pois com mais grupos, maior a similaridade dos componentes dentro de cada grupo. Em valores absolutos, a melhoria do esquema proposto em média de 44, bastante notável no gráfico, mostrando a habilidade do esquema proposto trabalhar com quantidades de grupos distintas.

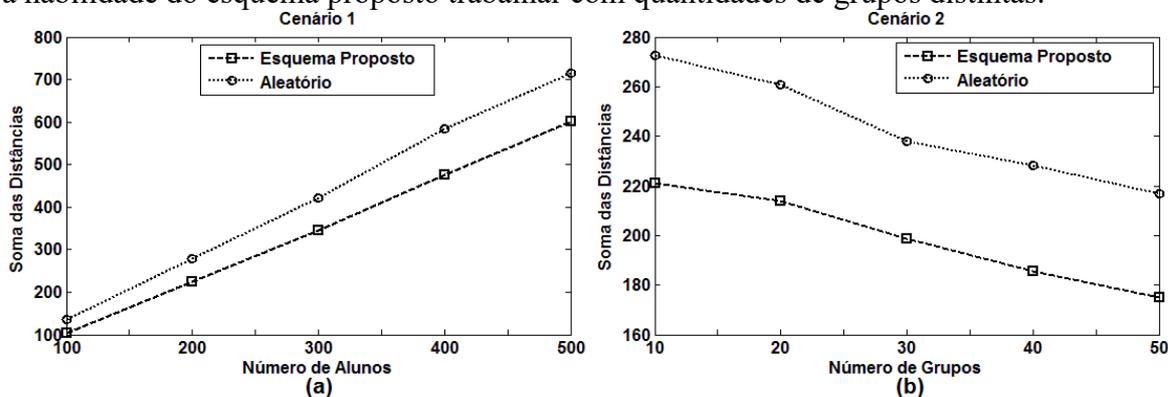


Figura 6. Resultados dos esquemas quando o número de alunos (parte a) e o número de grupos (parte b) variam

A Fig. 7 apresenta os resultados, quando diferentes níveis de similaridades dos alunos são considerados. O esquema proposto foi superior em ambos os casos avaliados, alcançando uma melhoria de até 25,49% e 23,62% em relação ao esquema aleatório, no caso homogêneo e heterogêneo, respectivamente. Nota-se que ambos os esquemas não apresentam grande variação de desempenho quando se adota níveis diferentes de similaridades dos alunos. Isso decorre da normalização das características. Um leve aumento de soma de distâncias no caso homogêneo é notado, decorrente da maior dificuldade em determinar qual o grupo mais adequado (entre os  $k$  existentes) para cada aluno, uma vez que eles são mais similares entre si.

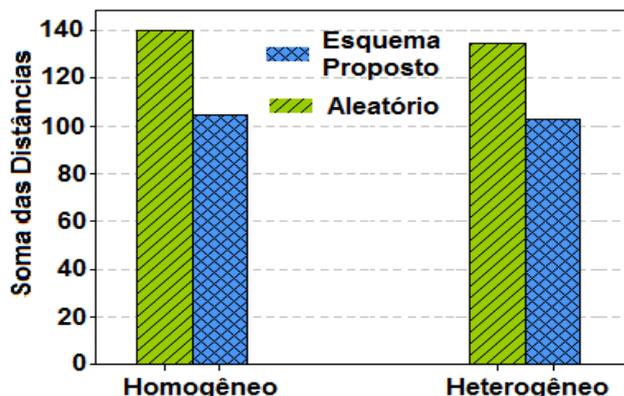


Figura 7. Resultados considerando alunos com diferentes níveis de similaridades

## 6. Conclusões e Trabalhos Futuros

Neste artigo formulou-se o problema de formação de grupos em AVAs. Este é um problema NP-difícil, que considera características diferentes daquelas adotadas em ambientes presenciais. Para este problema, um esquema baseado em algoritmo genético foi proposto. Este esquema considerou tanto informações estáticas como informações dinâmicas dos alunos para compor grupos com maior similaridade entre seus componentes. Resultados obtidos mostraram que o esquema proposto é superior ao esquema aleatório.

Para trabalhos futuros, pretende-se avaliar o esquema proposto, considerando

dados gerados por outras distribuições de probabilidades (ex. exponencial), bem como dados reais obtidos de AVAs. Desenvolver e avaliar outros esquemas (ex. baseado em k-médias ou inteligência computacional) em comparação com o proposto.

## Referências

- Balieiro, A. et. al. “A multi-objective genetic optimization for spectrum sensing in cognitive radio”, *Expert Systems with Applications*, v. 41, issue 8, p. 3640–3650, 2014.
- BRASIL. “Lei de Diretrizes e Bases da Educação Nacional”. Lei número 9394, 20 de dezembro de 1996.
- ABED.org. (2015). “Relatório Analítico da Aprendizagem a distância no Brasil”. Disponível em < [http://abed.org.br/arquivos/Censo\\_EAD\\_2015\\_POR.pdf](http://abed.org.br/arquivos/Censo_EAD_2015_POR.pdf) >. Acesso em 12 de abr, 2017.
- Chen, S. et. al. (2010). “Genetic algorithm-based optimization for cognitive radio networks”, *IEEE Sarnoff Symposium*, 2010.
- GLOBO.COM. (2014). “Matrículas em Cursos à Distância crescem 50% em um ano”. Disponível em: Acesso em: < <http://glo.bo/1pm9phB> > 29 de abril, 2016.
- Goldberg, D.E., “Genetic Algorithms in Search, Optimization and Machine Learning”, Addison-Wesley, 1989.
- Jácome Júnior, L.; Mendes Neto, F. M.; Silva, L. C. N. (2012). “Uma Abordagem Baseada em Algoritmo Genético para Recomendação de Objetos de Aprendizagem Sensível ao Contexto do Estudante”. *XXIII Simpósio Brasileiro de Informática na Educação (SBIE)*, 2012.
- Lin, Y. T.; Huang, Y. M.; Cheng, S. C. (2010). “An automatic group composition system for composing collaborative learning groups using enhanced particle swarm optimization”. *Computers & Education*, 55(4), 1483–1493, 2010.
- Moreno, J.; Ovalle, D. A.; Vicari, R. M. (2012). “A genetic algorithm approach for group formation in collaborative learning considering multiple student characteristics”. *Computers & Education*. 58(1), 560–569 (2012).
- Rocha, F. E. L. et al. (2004) . “Especificação de um Algoritmo Genético para auxiliar na Avaliação da Aprendizagem Significativa com Mapas Conceituais”. *XXV Simpósio Brasileiro de Informática na Educação (SBIE)*, 2004.
- Silveira, S. R. (2006). “Formação de grupos colaborativos em um Ambiente Multiagente Interativo de Aprendizagem na Internet: um estudo de caso utilizando sistemas multiagentes e algoritmos genéticos”. Universidade Federal do Rio Grande do Sul (UFRS), Tese, 2006.
- Thomaz, Cintia. (2014). “Os bons resultados do estudo em grupos – agora na internet”. Disponível em: < <http://veja.abril.com.br/educacao/os-bons-resultados-do-estudo-em-grupo-agora-na-internet/> > Acesso em 15 de maio, 2016.
- Viana, W. S.; Oliveira, E. H. T. (2015). “Uma abordagem ciente de contexto para Criação e Realização de atividades restritas por localização em EAD. *XXVI Simpósio Brasileiro de Informática na Educação (SBIE 2015)*, 2015. p. 907-916.
- Yannibelli, V.; Amandi, A. (2012). “A deterministic crowding evolutionary algorithm to form learning teams in a collaborative learning context”. *Expert Syst. Appl.* 39(10), 8584–8592, 2012.