

ANO
2015

JUCILANE ROSA CITADIN | UM ALGORITMO GENÉTICO PARA FORMAÇÃO DE GRUPOS HETEROGÊNEOS NA APRENDIZAGEM COLABORATIVA

Este trabalho avalia o uso de Algoritmo Genéticos (AGs) para formação de grupos heterogêneos na aprendizagem colaborativa, considerando um contexto massivo de estudantes. A formação dos grupos teve como objetivo maximizar a heterogeneidade dos conhecimentos e as interações dos estudantes no grupo, além de gerar grupos balanceados entre si (inter-homogêneos). Os resultados apontam a eficácia e a eficiência do AG em comparação com o algoritmo randômico (método base de comparação).

Orientadora: Profa. Dra. Avaniide Kemczinski

Joinville, 2015



UDESC

UNIVERSIDADE DO ESTADO DE SANTA CATARINA – UDESC
CENTRO DE CIÊNCIAS TECNOLÓGICAS – CCT
CURSO DE MESTRADO ACADÊMICO EM COMPUTAÇÃO APLICADA

DISSERTAÇÃO DE MESTRADO

UM ALGORITMO GENÉTICO PARA FORMAÇÃO DE GRUPOS HETEROGÊNEOS NA APRENDIZAGEM COLABORATIVA

JUCILANE ROSA CITADIN

JOINVILLE, 2015

JUCILANE ROSA CITADIN

**UM ALGORITMO GENÉTICO PARA FORMAÇÃO DE GRUPOS
HETEROGÊNEOS NA APRENDIZAGEM COLABORATIVA**

Dissertação apresentada ao Programa de Pós-Graduação em Computação Aplicada da Universidade do Estado de Santa Catarina, como requisito parcial para obtenção do grau de Mestre em Computação Aplicada.

Orientadora: Profa. Dra. Avanilde Kemczinski

JOINVILLE, SC

2015

C581u

Citadin, Jucilane Rosa

Um algoritmo genético para formação de grupos heterogêneos na aprendizagem Colaborativa / Jucilane Rosa Citadin. – 2015.

172 p. : il. ; 21 cm

Orientadora: Avanilde Kemczinski

Bibliografia: p. 143-148

Dissertação (mestrado) – Universidade do Estado de Santa Catarina, Centro de Ciências Tecnológicas, Programa de Pós-Graduação em Computação Aplicada, Joinville, 2015.

1. Computação aplicada. 2. Algoritmos genéticos. 3. Aprendizagem colaborativa.
4. Computer-Supported Collaborative Learning (CSCL). 5. Formação de grupos.
6. Grupos heterogêneos.

I. Kemczinski, Avanilde. II. Universidade do Estado de Santa Catarina. Programa de Pós-Graduação em Computação Aplicada. III. Título.

CDD 004 – 23. ed.

**“UM ALGORITMO GENÉTICO PARA FORMAÇÃO DE GRUPOS HETEROGÊNEOS
NA APRENDIZAGEM COLABORATIVA”**

por

JUCILANE ROSA CITADIN

Esta dissertação foi julgada adequada para a obtenção do título de

MESTRA EM COMPUTAÇÃO APLICADA

área de concentração em “Ciência da Computação”,
e aprovada em sua forma final pelo

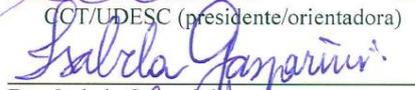
MESTRADO ACADÊMICO COMPUTAÇÃO APLICADA
CENTRO DE CIÊNCIAS TECNOLÓGICAS DA
UNIVERSIDADE DO ESTADO DE SANTA CATARINA.

Banca Examinadora:

Joinville, 31 de agosto de 2015.



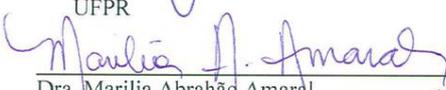
Dra. Avanilde Kemczinski
CCT/UDESC (presidente/orientadora)



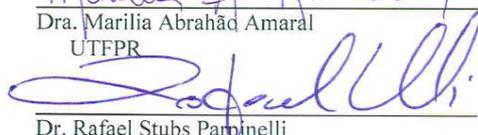
Dra. Isabela Gasparini
CCT/UDESC



Dra. Maria do Carmo Duarte Freitas
UFPR



Dra. Marília Abrahão Amaral
UTFPR



Dr. Rafael Stubs Parpinelli
CCT/UDESC

Este trabalho é dedicado ao meu esposo, ao meu filho e aos meus pais. Eu amo vocês.

AGRADECIMENTOS

Agradeço em primeiro lugar a Deus, pela graça diária da vida.

Muitas pessoas, direta e/ou indiretamente, me apoiaram na realização desta pesquisa, contudo algumas são especiais, entre elas destaco: minha orientadora, professora Avani de Kemczinski e meu coorientador professor Alexandre Veloso de Matos, sem os quais não teria chegado até aqui. Obrigada de coração, pelo apoio, aprendizado, contribuições e orientações para que eu alcançasse o sucesso da minha pesquisa!

A meu esposo Edson e a meu filho Lucas, pelo apoio incondicional em todas as circunstâncias e pela compreensão em todos os momentos de ausência, palavras são pouco para expressar meu amor e meu agradecimento a vocês!

Aos meus pais, Renato e Raquel, pelo apoio constante, sempre “cuidando” para que estivesse tudo bem. Obrigada pelo amor e dedicação de vocês.

Agradeço em especial meu colega de mestrado, Leanderson André, por todo o apoio e colaboração imprescindíveis para o sucesso da minha pesquisa.

Aos professores Rafael Rodrigues Obelheiro e Guilherme Piêgas Koslovski pela prontidão e disponibilidade nos momentos em que solicitei ajuda. Agradeço também ao pessoal do LABP2D (Laboratório de Processamento Paralelo e Distribuído), em especial ao Allan Krueger pela ajuda prestada.

À professora Elisa Henning, do DMAT-UDESC (Departamento de Matemática), pelas expressivas contribuições dadas a essa pesquisa em relação à estatística dos experimentos.

Aos professores e colegas do GPIE (Grupo de Pesquisa de Informática na Educação) pelas contribuições feitas ao trabalho em cada etapa apresentada.

Aos demais colegas de mestrado e professores, pela parceria, ensino e pelas boas horas de companhia que passamos juntos.

À FAPESC pelo apoio financeiro concedido.

A todos que de alguma forma contribuíram e me apoiaram e que porventura eu possa ter esquecido de mencionar especificamente aqui.

“Education is the most powerful weapon which you can use to change the world.”

Nelson Mandela

“Se a educação sozinha não transforma a sociedade, sem ela tampouco a sociedade muda.”

Paulo Freire

RESUMO

A formação de grupos para a aprendizagem colaborativa é uma tarefa importante, pois deve permitir a efetiva interação dos membros de um grupo. No entanto, a explosão combinatória relativa ao número de estudantes torna a formação de grupos um problema de otimização combinatória, que é *NP-hard*, isto é, um problema que não tem solução ótima em tempo polinomial. A resolução desses problemas de complexidade não polinomial requer um grande esforço computacional e aplicação de heurísticas ou meta-heurísticas para chegar a soluções viáveis em tempos adequados. Uma meta-heurística que tem provado ser eficiente na resolução de problemas de otimização *NP-hard* são os algoritmos genéticos (AG). Por isso, esta pesquisa avalia a adoção de algoritmos genéticos para a formação de grupos na aprendizagem colaborativa, considerando um contexto massivo de dados (milhares de estudantes). Pesquisas bibliográficas e um mapeamento sistemático da literatura sobre formação de grupos para aprendizagem colaborativa foram realizados. Constituir grupos de forma automática, considerando a abordagem selecionada e o critério heterogêneo, utilizando como recurso algoritmos genéticos, foram os resultados apontados pelo mapeamento. Assim, esta pesquisa se propôs a gerar grupos heterogêneos de forma automática, utilizando como abordagem grupo selecionado, considerando os conhecimentos e as interações dos estudantes. Algoritmos genéticos foram utilizados para a formação dos grupos, que teve como objetivo maximizar a heterogeneidade dos conhecimentos e as interações dos estudantes no grupo, além de gerar grupos balanceados entre si (inter-homogêneos). Foram desenvolvidos dois algoritmos, um AG e um randômico como método base de comparação. Ambos foram testados considerando dados de mil até dez mil estudantes, com diferentes configurações dos conhecimentos e interações. A comparação foi feita considerando o *fitness*, o percentual de grupos heterogêneos e/ou balanceados gerados pelos algoritmos e o tempo de processamento gasto por grupo gerado. Os resultados apontam que AG é eficaz para a formação de grupos heterogêneos na aprendizagem colaborativa. Os resultados do AG foram mais eficientes que os resultados do algoritmo randômico, para conhecimentos heterogêneos e conhecimentos aleatórios. Para conhecimentos homogêneos o AG teve a mesma eficiência que o algoritmo randômico, porém é mais eficaz, pois gera mais grupos balanceados do que o algoritmo randômico.

Palavras-chave: Algoritmos Genéticos. Aprendizagem Colaborativa. Computer-Supported Collaborative Learning (CSCL). Formação de Grupos. Grupos Heterogêneos.

ABSTRACT

Group formation for collaborative learning is an important task, because it should allow the effective interaction of members in a group. However, the combinatorial explosion on the number of students makes the group formation a combinatorial optimization problem, which is *NP-hard*, i.e., a problem that has not optimum solution in polynomial time. The resolution of these non-polynomial complexity problems requires a large computational effort and application of heuristics or meta-heuristics to reach viable solutions at appropriate times. A meta-heuristic that has proven to be efficient in solving *NP-hard* optimization problems are genetic algorithms (GA). Therefore, this research evaluates the adoption of genetic algorithms for the group formation in collaborative learning, considering a massive context data (thousands of students). Literature searches and a systematic mapping of the literature on group formation for collaborative learning were conducted. Form groups automatically considering the selected approach and the heterogeneous criteria, using as a resource genetic algorithms, were the results highlighted by mapping. Thus, this research aims to generate heterogeneous groups automatically, using a selected group approach, considering the students' knowledge and students' interactions. Genetic algorithms have been used for group formation, which aimed to maximize the diversity of knowledge and interactions of students in the group, besides generate balanced groups (inter-homogeneous groups). Two algorithms were developed, a GA and a random method for comparative purposes. Both were tested considering data from thousand to ten thousand students, with different configurations of knowledge and interaction. The comparison was made considering the *fitness*, the percentage of heterogeneous groups and / or balanced groups generated by algorithms and the processing time for generated group. The results indicate that AG is effective for the formation of heterogeneous groups in collaborative learning. The GA results were more efficient than the results of random algorithm for heterogeneous knowledge and random knowledge. To the homogeneous knowledge GA had the same efficiency as the random algorithm, but it is more effective because it generates more balanced groups than the random algorithm.

Key-words: Genetic Algorithm. Collaborative Learning. Computer-Supported Collaborative Learning (CSCL). Group Formation. Heterogeneous Group.

LISTA DE FIGURAS

Figura 1 – Etapas mapeamento sistemático.....	21
Figura 2 – Metodologia para a validação dos algoritmos.....	22
Figura 2 – Metodologia para a validação dos algoritmos.....	88
Figura 3 – Linha do tempo dos MOOCs	27
Figura 4 – Classificações adotadas para a formação de grupos	31
Figura 5 – Expressão de busca utilizada para massa	33
Figura 6 – Expressão de busca utilizada para LMSs	33
Figura 7 – Formação de grupos por <i>forma</i>	35
Figura 8 – Formação de grupos por <i>abordagem</i>	35
Figura 9 – Formação de grupos por <i>critério</i>	36
Figura 10 – Critérios complementares utilizados	36
Figura 11 – Relação algoritmos x dados	38
Figura 12 – Fluxograma genérico de algoritmos evolutivos	40
Figura 13 – Representação binária de um cromossomo	44
Figura 14 – Representação inteira de um cromossomo.....	44
Figura 15 – Representação por permutação de um cromossomo	45
Figura 16 – Exemplo para montar a roleta	48
Figura 17 – Roleta para o exemplo dado.....	48
Figura 18 – Exemplo do método de torneio	49
Figura 19 – Exemplo de amostragem estocástica uniforme	50
Figura 20 – Exemplo de seleção por <i>ranking</i>	51
Figura 21 – <i>Crossover</i> de um ponto	52
Figura 22 – <i>Crossover</i> de dois pontos	52
Figura 23 – <i>Crossover</i> uniforme.....	53
Figura 24(a) – <i>Crossover</i> mapeamento parcial (PMX)	54
Figura 24(b) – <i>Crossover</i> mapeamento parcial (PMX).....	54
Figura 24(c) – <i>Crossover</i> mapeamento parcial (PMX)	54
Figura 25(a) – <i>Crossover</i> de ordem.....	55
Figura 25(b) – <i>Crossover</i> de ordem.....	55
Figura 26(a) – <i>Crossover</i> de ciclo	55
Figura 26(b) – <i>Crossover</i> de ciclo	56
Figura 27 – Mutaç�o <i>bit a bit</i>	56
Figura 28 – <i>Swap mutation</i>	57
Figura 29 – <i>Insert mutation</i>	57
Figura 30 – <i>Inversion mutation</i>	58
Figura 31 – Exemplo de problema de formaç�o de grupos.....	71
Figura 32 – Exemplo de dados dos estudantes	71
Figura 33 – Exemplo de dados dos estudantes (normalizado)	72
Figura 34 – Exemplo da representaç�o adotada	79
Figura 35(a) – Exemplo de populaç�o para o problema proposto.....	83
Figura 35(b) – Exemplo de torneios para o problema proposto	83
Figura 36(a) – Exemplo <i>crossover</i> de ordem	84
Figura 36(b) – Exemplo <i>crossover</i> de ordem	84
Figura 37 – Exemplo <i>Swap mutation</i>	85
Figura 38 – Esquema do algoritmo gen�tico desenvolvido.....	86
Figura 39 – Definiç�es dos casos de testes.....	91
Figura 40 – Esquema do algoritmo rand�mico	92
Figura 41 – Converg�ncia <i>fitness</i> (melhor caso dos dados)	94

Figura 42 – Distribuição dos testes nas VMs	96
Figura 43 – Processo de análise estatística e apresentação dos resultados.....	99
Figura 44 – Resultados conhecimentos heterogêneos-interações uniformes	100
Figura 45(a) – <i>Box-plot Fitness</i> para conhecimentos heterogêneos-interações uniformes, $n = 1000$	102
Figura 45(b) – <i>Box-plot</i> Grupos Heterogêneos para conhecimentos heterogêneos-interações uniformes, $n = 1000$	102
Figura 46(a) – <i>Box-plot Fitness</i> para conhecimentos heterogêneos-interações uniformes, $n = 10000$	103
Figura 46(b) – <i>Box-plot</i> Grupos Heterogêneos para conhecimentos heterogêneos-interações uniformes, $n = 10000$	103
Figura 47 – Resultados conhecimentos heterogêneos-interações polarizadas	104
Figura 48 – Resultados conhecimentos heterogêneos-interações polarizadas ($\omega_2 = 2$)	106
Figura 49 – Resultados conhecimentos heterogêneos-interações aleatórias	108
Figura 50 – Resultados conhecimentos heterogêneos-interações aleatórias ($\omega_2 = 2$)	109
Figura 51 – Resultados conhecimentos homogêneos-interações uniformes	112
Figura 52 – Resultados conhecimentos homogêneos-interações polarizadas	114
Figura 53 – Resultados conhecimentos homogêneos-interações polarizadas ($\omega_2 = 2$)	115
Figura 54 – Resultados conhecimentos homogêneos-interações aleatórias	117
Figura 55 – Resultados conhecimentos homogêneos-interações aleatórias ($\omega_2 = 2$).....	118
Figura 56 – Resultados conhecimentos aleatórios-interações uniformes.....	121
Figura 57 – Resultados conhecimentos aleatórios-interações polarizadas.....	123
Figura 58 – Resultados conhecimentos aleatórios-interações polarizadas ($\omega_2 = 2$).....	124
Figura 59 – Resultados conhecimentos aleatórios-interações aleatórias.....	125
Figura 60 – Resultados conhecimentos aleatórios-interações aleatórias ($\omega_2 = 2$).....	126
Figura 61 – Tempo de processamento dos algoritmos (primeiro caso de teste).....	129
Figura 62 – Relação Tempo / Grupos Heterogêneos (custo) - caso 1	130
Figura 63 – Relação Tempo / Grupos Heterogêneos (custo) - caso 2	131
Figura 64 – Relação Tempo / Grupos Heterogêneos (custo) - caso 3	132
Figura 65 – Relação Tempo / Grupos Balanceados (custo) - caso 4.....	134
Figura 66 – Relação Tempo / Grupos Heterogêneos (custo) - caso 5	135
Figura 67 – Relação Tempo / Grupos Heterogêneos (custo) - caso 6	136

LISTA DE TABELAS

Tabela 1 – Resumo da revisão de literatura.....	65
Tabela 2 – Contribuições dos trabalhos relacionados para esta pesquisa.....	66
Tabela 3 – <i>Wilcoxon test</i> conhecimentos heterogêneos-interações polarizadas.....	105
Tabela 4 – <i>Wilcoxon test</i> conhecimentos heterogêneos-interações polarizadas, $\omega_2 = 2$, menores diferenças ocorridas	107
Tabela 5 – <i>Wilcoxon test</i> conhecimentos heterogêneos-interações aleatórias.....	108
Tabela 6 – <i>Wilcoxon test</i> conhecimentos heterogêneos-interações aleatórias, $\omega_2 = 2$, menores diferenças ocorridas	110
Tabela 7 – <i>Wilcoxon test</i> conhecimentos heterogêneos, $\omega_1 = 2$, menores diferenças ocorridas	111
Tabela 8 – <i>Wilcoxon test</i> conhecimentos homogêneos-interações uniformes.....	113
Tabela 9 – <i>Wilcoxon test</i> conhecimentos homogêneos-interações polarizadas.....	114
Tabela 10 – <i>Wilcoxon test</i> conhecimentos homogêneos-interações polarizadas ($\omega_2 = 2$).....	116
Tabela 11 – <i>Wilcoxon test</i> conhecimentos homogêneos-interações aleatórias.....	117
Tabela 12 – <i>Wilcoxon test</i> conhecimentos homogêneos-interações aleatórias ($\omega_2 = 2$)	119
Tabela 13 – <i>Wilcoxon test</i> conhecimentos homogêneos, $\omega_1 = 2$, menores diferenças ocorridas	119
Tabela 14 – <i>Wilcoxon test</i> conhecimentos aleatórios-interações uniformes	122
Tabela 15 – <i>Wilcoxon test</i> conhecimentos aleatórios-interações polarizadas	122
Tabela 16 – <i>Wilcoxon test</i> conhecimentos aleatórios-interações polarizadas ($\omega_2 = 2$)	124
Tabela 17 – <i>Wilcoxon test</i> conhecimentos aleatórios-interações aleatórias	126
Tabela 18 – <i>Wilcoxon test</i> conhecimentos aleatórios-interações aleatórias ($\omega_2 = 2$)	127
Tabela 19 – <i>Wilcoxon test</i> conhecimentos aleatórios, $\omega_1 = 2$, menores diferenças ocorridas	128
Tabela 20 – <i>Wilcoxon test</i> para a relação custo/benefício (tempo/grupos)	137

LISTA DE QUADROS

Quadro 1 – Terminologias dos AGs	42
Quadro 2 – Pseudocódigo de um AG básico	43
Quadro 3 – Fórmula método linear para o ranking	50
Quadro 4 – Fórmula normalização dos dados	72
Quadro 5 – Métrica de Minkowski.....	73
Quadro 6 – Distância Euclidiana	73
Quadro 7 – Máximas distâncias em relação ao conhecimento	75
Quadro 8 – Matriz com as interações máximas.....	76
Quadro 9 – Analogia dos elementos no AG	78

LISTA DE ABREVIATURAS E SIGLAS

ACM	Association for Computing Machinery
AEs	Ambientes E-Learning
AGs	Algoritmos Genéticos
AVAs	Ambientes Virtuais de Aprendizagem
CAI	Computer-Assisted Instruction
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
cMOOCs	Connectivists Massive Open Online Courses
CSCL	Computer-Supported Collaborative Learning
EaD	Educação a Distância
IEEE	Institute of Electrical and Electronics Engineers
ISO	International Organization for Standardization
LMSs	Learning Management Systems
LTS	Long Term Support
MBA	Mecanismo de Busca Acadêmica
MIT	Massachusetts Institute of Technology
MOOCs	Massive Open Online Courses
NP-hard	Não Polinomiais Difíceis
OU	Open University
PMX	Partially Mapped Crossover
REAs	Recursos Educacionais Abertos
TICs	Tecnologias de Informação e Comunicação
VMs	Virtual Machines
xMOOCs	Behaviorists Massive Open Online Courses

LISTA DE SÍMBOLOS

%	Porcentagem
(*)	Caractere coringa na <i>query</i> de busca
μ	Mi
λ	Lambda
*	Multiplicação quando usado nas fórmulas e equações
Σ	Somatória
$\sqrt{\quad}$	Raíz Quadrada
	Módulo (valor absoluto)
\in	Pertence (elemento <i>pertence</i> ao conjunto)
[]	Intervalo de valor
[]	Limite Superior (arredondar para cima)
[]	Limite inferior (não arredondar)
α	Alfa
β	Beta
\ll	Muito menor que
\approx	Aproximadamente

SUMÁRIO

1	INTRODUÇÃO	16
1.1	PROBLEMA	19
1.2	OBJETIVOS.....	19
1.2.1	Objetivo Geral	19
1.2.2	Objetivos Específicos	19
1.3	ESCOPO.....	20
1.5	METODOLOGIA	21
1.6	ESTRUTURA	22
2	FUNDAMENTAÇÃO	23
2.1	APRENDIZAGEM COLABORATIVA	23
2.1.1	Computer-Supported Collaborative Learning (CSCL)	24
2.1.2	Massive Open Online Courses (MOOCs)	26
2.2	FORMAÇÃO DE GRUPOS.....	28
2.2.1	Formação de Grupos para Aprendizagem	30
2.2.2	Mapeamento Sistemático sobre Formação de Grupos na Aprendizagem Colaborativa	32
2.3	ALGORITMOS EVOLUTIVOS.....	39
2.3.1	Visão Geral	39
2.3.2	Algoritmos Genéticos	41
2.3.2.1	Terminologias e Características dos Algoritmos Genéticos.....	41
2.3.2.2	Representação dos Indivíduos	43
2.3.2.3	População	45
2.3.2.4	Função de Avaliação (<i>Fitness Function</i>).....	47
2.3.2.5	Seleção de Pais	47
2.3.2.6	Recombinação ou Cruzamento (<i>Crossover</i>).....	51
2.3.2.7	Mutação	56
2.4	CONSIDERAÇÕES DO CAPÍTULO.....	58
3	TRABALHOS RELACIONADOS	60
3.1	ALGORITMOS GENÉTICOS PARA FORMAÇÃO DE GRUPOS	61
3.2	ANÁLISES E CONSIDERAÇÕES DO CAPÍTULO	64
4	UM ALGORITMO GENÉTICO PARA FORMAÇÃO DE GRUPOS HETEROGÊNEOS NA APRENDIZAGEM COLABORATIVA	68
4.1	MODELO PROPOSTO.....	68
4.1.1	Características adotadas	68
4.1.2	Formulação do Problema	70
4.2	MODELAGEM DO ALGORITMO GENÉTICO	77
4.2.1	Representação Adotada	78
4.2.2	População	80

4.2.3	Função de Avaliação (<i>Fitness Funtion</i>)	81
4.2.4	Seleção de Pais	82
4.2.5	Recombinação ou Cruzamento (<i>Crossover</i>)	83
4.2.6	Mutação	85
4.2.7	Configuração de Parâmetros	85
4.3	CONSIDERAÇÕES DO CAPÍTULO	87
5	VALIDAÇÃO, RESULTADOS OBTIDOS E DISCUSSÃO	88
5.1	VALIDAÇÃO DOS ALGORITMOS	89
5.1.1	Configuração dos dados	89
5.1.2	Método de Comparação	91
5.2	TESTES-PILOTO	93
5.3	TESTES MASSIVOS.....	95
5.4	MÉTODOS ESTATÍSTICOS UTILIZADOS.....	97
5.4.1	Verificação da normalidade dos dados.....	97
5.4.2	Testes de Hipóteses	98
5.5	RESULTADOS OBTIDOS E DISCUSSÃO.....	98
5.5.1	Conhecimentos heterogêneos.....	100
5.5.2	Conhecimentos homogêneos	111
5.5.3	Conhecimentos aleatórios	120
5.5.4	Tempo de Processamento.....	129
5.6	CONSIDERAÇÕES DO CAPÍTULO	137
6	CONCLUSÃO	139
	REFERÊNCIAS	143
	APÊNDICE A – TRABALHOS PUBLICADOS	149
	APÊNDICE B – TABELAS DOS RESULTADOS	150

1 INTRODUÇÃO

Desde os primórdios, a humanidade tem se organizado em grupos, pois percebe que obtém vantagens dessa forma. À medida que o ser humano evoluiu, aperfeiçoou a comunicação, acumulou conhecimento e foi produzindo cada vez mais artefatos, denominados tecnologias, que vão sendo transmitidas para as próximas gerações (DE PAIVA; PERNAMBUCO, 2005), até chegar ao que se tem hoje como computadores, Internet e uma série de recursos tecnológicos, aos quais se denominam Tecnologias da Informação e Comunicação (TICs). As TICs são utilizadas das mais diversas formas: na indústria (no processo de automação e controle), no comércio (no gerenciamento, vendas e publicidade), no setor de investimentos (informação simultânea, comunicação imediata) e na educação (no processo de ensino-aprendizagem, na Educação a Distância - EaD).

A sociedade dá forma às tecnologias de acordo com suas necessidades, valores e interesses, e as TICs são sensíveis ao efeito do uso social da própria tecnologia (CASTELLS; GERHARDT, 2000). Os desdobramentos da Internet, o aumento do acesso a recursos de computação e dispositivos móveis, a prevalência de conteúdos multimídia, seguidos por mudanças sociais, econômicas e culturais, impactaram as TICs e suas aplicações fazendo-as evoluir além da computação pessoal para facilitar a colaboração e as interações sociais em geral (WANG et al., 2007).

No processo de ensino-aprendizagem, a colaboração tem um papel importante para a construção do conhecimento. No entanto, a colaboração por si só não produz resultados na aprendizagem, ela depende da efetiva interação dos envolvidos em um grupo. Por isso, criar condições para uma efetiva interação em um grupo é uma preocupação da aprendizagem colaborativa (DILLENBOURG; JÄRVELÄ; FISCHER, 2009). Aprendizagem colaborativa apoiada por computador (do inglês, *Computer-Supported Collaborative Learning - CSCL*) estuda como as pessoas aprendem juntas com o auxílio dos computadores. Em CSCL, os estudantes não são apenas indivíduos, mas são parte de um grupo, e o processo de aprendizagem se dá no processo do grupo (STAHL; KOSCHMANN; SUTHERS, 2006). Por isso, a formação de grupos é importante para a aprendizagem colaborativa.

A formação de grupos pode ser feita de diferentes maneiras (WESSNER; PFISTER, 2001), baseando-se em diferentes abordagens (OUNNAS; DAVIS; MILLARD, 2007), critérios e características dos estudantes. Por exemplo, um grupo pode ser formado manualmente pelo professor ou automaticamente por um sistema; pode ser um grupo homogêneo ou heterogêneo, levando em consideração diferentes características dos

estudantes; ou pode ainda ser um grupo misto, considerando tanto características homogêneas quanto heterogêneas, entre outros.

Parece uma tarefa simples, mas torna-se complexa à medida que cresce o número de estudantes e/ou as características definidas para o agrupamento.

A relação entre essas características e a explosão combinatória relativa ao número de estudantes, que é fatorial, torna o problema da formação de grupos um problema de otimização combinatória, que é *NP-hard*, isto é, um problema que não tem solução ótima em tempo polinomial (CORMEN et al., 2011; MORENO; OVALLE; VICARI, 2012; TOSCANI; VELOSO, 2009; YANNIBELLI; AMANDI, 2012). A resolução desses problemas de complexidade não polinomial requer um grande esforço computacional e aplicação de heurísticas ou meta-heurísticas para chegar a soluções viáveis em tempos adequados. Heurísticas são procedimentos que, aplicando o conhecimento que se tem sobre um problema e técnicas aplicáveis para a sua solução, encontra soluções viáveis em tempos razoáveis. Meta-heurísticas são estratégias mais genéricas para melhorar os procedimentos heurísticos (MELIÁN; PÉREZ; VEGA, 2003). Uma heurística não garante a solução ótima, mas tende a achar uma solução próxima da ótima; e uma meta-heurística é uma heurística que pode ser aplicada a diferentes problemas de otimização (LINDEN, 2012; MELIÁN; PÉREZ; VEGA, 2003).

Uma meta-heurística que tem provado ser eficiente na resolução de problemas de otimização *NP-hard* são os algoritmos evolutivos (EIBEN; SMITH, 2007; LINDEN, 2012). A Computação Evolutiva é uma área da Ciência da Computação que, como o nome sugere, baseia-se nos processos naturais de evolução, propostos por Darwin, na sua Teoria da Evolução das Espécies (DARWIN, 1859). Os algoritmos evolutivos simulam a evolução das espécies através de processos de seleção, reprodução e mutação aplicados a uma população inicial. A computação evolutiva envolve algoritmos genéticos, estratégias evolutivas, programação evolucionária, programação genética, entre outros (EIBEN; SMITH, 2007; LINDEN, 2012).

Os algoritmos genéticos (AGs) são um ramo dos algoritmos evolutivos, considerados uma técnica de busca aleatória de otimização global, que vem sendo usada com sucesso para resolver problemas complexos de otimização (EIBEN; SMITH, 2007; LINDEN, 2012). Além disso, o mapeamento sistemático sobre formação de grupos para aprendizagem colaborativa realizado nesta pesquisa, apontou AG como um dos algoritmos mais utilizados. Por isso, para esta pesquisa foram adotados os algoritmos genéticos para a formação de grupos na aprendizagem colaborativa.

Apesar de serem considerados genéricos, e o Teorema da Inexistência do Almoço Grátis (WOLPERT; MACREADY, 1997) implicar que nenhum algoritmo genérico é melhor que um algoritmo desenhado especificamente para o problema, Linden (2012) destaca a importância de se embutir o máximo de conhecimento específico do problema no AG. Além disso, Linden (2012) destaca vantagens dos AGs sobre outras técnicas de otimização:

- Não usam apenas informações locais, assim, não necessariamente ficam presos em máximos locais;
- Não são totalmente aleatórios, pois usam informações da população corrente para melhorar a busca;
- Não são afetados por descontinuidade de função, podendo lidar com funções discretas e contínuas e inclusive com um misto de ambas;
- São apropriados para solucionar problemas de busca em grandes espaços de busca, que não podem ser solucionados por técnicas tradicionais de otimização (que envolvem processos mais determinísticos, como programação linear por exemplo, entre outras).

Desta forma, a utilização dos AGs para otimizar a formação de grupos na aprendizagem colaborativa torna-se uma opção. Como dito anteriormente, as TICs são utilizadas para o processo de ensino-aprendizagem e a CSCL estuda como as pessoas aprendem juntas com o auxílio de computadores. Ambientes Virtuais de Aprendizagem (AVAs) ou *E-learning* (AEs) são sistemas computacionais disponíveis na Internet, destinados ao suporte de atividades de ensino-aprendizagem, mediados pelas TICs (KEMCZINSKI, 2005).

A utilização dos AEs no processo de ensino-aprendizagem já não é mais uma novidade, o que se comprova com o recente fenômeno na aprendizagem *online*: os MOOCs (*Massive Open Online Courses* ou Cursos *Online* Abertos e Massivos). Segundo Dave Cormier, MOOCs podem ser definidos como: “Um fenômeno *online* que ganhou impulso nos últimos anos. Um MOOC integra a conectividade das redes sociais, a facilitação de um reconhecido especialista em um campo de estudo, e uma coleção de recursos educacionais de livre acesso *online*” (MCAULEY et al., 2010). Em relação aos AEs tradicionais, os MOOCs têm duas características-chave: o acesso aberto - qualquer pessoa pode participar de um curso *online* gratuito e a escalabilidade - são projetados para suportar um número indefinido de participantes (YUAN; POWELL, 2013).

Dada a escalabilidade, faz-se ainda mais necessária a formação de grupos para promover a aprendizagem colaborativa, entretanto, observou-se nesta pesquisa sobre a formação de grupos, que os trabalhos geralmente fazem testes com uma quantidade bem mais modesta de estudantes, chegando à casa das centenas, enquanto a escalabilidade dos MOOCs passa da casa das centenas para a casa dos milhares (MCAULEY et al., 2010).

1.1 PROBLEMA

Segundo Litto e Formiga (2009), a Educação a Distância (EaD) sempre esteve associada ao emprego de meios tecnológicos para armazenar e transmitir informação entre os participantes do processo de ensino-aprendizagem. Os autores destacam que o novo paradigma da educação *online* é a *Aprendizagem Aberta e Flexível*. Os Recursos Educacionais Abertos (REAs) fazem parte do conceito de abertura em EaD. Os REAs são tipicamente objetos de aprendizagem, que podem ser cursos inteiros ou parte deles e seus componentes de áudio, texto, imagens e vídeos. Os materiais são disponibilizados gratuitamente na Internet, geralmente com a licença *Creative Commons* (LITTO; FORMIGA, 2009). Do livre acesso aos REAs e, mais recentemente, em decorrência dos cursos *online* abertos e massivos (MOOCs), há um crescente impulso entre instituições de ensino superior a participar nesse movimento "aberto" (DANIEL, 2012; YUAN; POWELL, 2013).

Assim, com base na contextualização apresentada, o problema a ser tratado é: **É eficaz utilizar AG para formar grupos na aprendizagem colaborativa considerando dados massivos em MOOCs (milhares de estudantes)?**

A seguir discutem-se os objetivos estabelecidos para tratar o problema apresentado.

1.2 OBJETIVOS

A fim de contribuir com a evolução das pesquisas sobre formação de grupos, esta pesquisa estabelece os objetivos gerais e específicos destacados a seguir.

1.2.1 Objetivo Geral

Avaliar a adoção de algoritmos genéticos para a formação de grupos na aprendizagem colaborativa, no contexto massivo de dados.

1.2.2 Objetivos Específicos

Como objetivos específicos, tem-se:

- a) Investigar o estado da arte de formação de grupos na aprendizagem colaborativa;
- b) Especificar o algoritmo genético a ser desenvolvido e a função *fitness* (modelo matemático) para o problema de formação de grupos na aprendizagem colaborativa;
- c) Desenvolver o algoritmo genético (AG) e o algoritmo randômico, para a formação de grupos na aprendizagem colaborativa, considerando uma quantidade massiva de dados;
- d) Comparar o desempenho do AG desenvolvido com o desempenho do algoritmo randômico (base de comparação), considerando o *fitness* alcançado, o percentual de grupos heterogêneos e o tempo de processamento.

1.3 ESCOPO

A formação de grupos para aprendizagem colaborativa desenvolvida nesta pesquisa considera duas características dos estudantes (conhecimentos e interações). Como mencionado anteriormente, o problema de formação de grupos já é um problema *NP-hard*, de otimização combinatória, que cresce em complexidade à medida que aumenta o tamanho da entrada (número de estudantes) e/ou as características a serem consideradas. Assim, a fim de simplificar o problema devido à grande quantidade de estudantes, consideram-se apenas duas características. Essas características foram escolhidas por se tratarem de características geralmente disponíveis nos AVAs. No contexto desta pesquisa, os conhecimentos referem-se aos resultados das avaliações de aprendizagem, expressa por um indicativo de conhecimento com base no nível de qualidade do resultado alcançado, geralmente uma média dos resultados alcançados pelos estudantes em diferentes avaliações; e as interações referem-se às interações dos estudantes no fórum de discussão, isto é, às mensagens trocadas entre os estudantes no fórum de discussão, como perguntas e respostas às perguntas dos colegas. Esses conceitos são discutidos na seção 4.1.1.

Além de considerar as duas características supracitadas, esta pesquisa foca na adoção de AG como recurso para a formação de grupos na aprendizagem colaborativa, considerando o novo cenário da educação *online* (os MOOCs), através da avaliação com quantidades massivas de estudantes (milhares de estudantes).

Esta pesquisa é exploratória, pois o uso dos MOOCs ainda é um assunto novo e carente de estudos. Pretende-se contribuir com respostas iniciais para a formação de grupos

para aprendizagem colaborativa num contexto massivo de estudantes, de forma a permitir o desenvolvimento de trabalhos futuros nessa temática.

1.5 METODOLOGIA

A pesquisa, pela sua natureza, é classificada como aplicada, pelo fato de gerar conhecimento com aplicação prática imediata para solução de problema específico, baseados em conhecimentos prévios e com base na fundamentação teórica extraída da bibliografia que aborda os assuntos tratados (SILVA, 2001).

Do ponto de vista dos objetivos esta pesquisa é classificada como exploratória, pois tem como finalidade desenvolver, esclarecer e modificar conceitos e ideias, tendo em vista a formulação de problemas mais precisos ou hipóteses pesquisáveis para estudos posteriores (GIL, 1991).

A metodologia aplicada foi uma investigação planejada e desenvolvida em conformidade com as normas da metodologia científica. Pesquisas bibliográficas foram realizadas nas etapas iniciais do projeto de pesquisa, a fim de conhecer o estado da arte sobre formação de grupos na aprendizagem colaborativa, sobre *Massive Open Online Courses* (MOOCs) e sobre algoritmos genéticos (AGs). Além disso, um mapeamento sistemático da literatura foi realizado, a fim de buscar respostas para perguntas específicas sobre a formação de grupos na aprendizagem colaborativa. Para o mapeamento utilizou-se a metodologia proposta por Petersen et al. (2008), composta por cinco etapas: 1) Questões de Pesquisa, 2) Condução da Busca, 3) Análise dos Artigos, 4) Classificação dos Artigos e 5) Extração dos dados e Mapeamento, conforme mostra a Figura 1; e o mesmo está descrito na seção 2.2.2.

Figura 1 – Etapas mapeamento sistemático



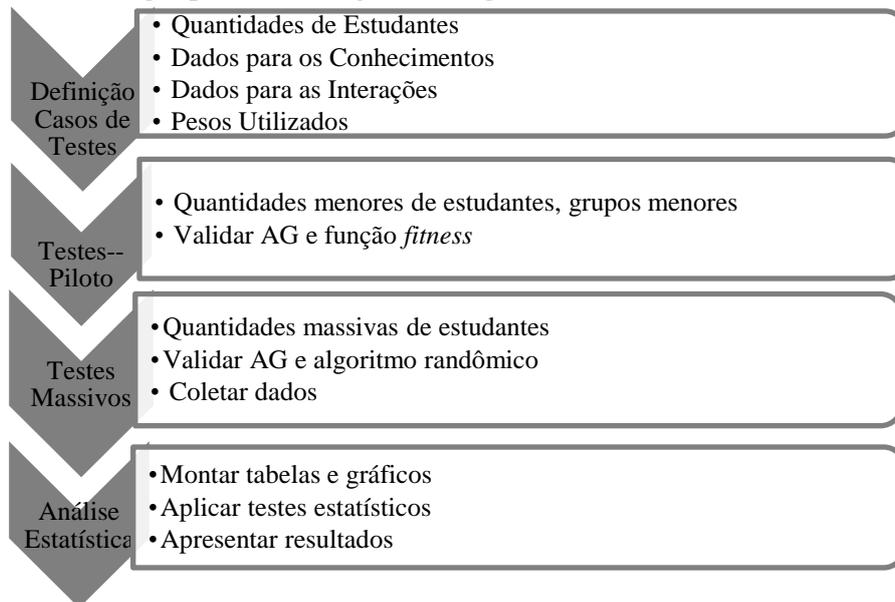
Fonte: Adaptado de Petersen et al. (2008)

Com base no material e no conhecimento adquirido no levantamento bibliográfico, foi definido o desenvolvimento do algoritmo genético para a formação de grupos na aprendizagem colaborativa.

Após a conclusão do AG, o mesmo foi executado com diferentes quantidades de estudantes (milhares), mensurando o desempenho do mesmo, através do *fitness*, do percentual de grupos heterogêneos e do tempo de processamento, para posterior comparação com o método randômico. A validação dos algoritmos foi realizada utilizando a metodologia descrita

na Figura 2. O processo iniciou-se com a definição das quantidades de estudantes a serem utilizadas, bem como a definição dos dados para as características (conhecimentos e interações) e pesos utilizados. A combinação de todos esses dados gerou 270 casos de testes. Foram realizados testes-piloto para as validações iniciais do AG e após estes, foram executados os 270 casos de testes tanto para o AG como para o algoritmo randômico, com as quantidades massivas de estudantes. Os dados coletados foram compilados e analisados estatisticamente, conforme descrito no capítulo 5.

Figura 2 – Metodologia para a validação dos algoritmos



Fonte: Próprio autor

No capítulo 5, a Figura 2 é melhor descrita e detalhada, conforme seguem as etapas desenvolvidas.

1.6 ESTRUTURA

Para apresentar o desenvolvimento da pesquisa realizada, esta dissertação apresenta a seguinte estrutura: no segundo capítulo é apresentada a fundamentação teórica sobre aprendizagem colaborativa, formação de grupos e algoritmos evolutivos; no terceiro capítulo são apresentados os trabalhos relacionados, com a revisão mais aprofundada dos mesmos; no quarto capítulo são apresentadas as definições em relação ao algoritmo genético desenvolvido para a formação de grupos; no quinto capítulo são apresentadas as definições para a validação dos algoritmos desenvolvidos, os testes realizados, os resultados obtidos e as discussões; e por fim no sexto capítulo é apresentada a conclusão, com as considerações finais e os trabalhos futuros.

2 FUNDAMENTAÇÃO

Para fundamentar a pesquisa aqui apresentada, pontos importantes para o entendimento são discutidos nas seções a seguir: 2.1 Aprendizagem Colaborativa, 2.2 Formação de Grupos e 2.3 Algoritmos Evolutivos.

2.1 APRENDIZAGEM COLABORATIVA

A aprendizagem colaborativa é um conceito complexo e não é definido de forma clara na literatura. Há diferentes significados em relação à adoção dos termos “colaborativa” (do inglês, *collaborative*) ou “cooperativa” (do inglês, *cooperative*), com divergentes opiniões sobre quais sejam suas diferenças ou similaridades (RESTA; LAFERRIÈRE, 2007). Roschelle e Teasley (1995) afirmam que a aprendizagem “cooperativa” diz mais respeito à divisão do trabalho entre os participantes, enquanto a aprendizagem “colaborativa” envolve um engajamento mútuo dos participantes para resolver o problema juntos. Panitz (1996) considera a colaboração como “uma filosofia de interação e estilo de vida pessoal”, enquanto a cooperação é vista mais como a estrutura da interação, que é projetada para facilitar a realização do trabalho em grupo.

Apesar da falta de consenso em relação aos termos e seus significados, Kirschner (2001) aponta que ambas (*cooperative* e *collaborative*) têm elementos em comum:

- A aprendizagem é ativa;
- O professor é mais um facilitador;
- Os processos de ensino-aprendizagem são experiências compartilhadas;
- As atividades são em pequenos grupos;
- Os estudantes assumem a responsabilidade pela aprendizagem;
- Os estudantes refletem sobre suas próprias suposições e processos de pensamento;
- Habilidades sociais e de equipe são desenvolvidas através da construção de consensos.

Assim, é possível afirmar que a aprendizagem colaborativa é uma abordagem centrada no estudante e orientada ao grupo, na qual os estudantes são ativos e responsáveis pela sua própria aprendizagem e o professor deixa de ser o centro das atenções e passa a promover ações para que o estudante possa progredir por seus próprios esforços (CASTRO; MENEZES, 2011).

Com a abordagem centrada no estudante e orientada ao grupo, destacam-se as interações entre os pares que constituem o fator mais importante na aprendizagem colaborativa, embora sem excluir outros fatores, tais como o material de aprendizagem e interação com os professores (DILLENBOURG; JÄRVELÄ; FISCHER, 2009).

Dillenbourg (1999) numa definição mais ampla descreve a aprendizagem colaborativa como uma situação em que duas ou mais pessoas aprendem ou tentam aprender algo juntas.

2.1.1 Computer-Supported Collaborative Learning (CSCL)

A aprendizagem colaborativa apoiada por computador (do inglês, *Computer-Supported Collaborative Learning* - CSCL) estuda como as pessoas aprendem juntas com o auxílio dos computadores. A CSCL surgiu a partir de três projetos preliminares: o Projeto ENFI, da Universidade Gallaudet, o projeto CSILE, da Universidade de Toronto e o Projeto *Fifth Dimension*, da Universidade da Califórnia em San Diego. Esses três projetos trabalharam com o uso da tecnologia para apoio a aprendizagem, o que futuramente veio a compor o que é conhecido por CSCL (STAHL; KOSCHMANN; SUTHERS, 2006).

Na década de 1990 ocorreu a ascensão do CSCL, estimulada pela intensificação do uso da Internet para conectar pessoas de formas inovadoras, incluindo a comunicação síncrona e assíncrona, e a possibilidade dos estudantes estarem geograficamente distribuídos (RESTA; LAFERRIÈRE, 2007). Em 1995 ocorreu a primeira conferência propriamente de CSCL, que foi organizada na Universidade de Indiana. A partir dessa primeira conferência, foram surgindo outras conferências pelo menos a cada dois anos (STAHL; KOSCHMANN; SUTHERS, 2006).

A CSCL é um campo de pesquisa interdisciplinar, e vem emergindo como um processo dinâmico, focado em como a tecnologia pode facilitar a criação e o compartilhamento de conhecimentos e competências através da interação entre pares e dos processos de aprendizagem em grupo. Pesquisadores de áreas como ciência cognitiva, ciências da aprendizagem (psicologia, ciência da computação, educação), psicologia educacional, comunicação, inteligência artificial, e tecnologia da informação (sistemas de apoio de grupo), entre outros, têm se envolvido em pesquisas de CSCL. As pesquisas em CSCL são variadas, utilizando diferentes abordagens e metodologias, considerando uma diversidade de contextos de aprendizagem e domínios de conhecimento, a complexidade e a duração das tarefas de aprendizagem, o tipo e tamanho dos grupos, o número de participantes, entre outros (RESTA; LAFERRIÈRE, 2007).

O objetivo básico da CSCL é investigar como utilizar as tecnologias computacionais para melhorar os processos de aprendizagem, facilitando a aprendizagem colaborativa de maneira eficaz (STAHL; KOSCHMANN; SUTHERS, 2006).

Resta e Laferrière (2007) identificaram quatro motivos para o uso da CSCL:

- Preparar os estudantes para a sociedade do conhecimento (habilidades de colaboração e criação de conhecimento);
- Melhorar o desempenho cognitivo do estudante ou promover uma compreensão mais profunda;
- Adicionar flexibilidade de tempo e espaço para a aprendizagem cooperativa/colaborativa;
- Promover o envolvimento dos estudantes e manter o controle do trabalho cooperativo/colaborativo do estudante (discurso *online*).

A prática da CSCL tem demonstrando ser benéfica em diversos contextos educacionais (STAHL; KOSCHMANN; SUTHERS, 2006). Vários benefícios são atribuídos a CSCL: o desenvolvimento de habilidades de pensamento de ordem superior, a satisfação dos estudantes com a experiência de aprendizagem e melhoria da produtividade (RESTA; LAFERRIÈRE, 2007), qualidade de argumentação em discussões *online* e elaboração cognitiva (STEGMANN et al., 2012), entre outros. Pesquisas em relação ao desempenho acadêmico (JOHNSON; JOHNSON, 1989; JOHNSON; JOHNSON; SMITH, 1998) descobriram que o uso da CSCL teve maior rendimento em quantidade e qualidade da conquista diária, maior domínio da informação factual, e maior sucesso na resolução de problemas de aprendizagem do que a aprendizagem individual com suporte por computador.

No ensino superior, estudantes de especialização que se envolveram em aprendizagem colaborativa utilizando um sistema de apoio à decisão do grupo obtiveram notas mais elevadas do que os estudantes do outro grupo que participou do experimento sem se envolver em atividades de aprendizagem colaborativa (ALAVI, 1994). Pesquisas em relação aos produtos gerados individualmente e os produtos gerados pelo grupo afirmam que grupos *online* entregam trabalhos mais completos, tomam decisões de maior qualidade e têm melhor desempenho em tarefas que exigem geração de ideias em grupo (BENBUNAN-FICH; HILTZ; TUROFF, 2003; FJERMESTAD, 2004).

Um estudo explorou a relação entre a argumentação em discussões *online*, a elaboração cognitiva e a aquisição de conhecimento individual. Esse estudo, num experimento com 48 participantes descobriu que a CSCL pode promover a qualidade formal da

argumentação, bem como a correspondente elaboração cognitiva para a construção dos argumentos. A construção de argumentos está positivamente relacionada à elaboração cognitiva profunda e à aquisição individual de conhecimentos envolvidos na argumentação (STEGMANN et al., 2012).

Portanto, observa-se que uso da tecnologia em ambientes educacionais para facilitar a aprendizagem colaborativa favorece o processo de ensino-aprendizagem. Vários sistemas de aprendizagem colaborativa foram desenvolvidos buscando melhorar a interação das pessoas, a comunicação, a coordenação das atividades e o processo de ensino-aprendizagem (APARICIO; BACAO, 2013). Aparicio e Bacao (2013) consideram que os AEs são um conceito em evolução e apontam os diferentes focos dados aos sistemas desde os primeiros *Computer-Assisted Instruction* (CAIs) até os mais recentes *Massive Open Online Course* (MOOCs).

2.1.2 Massive Open Online Courses (MOOCs)

Seguindo a trilha do desenvolvimento tecnológico, a utilização das TICs na educação, além da adoção da EaD, que proporciona a flexibilidade tempo-espço, e do movimento de abertura da EaD, surgem os *Massive Open Online Courses* (MOOCs) ou Cursos Online Abertos e Massivos.

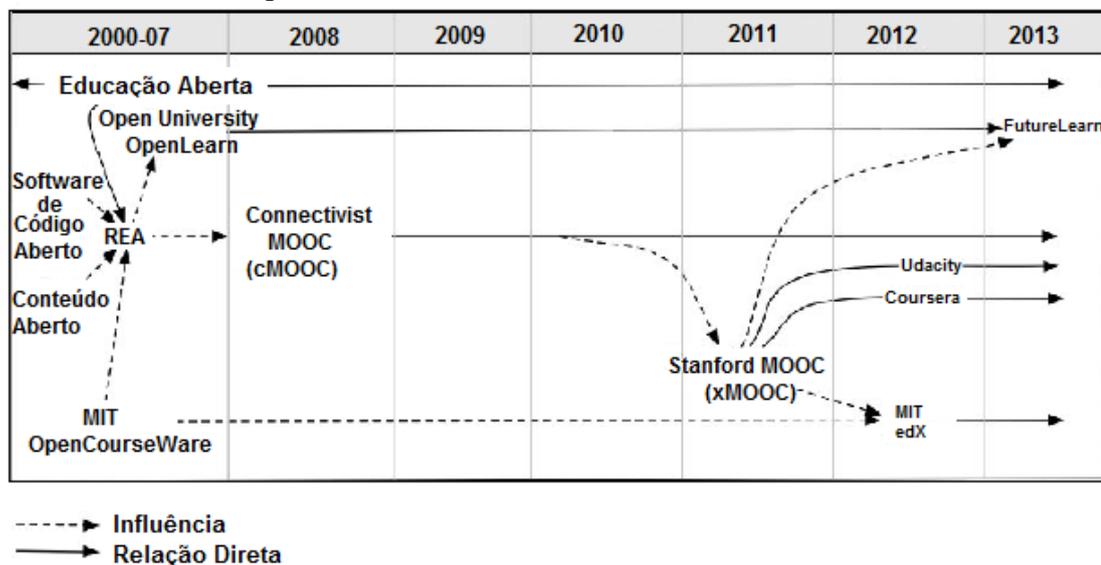
Os MOOCs têm recebido atenção da mídia, profissionais da educação e empresas de tecnologia. A promessa dos MOOCs é que eles vão proporcionar livre acesso ao ensino superior, podendo reduzir o custo da educação nesse nível e romper com os modelos pedagógicos existentes. Há um crescente impulso para as instituições de nível superior participarem do movimento de abertura da educação. A rápida expansão dos MOOCs despertou o interesse comercial dos analistas de risco e das grandes corporações que querem entrar no mercado do ensino superior através de uma abordagem MOOC. Mais do que isso, abriu discussões estratégicas sobre a potencial ruptura dos MOOCs no ensino superior e forçou as instituições de ensino superior a considerarem a aprendizagem *online* e a educação aberta como opções estratégicas para o futuro (YUAN; POWELL, 2013).

O termo MOOC surgiu em 2008, no Canadá, quando Dave Cormier e Bryan Alexander cunharam a sigla para descrever um curso *online* aberto na Universidade de Manitoba, desenhado por George Siemens e Stephen Downes. O curso, Conectivismo e Conhecimento Conectivo (CCK08), foi apresentado a 25 estudantes matriculados na universidade e a 2.300 outros estudantes do público em geral *online* gratuitamente (DANIEL, 2012; YUAN; POWELL, 2013). Os MOOCs têm raízes no movimento dos REAs e do

Conectivismo, modelo de aprendizagem que reconhece que a aprendizagem na sociedade atual não é uma atividade individualista interna, mas a aprendizagem, o conhecimento e a compreensão passam pela conexão em redes pessoais de aprendizagem (DOWNES, 2006; SIEMENS, 2004).

A Figura 3 mostra como o conceito de abertura na educação evoluiu fora do Brasil. O *Massachusetts Institute of Technology* (MIT) criou o *OpenCourseWare* em 2002 e a *Open University* (OU) criou o *OpenLearn* em 2006, seguindo em um desenvolvimento contínuo do movimento de educação aberta e influenciado pelos MOOCs, criados a partir de 2008, várias plataformas de aprendizagem abertas foram criadas para ofertar cursos MOOCs, como por exemplo, *FutureLearn*, *Udacity*, *Coursera*, *MIT EdX*, entre outras (YUAN; POWELL, 2013).

Figura 3 – Linha do tempo dos MOOCs



Fonte: Adaptado de Yuan e Powell (2013)

As plataformas geralmente fazem parcerias com as universidades para a distribuição dos cursos. Esta parceria é firmada baseada em requisitos definidos pelas plataformas e contratos firmados, pois a maioria delas são empresas privadas. No entanto, a plataforma *EdX*¹ disponibilizou uma versão *open source*, que pode ser utilizada por qualquer universidade para disponibilizar seus cursos, desde que possua os recursos necessários para a instalação e manutenção da mesma.

A Figura 3 também mostra duas distintas classificações dos MOOCs, que seguem direções pedagógicas distintas e que são divididos em dois tipos, conhecidos como cMOOCs e xMOOCs (DANIEL, 2012). Os MOOCs conectivistas são baseados na teoria do conectivismo, da aprendizagem em redes desenvolvidas informalmente, enfatizam o estar

¹ Disponível em <http://code.edx.org/>

conectado, a aprendizagem colaborativa e os cursos são construídos em torno do grupo. Os xMOOCs são baseados em conteúdo, seguindo uma abordagem mais behaviorista, com apresentações de vídeo, questionários curtos e testes (YUAN; POWEL, 2013).

Em relação aos tradicionais cursos *online*, os MOOCs têm duas características-chave: o acesso aberto - qualquer pessoa pode participar de um curso *online* gratuito e a escalabilidade - são projetados para suportar um número indefinido de participantes (YUAN; POWELL, 2013).

Em pesquisas iniciais fez-se uma análise da colaboração em AEs, inclusive em MOOCs, e observou-se que há uma deficiência no suporte à colaboração tanto nos AEs selecionados quanto nos MOOCs, apontando a necessidade desses ambientes de ensino--aprendizagem promoverem mais a colaboração. No Apêndice A encontram-se os trabalhos desenvolvidos sobre a colaboração em *Massive Open Online Courses* (MOOCs). Para promover a colaboração e a aprendizagem colaborativa, a primeira necessidade é a formação de grupos, discutida na seção 2.2.

2.2 FORMAÇÃO DE GRUPOS

Com a abordagem centrada no estudante e orientada ao grupo, na aprendizagem colaborativa as interações entre os integrantes do grupo destacam-se como um dos fatores mais importantes (CASTRO; MENEZES, 2011; DILLENBOURG; JÄRVELÄ; FISCHER, 2009). Kirschner (2001) destaca que um dos elementos da aprendizagem colaborativa são as atividades desenvolvidas em pequenos grupos. Assim, uma das primeiras necessidades e um tema que vem sendo estudado em CSCL é a formação dos grupos para aprendizagem colaborativa, pois o processo de aprendizagem se dá no processo do grupo (STAHL; KOSCHMANN; SUTHERS, 2006).

A formação de grupos é uma tarefa importante, pois deve permitir a efetiva interação dos membros de um grupo. Por isso, criar condições para que uma efetiva interação ocorra no grupo é um dos propósitos da CSCL (DILLENBOURG; JÄRVELÄ; FISCHER, 2009), na qual os estudantes não são apenas indivíduos, mas sim membros de um grupo (STAHL; KOSCHMANN; SUTHERS, 2006).

Perguntas como: Qual o tamanho ideal de um grupo? Grupos homogêneos são melhores que grupos heterogêneos? Que dados considerar para a formação de grupos? Deve-se juntar meninos e meninas?, entre outras, têm inspirado diversas pesquisas em aprendizagem colaborativa. Essa questão em relação à formação do grupo “ideal” para que ocorra efetivamente a colaboração e a aprendizagem, no entanto, é uma questão difícil, visto

que as características e condições iniciais do grupo interagem entre si de forma complexa. A heterogeneidade do grupo, por exemplo, tem efeito diferente para diferentes tarefas (DILLENBOURG, 1999).

A formação do grupo influencia diretamente a capacidade de realização das tarefas designadas ao grupo. Segundo Minicucci (1997), a formação intelectual é privilegiada quando existe colaboração entre os participantes de um grupo. A alocação adequada, focando no aumento da aprendizagem entre os envolvidos é o que se busca com uma formação adequada do grupo (ISOTANI, 2009). Uma formação de grupos inadequada pode causar falha no processo de aprendizagem (GRAF; BEKELE, 2006), tais como: participação desigual, desinteresse em efetuar as tarefas designadas, desigualdade de ritmo entre os membros do mesmo grupo e resistência ao trabalhar em grupos são apontados em diversas pesquisas (BARKLEY; CROSS; MAJOR, 2005; DILLENBOURG, 2002; ISOTANI; MIZOGUCHI, 2008).

A formação de grupos tem sido estudada por décadas, não apenas em CSCL, mas também em áreas como psicologia, sociologia e educação (OWENS; MANNIX; NEALE, 1998). Na literatura, estudos indicam que formar grupos heterogêneos é mais vantajoso que grupos homogêneos (BROPHY, 1998; CHAN, 2006; JANIS, 1971, WEBB, 1983; WEBB, 1992).

Janis (1971) afirma que em grupos homogêneos, cujos membros têm pontos de vista parecidos, as decisões fluem mais facilmente. No entanto, a efetividade das decisões é prejudicada uma vez que diferentes opiniões, perspectivas e pontos de vista deixaram de ser analisados.

Webb (1992) aponta que o rendimento de estudantes com menor habilidade, isto é, aqueles mais fracos em um determinado assunto, melhora em grupos heterogêneos. Isso ocorre porque estes estudantes recebem explicações mais elaboradas de seus colegas mais hábeis. Além disso, os estudantes mais hábeis também se beneficiam em grupos heterogêneos, pois ao explicarem o conteúdo aos colegas que têm maior dificuldade, eles precisam reorganizar as ideias de modo a tornar as informações mais claras em diferentes aspectos, o que leva ao desenvolvimento da metacognição (WEBB, 1983).

Brophy (1998) sugeriu que grupos com diferentes preferências e habilidades, conhecimentos e configurações serão mais eficientes durante a resolução de problemas.

Especialmente em aprendizagem colaborativa, os grupos heterogêneos são frequentemente formados com vistas a criar interdependência mútua entre os membros,

fazendo com que eles aprendam ajudando uns aos outros através das interações (CHAN, 2006).

2.2.1 Formação de Grupos para Aprendizagem

A formação de grupos é feita de diferentes maneiras e considerando diferentes abordagens, critérios e características ou dados dos estudantes. Por exemplo, o professor determina os grupos conforme algum critério estabelecido, ou aleatoriamente. Ou ainda delega a formação dos grupos para os próprios estudantes.

Wessner e Pfister (2001) falam que a formação de grupos para aprendizagem colaborativa pode ser feita “fora” ou “dentro” do sistema. De forma geral, se o grupo é formado “fora”, então deverá ser informado ao sistema quem pertence a cada grupo. Se o grupo é formado “dentro”, então o sistema identifica quem pertence a cada grupo. Adaptou-se essa classificação para a *forma* como o grupo é gerado, utilizando *manual* (“fora”) ou *automático* (“dentro”). Se o grupo for *manual*, então é gerado externamente e informado posteriormente, ou é gerado diretamente no sistema, no entanto, de forma manual pelo professor ou instrutor. Se o grupo é *automático*, então é gerado automaticamente pelo sistema, que identifica os seus integrantes através de parâmetros e critérios definidos anteriormente.

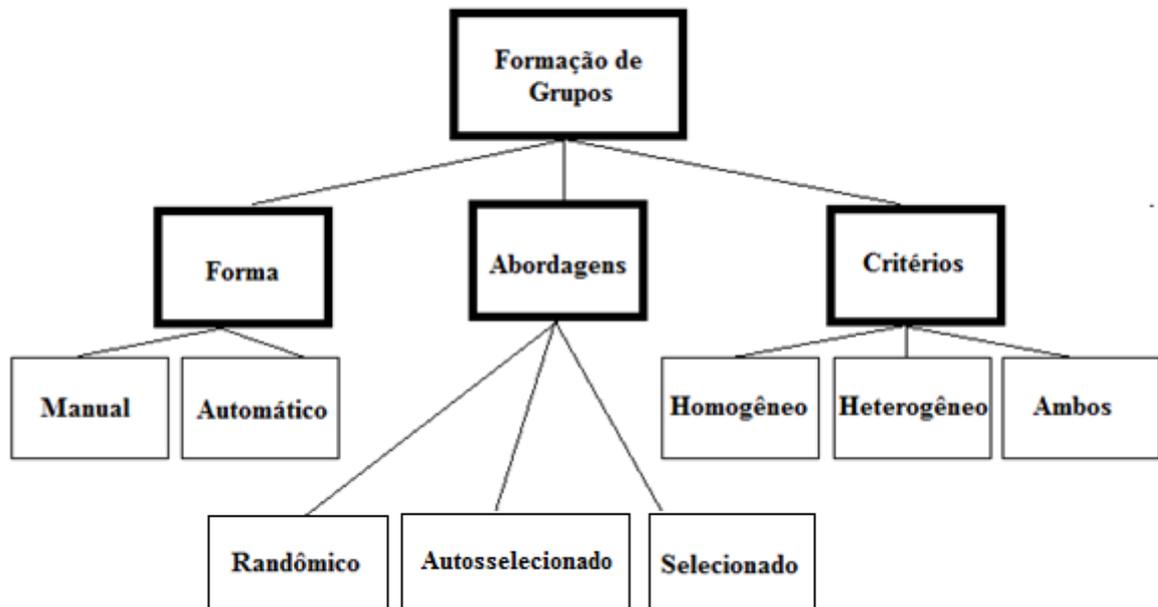
Ounnas, Davis e Millard (2007) definem três diferentes *abordagens* para a formação de grupos: *randômico*, *autosselecionado* e *selecionado* pelo professor. *Randômico*, quando o professor ou o sistema gera o grupo aleatoriamente, distribuindo os estudantes em cada grupo sem algum critério específico definido. *Autosselecionado*, quando os estudantes negociam a formação de grupos entre si, ou quando o grupo é formado de acordo com o interesse do estudante em algum tópico, por exemplo. *Selecionado*, quando o professor ou sistema determina os parâmetros e critérios para a formação de grupos. Essas definições são compartilhadas na literatura, com algumas variações. Bekele (2005), por exemplo, considera que existem quatro maneiras de formar grupos: randômico, delegada aos estudantes (autosselecionado), baseada em critério definido (selecionado) e designação por tarefa. Na designação por tarefas, o professor oferece aos estudantes uma lista de tópicos a serem escolhidos e, de acordo com as preferências individuais de cada um, os grupos são gerados, reunindo os estudantes que escolheram o mesmo tópico para ser abordado. Nesta pesquisa adotaram-se as três abordagens definidas por Ounnas, Davis e Millard (2007), considerando-se que a escolha de um tópico de interesse do estudante é também uma forma de autosseleção.

Além dessas abordagens, Ounnas (2010) refere-se aos seguintes critérios para a formação de grupos: homogêneo, heterogêneo ou baseado em alguma regra específica.

Gogoulou et al. (2007) apresenta um algoritmo genético para a formação de grupos homogêneos, heterogêneos ou mistos. Adaptando essas classificações, utilizam-se como *critério* grupos *homogêneos, heterogêneos ou ambos (mistos)*. *Homogêneo*, quando o grupo adota estudantes de características similares para a formação de grupos. *Heterogêneo*, quando o grupo adota estudantes de características diferentes para a formação de grupos. *Ambos*, quando o grupo for *misto*, isto é, adotar estudantes homogêneos em alguma(s) característica(s) e heterogêneos em outras, ou quando permite formar tanto grupos homogêneos quanto grupos heterogêneos. As características ou dados dos estudantes utilizadas para a formação de grupos dependem da disponibilidade de dados existentes no sistema, ou dos dados aos quais o professor tem acesso, no caso de grupo gerado de forma manual.

A Figura 4 apresenta graficamente as classificações discutidas nesta seção, de forma a facilitar o entendimento do mapeamento sistemático, que utilizou essas classificações de *forma, abordagens e critérios* para o mapeamento sistemático sobre formação de grupos na aprendizagem colaborativa, discutido a seguir.

Figura 4 – Classificações adotadas para a formação de grupos



Fonte: Próprio autor

2.2.2 Mapeamento Sistemático sobre Formação de Grupos na Aprendizagem Colaborativa

No decorrer dos estudos sobre formação de grupos para aprendizagem colaborativa, algumas perguntas apareceram. Além da *forma*, das *abordagens* e dos *critérios* adotados para a formação do grupo (conforme descrito na seção 2.2.1), posto que o problema da formação de grupos é um problema *NP-hard*, foi necessário identificar quais abordagens de solução os algoritmos utilizavam para esse problema e que dados estavam sendo utilizados para a formação de grupos. Desta forma, seguiu-se o mapeamento sistemático da literatura para identificar também estas questões, publicado em Citadin, Kemczinski e De Matos (2014). Para conduzir o mapeamento sistemático foi utilizado o modelo proposto por Petersen et al. (2008). Esse modelo é composto por cinco etapas: 1) Questões de Pesquisa, 2) Condução da Busca, 3) Análise dos Artigos, 4) Classificação dos Artigos e 5) Extração dos dados e Mapeamento, conforme apresentado na Figura 1, na seção 1.5.

Na primeira etapa, as cinco questões de pesquisa propostas foram:

- **QP1:** Quais as maneiras ou *formas* utilizadas para a formação de grupos?
- **QP2:** Quais as *abordagens* utilizadas?
- **QP3:** Quais os *critérios* utilizados?
- **QP4:** Quais os *algoritmos* utilizados?
- **QP5:** Quais os *dados* ou características dos estudantes utilizados para a formação de grupos?

Na segunda etapa, seguiu-se a condução da busca nos seguintes Mecanismos de Busca Acadêmica (MBA): *Web of Science*, *Scopus*, *IEEE Xplore*, *ACM*, *Science Direct*, *Springer* e *Periódicos da CAPES*, de forma a identificar o maior número possível de artigos para o estudo. Foram utilizadas duas expressões principais para a condução da busca, ilustradas a seguir nas Figuras 5 e 6, respectivamente, onde o asterisco (*) é o caractere coringa; e com o intuito de buscar artigos referentes a trabalhos desenvolvidos com quantidades maiores de estudantes (massa) e aplicados em ambientes virtuais de aprendizagem (*Learning Management System – LMS*).

Devido à condução da busca em sete diferentes MBAs, a expressão foi adaptada em cada um deles, de acordo com as necessidades e a disponibilidade de cada ferramenta em relação à expressão adotada, sem, no entanto, perder a ideia principal da formação de grupos para aprendizagem colaborativa. Além disto, a busca foi realizada considerando campos específicos como títulos, resumos e palavras-chave, de acordo com os campos disponíveis em

cada MBA; e limitada ao período de 2004 a 2014. Mais detalhes são encontrados em Citadin, Kemczinski e De Matos (2014), no Apêndice A desse documento.

Figura 5 – Expressão de busca utilizada para massa

```
( ((team OR group) NEAR/AND (formation OR composition)) AND
("collaborative learning" OR CSCL) AND ((massi* OR large*) NEAR/AND (team
OR group OR class)) )
```

Fonte: Próprio autor

Figura 6 – Expressão de busca utilizada para LMSs

```
( ((team OR group) NEAR/AND (formation OR composition)) AND
("collaborative learning" OR CSCL) AND (("learning management system" OR
LMS) OR ("virtual learning environment" OR VLE) OR ("elearning" OR "e-
learning" OR "online learning" OR "distance learning")) )
```

Fonte: Próprio autor

Na terceira etapa, seguiu-se a análise dos artigos selecionados, seguindo os critérios de inclusão:

- **CI1:** Artigos escritos em inglês;
- **CI2:** Artigos disponíveis para *download* (de acordo com a assinatura da universidade) na íntegra em PDF (*Portable Document Format*);
- **CI3:** Artigos primários (mapeamentos e/ou revisões sistemáticas e artigos teóricos não foram considerados como primários);
- **CI4:** Artigos completos com mais de 4 páginas (notas ou artigos não completos - *short papers* - foram desconsiderados).

Dos artigos selecionados, foram excluídos todos aqueles que satisfizeram um ou mais dos critérios de exclusão, definidos a seguir:

- **CE1:** Artigos duplicados;
- **CE2:** Artigos em que a sigla CSCL não significar *Computer-Supported Collaborative Learning*;
- **CE3:** Artigos nos quais a expressão de busca seja satisfeita com os termos presentes apenas nas referências, biografia do autor, agradecimentos ou conclusão;
- **CE4:** Artigos que não descrevam a formação de grupos como suporte para aprendizagem colaborativa.

Após estas etapas, com um total de 27 trabalhos selecionados, seguiram-se as quarta e quinta etapas, com a classificação, extração dos dados e mapeamento. Considerando as classificações mencionadas de *forma*: manual ou automático; de *abordagens*: randômico, autosselecionado e selecionado; e de *critérios*: homogêneo, heterogêneo, ambos (misto), foram tabulados os artigos selecionados no mapeamento. Para *critérios*, observou-se a necessidade de estabelecer ainda a opção *outros* para quando não se pôde classificar ou identificar o critério adotado (por exemplo, gerado randomicamente pelo professor).

Além disso, percebeu-se que além do critério, conforme a classificação de homogêneo ou heterogêneo adotada, outros critérios são utilizados para a formação de grupos, em conjunto com esta classificação. Por exemplo, homogêneo ou heterogêneo em relação a outro critério como estilo de aprendizagem ou personalidade, entre outros. Assim, esses dados foram classificados como *critérios complementares*. Também foram tabulados os algoritmos e os dados relacionados utilizados para o agrupamento dos estudantes (por exemplo, modelo do estudante ou dados do sistema).

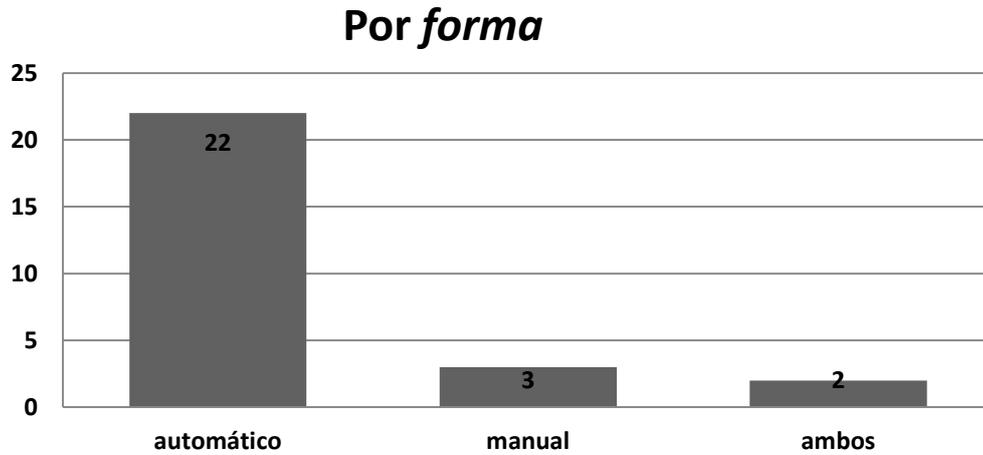
Os resultados obtidos pelo mapeamento apontam que a *forma* mais utilizada é o grupo **automático**. Assim como a *abordagem* mais utilizada é o grupo **selecionado**, e o *critério* mais utilizado é o **heterogêneo** e ambos (misto - tanto homogêneo quanto heterogêneo).

Observou-se também que uma variedade de critérios complementares foi utilizada, como, por exemplo, estilo de aprendizagem, papéis designados aos estudantes, conhecimento, competências, preferências, interações, entre outros. Esses critérios complementares dependem dos dados que se têm disponível no sistema para a formação automática dos grupos.

Em relação aos algoritmos, houve também uma diversidade de aplicações (algoritmos genéticos, enxame de partículas, colônia de abelhas, entre outros), no entanto, observou-se que algoritmos genéticos, que são uma categoria de algoritmos inseridos no escopo de algoritmos evolutivos, foram um dos mais utilizados, resultado corroborado por Cruz e Isotani (2014) que também fizeram um mapeamento sistemático.

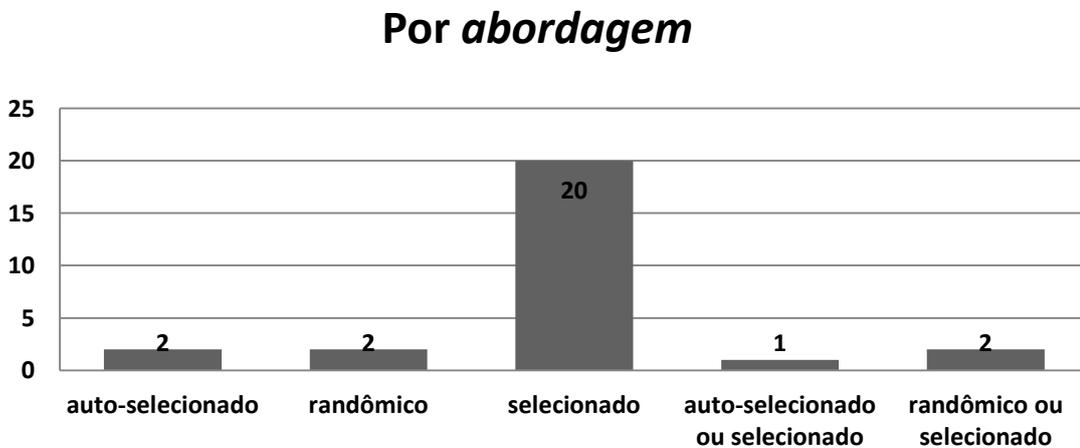
As Figuras 7, 8, 9, 10 e 11 apresentam os resultados do mapeamento e são discutidas a seguir.

Na Figura 7, observa-se que a forma de grupo **automático** foi a mais utilizada, com mais de 80% dos trabalhos classificados desta forma.

Figura 7 – Formação de grupos por *forma*

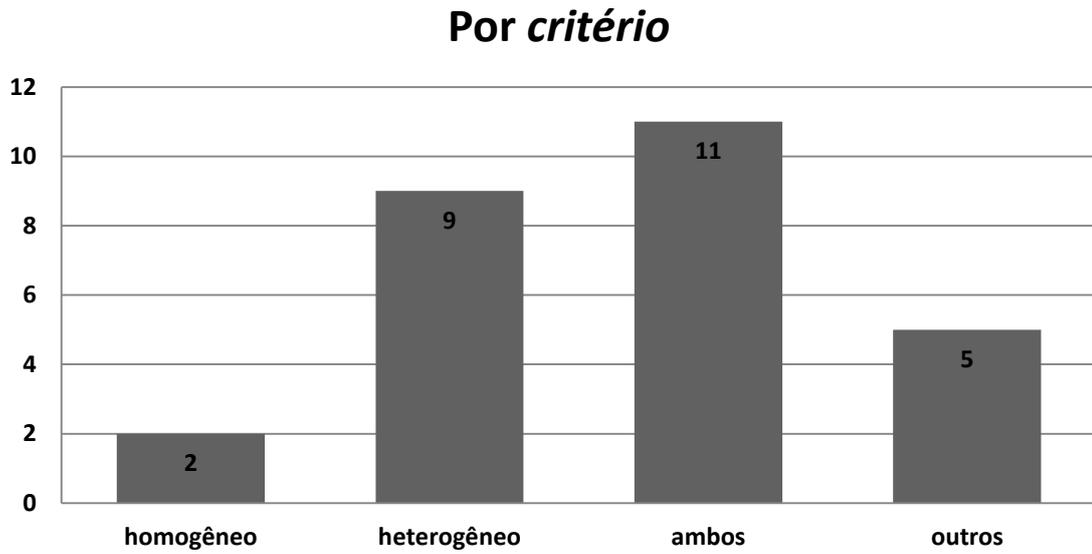
Fonte: Próprio autor

Na Figura 8, observa-se que a abordagem de grupo **selecionado** foi a mais utilizada, sendo que 74% dos trabalhos utilizaram essa abordagem. E outros três trabalhos utilizaram mais de uma abordagem, além de selecionado, totalizando 85% dos trabalhos.

Figura 8 – Formação de grupos por *abordagem*

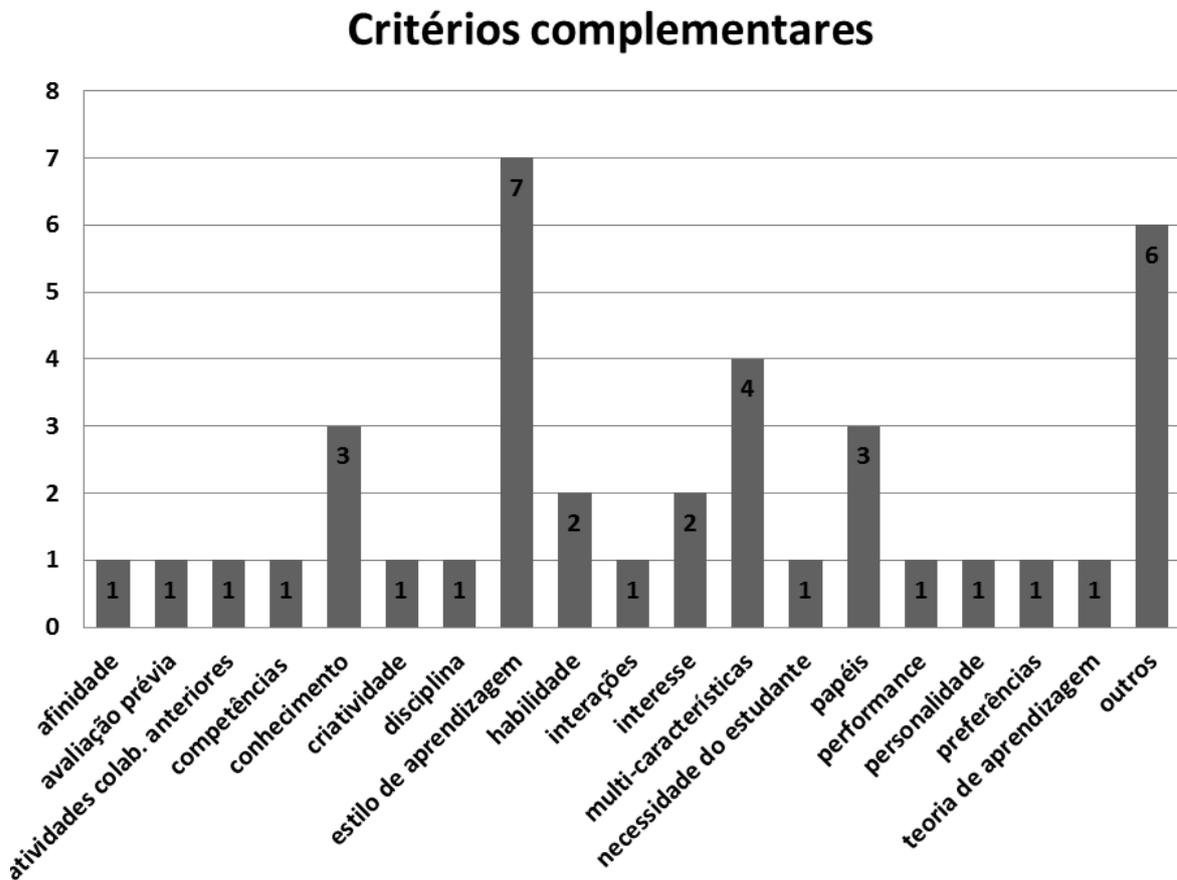
Fonte: Próprio autor

Em relação ao critério utilizado para a formação de grupos, observa-se que **heterogêneo** e ambos (tanto homogêneo quanto heterogêneo) foram os critérios mais utilizados, com 74% dos trabalhos classificados nesses critérios. Outros cinco trabalhos foram classificados como outros, por não ser possível identificar o critério utilizado na formação do grupo, conforme mostra a Figura 9.

Figura 9 – Formação de grupos por *critério*

Fonte: Próprio autor

Figura 10 – Critérios complementares utilizados



Fonte: Próprio autor

A Figura 10 ilustra a variedade de critérios complementares possíveis para a formação de grupos, conforme os trabalhos analisados. Os termos utilizados foram utilizados de acordo

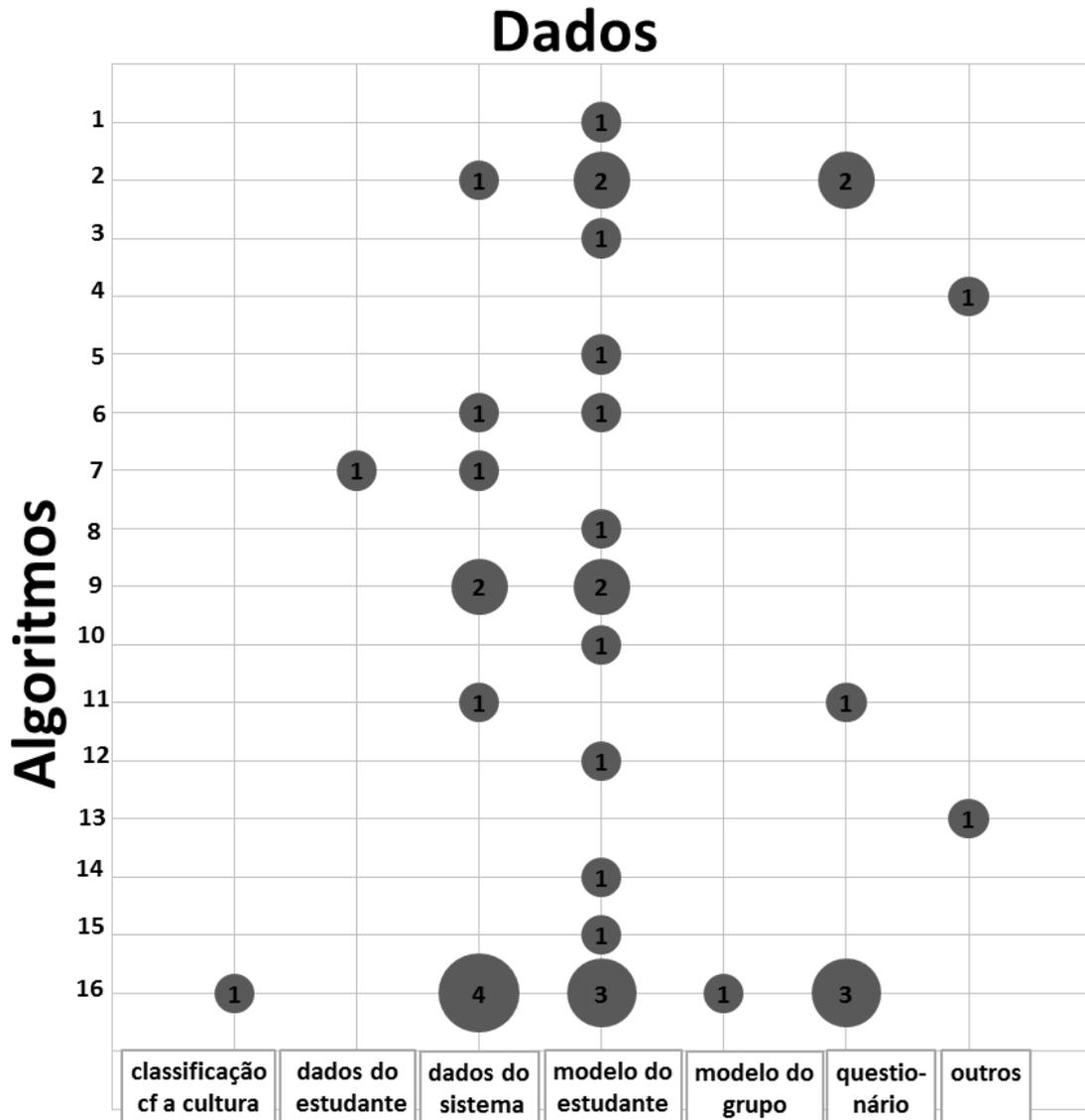
com os termos mencionados nos trabalhos. Isto é, se o trabalho analisado mencionava conhecimento, por exemplo, foi tabulado como conhecimento; quando outro trabalho mencionava conhecimento também foi tabulado como conhecimento, independentemente das definições que cada um adotou para conhecimento. Observa-se que estilo de aprendizagem destaca-se entre os trabalhos analisados, com sete dos trabalhos mencionando esse critério ($\approx 26\%$). Em segundo lugar estão quatro trabalhos que mencionam ser multicaracterísticas, isto é, utilizam várias características para a formação de grupos; e em terceiro lugar têm-se três trabalhos que mencionam o conhecimento prévio do estudante, e três trabalhos que utilizam papéis designados aos estudantes no grupo como critérios para a formação de grupos. Desta forma, observa-se que a formação de grupos pode ser bastante variada em relação aos critérios complementares utilizados. Em *outros* estão contados os trabalhos em que não foi possível identificar o critério complementar utilizado.

A Figura 11 ilustra a relação de algoritmos e dados, sendo que os algoritmos foram numerados de 1 a 16, representados nas linhas horizontais do gráfico; e os dados, representados pelas sete linhas verticais do gráfico, são descritos nos sete quadros das linhas verticais, respectivamente. Os termos utilizados foram adotados de acordo com o mencionado nos trabalhos analisados, por exemplo, classificação conforme a cultura (quadro 1) foi utilizada no trabalho de Popov et al. (2014). Em *outros* foram classificados os trabalhos em que não foi possível identificar a origem dos dados claramente.

Abaixo da Figura 11 segue a legenda dos algoritmos para facilitar a identificação. Em relação aos algoritmos, observa-se que **algoritmos genéticos** e **otimização por enxame de partículas** se destacam com mais de uma ocorrência e também a classificação como *outros algoritmos*, para quando não foi especificado o algoritmo utilizado.

Em relação aos dados utilizados, destacam-se as colunas de **dados do sistema**, onde foram tabulados os trabalhos que mencionavam coletar dados do sistema ou plataforma, como o *Moodle*, por exemplo, e/ou da navegação do estudante no sistema, ou atividades do estudante na plataforma, entre outros; e de **modelo do estudante**, onde foram tabulados dados do modelo do usuário e/ou do estudante como perfil do estudante, preferências, nível de conhecimento, competências, habilidades, dados acadêmicos, portfólio, entre outros, de acordo com o mencionado em cada trabalho.

Figura 11 – Relação algoritmos x dados



Legenda Algoritmos:

1. algoritmo evolutivo (aglomeração determinística)
2. algoritmo genético
3. autoaprendizado - regressão linear múltipla
4. clusterização
5. colônia de abelhas
6. framework
7. grafo
8. Hete-A
9. otimização por enxame de partículas
10. profundidade dos dados
11. protótipo 3D
12. random
13. scripts IMS-LD
14. semipareto optimal set
15. VALCAM
16. Outros algoritmos

Fonte: Próprio autor

2.3 ALGORITMOS EVOLUTIVOS

Um campo importante de pesquisa relacionado à otimização são os Algoritmos Probabilísticos (*Probabilistic Algorithms*). Trabalhos na literatura utilizam algoritmos probabilísticos para a formação de grupos (BEKELE, 2005; GRAF; BEKELE, 2006; MORENO; OVALLE; VICARI, 2012; WANG; LIN; SUN, 2007; YANNIBELLI; AMANDI, 2012). A Figura 11 apresenta também uma variedade desses algoritmos (algoritmos evolutivos, algoritmos genéticos, enxame de partículas, entre outros) utilizados para a formação de grupos, conforme identificado nos trabalhos analisados pelo mapeamento sistemático. Esses algoritmos efetuam, em tempo de execução, trocas na solução atual para uma que seja considerada melhor. A nova solução encontrada não é a melhor possível, mas essa mudança, de uma solução inferior para uma solução superior, pode ser considerada como uma melhoria (WEISE, 2009). Em sua taxonomia, Weise (2009) aponta a Computação Evolutiva como uma das áreas que estudam esses algoritmos probabilísticos. A Computação Evolutiva é uma área da Ciência da Computação que, como o nome sugere, se baseia nos processos naturais de evolução, propostos por Darwin, através da sua Teoria da Evolução das Espécies (DARWIN, 1859; EIBEN; SMITH, 2007).

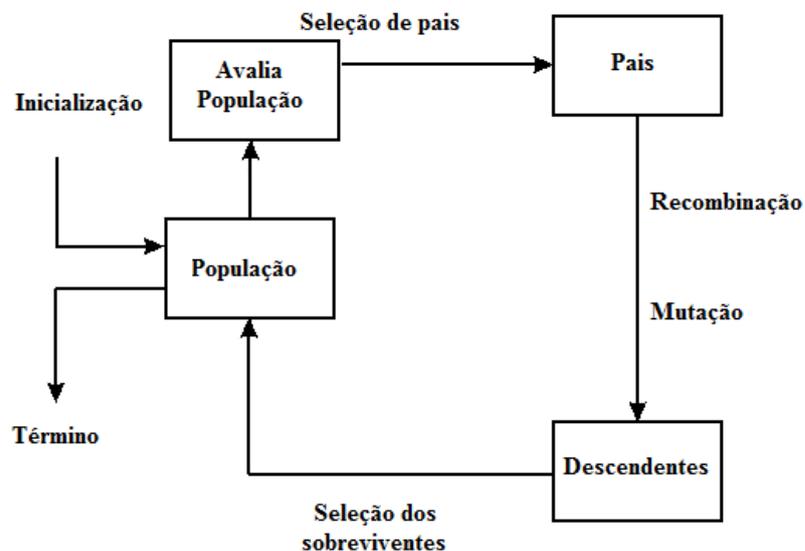
2.3.1 Visão Geral

Usar os princípios de Darwin para solução de problemas não é uma novidade. Em torno de 1948 Turing propôs uma “pesquisa genética ou evolutiva” e, por volta de 1962, Bremermann executou experimentos computacionais de “otimização através de evolução e recombinação” (EIBEN; SMITH, 2007). Durante a década de 1960, três iniciativas em diferentes lugares formaram a base da Computação Evolutiva. Nos Estados Unidos (USA), Fogel, Owens e Walsh introduziram a ideia de **programação evolutiva**, enquanto Holland chamou seu método de **algoritmo genético**. Na Alemanha, Rechenberg e Schwefel inventaram as **estratégias evolutivas** (EIBEN; SMITH, 2007). Durante anos essas áreas desenvolveram-se separadamente, mas, em torno de 1990, passaram a ser parte de uma área conhecida como Computação Evolutiva. Ainda em torno de 1990, uma quarta abordagem surgiu, denominada **programação genética**. Portanto, essas quatro vertentes formam a computação evolutiva como um todo, e os algoritmos são chamados algoritmos evolutivos (EIBEN; SMITH, 2007).

A ideia geral dos algoritmos evolutivos, baseada na teoria da evolução, é que dada uma população de indivíduos em um ambiente com recursos limitados, a competição por estes

recursos causa a sobrevivência dos mais aptos (seleção natural) (EIBEN; SMITH, 2007; LINDEN, 2012). Os “mais aptos” são determinados pela função de avaliação, que quantifica a qualidade dos indivíduos como solução para o problema em questão. Dada uma função a ser maximizada, são aplicados os operadores genéticos sobre a população atual para simular a evolução e o processo de sobrevivência dos mais aptos. Esses operadores são a recombinação (ou cruzamento, do inglês, *crossover*) e a mutação, que são aplicados para gerar os descendentes (filhos) da geração atual. A avaliação simula o processo de sobrevivência, em que os “mais aptos” sobrevivem e seus filhos passam a ser a nova população. (EIBEN; SMITH, 2007; LINDEN, 2012). A Figura 12 mostra um fluxograma genérico que representa a estrutura de funcionamento de algoritmos evolutivos. Esses algoritmos iniciam gerando uma população inicial, que passa pelo processo de avaliação para identificar os pais mais aptos. Esses pais selecionados combinam-se no processo de recombinação e mutação, simulando os processos genéticos, de onde nascem os descendentes, que passam a ser a nova população. Esse processo se repete até a condição de término determinada.

Figura 12 – Fluxograma genérico de algoritmos evolutivos



Fonte: Adaptado de Eiben e Smith (2007)

Para realizar esse processo, os algoritmos evolutivos necessitam de componentes básicos como: Representação, População, Função de Avaliação, Seleção dos Pais, Recombinação, Mutação e Seleção dos Sobreviventes (EIBEN; SMITH, 2007). Os algoritmos genéticos são um ramo dos algoritmos evolutivos e o tipo mais conhecido de algoritmo evolutivo (EIBEN; SMITH, 2007; LINDEN, 2012). Os AGs e seus componentes são discutidos a seguir.

2.3.2 Algoritmos Genéticos

Os algoritmos genéticos (AGs) são um ramo dos algoritmos evolutivos, considerados uma técnica de busca aleatória de otimização global, baseados nos processos de evolução natural, que vem sendo usado com sucesso para resolver problemas complexos de otimização (EIBEN; SMITH, 2007; LINDEN, 2012).

O estudo de problemas de otimização é aquele no qual se busca minimizar ou maximizar uma função através da escolha sistemática dos valores de variáveis reais ou inteiras dentro de um conjunto viável. Para isso, usam-se os algoritmos de otimização, que são métodos de busca cujo objetivo é encontrar uma solução ótima (ENGELBRECHT, 2007), seja ela um ótimo local ou global. Para problemas de otimização combinatória, isto é, problemas que não tem solução ótima em tempo polinomial, usam-se heurísticas ou meta-heurísticas, que tendem a achar soluções próximas da ótima (EIBEN; SMITH, 2007; LINDEN, 2012; MELIÁN; PÉREZ; VEGA, 2003).

Segundo Engelbrecht (2007) os AGs foram propostos primeiramente por Fraser e mais tarde por Bremermann e Reed et al., porém um trabalho intensivo sobre o assunto foi realizado por Holland (1975), o que o tornou responsável pela popularização dos AGs, e por isso Holland é geralmente considerado o pai dos AGs.

Por se tratar de um tipo de algoritmo evolutivo, a ideia principal de um AG é gerar indivíduos (cromossomos) dentro de uma população (grupos de soluções). Em seguida, usando operadores genéticos esses são recombinados para produzirem descendentes. Então é feita a avaliação, de modo que os melhores indivíduos tenham maior probabilidade de sobreviverem e gerarem descendentes, tornando as soluções melhores à medida que novos descendentes são gerados (EIBEN; SMITH, 2007; LINDEN, 2012).

2.3.2.1 Terminologias e Características dos Algoritmos Genéticos

Como AGs são inspirados na genética e na teoria da evolução das espécies, há uma analogia entre os termos da biologia e os termos usados para os AGs (LINDEN, 2012). Em sistemas naturais, um grupo de **cromossomos** se combina para formar as características genéticas básicas do **indivíduo**. Em genética, os cromossomos são formados por **genes**, que podem ter valores, denominados de **alelos**. Por sua vez, a posição do gene é chamada **locus**. Em AGs, os termos indivíduos e cromossomos podem ser intercambiáveis, significando a mesma coisa (EIBEN; SMITH, 2007; LINDEN, 2012).

Outros termos importantes são **genoma**, **genótipo** e **fenótipo**. Genoma relaciona-se ao significado do pacote genético (não possui análogo em AG), genótipo é a estrutura do cromossomo e um fenótipo é a interação do conteúdo genético com o ambiente, o que corresponde à avaliação do indivíduo, feita pela função *fitness* do algoritmo. O Quadro 1 apresenta um resumo dessas terminologias.

Quadro 1 – Terminologias dos AGs

Linguagem da ciência natural	AG	Observação
Cromossomo	Indivíduo / string	Representar uma solução possível.
Gene	Característica	Característica apresentada pela solução
Alelo	Valor	Em strings binárias, 0 ou 1, por exemplo.
Locus	Posição	Índice do vetor do cromossomo
Genótipo	Estrutura	Estrutura dentro do GA
Fenótipo	Expressão do Indivíduo no Ambiente	Função de Avaliação
Genoma	Não há correspondência	Não se aplica

Fonte: Adaptado de Linden (2012)

Além dessas terminologias específicas, algumas características dos AGs são citadas. Segundo Linden (2012), por serem técnicas probabilísticas, os AGs podem encontrar soluções diferentes a cada execução do algoritmo. Além disso, os AGs são programas simples, que necessitam somente de informações locais, não sendo necessárias informações derivadas, o que faz com que possam ser aplicados a problemas que os dados são inteiros ou discretos (descontinuidades duras). No caso desta pesquisa, por exemplo, não é possível alocar meio estudante em um grupo; logo, este problema não admite solução real, apenas inteira. Outra característica é que AGs trabalham com grandes populações e seu diferencial é que eles não realizam a busca em todos os pontos possíveis (exaustiva), apenas em um subconjunto deste. Os AGs buscam informações pertinentes ao problema, não realizando apenas buscas aleatórias pelo espaço de soluções, mas seguindo um direcionamento do mecanismo de seleção. E por último, AGs trabalham com uma forma codificada de parâmetros, por isso o esquema de codificação e decodificação desses parâmetros é importante. Isto equivale à representação cromossomial do AG (LINDEN, 2012).

Quadro 2 – Pseudocódigo de um AG básico

1. Inicialize a população de cromossomos (soluções possíveis);
2. Avalie cada cromossomo na população;
- Repita até a condição de término:**
3. Selecione os pais para gerar novos cromossomos;
4. Aplique os operadores de recombinação e mutação a estes pais de forma a gerar os indivíduos da nova geração;
5. Apague os velhos membros da população;
6. Avalie todos os novos cromossomos e insira-os na população;
- Fim condição de término**
7. Retorne a melhor solução encontrada.

Fonte: Adaptado de Eiben e Smith (2007) e Linden (2012)

O funcionamento de um AG pode ser resumido no pseudocódigo apresentado no Quadro 2 (EIBEN; SMITH, 2007; LINDEN, 2012). O processo inicia com a geração da população inicial e a avaliação dessa população. A partir daí começa o processo de evolução do AG, onde os melhores pais são selecionados para que sejam recombinados gerando seus descendentes. Esses descendentes formam a nova população que é avaliada. O processo se repete até a condição de término definida.

2.3.2.2 Representação dos Indivíduos

A representação do indivíduo é fundamental e é o primeiro passo a ser dado para a definição do AG. Ela consiste na maneira como o problema é representado para o AG e deve ser uma maneira viável de o problema ser tratado pelo computador, pois é o *link* entre o problema do “mundo real” com o “mundo do AG” (EIBEN; SMITH, 2007; LINDEN, 2012).

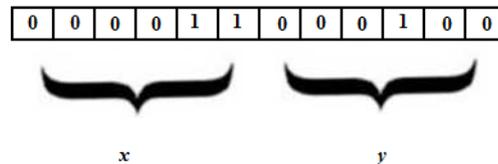
Cada parte indivisível da representação é o **gene**, que são as partes fundamentais que compõem o **cromossomo** (em analogia com a biologia). Linden (2012) também aponta que: a) a representação deve ser a mais simples possível; b) se houver soluções proibidas, é preferível que elas não tenham uma representação e c) se o problema impuser alguma condição, esta deve estar implícita na representação.

A representação dos indivíduos em um AG pode ser feita de diferentes maneiras e é importante escolher a maneira certa para representar o problema desejado (EIBEN; SMITH, 2007; LINDEN, 2012). Dentre as representações possíveis, destacam-se as seguintes:

- **Representação Binária:** essa é a maneira mais utilizada e foi adotada inicialmente por Holland (1975), onde um cromossomo é uma sequência de *bits* e um gene é apenas um *bit*. É possível utilizar a representação binária para representar quantos números se quer, estendendo a representação para mais *bits* ou utilizando menos *bits* para cada número (LINDEN, 2012). Se for

necessário representar duas variáveis x e y utilizando 6 *bits*, por exemplo; e se tem o cromossomo 000011000100, então se pode chegar aos valores de x e y pela decodificação do cromossomo, conforme mostra a Figura 13:

Figura 13 – Representação binária de um cromossomo

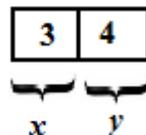


Fonte: Próprio autor

Entretanto, a representação binária não atende certas demandas como, por exemplo, quando os valores possíveis são uma lista de valores. Nesses casos, usa-se a representação inteira (EIBEN; SMITH, 2007).

- **Representação Inteira:** essa representação é usada para dados cujos valores sejam inteiros, incluindo qualquer valor inteiro possível, ou para valores inteiros limitados a certo conjunto de valores como, por exemplo, uma lista finita de valores. Pode inclusive ser usada para representar variáveis categóricas (valores não numéricos) que possam ser representadas por números inteiros; por exemplo, se quiser identificar valores como “alto”, “médio” e “baixo”, utilizam-se as representações {2,1,0} respectivamente. Cada gene será o valor que representa, formando o cromossomo de acordo com a codificação desejada. Usando o mesmo exemplo anterior de x e y , tem-se então o cromossomo conforme mostra a Figura 14:

Figura 14 – Representação inteira de um cromossomo



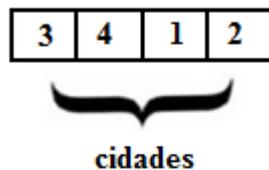
Fonte: Próprio autor

Problemas que, além de considerar valores inteiros, envolvem a ordem em que esses valores devem estar, ou a não repetição dos mesmos na representação, são problemas de permutação, discutidos a seguir.

- **Representação por Permutação:** essa representação envolve duas classes de problemas: uma em que a ordem dos elementos importa e outra em que a ordem não importa. Além disso, a representação por permutação deve manter a

propriedade da permutação de que cada elemento ocorre apenas uma vez (EIBEN; SMITH, 2007). Cada gene é representado por um elemento da permutação. Por exemplo, no caso do problema do caixeiro viajante (do inglês, *Travelling Sales Problem*), onde o caixeiro deve visitar cada cidade apenas uma vez, buscando percorrer a menor distância possível. Nesse caso, se houver 4 cidades para visitar, as cidades podem ser consideradas como 1, 2, 3 e 4. Então se têm possíveis soluções como {1,2,3,4}, {1,3,4,2}, {3,4,1,2}, entre outras. A Figura 15 ilustra essa situação.

Figura 15 – Representação por permutação de um cromossomo



Fonte: Próprio autor

Após definida a representação a ser adotada, o algoritmo inicia a geração da população, discutida na seção 2.3.2.3.

2.3.2.3 População

A geração da população inicial é feita geralmente de forma aleatória. A distribuição aleatória permite uma boa distribuição do espaço de busca (LINDEN, 2012). Analogamente à biologia, a população é um conjunto de indivíduos ou cromossomos, sobre a qual incidirão os operadores genéticos para que ocorra a evolução do AG (EIBEN; SMITH, 2007; LINDEN, 2012). Define-se um tamanho para a população inicial, que geralmente não muda no decorrer da evolução do algoritmo. Esse tamanho deve ser determinado com cuidado, pois a quantidade de indivíduos que o AG irá avaliar será o tamanho da população vezes o número de iterações definidas para o algoritmo. Portanto, se for um número pequeno, a busca pode não chegar a avaliar melhores soluções e se for um número grande, deve-se cuidar com a complexidade da função de avaliação da população, devido ao custo computacional. A diversidade da população é medida pelas diferentes opções de solução presentes na população, no entanto não há uma medida única definida para isso (EIBEN; SMITH, 2007; LINDEN, 2012).

Dada uma população inicial, após aplicar os operadores genéticos de recombinação e mutação, será criada uma nova “geração” de indivíduos (descendentes ou filhos). Dois

modelos de geração de população são usados: *Generational Model* e *Steady-state Model*. O *Generational Model* é utilizado quando todos os indivíduos da população atual são substituídos pelos seus descendentes, isto é, cada indivíduo existe por apenas um ciclo e todos os pais são substituídos pelos descendentes gerados. O *Steady-state Model* não substitui a geração atual toda de uma vez, mas apenas uma parte dela, utilizando-se de alguma estratégia (EIBEN; SMITH, 2007).

No processo de substituição (do inglês, *replacement*), existem algumas estratégias para selecionar os “sobreviventes”, pois substituir os pais por seus descendentes aleatoriamente funciona, mas algumas estratégias funcionam melhor como, por exemplo (EIBEN; SMITH, 2007; LINDEN, 2012):

- o **Elitismo**, em que se define um número k de indivíduos que não devem “morrer”, mas sim passar para a próxima geração. Desta forma, garante-se que os k melhores indivíduos da geração atual t , passem para a próxima geração $t + 1$, garantindo que seus genomas sejam preservados e que o desempenho do AG sempre cresça com o decorrer das gerações;
- a ***Fitness-based*** ou “ $\mu + \lambda$ ”, isto é, uma estratégia baseada na função de avaliação, em que um número definido μ de indivíduos, considerando os μ indivíduos da geração atual mais os λ descendentes gerados, sobrevive para a próxima geração. Geralmente se usa a função de avaliação (*fitness*) para selecionar os melhores de ambas as gerações, isto é, os melhores pais e os melhores filhos sobrevivem, até o limite de μ indivíduos, que é o tamanho da população;
- a ***Replace Worst***, em que um número λ de “piores” indivíduos da geração atual é escolhido para ser substituído pelos descendentes gerados. Geralmente utilizado para populações grandes.

Um cuidado que se precisa ter em relação à substituição dos indivíduos, em manter “melhores” ou substituir “piores”, é a convergência genética, isto é, quando acontece de os indivíduos da população ficarem “parecidos”, dependendo do problema e da função de avaliação. Assim, é preciso manter certa diversidade genética, para que seja garantida a exploração do espaço de busca (possíveis soluções). O operador genético de mutação ajuda nessa tarefa (EIBEN; SMITH, 2007; LINDEN, 2012).

Após a representação do indivíduo e a geração da população inicial, o processo segue com a avaliação dos indivíduos, que é usado inclusive para determinar as substituições (*replacement*) da população, como foi descrito. Discute-se a seguir a função de avaliação.

2.3.2.4 Função de Avaliação (*Fitness Function*)

Em otimização, o principal objetivo é maximizar ou minimizar uma função que representa o problema, dita função objetivo. Algoritmos genéticos utilizam uma função para determinar a qualidade de um indivíduo como solução do problema que está sendo analisado pelo AG. Esta é a função de avaliação (do inglês, *fitness function*). No caso de otimização, a função de avaliação pode ser a própria função objetivo, ou uma transformação dessa função (EIBEN; SMITH, 2007). Por isso, a função de avaliação deve ser escolhida com cuidado. Ela deve materializar todo o conhecimento que se possui sobre o problema, tanto suas restrições quanto seus objetivos de qualidade. O papel da função de avaliação é representar os requisitos aos quais a população deve se adaptar, de forma que os mais “adaptados” sejam os melhores avaliados (EIBEN; SMITH, 2007; LINDEN, 2012).

A função de avaliação calcula um valor numérico (*fitness*) que reflete o quão bom é aquele indivíduo da população. Isto é, a função de avaliação é a “nota” dada ao indivíduo como uma possível solução para o problema (EIBEN; SMITH, 2007; LINDEN, 2012). Por exemplo, se o objetivo é minimizar uma função objetivo, então o indivíduo que tem menor valor é melhor que o indivíduo que tem maior valor, independentemente de quantas variáveis se está considerando nessa função. De maneira análoga, se o objetivo é maximizar a função, então os “melhores” indivíduos serão os que tiverem um valor maior.

Assim, a função de avaliação é usada para diferenciar as boas e as más soluções do problema, usando os valores (parâmetros) armazenados em cada cromossomo. Ela também deve diferenciar entre duas soluções sub-ótimas (s_1 e s_2), de sorte que a solução mais próxima da ótima (s_1) deve ser melhor avaliada do que a outra solução (s_2), ainda que esta também seja uma boa solução (LINDEN, 2012).

A função de avaliação é utilizada também na seleção dos pais, para que se inicie o processo de evolução das soluções possíveis.

2.3.2.5 Seleção de Pais

O processo de evolução do AG inicia-se com a seleção dos pais, isto é, com a seleção dos melhores indivíduos (os mais bem avaliados pela função *fitness*) para gerar descendentes (do inglês, *offspring*). Analogamente à biologia, a seleção de pais simula o processo de seleção natural, onde os mais aptos têm mais chances de gerar filhos. Os indivíduos de maior qualidade têm mais chances de se reproduzir, mas os indivíduos de menor qualidade também terão uma chance para isso, embora pequena (EIBEN; SMITH, 2007; LINDEN, 2012).

É importante que os indivíduos com menor qualidade também possam se reproduzir, pois estes podem ter características que ao se reproduzirem gerem uma solução melhor para o problema. Se apenas os melhores puderem se reproduzir, a população tenderá a ficar semelhante e faltará diversidade para a população, o que se denomina convergência genética (LINDEN, 2012).

Existem diferentes métodos para selecionar os pais, dentre eles (EIBEN; SMITH, 2007; LINDEN, 2012):

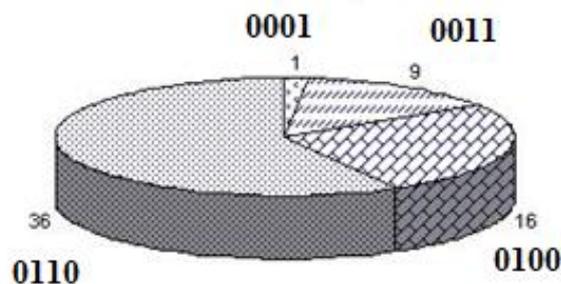
- **Roleta:** nesse método cria-se uma roleta (virtual) na qual cada indivíduo recebe um “pedaço” proporcional a sua avaliação (o todo não pode ultrapassar 100%). Depois roda-se a roleta e o selecionado será o indivíduo indicado. A roleta é gerada fazendo-se a proporção da avaliação do indivíduo pelo total e uma regra de três para definir os graus (360° - posição do círculo) que cada indivíduo representará. As Figuras 16 e 17 ilustram a situação, para 4 indivíduos diferentes, representados de forma binária com 4 *bits* cada um. Para rodar a roleta escolhe-se aleatoriamente um número entre 0 e 100 (percentual de cada indivíduo) ou entre 0 e 360 (posição no círculo) ou entre 0 e a soma total das avaliações (pedaço do somatório). Assim, cada indivíduo terá uma chance proporcional a sua avaliação de ser sorteado, sem desconsiderar as chances dos menos favorecidos.

Figura 16 – Exemplo para montar a roleta

<i>Indivíduo</i>	<i>Avaliação</i>	Pedaço da roleta	
		(%)	(°)
0001	1	1,61	5,8
0011	9	14,51	52,2
0100	16	25,81	92,9
0110	36	58,07	209,1
Total	62	100	360

Fonte: Adaptado de Linden (2012)

Figura 17 – Roleta para o exemplo dado

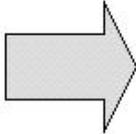


Fonte: Adaptado de Linden (2012)

- **Torneio:** o método do torneio seleciona um k número de indivíduos da população e faz com que eles concorram entre si pelo direito de ser pai, usando a avaliação (*fitness*) de cada um como parâmetro. O tamanho do torneio (k) é de no mínimo 2, pois ao contrário não haverá competição. Diferente da roleta, os indivíduos são selecionados aleatoriamente, conforme o k definido, e quem dentre eles possui a melhor avaliação é selecionado para a reprodução (recombinação ou cruzamento), gerando seus descendentes. A Figura 18 mostra um exemplo considerando $k = 3$ para o torneio.

Figura 18 – Exemplo do método de torneio

Indivíduo	Fitness
x_1	200
x_2	100
x_3	9500
x_4	100
x_5	100
x_6	10000
x_7	1
x_8	40



Torneios		
x_1	x_7	x_8
x_2	x_3	x_5
x_6	x_4	x_4
x_2	x_7	x_1
x_5	x_5	x_5
x_3	x_4	x_2
x_4	x_2	x_6
x_4	x_6	x_5

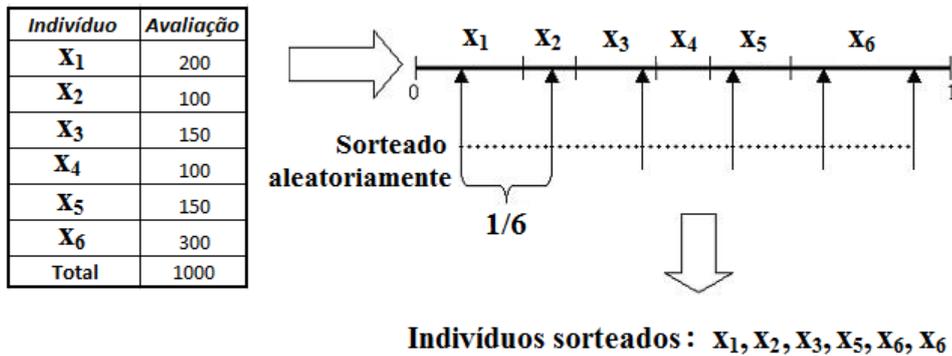
Fonte: Linden (2012)

A Figura 18 mostra uma situação em que dois indivíduos se destacam (são os chamados “superindivíduos” – x_3 e x_6). Na roleta, eles dominariam a população, mas no caso do torneio eles não dominam. A única vantagem que têm é que quando selecionados eles vencem o torneio. Um cuidado que se precisa ter com esse método é o tamanho do torneio, pois quanto maior o tamanho do torneio maior a dominância do primeiro colocado, sendo mínimas as chances dos menos bem colocados de vencerem o torneio. Assim, para minimizar essa situação usa-se k com valor baixo (2 ou 3, por exemplo).

- **Amostragem Estocástica Uniforme:** nesse método, os indivíduos são mapeados como segmentos de uma reta, cada um de tamanho proporcional à avaliação de cada indivíduo. É possível normalizar para que a soma total seja igual a 1. Para selecionar os pais é sorteado um número i entre 0 e $1/n$ (número de pais a ser selecionado). Depois ponteiros apontam para segmentos de reta nas posições i , $i + 1/n$, $i + 2/n$, $i + 3/n$, ..., $i + n-1/n$. Os indivíduos aos quais o segmento pertencer são os selecionados para a reprodução. Por exemplo, em uma população de 6 indivíduos sorteia-se um número i entre 0 e $1/6$ (visto que

se deseja pegar 6 pais – número de indivíduos da população). Assim, após este número, mais 5 ponteiros ($n-1$) serão marcados, todos a $1/6$ de distância de i . Desta forma, os cinco ponteiros marcados indicarão os indivíduos selecionados. A Figura 19 ilustra melhor a situação.

Figura 19 – Exemplo de amostragem estocástica uniforme



Fonte: Linden (2012)

- **Ranking:** a seleção por ranking consiste em ordenar os indivíduos de acordo com a função de avaliação e usar esse *ranking* como base para a seleção. Em vez de usar o valor da avaliação (*fitness*) usa-se o *ranking*, evitando a dominância de algum “superindivíduo” e a convergência genética prematura. Cada indivíduo recebe um valor no *ranking*, geralmente calculado pelo método linear (ver Quadro 3).

Quadro 3 – Fórmula método linear para o *ranking*

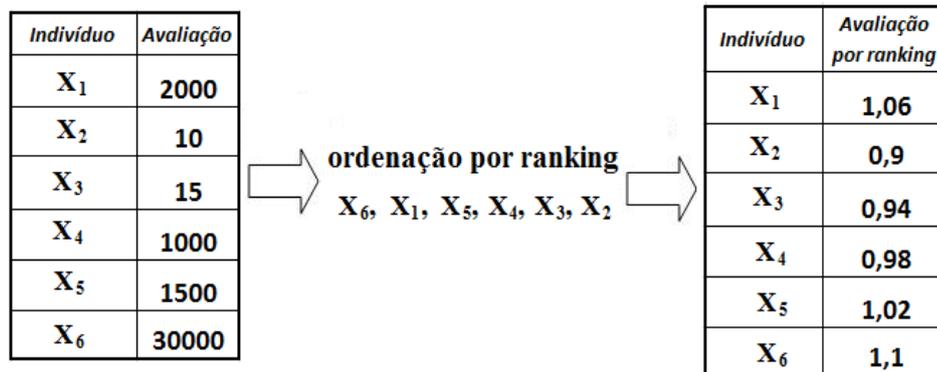
$$E(i,t) = Min + (Max - Min) * \frac{rank(i,t) - 1}{n-1}$$

Fonte: Linden (2012)

Onde:

- *Min* é o valor do pior indivíduo no *ranking*;
- *Max* é o valor do melhor indivíduo no *ranking*;
- *n* é o número de indivíduos da população;
- *rank(i,t)* é o *ranking* do indivíduo *i* na geração *t* do AG.

Bons resultados foram encontrados para os parâmetros $Max = 1,1$ e $Min = 0,9$, pois esses valores garantem que a soma das avaliações da geração *t* seja sempre igual a *n* (número de indivíduos da população) (LINDEN, 2012). A Figura 20 mostra um exemplo.

Figura 20 – Exemplo de seleção por *ranking*

Fonte: Linden (2012)

A seleção de pais é apenas o primeiro passo para o processo de evolução do AG. A partir dos pais selecionados, inicia-se o processo de simulação da reprodução que acontece no processo natural. O papel da seleção de pais é distinguir, entre os indivíduos, os de maior qualidade, para que sobre eles seja aplicado o operador genético de recombinação ou cruzamento (do inglês, *crossover*), simulando a procriação, e gerando descendentes de qualidade.

2.3.2.6 Recombinação ou Cruzamento (*Crossover*)

A recombinação ou cruzamento (do inglês, *crossover*) simula a reprodução sexuada que acontece no processo natural, onde dois pais combinados geram filhos com parte das suas características. O papel da recombinação é gerar novas soluções (indivíduos), fazendo com que a evolução aconteça.

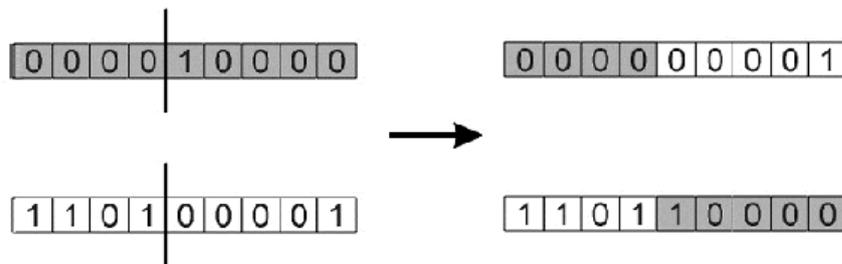
O princípio é o mesmo da natureza onde dois pais com características desejáveis são combinados para gerar descendentes combinando suas características. Isso não garante que os descendentes serão melhores, pois existe a possibilidade de que sejam piores, mas há uma boa chance de que alguns sejam melhores (EIBEN; SMITH, 2007; LINDEN, 2012).

A recombinação é um dos mais importantes componentes de um AG e é aplicada de acordo com uma probabilidade, chamada taxa de *crossover* (p_c), usualmente definida entre [0.5,1.0], sendo comum utilizar uma probabilidade maior ou igual a 0.8 (EIBEN; SMITH, 2007; LINDEN, 2012). É importante ressaltar que os operadores de recombinação dependem da representação adotada para o problema. Para representação binária e inteira, o mais simples deles é o *crossover* de um ponto, mas tem-se também o de n pontos e o uniforme. Para representação por permutação tem-se o *crossover* mapeamento parcial, o de ordem e o de ciclo, entre outros.

Alguns exemplos de operadores de recombinação (*crossover*) são apresentados a seguir (EIBEN; SMITH, 2007; LINDEN, 2012):

- **Um Ponto:** nesse operador um ponto de corte é estabelecido para que os genes dos pais sejam recombinados. Cada indivíduo de n genes tem $n-1$ pontos de corte possíveis e o ponto de corte estabelece o ponto de separação dos genes. O ponto de corte é um número aleatório entre $[1, n-1]$, de forma que permita a separação do indivíduo em duas partes, uma parte a direita do corte e outra a esquerda. A recombinação junta a parte a direita de um pai com a parte a esquerda do outro, e a parte a esquerda de um com a parte a direita do outro. Por exemplo, para um indivíduo de 9 genes e sorteia-se como ponto de corte o número 4. Assim, os indivíduos serão separados no ponto entre o quarto e o quinto gene para serem recombinados. A Figura 21 facilita o entendimento.

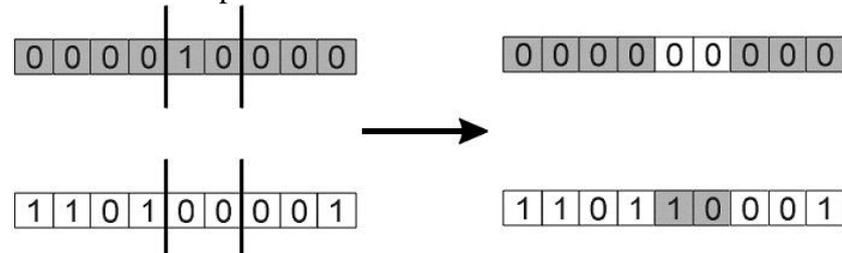
Figura 21 – *Crossover* de um ponto



Fonte: Eiben e Smith (2007)

- **N Pontos:** esse operador é uma variação do anterior, podendo ser estabelecido 2 pontos de corte ou mais. Neste caso os indivíduos são segmentados em vários pedaços de acordo com o número de pontos de corte. A recombinação utiliza estas partes para gerar os descendentes (filhos). A Figura 22 mostra um exemplo, com 2 pontos de corte, onde os filhos foram recombinados trocando as partes entre os pontos de corte estabelecidos.

Figura 22 – *Crossover* de dois pontos

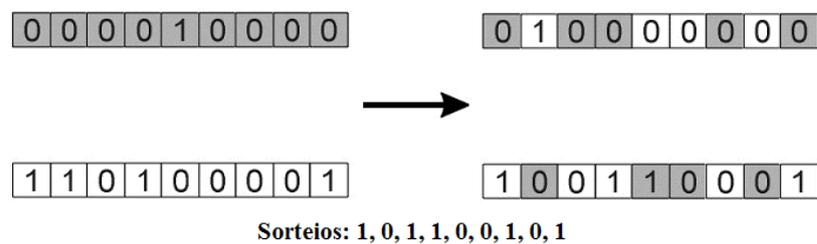


Fonte: Eiben e Smith (2007)

- **Uniforme:** os operadores anteriores combinam esquemas limitados dos pais, pois os pais são sempre separados em pedaços, e esses pedaços são

recombinados; no caso desse operador, cada gene dos pais pode ser recombinado. Para cada gene é sorteado um número 0 ou 1. Se o valor sorteado for 1, o primeiro filho recebe o gene da posição corrente do primeiro pai, e o segundo filho o gene da mesma posição do segundo pai. Se o valor for 0, inverte-se o processo, sendo que o primeiro filho recebe o gene do segundo pai e o segundo filho recebe o gene do primeiro pai. O único cuidado que se precisa ter é que esse operador pode “estragar” um esquema considerado bom, mas seu desempenho é superior ao dos anteriores. A Figura 23 mostra um exemplo.

Figura 23 – *Crossover* uniforme

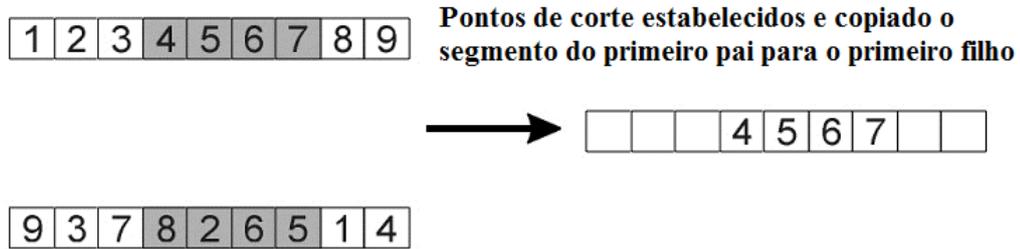


Fonte: Adaptado de Eiben e Smith (2007)

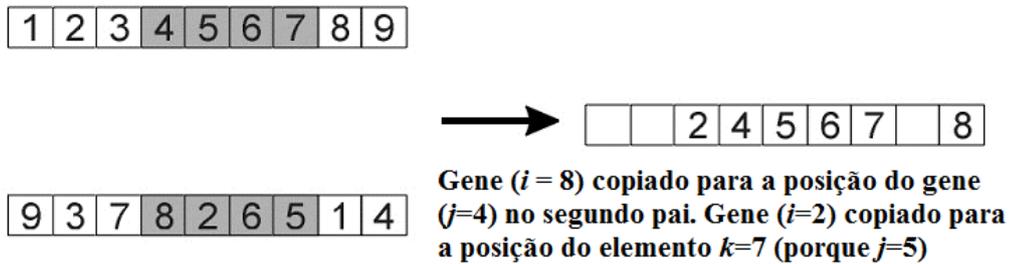
Para a representação por permutação, alguns dos operadores possíveis são:

- **Mapeamento Parcial (PMX):** esse operador troca uma sequência intermediária entre dois pais, garantindo que os filhos receberão o conjunto completo de genes na troca. Estabelece dois pontos de corte nos pais, e copia a parte intermediária entre esses pontos do primeiro pai para o primeiro filho. Depois segue os seguintes passos para copiar os demais genes:
 - A partir do primeiro ponto de corte, busca por genes que não foram copiados no segundo pai;
 - Para cada gene (i), busca o gene (j) que foi copiado do primeiro pai; coloca o gene (i) no lugar ocupado pelo gene (j) no segundo pai;
 - Se o gene (j) já está em uma posição ocupada pelo gene k no filho, coloca o gene (i) no lugar do gene k no segundo pai;
 - Depois de tratar o segmento entre os pontos, copia os elementos do segundo pai que ainda não foram copiados.

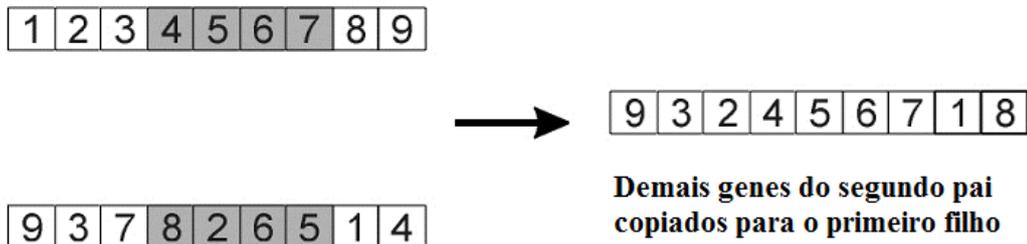
O segundo filho é gerado da mesma maneira, trocando-se as posições dos pais, isto é, o primeiro pai passa a ser o segundo e o segundo pai passa a ser o primeiro. As Figuras 24(a), 24(b) e 24(c) ilustram essas situações.

Figura 24(a) – *Crossover* mapeamento parcial (PMX)

Fonte: Adaptado de Eiben e Smith (2007)

Figura 24(b) – *Crossover* mapeamento parcial (PMX)

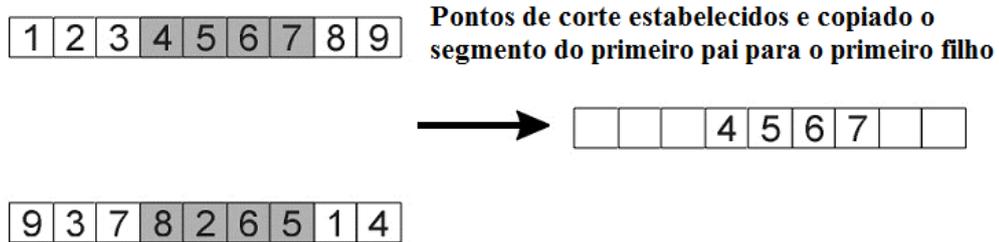
Fonte: Adaptado de Eiben e Smith (2007)

Figura 24(c) – *Crossover* mapeamento parcial (PMX)

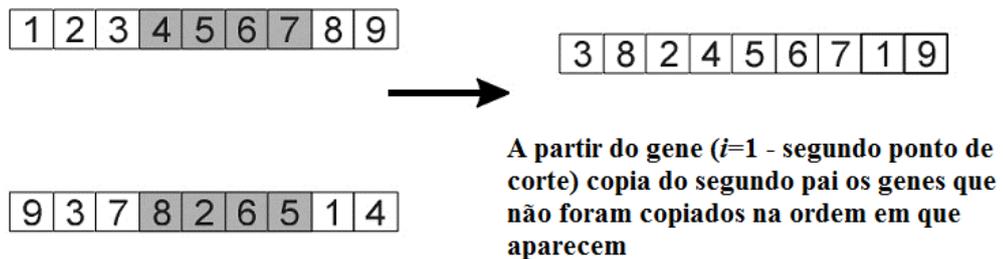
Fonte: Adaptado de Eiben e Smith (2007)

- **Ordem (*Order Crossover*):** é similar ao PMX descrito anteriormente, mas transmite uma relativa informação de ordem do segundo pai para o primeiro filho. Segue os seguintes passos:
 - Estabelece dois pontos de corte aleatoriamente e copia o segmento entre eles do primeiro pai para o primeiro filho;
 - A partir do segundo ponto de corte, copia os genes do segundo pai na ordem em que aparecem para o primeiro filho; se o gene já foi copiado (pelo segmento do primeiro pai) passa para o próximo.

Como no PMX, o segundo filho é gerado da mesma maneira, trocando-se as posições dos pais, isto é, o primeiro pai passa a ser o segundo e o segundo pai passa a ser o primeiro. As Figuras 25(a) e 25(b) ilustram essas situações.

Figura 25(a) – *Crossover* de ordem

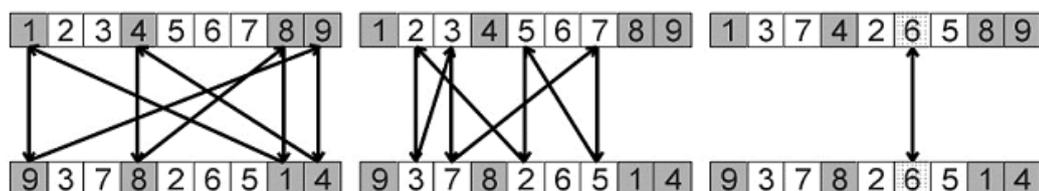
Fonte: Adaptado de Eiben e Smith (2007)

Figura 25(b) – *Crossover* de ordem

Fonte: Adaptado de Eiben e Smith (2007)

- **Ciclo (Cycle Crossover):** esse operador trabalha dividindo os genes em ciclos. Um ciclo é um subconjunto de genes onde cada gene sempre aparece emparelhado com outro gene quando os pais estão alinhados. Para estabelecer os ciclos, o operador segue os seguintes passos:
 - Inicia pelo primeiro gene do primeiro pai;
 - Busca o gene na mesma posição no segundo pai;
 - Busca o alelo (valor) no primeiro pai;
 - Adiciona este alelo (valor) ao ciclo;
 - Repete os passos 2-4 até encontrar o primeiro alelo (valor).

Os filhos são gerados por ciclos alternados de cada pai. As Figuras 26(a) e 26(b) ilustram essas situações.

Figura 26(a) – *Crossover* de ciclo

Definição dos ciclos

Fonte: Adaptado de Eiben e Smith (2007)

Figura 26(b) – *Crossover* de ciclo

Fonte: Adaptado de Eiben e Smith (2007)

A evolução utiliza-se dos operadores genéticos para fazer as modificações nas soluções e avaliá-las posteriormente. Além dos operadores de recombinação (*crossover*), tem-se o operador genético de mutação, que está descrito a seguir.

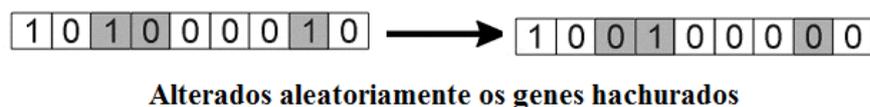
2.3.2.7 Mutação

O operador de mutação é aplicado sobre um indivíduo e gera um descendente (filho) mutante, com alguma modificação que foi aplicada. Pode ser aplicado a um pai para gerar um novo filho, ou pode ser aplicado aos filhos gerados para gerar alguma modificação. (EIBEN; SMITH, 2007). Assim como no *crossover*, a mutação tem associado a ela uma probabilidade, chamada taxa de mutação (p_m), que é geralmente baixa, na ordem de 0.05. Depois sorteia-se um número entre 0 e 1, se for menor ou igual que a probabilidade, então o operador atua sobre o gene em questão (EIBEN; SMITH, 2007; LINDEN, 2012).

A mutação também está ligada à representação adotada para o problema. Para representação binária, o mais comum é usar a mutação *bit a bit*. Para a representação inteira, têm-se duas principais formas: *random resetting* e *creep mutation*. Para representação por permutação têm-se: *swap mutation*, *insert mutation*, *inversion mutation*, entre outros (EIBEN; SMITH, 2007).

Esses operadores são apresentados a seguir (EIBEN; SMITH, 2007; LINDEN, 2012):

- **Bit a bit:** é a forma mais simples, utilizada para representações binárias, onde cada gene é um *bit*. A mutação fará a troca de 1 para 0, ou de 0 para 1, aleatoriamente, de acordo com a taxa de mutação estabelecida. A Figura 27 ilustra esta situação.

Figura 27 – Mutação *bit a bit*

Fonte: Adaptado de Eiben e Smith (2007)

- **Random Resetting:** é o *bit a bit* estendido para valores inteiros. Isto é, de acordo com a taxa de mutação estabelecida, cada gene é atualizado com um novo valor escolhido aleatoriamente de um conjunto de valores possíveis. É o mais indicado para números cardiais (valores absolutos).
- **Creep Mutation:** foi desenvolvido para valores ordinais, isto é, os que indicam uma ordem, e funciona adicionando um valor pequeno (positivo ou negativo) a cada gene, de acordo com uma probabilidade p . Este valor é escolhido aleatoriamente de uma distribuição simétrica a zero (por exemplo, de $[-2,2]$) e tende a gerar pequenas mudanças. Escolher esses parâmetros para a distribuição não é uma tarefa fácil, por isso, em alguns casos, combina-se mais de um operador para problemas com números inteiros (por exemplo, *random resetting* e *creep mutation* conjuntamente).

Para a representação por permutação, alguns operadores possíveis são:

- **Swap Mutation:** funciona selecionando aleatoriamente dois genes do indivíduo e trocando os valores (alelos) de suas posições. A Figura 28 ilustra esta situação.

Figura 28 – *Swap mutation*



Troca entre os genes 2 e 5

Fonte: Adaptado de Eiben e Smith (2007)

- **Insert Mutation:** funciona selecionando aleatoriamente dois genes e movendo-os de forma que fiquem juntos, enquanto os demais genes são reorganizados. A Figura 29 ilustra esta situação.

Figura 29 – *Insert mutation*



Genes 2 e 5 colocados juntos. Demais genes reorganizados

Fonte: Adaptado de Eiben e Smith (2007)

- **Inversion Mutation:** funciona selecionando aleatoriamente duas posições e invertendo a ordem dos genes entre elas. Essa é a menor alteração que pode ser feita no indivíduo, pois quebra o mesmo em três partes, mas mantém duas

partes intactas e apenas inverte os valores de uma delas. A Figura 30 ilustra esta situação.

Figura 30 – *Inversion mutation*



Genes entre 2 e 5 alterados por inversão

Fonte: Adaptado de Eiben e Smith (2007)

Assim, os pontos importantes para o entendimento da pesquisa foram descritos. A seguir, as considerações do capítulo destacam alguns aspectos.

2.4 CONSIDERAÇÕES DO CAPÍTULO

Com base na fundamentação de aprendizagem colaborativa apresentada, destacam-se os seguintes aspectos:

- A aprendizagem colaborativa é uma abordagem centrada no estudante e orientada ao grupo, onde as interações entre os pares destacam-se como o fator mais importante na aprendizagem;
- O uso da CSCL traz benefícios para o processo de ensino-aprendizagem, desenvolvendo habilidades de pensamento de ordem superior, aumentando a satisfação dos estudantes com a experiência de aprendizagem e melhorando a produtividade dos estudantes;
- Além dos benefícios comprovados da CSCL, outros motivos são citados para o uso da CSCL, entre eles, especialmente preparar os estudantes para a sociedade do conhecimento (habilidades de colaboração e criação de conhecimento);
- Na sociedade do conhecimento, um novo paradigma da educação (MOOCs), abriu discussões estratégicas sobre a potencial ruptura dos MOOCs no ensino superior, considerando tanto a abertura da educação quanto a escalabilidade, que são os diferenciais dos MOOCs em relação aos AEs tradicionais;

Em relação à formação de grupos para aprendizagem colaborativa, destacam-se os resultados obtidos no mapeamento sistemático. Tais resultados apontaram a importância de esta pesquisa adotar a formação de grupos de *forma automática*, utilizando a *abordagem selecionada* e como *critério* grupos **heterogêneos**. Além disso, como parte da revisão sistemática, esta pesquisa considera os conhecimentos (que estão relacionados aos resultados alcançados pelos estudantes nas avaliações de aprendizagem) e as interações dos estudantes

(que estão relacionadas às interações no fórum de discussão), utilizando **algoritmos genéticos**.

O uso de algoritmos genéticos na solução de problemas complexos de otimização tem se mostrado promissor, destacando-se especialmente a possibilidade de os AGs serem aplicados a problemas com dados inteiros ou discretos (descontinuidades duras) (LINDEN, 2012), como é o caso desta pesquisa. As escolhas em relação ao algoritmo genético baseiam-se, além da fundamentação dada neste capítulo, nos trabalhos relacionados, por isso estas escolhas são descritas após os trabalhos relacionados, na seção 4.2. Os trabalhos relacionados são descritos no capítulo 3, a seguir.

3 TRABALHOS RELACIONADOS

Para atender aos objetivos iniciais desta pesquisa, foram realizadas pesquisas bibliográficas e o mapeamento sistemático da literatura para entender melhor a formação de grupos na aprendizagem colaborativa. Entretanto, para o desenvolvimento do algoritmo genético, foi necessária uma Revisão Sistemática de Literatura (do inglês, *Systematic Literature Review*), que requer a revisão sistêmica dos materiais encontrados na pesquisa bibliográfica, a fim de fornecer uma visão mais detalhada em comparação com o mapeamento sistemático (PETERSEN et al., 2008).

De acordo com os resultados do mapeamento sistemático e com as definições previamente adotadas nesta pesquisa, a revisão focou nos trabalhos que utilizam **algoritmos genéticos** para a formação de grupos para aprendizagem colaborativa, utilizando a *forma automática*, a *abordagem selecionada* e o *critério heterogêneo*. Além disto, os trabalhos que adotam uma abordagem para aplicações como *mobile learning* ou redes sociais, foram excluídos desse processo, visto que o interesse desta pesquisa são os AEs e os MOOCs (massivo).

A revisão incluiu também as referências dos trabalhos selecionados, buscando por trabalhos dentro desta linha de pesquisa, isto é, trabalhos sobre formação de grupos para aprendizagem colaborativa, considerando os seguintes critérios de inclusão e de exclusão listados:

- **CI1:** Artigos escritos em inglês;
- **CI2:** Artigos disponíveis para download (de acordo com a assinatura da universidade) na íntegra em PDF (Portable Document Format);
- **CI3:** Artigos primários (mapeamentos e/ou revisões sistemáticas e artigos teóricos não foram considerados como primários);
- **CI4:** Artigos completos (notas ou artigos não completos foram desconsiderados);
- **CI5:** Artigos mais recentes (2004-2014);
- **CE1:** Artigos duplicados (trabalhos já considerados);
- **CE2:** Artigos que não utilizam algoritmos genéticos;
- **CE3:** Artigos que não geram grupos heterogêneos;

Seis trabalhos foram identificados. A seguir, em cada parágrafo, destacam-se os trabalhos decorrentes da revisão e que foram posteriormente revisados, a fim de identificar, sempre que possível, a quantidade de características consideradas, quais eram essas

características, a representação adotada para o AG, a quantidade de estudantes envolvidos na pesquisa, se houve comparação da aplicação desenvolvida com o método randômico (desempenho), e se houve algum estudo de caso.

3.1 ALGORITMOS GENÉTICOS PARA FORMAÇÃO DE GRUPOS

Gogoulou et al. (2007) apresentam uma ferramenta, chamada OmadoGenesis, para a formação de grupos homogêneos, heterogêneos ou mistos, considerando até três características dos estudantes disponíveis no sistema como, por exemplo, competências e índice de colaboração, entre outros. Foram criados três algoritmos: Homo-A, Hete-A e um algoritmo genético (AG) para ambas as opções, inclusive grupos mistos. Para gerar grupos homogêneos, o Homo-A se baseia no algoritmo *K-Means*, e gera o agrupamento utilizando *clusters*. A similaridade dos estudantes é calculada através da distância Euclidiana. O Hete-A se baseia em uma matriz com o número de estudantes com cada característica. O algoritmo distribui os estudantes com as mesmas características em grupos distintos, realizando esse processo do maior número de estudantes para o menor, até que todos os estudantes estejam alocados em algum grupo. O AG permite a criação tanto de grupos homogêneos, heterogêneos ou mistos, e calcula a qualidade de cada grupo e de cada solução proposta. A qualidade é calculada pela função de avaliação, conforme modelo proposto pelos autores, buscando maximizar a qualidade dos grupos. A representação adotada pelo AG é a representação inteira. Um experimento com dados reais é descrito resumidamente, considerando 52 estudantes, usando duas características, em grupos de quatro estudantes por grupo. Os autores consideram que o AG teve bons resultados, gerando grupos com maior qualidade em comparação com o Homo-A para grupos homogêneos e com uma boa qualidade para grupos mistos. No entanto, em comparação com o Hete-A, o resultado do AG foi inferior. Os autores não comparam os resultados do AG com o método randômico.

Wang, Lin e Sun (2007) apresentam a ferramenta DIANA para formar grupos heterogêneos, considerando variáveis psicológicas como, por exemplo, estilo de pensamento; dados estes que são coletados dos estudantes, utilizando-se questionários psicológicos definidos na literatura. Os autores projetaram DIANA para criar grupos que apresentam heterogeneidade interna e equilíbrio entre eles, evitando formar grupos extremamente fracos. A ferramenta DIANA utiliza um algoritmo genético para formar os grupos, no entanto, antes de gerar os grupos heterogêneos, os estudantes são categorizados e agrupados por similaridade em k categorias, conforme o número de estudantes por grupo. Posteriormente o algoritmo genético busca um de cada categoria para formar o grupo heterogêneo. A

representação adotada para o AG é a representação inteira. A ferramenta considera até sete variáveis (diferentes variáveis psicológicas) e aloca de 3 a 7 membros por grupo. Um experimento com casos reais foi feito, considerando estilos de pensamento/aprendizagem (STERNBERG, 1994), com 66 estudantes, e mostrou que os grupos gerados pelo DIANA tiveram melhor desempenho (de acordo com a avaliação definida pelos professores) do que os gerados aleatoriamente. Além disso, os autores descrevem que os grupos também tiveram menor variação entre si em relação aos resultados observados. O interessante desse trabalho é que além de analisar o resultado dos grupos com diferentes formações (DIANA e randômico) os autores aplicaram um questionário para entender a percepção dos estudantes em relação aos seus colegas de grupo, aos resultados alcançados pelo grupo e ao processo de aprendizagem colaborativa. Os níveis de satisfação com os colegas, o processo de colaboração e os resultados do grupo também foram maiores entre os membros dos grupos determinados pelo AG. A comparação se dá em relação aos resultados alcançados pelos grupos e a satisfação dos estudantes, no entanto não há comparação do desempenho do AG com o método randômico.

Ani et al. (2010) propõem um método para formar grupos baseado nas habilidades de programação dos estudantes. Os estudantes são classificados em “bom”, “moderado” e “fraco”. O AG busca criar grupos heterogêneos com representantes de ambas as classificações, levando em conta um número mínimo de estudantes de cada categoria calculado pelo algoritmo, e assim garantir o balanceamento entre os grupos. A representação adotada pelo AG é a representação inteira. Um experimento com dados reais é gerado, considerando 35 estudantes em sete grupos de cinco estudantes cada. O AG forma os grupos corretamente e os autores comparam os resultados dos grupos formados pelo AG com os formados manualmente no semestre anterior. Os autores afirmam que, dos sete grupos, apenas um não conseguiu entregar o trabalho proposto, assumindo que 85% deles teve sucesso. No entanto não há menção sobre os resultados alcançados anteriormente pelos grupos gerados manualmente. Não há comparação do desempenho do AG com o algoritmo randômico.

Abnar, Orooji e Taghiyareh (2012) propõem uma formação de grupos flexível em relação ao número (multi) e tipo de características (várias) consideradas pelo algoritmo, de acordo com a disponibilidade de dados do sistema e a escolha do professor. A abordagem proposta busca facilitar a formação de grupos, permitindo ao professor/instrutor adicionar, selecionar e ranquear as características desejadas; e permite formar grupos homogêneos, heterogêneos ou mistos. A representação adotada pelo AG foi a representação inteira. O

algoritmo foi aplicado no MOODLE² e busca maximizar a qualidade dos grupos formados. A qualidade é medida conforme modelo proposto pelos autores. Os autores fazem uma validação considerando dados reais extraídos do MOODLE, o estilo de aprendizagem dos estudantes, além do desempenho e da motivação. No entanto, não descrevem detalhes do experimento e não apresentam os dados obtidos, apenas citam que o algoritmo proposto foi trinta por cento melhor que o método randômico. Não informam a quantidade de estudantes e não há estudo de caso.

Moreno, Ovalle e Vicari (2012) propõem uma formação de grupos inter-homogêneo e intra-heterogêneo, considerando múltiplas características dos estudantes como, por exemplo conhecimento e habilidade de comunicação, entre outros. Isto é, consideram que a situação ideal é ter grupos que são semelhantes entre si (inter-homogêneo), mas que tenham estudantes com diferenças individuais internamente (intra-heterogêneo). Esse trabalho é interessante devido ao número arbitrário de características que considera (multi), no entanto, o modelo proposto pelo autor não garante a intra-heterogeneidade, mas apenas a inter-homogeneidade. Ao minimizar a soma dos quadrados das diferenças das características considerando a média total da turma o modelo garante a inter-homogeneidade; de modo que a formação de grupos com menor diferença com a média total da turma, terá a menor diferença possível. A representação adotada para o AG foi a representação inteira, utilizando uma matriz grupos x estudantes. A validação do algoritmo foi feita em seis diferentes bases de dados, contendo de 6 a 180 estudantes, e os resultados obtidos apresentam bom desempenho do AG comparado ao método exaustivo e ao método randômico. Outro ponto interessante desse trabalho, diz respeito a um experimento com dados reais, de 135 estudantes, considerando três características: uma estimativa do nível de conhecimento do estudante, uma estimativa das habilidades comunicativas dos estudantes e uma estimativa da capacidade de liderança do estudante. Esse experimento permitiu a validação, não só do ponto de vista computacional, mas também do ponto de vista pedagógico, medindo os resultados dos estudantes, e comparando-os com duas outras abordagens: randômica e autosselecionada. O experimento considerou três momentos distintos: o questionário individual realizado antes da atividade de aprendizagem colaborativa, as notas do grupo obtidas durante a atividade, e as notas individuais de um exame individual aplicado após a atividade colaborativa. Os resultados mostraram que os grupos formados pelo AG tiveram melhor desempenho que os formados pela abordagem aleatória ou autosselecionada.

² MOODLE é um LMS (*Learning Management System* – Sistema Gerenciador de Aprendizagem) *open source*.

Yannibelli e Amandi (2012) propõem a formação de grupos heterogêneos baseados em papéis preferidos dos estudantes. As autoras utilizam o modelo proposto por Belbin (1981 e 1993) e os dados foram coletados através do *Belbin Team-Papel Self-Perception Inventory* (BTRSPI), que determina os papéis preferidos dos estudantes através de questionário de auto-avaliação. O algoritmo genético busca maximizar o equilíbrio dos papéis nos grupos, isto é, busca que cada grupo tenha um membro desempenhando cada um dos papéis propostos (9 diferentes papéis). A representação adotada foi a representação inteira, utilizando a permutação, o que é interessante uma vez que é uma forma simples e adequada ao problema, pois os estudantes não podem se repetir em mais de um grupo e são rearranjados em diferentes grupos para avaliar qual a melhor formação. Outro ponto interessante do trabalho é a avaliação feita em dez diferentes bases de dados, contendo de 18 a 3000 estudantes. Os resultados obtidos são promissores, pois o algoritmo encontrou a solução ótima nas quatro primeiras bases de dados e a solução quase ótima nas seis demais, em tempo considerado aceitável pelas autoras. Além disso, em comparação com o método exaustivo e com o método randômico, o AG também obteve melhores resultados. Esse trabalho não realiza um estudo de caso com dados reais, no entanto, considera-se o mais importante desta revisão, visto que é o único trabalho a considerar uma quantidade maior de estudantes.

A seguir, discutem-se os resultados alcançados com a revisão sistemática.

3.2 ANÁLISES E CONSIDERAÇÕES DO CAPÍTULO

Como o objetivo desta pesquisa é avaliar a adoção de algoritmos genéticos para a formação de grupos na aprendizagem colaborativa, considerando um contexto massivo de dados (milhares de estudantes), e como propõe-se validar o AG desenvolvido comparando seu desempenho com o algoritmo randômico (base de comparação), é importante que os trabalhos revisados apresentem essas vertentes (quantidade de estudantes e comparação do desempenho do AG). Outro aspecto importante é a representação do indivíduo adotada pelo AG, visto que a representação deve ser a mais simples possível e já considerar as restrições do problema, conforme sugerido por Linden (2012).

A Tabela 1 sintetiza a revisão realizada, onde as colunas representam os trabalhos revisados e as linhas representam os dados coletados na revisão: a quantidade de características consideradas em cada trabalho, as características utilizadas, a representação adotada para o AG, a quantidade de estudantes envolvidos na pesquisa, se se comparou o desempenho do AG com o método randômico, e se houve algum estudo de caso.

Tabela 1 – Resumo da revisão de literatura

Trabalho	<i>Gogoulou et al. (2007)</i>	<i>Wang, Lin e Sun (2007)</i>	<i>Ani et al. (2010)</i>	<i>Abnar, Orooji e Taghiyareh (2012)</i>	<i>Moreno, Ovalle e Vicari (2012)</i>	<i>Yannibelli e Amandi (2012)</i>	
Qtd característica	até 3	até 7	1	multi	multi	1	
Característica	várias	PS	HP	várias	várias	papéis	
Repres. AG	inteira	inteira	inteira	inteira	inteira (matriz)	inteira (permutação)	
Qtd estudantes	52	66	35	?	6-180	18-3000	
Comparou AG?	Não	Não	Não	Sim	Sim	Sim	
	–	–	–	RM	EM e RM	EM e RM	
	Sim	Sim	Sim	Não	Sim	Não	
Estudo caso?	entre as soluções propostas	Comparou com				RAND	–
		RAND	MAN	–	AUTO	–	

Legenda: PS = Psicológicas; HP = Habilidades Programação

Multi = Múltiplas; Várias = várias características podem ser consideradas

EM = Método Exaustivo; RM = Método Randômico

RAND = Randômico; MAN = Manual; AUTO = Autosselecionado

Fonte: Próprio autor

Considerando o objetivo desta pesquisa e a revisão realizada, destacam-se na Tabela 1 os três últimos trabalhos, que compararam o desempenho do AG proposto em cada trabalho com o método randômico. No entanto, o trabalho de Abnar, Orooji e Taghiyareh (2012) não menciona a quantidade de estudantes envolvidas no estudo.

O trabalho de Moreno, Ovalle e Vicari (2012) avaliou o desempenho do AG proposto em seu trabalho, comparando-o com dois outros métodos (exaustivo e randômico), considerando de 6 até 180 estudantes. Além disso, realizou um estudo de caso com 135 estudantes comparando o desempenho dos estudantes agrupados pelo AG com os agrupados com duas outras *abordagens*: randômico e autosselecionado. Esse trabalho considera múltiplas características, o que é interessante; e propõe a formação de grupos inter-homogêneos e intra-heterogêneos. No entanto, observou-se que o modelo proposto garante apenas a inter-homogeneidade, sem garantir a intra-heterogeneidade. A ideia de inter-homogeneidade proposta por Moreno, Ovalle e Vicari (2012) foi adotada nesta pesquisa, isto é, o balanceamento entre os grupos, mas garantindo a intra-heterogeneidade dos mesmos, conforme descrito no capítulo 4, a seguir. Em relação à quantidade de características, esta pesquisa adotou a utilização de duas características, a fim de simplificar a formação de grupos, posto que já se vai considerar uma quantidade massiva de estudantes.

O trabalho de Yannibelli e Amandi (2012) avaliou o desempenho do AG proposto comparando-o com dois outros métodos (exaustivo e randômico), considerando de 18 até 3000 estudantes. Esse é o único trabalho revisado que apresenta uma quantidade maior de estudantes, na casa dos milhares. Além disso, a representação dos indivíduos adotada pelo AG utiliza a permutação, o que se pode dizer que é a representação mais apropriada para o problema, visto que a distribuição dos estudantes de uma classe em grupos é uma permutação (os estudantes devem pertencer a um grupo apenas, sem se repetir na solução). Essa escolha implica em todo o processo de evolução do AG, visto que, em conformidade com a representação adotada, são escolhidos os componentes de seleção de pais, *crossover* e mutação. Desta forma, dentre os trabalhos revisados, o trabalho de Yannibelli e Amandi (2012) destaca-se como o principal trabalho relacionado a esta pesquisa.

Apesar dos trabalhos de Moreno, Ovalle e Vicari (2012) e de Yannibelli e Amandi (2012) se destacarem nas contribuições dadas a esta pesquisa, todos os trabalhos contribuíram direta ou indiretamente para as definições adotadas, conforme mostram os balões na Tabela 2.

Tabela 2 – Contribuições dos trabalhos relacionados para esta pesquisa

Trabalho	Gogoulou et al. (2007)	Wang, Lin e Sun (2007)	Ani et al. (2010)	Abnar, Orooji e Taghiyareh (2012)	Moreno, Ovalle e Vicari (2012)	Yannibelli e Amandi (2012)	Citadin (2015)
Qtd característica	3	até 1	1	multi	multi	1	2
Característica			HP	várias	várias	Escolhas AG	conhecimentos e interações
Repres. AG			inte	inte	inte	inteira (permutação)	inteira (permutação)
Qtd estudantes	32	30	30	?	?	18-3000	1000-10000
Comparou AG?	Não	Não	Não	Sim	Sim	Sim	Sim
Estudo caso?	Sim			Sim	Sim	Não	Não
	entre as soluções propostas			Comparou com RAND AUTO	EM e RM	EM e RM	RM

Legenda: PS = Psicológicas; HP = Habilidades Programação

Multi = Múltiplas; Várias = várias características podem ser consideradas

EM = Método Exaustivo; RM = Método Randômico

RAND = Randômico; MAN = Manual; AUTO = Autosseleccionado

Fonte: Próprio autor

Observa-se na Tabela 2, por exemplo, o uso da distância Euclidiana como medida de similaridade (GOGOULOU et al., 2007) e as visões de grupos balanceados (ANI et al., 2010; MORENO; OVALLE; VICARI, 2012; WANG, LIN; SUN, 2007), além da normalização utilizada por Moreno, Ovalle e Vicari (2012). O trabalho de Yannibelli e Amandi (2012) destacou-se em relação as definições adotadas para o AG. A coluna Citadin (2015) refere-se às definições desta pesquisa, discutidas a seguir, no capítulo 4.

4 UM ALGORITMO GENÉTICO PARA FORMAÇÃO DE GRUPOS HETEROGÊNEOS NA APRENDIZAGEM COLABORATIVA

Esta pesquisa tem por objetivo geral avaliar a adoção de algoritmos genéticos para a formação de grupos na aprendizagem colaborativa, considerando um contexto massivo de dados (milhares de estudantes), através da comparação do desempenho do algoritmo genético (AG) com o algoritmo randômico. Assim, para alcançar esse objetivo, este capítulo discute o modelo proposto e a modelagem definida para o algoritmo genético desenvolvido. O modelo proposto detalha as características consideradas para a formação do grupo e a formulação adotada para o problema. A modelagem do algoritmo genético detalha as escolhas feitas em relação ao desenvolvimento do algoritmo, desde a representação dos indivíduos até o operador de mutação.

4.1 MODELO PROPOSTO

Nos trabalhos relacionados, observou-se uma variedade de propostas para a formação de grupos para aprendizagem colaborativa, considerando uma ou várias características dos estudantes, criando grupos homogêneos, heterogêneos ou mistos, permitindo a geração manual ou automática, entre outras. Entretanto, para esta pesquisa foi adotada a formação de grupos de *forma automática*, utilizando a *abordagem selecionada* e como *critério* grupos **heterogêneos**. Essa definição está embasada em toda a fundamentação teórica apresentada, especialmente nos resultados do mapeamento sistemático realizado. No mapeamento sistemático, observou-se a necessidade de *critérios complementares*, ou seja, as características que a formação dos grupos vai considerar junto ao *critério* heterogêneo. Essas características são discutidas na seção 4.1.1.

4.1.1 Características adotadas

O problema de formação de grupos é um problema *NP-hard*, de otimização combinatória, que cresce em complexidade, à medida que aumenta o tamanho da entrada (número de estudantes) e as características a serem consideradas. Como esta pesquisa considera um contexto massivo de dados (milhares de estudantes), a fim de simplificar o problema devido à grande quantidade de estudantes já considerada, adotou-se **duas características** dos estudantes (**conhecimentos** e **interações**).

Essas características foram definidas com base na disponibilidade de dados em AEs como o Moodle, TelEduc, Sakai, Claroline, entre outros, e na plataforma *EdX*³, uma plataforma MOOC *open source*. Essa escolha se deu não apenas pela disponibilidade dos dados, que poderiam ser outros, como gênero ou idade, por exemplo; mas também pela importância de se criar grupos heterogêneos considerando os conhecimentos dos estudantes (CHAN, 2006; WEBB, 1983; WEBB, 1992) e a importância das interações para a aprendizagem colaborativa ((DILLENBOURG; JÄRVELÄ; FISCHER, 2009).

Não há um consenso para definir o que é conhecimento mesmo em estudos de Gestão do Conhecimento ou Ciência da Informação, em que os conceitos de dado, informação e conhecimento são fundamentais (CASTILLO; CAZARINI, 2010; ZINS, 2007). Zins (2007), em painel internacional com 54 cientistas da informação de 16 países, documenta 130 definições para dado, informação e conhecimento; 43 delas para conhecimento. Castillo e Cazarini (2010) além de discutir diferentes definições de conhecimento, discutem uma série de taxonomias do conhecimento encontradas na literatura. Os autores também citam Muller-Merbach, que cita Eysenck (1979, *apud* Muller-Merbach, 2006 *apud* CASTILLO; CAZARINI, 2010) e sugere que a definição de conhecimento deve ser algo como “conhecimento neste contexto deve ser entendido como...” e não uma expressão do tipo “conhecimento é...”. Por isso, define-se conhecimento no contexto desta pesquisa, que é o contexto do processo de ensino-aprendizagem. Nesse contexto, a avaliação da aprendizagem é um agente de controle, em que diferentes tipos de avaliação (diagnóstica, formativa ou somativa) auxiliam no controle e são aplicadas em diferentes períodos do processo de ensino-aprendizagem (início, durante e final) (ALVES et al., 2008). Por exemplo, uma avaliação diagnóstica é aplicada no início do curso para que o professor identifique o nível de conhecimento dos estudantes sobre determinado assunto (pré-teste). No decorrer do curso, avaliações formativas são aplicadas para acompanhar o aprendizado dos estudantes, enquanto que, ao final, a avaliação somativa considera todo o aprendizado de determinado conteúdo ou curso. Para avaliar é preciso medir, e a avaliação passa pelo julgamento de resultados (SILVA; JOYE, 2015). Os resultados das avaliações de aprendizagem expressam um indicativo do conhecimento do estudante com base no nível de qualidade dos resultados alcançados pelo estudante frente aos objetivos estabelecidos. Esses resultados são geralmente expressos por uma “nota” (medida). Assim, com base no contexto apresentado, considerando-se as notas como uma medida dos conhecimentos dos estudantes e que os estudantes com melhores notas podem contribuir para a aprendizagem dos estudantes com piores notas, a

³ Disponível em <http://code.edx.org/>

primeira característica considerada foi **conhecimentos**. Dessa forma, para esta pesquisa definem-se conhecimentos como a média das notas alcançadas pelo estudante nas avaliações propostas no curso.

Como criar condições para uma efetiva interação no grupo é uma das preocupações da aprendizagem colaborativa (DILLENBOURG; JÄRVELÄ; FISCHER, 2009), a outra característica considerada foram as **interações** dos estudantes. Medir a qualidade das interações não é uma tarefa simples, por isso, nesta pesquisa foi considerada apenas a quantidade de interações entre os estudantes do mesmo grupo. Significa que neste momento não é considerada a qualidade da mensagem, através de julgamento de valor feito pelos estudantes em avaliações das mensagens como curtir/não curtir, gostei/não gostei, ajudou/não ajudou, enfim. Apenas as quantidades são consideradas neste momento, pois dessa forma é garantido que os estudantes pelo menos já tenham interagido de alguma forma entre si. Essa definição se dá também em função de que nem todos os AEs têm a opção de avaliar as postagens feitas no fórum de discussão. Portanto, para esta pesquisa definem-se interações as quantidades das interações realizadas entre os estudantes no fórum de discussão do curso, considerando-se as mensagens trocadas entre os estudantes como perguntas e respostas às perguntas dos colegas.

Assim, a formulação do problema de formação de grupos de *forma automática*, utilizando a *abordagem selecionada* e como *critério* grupos **heterogêneos** considerando os **conhecimentos** e as **interações** dos estudantes, é discutida a seguir.

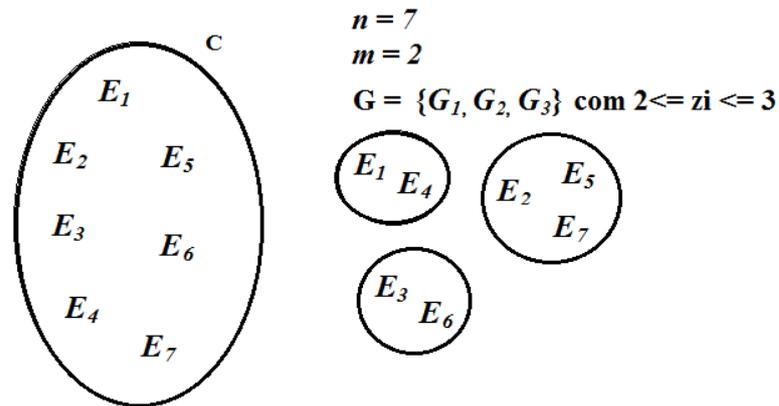
4.1.2 Formulação do Problema

O problema proposto consiste na formação de grupos de estudantes para realizar tarefas de aprendizagem colaborativa. Como parte do problema, considera-se que a heterogeneidade e as interações nos grupos devem ser maximizadas, isto é, o objetivo é gerar grupos **heterogêneos** em relação aos **conhecimentos** e com o **maior número de interações** possível entre os membros do grupo, gerando ainda grupos balanceados entre si (**inter-homogêneos**).

O problema pode ser formalmente formulado como: Dada uma classe C constituída por n estudantes E , $C = \{E_1, E_2, \dots, E_n\}$, o professor deve dividir os estudantes em grupos de m estudantes, criando assim g grupos, $G = \{G_1, G_2, \dots, G_g\}$. Cada grupo G_i é composto de um número z_i de estudantes membros; e cada estudante só pode pertencer a um grupo. Em relação ao tamanho dos grupos, os estudantes devem ser divididos de tal forma que os g grupos tenham um número semelhante de estudantes, isto é, a diferença entre o tamanho de um grupo

e o tamanho dos outros grupos não deve exceder a um estudante ($z_i = m$ ou $z_i = m + 1$). Os valores de C , n e m são conhecidos. A Figura 31 ilustra um exemplo do problema: Uma classe com $n = 7$ estudantes deve ser dividida em grupos (G) de $m = 2$ estudantes. Assim, como n e m não são múltiplos, G_1 e G_2 têm $z_i = 2$ estudantes e G_3 tem $z_i = 3$ estudantes.

Figura 31 – Exemplo de problema de formação de grupos



Fonte: Próprio autor

Cada estudante possui uma média de suas avaliações que será considerada como seu conhecimento (*conhect*), com valor no intervalo $[0,10]$, conforme as notas atribuídas às avaliações.

Além disso, cada estudante possui um número de interações (*int*) efetuadas com os demais colegas da classe no fórum de discussão, que serão tabuladas em uma matriz $n \times n$, contendo as interações de cada um com o respectivo colega. A Figura 32 ilustra estes dados, com as notas dos estudantes (Conhecimentos) e os pares de interações (Matriz de Interações).

Figura 32 – Exemplo de dados dos estudantes

Estudante	Conhecimentos
E_1	0
E_2	5
E_3	8
E_4	7
E_5	9
E_6	3
E_7	10
Total	42
Média	6

Matriz de Interações							
	E_1	E_2	E_3	E_4	E_5	E_6	E_7
E_1	0	2	1	2	0	3	1
E_2	2	0	1	1	2	0	0
E_3	1	1	0	1	0	1	2
E_4	2	1	1	0	2	1	1
E_5	0	2	0	2	0	0	1
E_6	3	0	1	1	0	0	1
E_7	1	0	2	1	1	1	0

Fonte: Próprio autor

Dessa forma, com o objetivo de maximizar a heterogeneidade dos conhecimentos no grupo e as interações do mesmo, considerando ainda a inter-homogeneidade entre os grupos, o problema é descrito na seguinte função objetivo:

$$\text{Maximizar } f(x) = \omega_1 \sum_{i=1}^g f(\text{het_conhect}_i) + \omega_2 \sum_{i=1}^g f(\text{int}_i) + (1 - \sum_{i=1}^g f(\text{inter_homo}_i)) \quad (1)$$

Onde:

- ω_1 e ω_2 são pesos para os conhecimentos e para as interações respectivamente;
- i, g indicam os índices dos grupos ($g = \text{qtd de grupos}$);
- $f(\text{het_conhect}_i)$ é a função de cálculo da heterogeneidade dos conhecimentos, com i indicando cada grupo;
- $f(\text{int}_i)$ é a função de cálculo das interações, com i indicando cada grupo;
- $f(\text{inter_homo}_i)$ é a função de cálculo da inter-homogeneidade, com i indicando cada grupo.

Como a função objetivo prevê a aplicação de pesos, os dados foram normalizados para o intervalo $[0,1]$, tanto os conhecimentos quanto as interações, utilizando os máximos (C_{max}) e mínimos (C_{min}) de cada característica (MORENO; OVALLE; VICARI, 2012), conforme a fórmula:

Quadro 4 – Fórmula normalização dos dados

$$C^* = \left(\frac{C - C_{min}}{C_{max} - C_{min}} \right)$$

Fonte: Moreno, Ovalle e Vicari (2012)

A Figura 33 mostra os exemplos da Figura 32 com os dados já normalizados pela fórmula do Quadro 4, de forma que os conhecimentos e as interações ficam no intervalo $[0,1]$.

Figura 33 – Exemplo de dados dos estudantes (normalizado)

Estudante	Conhect	Conhect (Normalizado)	Matriz de Interações (Normalizada)						
			E_1	E_2	E_3	E_4	E_5	E_6	E_7
E_1	0	0	0,00	0,67	0,33	0,67	0,00	1,00	0,33
E_2	5	0,5	0,67	0,00	0,33	0,33	0,67	0,00	0,00
E_3	8	0,8	0,33	0,33	0,00	0,33	0,00	0,33	0,67
E_4	7	0,7	0,67	0,33	0,33	0,00	0,67	0,33	0,33
E_5	9	0,9	0,00	0,67	0,00	0,67	0,00	0,00	0,33
E_6	3	0,3	1,00	0,00	0,33	0,33	0,00	0,00	0,33
E_7	10	1	0,33	0,00	0,67	0,33	0,33	0,33	0,00
Total	42	4,2							
Média	6	0,6							

Fonte: Próprio autor

A função de cálculo da heterogeneidade dos conhecimentos utiliza a distância de Manhattan entre os conhecimentos do grupo. A distância de Manhattan é uma versão da chamada métrica de Minkowski que é uma das principais medidas utilizadas pelos algoritmos de agrupamento para cálculo de similaridade (homogêneo) ou dissimilaridade (heterogeneidade) (LINDEN, 2009). A métrica de Minkowski é dada pela seguinte fórmula:

Quadro 5 – Métrica de Minkowski

$$d(i,j) = \sqrt[q]{(|x_{i1} - x_{j1}|^q + |x_{i2} - x_{j2}|^q + \dots + |x_{ip} - x_{jp}|^q)}$$

Fonte: Linden (2009)

A versão mais conhecida desta métrica é quando $q = 2$, a chamada distância Euclidiana (LINDEN, 2009). Como esta pesquisa considera apenas uma dimensão para o cálculo da distância, tem-se:

Quadro 6 – Distância Euclidiana

$$d(i,j) = \sqrt{|x_i - x_j|^2}$$

Fonte: Próprio autor

Nesse caso, a raiz quadrada e a potência eliminam-se, ficando apenas o módulo da diferença da característica escolhida. Este caso, quando $q = 1$, é a versão conhecida como distância de Manhattan. Dessa forma, quanto maior for esse número, mais heterogêneo é o grupo, pois se os estudantes tiverem conhecimentos parecidos, essa distância tende a ser pequena. Por exemplo, se em um mesmo grupo houver estudantes com o mesmo conhecimento, medido pela nota do estudante, a distância será zero.

Assim, a função de cálculo da heterogeneidade dos conhecimentos é descrita como:

$$f(het_conhect_i) = \sum_{a=1}^n \sum_{b=1}^n |(conhect_a - conhect_b)| \quad (2)$$

Onde:

- i indica o índice do grupo;
- n indica o total de estudantes da classe;
- a, b indicam estudantes pertencentes ao grupo, isto é, $a, b \in G_i$.

A função de cálculo das interações do grupo calcula a soma das interações, de forma que, quanto maior o número de interações ocorridas entre os membros, maior será esse número. Se os membros do grupo interagiram pouco, esse número tende a diminuir e se os membros não interagiram, esse número será zero. Como a matriz de interações é uma matriz simétrica, a soma é dividida por 2, para não considerar as interações em duplicidade.

Assim, a função de cálculo das interações é descrita como:

$$f(int_i) = \frac{\sum_{a=1}^n \sum_{b=1}^n (interações_{a,b})}{2} \quad (3)$$

Onde:

- i indica o índice do grupo;
- n indica o total de estudantes da classe;
- a, b indicam estudantes pertencentes ao grupo, isto é, a e $b \in G_i$.

A função de cálculo da inter-homogeneidade dos grupos tem como objetivo balancear os grupos entre si, em relação aos conhecimentos, de forma que os grupos sejam heterogêneos internamente (intra-heterogêneos), mas balanceados entre si (inter-homogêneos). Como os grupos podem ter tamanhos diferentes, variando em até um estudante por grupo, foi necessário calcular a distância média alcançada pelo grupo (distância alcançada dividida pelos membros do grupo (z_i)) e a distância média da classe (a soma das distâncias médias alcançadas, dividida pela quantidade de grupos (g)). Assim, verifica-se o balanceamento de cada grupo em relação à classe como um todo, utilizando-se a soma do quadrado da diferença entre a média da classe e a média de cada grupo. De forma que, quanto maior a diferença dos grupos em relação à média da classe, maior será esse número; e quanto menor for esse número, mais balanceados os grupos estão em relação à classe.

Assim, a função de cálculo da inter-homogeneidade é descrita como:

$$f(inter_homo_i) = \sum_{i=1}^g (média_c - média_i)^2 \quad (4)$$

Onde:

- i indica o índice do grupo;
- g indica o total de grupos;
- $média_c$ indica a média da classe;
- $média_i$ indica a média do grupo G_i .

Como a função objetivo é uma função de maximização, e o objetivo nesse caso é minimizar as diferenças entre os grupos, foi necessário incluir o 1 (um) menos a inter-homogeneidade na fórmula (1), a fim de permitir a maximização.

Para garantir que a função de inter-homogeneidade não possuísse um valor além de 1 (um), gerando assim um valor negativo na fórmula, foi necessária a normalização utilizando os máximos e mínimos, conforme feito com os dados. Como a função objetivo (1) é dada por uma única fórmula contendo as três demais, (2), (3) e (4), todas as medidas foram normalizadas.

Com a normalização, cada medida deve chegar a no máximo 1 (um), permitindo assim que o valor máximo da função objetivo seja estabelecido. No caso de os pesos (ω_1 e ω_2) serem iguais a 1 (um), este valor será 3 (três).

Para normalizar a heterogeneidade, fórmula (2), foi necessário calcular a distância máxima possível para cada grupo, considerando os valores dos dados já normalizados anteriormente no intervalo [0,1]. Assim, como o menor valor possível para os conhecimentos é 0 (zero) e o maior valor possível é 1 (um), o melhor caso de heterogeneidade dos conhecimentos é quando se têm estudantes com valor 0 (zero) e estudantes com valor 1 (um) em um mesmo grupo, pois dessa forma tem-se o máximo da heterogeneidade possível. No entanto, esse valor dependerá da quantidade de estudantes por grupo, sendo necessário estabelecer uma forma de calcular esse valor. Portanto, a distância máxima, para cada tamanho de grupo, foi calculada pela seguinte fórmula:

$$Dist_{max_{z_i}} = \left(\left\lfloor \frac{z_i}{2} \right\rfloor * \left\lfloor \frac{z_i}{2} \right\rfloor * |max - min| \right) \quad (5)$$

Onde:

- z_i indica o tamanho do grupo;
- max indica o valor máximo = 1;
- min indica o valor mínimo = 0.

O Quadro 7 mostra os valores máximos das distâncias para grupos de 4 e 5 estudantes, por exemplo, considerando o melhor caso de heterogeneidade, isto é, estudantes com valores 1 (um) e 0 (zero) para os conhecimentos.

Quadro 7 – Máximas distâncias em relação ao conhecimento

4 Estudantes						5 Estudantes						
E_n	1	2	3	4	Dist	E_n	1	2	3	4	5	Dist
Conhec	1	1	1	1	0	1	1	1	1	1	0	0
	1	1	1	0	3	1	1	1	1	0	0	4
	1	1	0	0	4	1	1	1	0	0	0	6
	1	0	0	0	3	1	1	0	0	0	0	6
	0	0	0	0	0	1	0	0	0	0	0	4
	0	0	0	0	0	0	0	0	0	0	0	0

Fonte: Próprio autor

No Quadro 7, encontram-se em destaque as distâncias máximas, sendo que para um grupo de 4 estudantes, o melhor caso de heterogeneidade seriam dois com valor 1 (um) e dois com valor 0 (zero), alcançando uma distância máxima de 4 (quatro). Para um grupo de 5 estudantes, duas situações permitem alcançar a distância máxima, sendo três com valor 1 (um)

e dois com valor 0 (zero); ou dois com valor 1 (um) e três com valor 0 (zero). Assim, se o grupo for totalmente homogêneo, a sua distância somada será 0 (zero), que normalizada pela $dist_{max_{z_i}}$ terá o valor 0 (zero); e um grupo que alcança a distância máxima, sendo esta normalizada pela $dist_{max_{z_i}}$ terá o valor 1 (um), ficando sempre a heterogeneidade no intervalo [0,1].

Para normalizar as interações, fórmula (3), foi necessário calcular o máximo de interações possíveis em um grupo, considerando os valores dos dados já normalizados anteriormente no intervalo [0,1]. Assim, como o menor valor possível para as interações é 0 (zero) e o maior valor possível é 1 (um), o melhor caso para as interações é quando todos os estudantes interagiram o máximo possível com seus colegas do grupo, isto é, todos têm o valor 1 (um) na matriz de interação. No entanto, esse valor dependerá da quantidade de estudantes por grupo, sendo necessário estabelecer uma forma de calcular esse valor. Portanto, a interação máxima, para cada tamanho de grupo, foi calculada pela seguinte fórmula:

$$Int_{max_{z_i}} = \frac{(z_i * (z_i - 1))}{2} \quad (6)$$

Onde:

- z_i indica o tamanho do grupo.

O Quadro 8 mostra a matriz de interação para um grupo de 5 estudantes, contendo o máximo de interação entre todos os membros do grupo.

Quadro 8 – Matriz com as interações máximas

	1	2	3	4	5
1	0	1	1	1	1
2	1	0	1	1	1
3	1	1	0	1	1
4	1	1	1	0	1
5	1	1	1	1	0

Fonte: Próprio autor

No Quadro 8, está em destaque a diagonal da matriz e observa-se que a soma das interações, conforme a fórmula (3), é igual a 10. Esse é também o valor para $int_{max_{z_i}}$, para um grupo de 5 estudantes, onde todos os integrantes do grupo interagem o máximo entre si. Assim, se o grupo não tiver interações, a soma das interações do grupo será 0 (zero), que normalizada pela $int_{max_{z_i}}$ terá o valor 0 (zero); e um grupo que alcança a interação máxima,

sendo esta normalizada pela int_max_{zi} terá o valor 1 (um), ficando sempre as interações no intervalo [0,1].

Para normalizar a inter-homogeneidade, fórmula (4), foi necessário calcular o quadrado da distância média da classe, utilizada para o cálculo do quadrado das diferenças com a média de cada grupo. Assim, como o menor valor possível para a diferença é 0 (zero), quando o grupo tiver a mesma média que a classe; e o maior valor possível para a diferença é a própria média, quando o grupo tiver média 0 (zero); a normalização é feita dividindo-se a inter-homogeneidade alcançada pela média da classe ao quadrado:

$$inter_homo_i^* = \frac{inter_homo_i}{(média_c)^2} \quad (7)$$

Onde:

- $inter_homo_i$ é a inter-homogeneidade calculada pela fórmula (4).

Dessa forma, um grupo que tenha a mesma média que a classe terá sua inter-homogeneidade igual a 0 (zero), sendo esta normalizada pela $inter_homo_i^*$ terá o valor 0 (zero); e se o grupo for totalmente homogêneo, a sua distância média será 0 (zero), sendo a sua inter-homogeneidade, pela fórmula (4), a própria média da classe ao quadrado, que normalizada pela $inter_homo_i^*$ terá o valor 1 (um), ficando sempre a inter-homogeneidade no intervalo [0,1].

Tem-se, portanto, o problema da formação de grupos para aprendizagem colaborativa, considerando as duas características definidas: conhecimentos e interações, e a inter-homogeneidade entre os grupos, definido nas fórmulas (1), (2), (3) e (4) descritas, e normalizadas pelas fórmulas (5), (6) e (7).

Assim, em relação à formação de grupos para aprendizagem colaborativa, esta pesquisa, fundamentada no mapeamento sistemático realizado, adotou a formação de grupos de *forma automática*, utilizando a *abordagem selecionada* e como *critério* grupos **heterogêneos** considerando os **conhecimentos** e as **interações** dos estudantes. Destaca-se em relação aos resultados obtidos pelo mapeamento sistemático, que algoritmos genéticos foram um dos mais utilizados, resultado corroborado por Cruz e Isotani (2014), e que fundamentam a escolha dessa meta-heurística para a solução do problema proposto. Por isso, a seguir discute-se a modelagem do problema para o algoritmo genético.

4.2 MODELAGEM DO ALGORITMO GENÉTICO

Para que seja possível aplicar algoritmos genéticos (AGs) na solução de problemas de otimização, é preciso modelar o problema de forma que o algoritmo possa tratá-lo. Esta etapa

é crucial para o êxito da solução, pois representará o problema de modo que o algoritmo possa compreender e recombinar as soluções, a fim de que as evoluções encontrem soluções melhores.

Para utilizar um AG de maneira eficiente, não basta simplesmente aplicar as formas genéricas desse algoritmo, é necessário encontrar uma representação adequada para o problema e aplicar os operadores genéticos adequados para esta representação. Além disso, é importante embutir o máximo de conhecimento específico do problema no AG (EIBEN; SMITH, 2007; LINDEN, 2012).

Assim, conforme as terminologias do AG, definem-se, no Quadro 9, os principais elementos para o problema proposto. Cada indivíduo é uma permutação dos E estudantes, distribuídos em g grupos (característica). O valor de cada alelo é o identificador dos estudantes, sendo um número inteiro no intervalo $[0, n-1]$, onde n é o número de estudantes. Cada estudante está em uma posição do vetor, que forma a estrutura do cromossomo. A função de avaliação calcula o *fitness* para cada indivíduo conforme a função objetivo descrita.

Quadro 9 – Analogia dos elementos no AG

Linguagem da ciência natural	AG	Problema Proposto
Cromossomo	Indivíduo	Permutação dos estudantes em g grupos
Gene	Característica	$g =$ Grupos formados
Alelo	Valor	Identificador dos estudantes, representados por números inteiros
Locus	Posição	Índice do vetor do cromossomo
Genótipo	Estrutura	Estrutura do cromossomo
Fenótipo	Expressão do Indivíduo no Ambiente	Função de Avaliação

Fonte: Adaptado de Linden (2012)

4.2.1 Representação Adotada

Para a representação dos indivíduos em um AG, é importante escolher a maneira certa para representar o problema desejado (EIBEN; SMITH, 2007; LINDEN, 2012). Entre as possíveis representações há: representação binária, representação inteira e representação inteira por permutação.

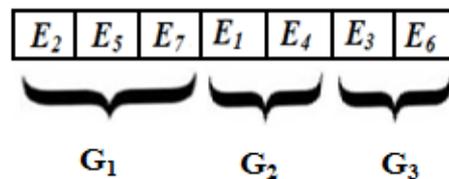
Linden (2012) também recomenda que a representação seja a mais simples possível, que as soluções proibidas preferencialmente não tenham uma representação e se houver

alguma condição específica para o problema, que esteja implícita na representação. No caso desta pesquisa, a representação inteira por permutação torna-se a mais adequada, e já atende as três recomendações citadas. Por exemplo, a formação de grupos definida como problema desta pesquisa requer que um estudante esteja em apenas um grupo, essa é uma condição imposta pelo problema, e a representação por permutação já inclui esta condição.

Portanto, adota-se como representação dos indivíduos a representação inteira por **permutação**. Cada indivíduo representa um conjunto de grupos formados (g) pelos estudantes (E_n) da classe (C). Isto é, cada indivíduo ou cromossomo é uma lista de tamanho n , conforme o número de estudantes da classe. A Figura 34 mostra um exemplo.

Figura 34 – Exemplo da representação adotada

Permutação dos estudantes em g grupos



Fonte: Próprio autor

Observa-se na Figura 34 que nem todos os grupos têm o mesmo tamanho, já que o número de estudantes (n) pode não ser múltiplo da quantidade de estudantes definida por grupo (m). No entanto, na formulação do problema essa questão está tratada de forma que os grupos podem variar em tamanho, sem que esta diferença exceda o número de um estudante. Assim, é necessário definir uma forma para a codificação e para a decodificação do problema.

Adota-se a codificação proposta por Yannibelli e Amandi (2012), que já considera as mesmas restrições impostas ao problema nesta pesquisa, isto é, cada estudante pertence apenas a um grupo, e os grupos variam em tamanho com no máximo um estudante de diferença. No entanto, a codificação diferencia-se da proposta pelas autoras, pois no caso de Yannibelli e Amandi (2012) o valor de g (número de grupos) é conhecido. No caso desta pesquisa, o valor de m (estudantes por grupo) é conhecido (entre 3 a 5 estudantes).

O primeiro passo da codificação é definir quantos grupos serão formados, sendo que $g = (n/m)$. Se g for um número inteiro, isto é, não houver resto desta divisão, então n é múltiplo de m e todos os grupos terão o mesmo tamanho $z_i = m$. O processo segue alocando os estudantes a cada grupo, de forma que cada grupo é composto pelos estudantes na sequência da lista de n estudantes, de m em m . Isto é, se $m = 2$ ($z_i = m = 2$), por exemplo, os estudantes

vão sendo alocados de 2 em 2 para cada grupo. Portanto, os grupos são compostos pelos estudantes da posição $[(((i-1) * z_i) + 1), \dots, (i * z_i)]$.

Se g não for um número inteiro, isto é, há resto na divisão de n/m , então n não é múltiplo de m , e os grupos terão tamanhos diferentes. Neste caso, é preciso considerar a restrição imposta de que os grupos devem variar em no máximo um estudante. Portanto, alguns grupos terão $z_1 = m + 1$, e alguns grupos terão $z_2 = m$. O processo define quantos grupos terão $z_1 = (m + 1)$ estudantes considerando $g_1 = (n - (m * \text{parte inteira de } g))$ e os demais grupos terão $z_2 = m$ estudantes. Por exemplo, se $n = 7$ e $m = 2$, então $g_1 = (7 - (2 * \text{parte inteira } g = 3)) = 1$. Portanto 1 grupo terá 3 estudantes (z_1) e os demais (2 grupos) terão 2 estudantes (z_2) cada um (exemplo dado na Figura 28). O processo segue alocando os estudantes em cada grupo, seguindo as regras:

- a) Os grupos de $i = 1$ até g_1 seguem alocando os estudantes da posição $[(((i-1) * z_1) + 1), \dots, (i * z_1)]$;
- b) O grupo $(g_1 + 1)$ segue alocando os estudantes da posição $[((g_1 * z_1) + 1), \dots, ((g_1 * z_1) + z_2)]$;
- c) Os grupos $f = (g_1 + 2)$ até g (total de grupos) seguem alocando os estudantes da posição $[(((g_1 * z_1) + z_2) + ((f - (g_1 + 2)) * z_2) + 1), \dots, (((g_1 * z_1) + z_2) + ((f - (g_1 + 1)) * z_2))]$.

Seguindo com o exemplo dado de $n = 7$ e $m = 2$, onde $g_1 = 1$ grupo terá 3 estudantes ($z_1 = 3$) e os outros 2 grupos terão 2 estudantes ($z_2 = 2$), tem-se:

- a) O grupo de $i = 1$ até $g_1 = 1$, aloca os estudantes da posição $[(((i=1-1) * z_1=3) + 1), \dots, (i=1 * z_1=3)] = [1, 3]$;
- b) O grupo $(g_1 + 1) = 2$ segue alocando os estudantes da posição $[((g_1=1 * z_1=3) + 1), \dots, ((g_1=1 * z_1=3) + z_2=2)] = [4,5]$;
- c) O grupo $f = (g_1 + 2) = 3$ até $g=3$ (total de grupos) segue alocando os estudantes da posição $[(((g_1=1 * z_1=3) + z_2=2) + ((f=3 - (g_1=1 + 2)) * z_2=2) + 1), \dots, (((g_1=1 * z_1=3) + z_2=2) + ((f=3 - (g_1=1 + 1)) * z_2=2))] = [6,7]$.

4.2.2 População

A geração da população inicial pode ser feita de forma aleatória, o que de maneira geral permite uma boa distribuição do espaço de busca (LINDEN, 2012). Sobre a população incidem os operadores genéticos para que ocorra a evolução do AG (EIBEN; SMITH, 2007; LINDEN, 2012).

Definir o tamanho da população não é uma tarefa fácil. A maioria dos estudos publicados utiliza 100, mas não há consenso quanto a isso (LINDEN, 2012). Uma sugestão inicial é 40*características do cromossomo (LINDEN, 2012), no caso desta pesquisa seria g – número de grupos. No entanto, como o objetivo é trabalhar com uma quantidade massiva de dados, 40*1.000 grupos, por exemplo, seria igual a 40.000 indivíduos iniciais, o que pode ficar pesado para a função de avaliação, gerando um elevado custo computacional. Portanto, assim como Yannibelli e Amandi (2012), adotam-se inicialmente 100 indivíduos, sendo que **a população inicial será composta de 100 diferentes permutações aleatórias de n estudantes em g grupos**, uniformemente distribuída.

Em relação ao modelo da população para a seleção dos sobreviventes, adota-se para esta pesquisa o *Generational Model*, onde a população é substituída toda de uma vez pelos seus descendentes. Assim, cada população existe por apenas um ciclo. Utiliza-se ainda a estratégia do **Elitismo**, onde os k melhores indivíduos da geração atual t , passam para a próxima geração $t + 1$, garantindo que o desempenho do AG sempre cresça com o decorrer das gerações. O valor de k foi definido como $k = 1$, isto é, 1 indivíduo (o melhor da geração atual) passa para a próxima geração. Esse valor foi escolhido para evitar a convergência genética e manter certa diversidade genética na população. Além disso, o operador genético de mutação será utilizado (EIBEN; SMITH, 2007; LINDEN, 2012).

4.2.3 Função de Avaliação (*Fitness Function*)

A fim de maximizar a função objetivo, que representa o problema de formação de grupos levando em conta a heterogeneidade dos conhecimentos e as interações dos estudantes do grupo, além da inter-homogeneidade entre os grupos, a qualidade de um indivíduo como solução do problema que está sendo analisado pelo AG (o *fitness*) deverá ter o maior valor possível.

Portanto, a função de avaliação será a função objetivo, fórmula (1), na qual estão embutidas as fórmulas (2), (3) e (4) definidas na seção 4.1.2 (Formulação do problema):

$$\text{Maximizar } f(x) = \omega_1 \sum_{i=1}^g f(\text{het_connect}_i) + \omega_2 \sum_{i=1}^g f(\text{int}_i) + (1 - \sum_{i=1}^g f(\text{inter_homo}_i)) \quad (1)$$

Onde:

$$f(\text{het_connect}_i) = \sum_{a=1}^n \sum_{b=1}^n |(\text{connect}_a - \text{connect}_b)| \quad (2)$$

e

$$f(\text{int}_i) = \frac{\sum_{a=1}^n \sum_{b=1}^n (\text{interações}_{a,b})}{2} \quad (3)$$

e

$$f(\text{inter_homo}_i) = \sum_{i=1}^g (\text{média}_C - \text{média}_i)^2 \quad (4)$$

Essa função calcula um valor numérico (*fitness*), normalizado pelas fórmulas (5), (6) e (7), que reflete o quão heterogêneo, o quão interacional e o quão balanceados são os grupos formados por aquele indivíduo (conjunto de g grupos formados pelos n estudantes). Dessa forma, a função de avaliação diferencia as boas e as más soluções do problema.

Como os valores são normalizados, a função de avaliação terá sempre um valor no intervalo $[0,1]$ para cada medida utilizada, conforme descrito na seção 4.1.2. Assim, se os pesos ω_1 e ω_2 forem iguais a 1 (um), o valor do *fitness* será sempre no intervalo $[0,3]$. Se os pesos forem diferente de 1 (um) o valor do *fitness* variará conforme os pesos adotados. Por exemplo, se o ω_1 for 2 (dois) e o ω_2 for 1 (um), o valor do *fitness* será no intervalo $[0,4]$ ($(2 * 1) + (1 * 1) + 1$).

A seguir, discute-se a seleção dos pais, na qual a função de avaliação é usada para identificar os “melhores” indivíduos.

4.2.4 Seleção de Pais

O processo de evolução do AG inicia-se com a seleção dos pais, quando os melhores indivíduos (os mais bem avaliados pela função de avaliação - *fitness*) são selecionados para gerar descendentes (filhos).

É importante que os indivíduos com menor qualidade (*fitness* menor) também possam se reproduzir. Como já foi adotado o Elitismo para a seleção dos sobreviventes (onde o “melhor” indivíduo passa para a próxima geração), adota-se, para a seleção de pais, o método do **torneio**. Esse método de seleção foi adotado para evitar a dominação dos melhores indivíduos, como é o caso da seleção pela roleta, utilizando o tamanho do torneio $k = 3$, reduzindo assim a possibilidade de convergência prematura e ampliando a diversidade da população (EIBEN; SMITH, 2007; LINDEN, 2012). Esta escolha é também fundamentada em Yannibelli e Amandi (2012), que adotaram esse método e tiveram bons resultados considerando um número maior de estudantes (milhares).

Três indivíduos são aleatoriamente selecionados, e seus *fitness* são comparados. O melhor deles (com maior *fitness*) é selecionado para ser o primeiro pai. Tira-se o indivíduo selecionado da população e os outros dois indivíduos não selecionados retornam para a população. Repete-se a operação para selecionar o segundo pai. Esse processo é aplicado sucessivamente até $M/2$ vezes, onde M é o tamanho da população. Dessa forma, obtém-se $M/2$ pares de indivíduos para iniciar o processo de recombinação (*crossover*). As Figuras 35(a) e 35(b) apresentam um exemplo do problema proposto.

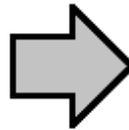
Figura 35(a) – Exemplo de população para o problema proposto

Indivíduo								<i>Fitness</i>
1	E ₁	E ₂	E ₃	E ₄	E ₅	E ₆	E ₇	2,667
2	E ₁	E ₃	E ₇	E ₄	E ₆	E ₂	E ₅	2,601
3	E ₁	E ₃	E ₄	E ₂	E ₅	E ₆	E ₇	2,856
4	E ₂	E ₃	E ₅	E ₁	E ₄	E ₆	E ₇	2,587
5	E ₁	E ₂	E ₆	E ₄	E ₅	E ₃	E ₇	2,598
6	E ₁	E ₂	E ₄	E ₃	E ₅	E ₆	E ₇	2,903
7	E ₁	E ₃	E ₆	E ₄	E ₅	E ₂	E ₇	2,314
8	E ₂	E ₃	E ₅	E ₄	E ₆	E ₁	E ₇	2,512

Fonte: Próprio autor

Figura 35(b) – Exemplo de torneios para o problema proposto

Indivíduo	<i>Fitness</i>	Torneios		
1	2,667	1	7	8
2	2,601	2	3	5
3	2,856	6	4	5
4	2,587	2	7	1
5	2,598	5	4	7
6	2,903	3	4	2
7	2,314	4	2	6
8	2,512	4	2	5



Torneios aplicados com $k = 3$. O indivíduo hachurado vence o respectivo torneio, até selecionar o último pai.

Fonte: Próprio autor

4.2.5 Recombinação ou Cruzamento (*Crossover*)

O processo de seleção de pais determinou pares de indivíduos que devem ser recombinados, fazendo com que a evolução aconteça.

A recombinação é aplicada de acordo com uma probabilidade, chamada taxa de *crossover* (p_c), definida para esta pesquisa em 0.8 (80%) (EIBEN; SMITH, 2007; LINDEN, 2012, YANNIBELLI; AMANDI, 2012). Para aplicar o *crossover*, sorteia-se um número entre 0 e 1, se for menor ou igual a probabilidade, então o operador atua sobre os pais selecionados.

Como a representação adotada para o problema foi a inteira por permutação, é necessário escolher um operador de *crossover* adequado a essa representação. Para essa representação tem-se o *crossover* mapeamento parcial, o de ordem e o de ciclo, entre outros. Portanto, adota-se nesta pesquisa o ***crossover* de ordem** (*order crossover*). Esta escolha é

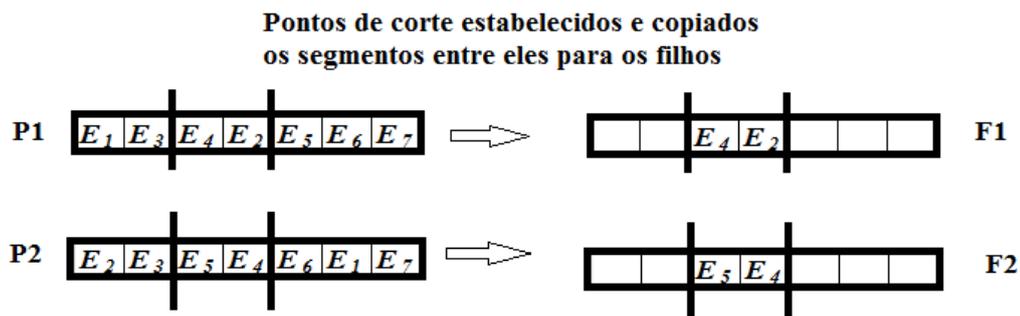
também fundamentada em Yannibelli e Amandi (2012), que adotaram esse operador e tiveram bons resultados considerando um número maior de estudantes (milhares).

Esse operador é aplicado aos dois pais selecionados, gerando dois novos filhos (descendentes), seguindo os seguintes passos:

- Estabelece dois pontos de corte aleatoriamente e copia o segmento entre eles do primeiro pai para o primeiro filho;
- A partir do segundo ponto de corte, copia os genes do segundo pai na ordem em que aparecem para o primeiro filho; se o gene já foi copiado (pelo segmento do primeiro pai) passa para o próximo.

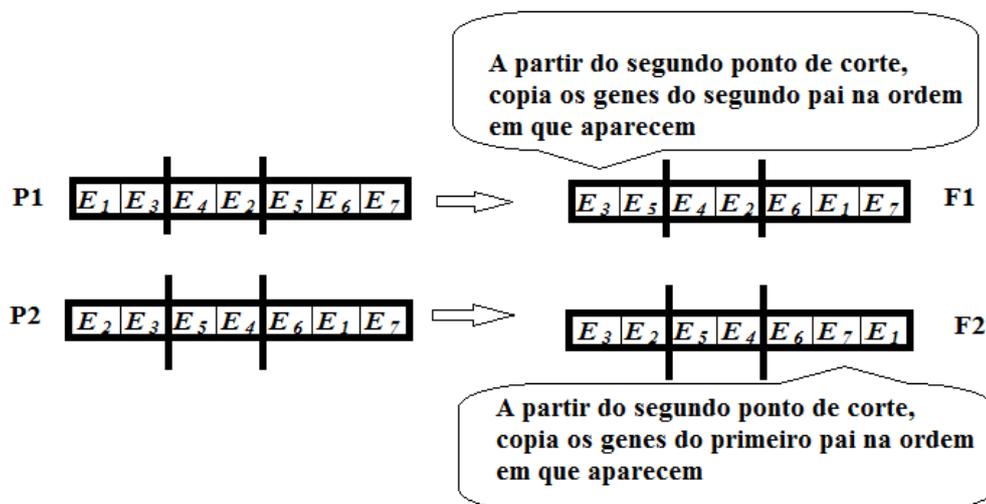
Para o segundo filho, trocam-se as posições dos pais, isto é, o primeiro pai passa a ser o segundo e o segundo pai passa a ser o primeiro. As Figuras 36(a) e 36(b) ilustram essa situação.

Figura 36(a) – Exemplo *crossover* de ordem



Fonte: Próprio autor

Figura 36(b) – Exemplo *crossover* de ordem



Fonte: Próprio autor

4.2.6 Mutação

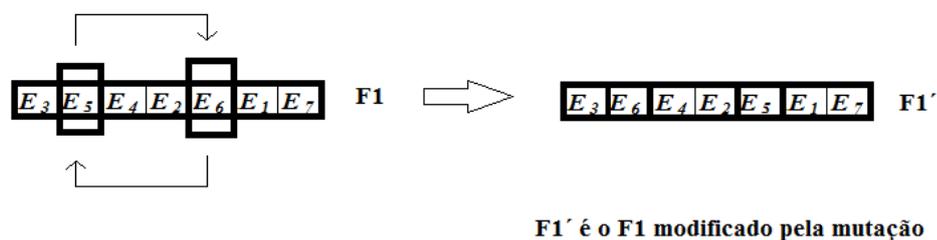
O operador de mutação é um operador aleatório, aplicado a um pai para gerar um novo filho ou aplicado aos filhos gerados para gerar alguma modificação. Esse operador pode ser usado também para evitar a convergência genética e introduzir diversidade na população (EIBEN; SMITH, 2007).

Assim como no *crossover*, a mutação tem associada a ela uma probabilidade, chamada taxa de mutação (p_m), que deve ser baixa (geralmente até 0.05). Assim, para esta pesquisa, essa taxa foi definida empiricamente em 0.03 (3%). Para aplicar a mutação, sorteia-se um número entre 0 e 1, se for menor ou igual a probabilidade, então o operador atua sobre o gene; caso contrário, a mutação não é aplicada (EIBEN; SMITH, 2007; LINDEN, 2012).

A mutação também está ligada a representação adotada para o problema e para a representação por permutação, adotada nesta pesquisa, têm-se: *swap mutation*, *insert mutation*, *inversion mutation*, entre outros (EIBEN; SMITH, 2007). Portanto, para esta pesquisa foi adotado o operador *swap mutation*. Esta escolha é fundamentada em Yannibelli e Amandi (2012), que adotaram esse operador e tiveram bons resultados considerando um número maior de estudantes (milhares).

Esse operador funciona selecionando aleatoriamente dois genes do indivíduo e trocando os valores (alelos) de suas posições, no caso desta pesquisa, os estudantes. Por exemplo, para um dos filhos gerados pelo *crossover*, dois genes são escolhidos e trocam-se os estudantes entre esses dois genes. A Figura 37 ilustra esta situação.

Figura 37 – Exemplo *Swap mutation*



Fonte: Próprio autor

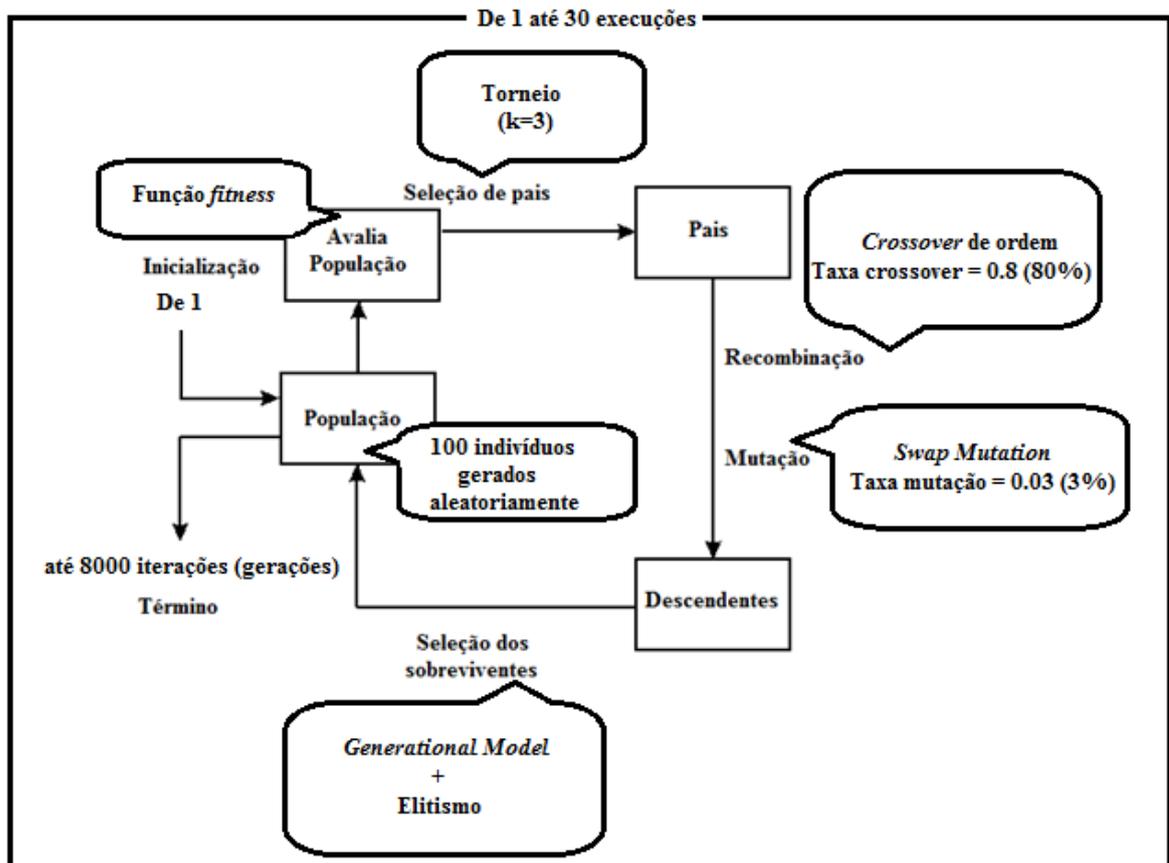
4.2.7 Configuração de Parâmetros

Para a execução do AG são necessários vários parâmetros como: o tamanho da população, a taxa de *crossover* e a taxa de mutação, entre outros. Assim, para esta pesquisa, foram adotados valores empiricamente, baseados na faixa de valores usualmente encontrados na literatura como parâmetros, sendo eles: População com 100 indivíduos; Taxa de *crossover*

de 80%; Taxa de mutação de 3% (EIBEN; SMITH, 2007; LINDEN, 2012; YANNIBELLI; AMANDI, 2012). O ajuste dos parâmetros de um AG tem um papel importante para qualidade das soluções encontradas. No entanto, não é uma tarefa fácil; além disso, a configuração adequada pode variar durante a execução do algoritmo (ANDRÉ; PARPINELLI, 2014). Portanto, o ajuste dos parâmetros é uma questão que será tratada em trabalhos futuros.

A Figura 38 apresenta um esquema gráfico resumindo a modelagem do algoritmo genético desenvolvido, em que, para cada componente do fluxograma do AG apresenta um balão correspondente, com as respectivas escolhas feitas em cada etapa.

Figura 38 – Esquema do algoritmo genético desenvolvido



Fonte: Próprio autor

Como o AG é um processo estocástico, o algoritmo foi executado 30 vezes, utilizando uma quantidade de 8.000 gerações (iterações), para se fazer a média dos resultados obtidos nas execuções. Dessa forma, multiplicando-se a quantidade de gerações pelos indivíduos da população, tem-se 800 mil avaliações executadas pela função de avaliação, o que permite uma boa exploração do espaço de busca (possíveis soluções). Esse número de gerações foi definido com base na convergência da função de avaliação, sendo que com esse número de

iterações o valor do *fitness* tende ao valor esperado, mesmo para os casos com maior número de estudantes.

Desta forma, a modelagem do AG baseou as escolhas tanto na fundamentação e mapeamento sistemático quanto nos trabalhos relacionados, fechando o ciclo de modelagem do algoritmo.

4.3 CONSIDERAÇÕES DO CAPÍTULO

Conforme o que foi descrito no capítulo 4, em resumo, a pesquisa gerou grupos de *forma automática*, utilizando como *abordagem* grupo **selecionado**, considerando os **conhecimentos** e as **interações** dos estudantes, e como *critério* grupos **heterogêneos**, com o **objetivo de maximizar a heterogeneidade dos conhecimentos no grupo e as interações dos estudantes do grupo**, além de gerar grupos balanceados entre si (**inter-homogêneos**).

Em relação ao AG definido para solucionar o problema, a representação adotada foi a inteira por **permutação**; com uma **população inicial de 100 diferentes indivíduos**, adotando-se o *generational model* como modelo para a população, e a estratégia do **Elitismo** ($k = 1$) para as gerações posteriores; utilizando o **torneio** ($k = 3$) para a seleção de pais; o *crossover* de **ordem** com uma **taxa de crossover – p_c de 80%** para a recombinação; e o *swap mutation* a uma **taxa de mutação – p_m de 3%** para a **mutação**.

Além dessas definições, o AG precisa do número de **iterações** como critério de parada do algoritmo. Esse número foi definido como **8.000**. Essa quantidade de iterações resulta em 800.000 avaliações por execução do AG.

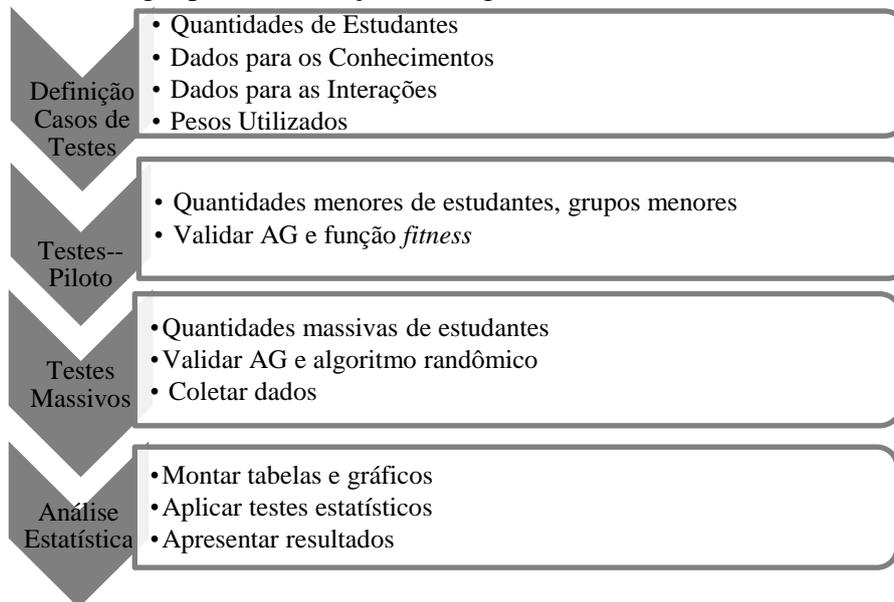
Dois algoritmos foram desenvolvidos para as devidas comparações: **AG** e **randômico** (base de comparação). No capítulo 5, a seguir, discutem-se a validação do AG, os testes realizados e os resultados obtidos.

5 VALIDAÇÃO, RESULTADOS OBTIDOS E DISCUSSÃO

Com o objetivo de avaliar a eficácia e a eficiência do AG desenvolvido nesta pesquisa, foram delineados diferentes casos de testes. Em avaliação de sistemas, qualidade de produto, a eficácia está geralmente associada à capacidade do sistema/programa de alcançar os objetivos estabelecidos (ISO 9126-1, 2003), neste caso, a formação de grupos heterogêneos na aprendizagem colaborativa, considerando uma quantidade massiva de estudantes. A eficiência está geralmente associada aos recursos necessários para o sistema/programa atingir os objetivos (ISO 9126-4, 2004), neste caso, a relação custo/benefício em função do tempo de processamento.

Na seção 5.1 são apresentados a validação e o método base de comparação; na seção 5.2 são apresentados os testes-piloto; na seção 5.3 são apresentados os testes massivos; na seção 5.4 são apresentados os métodos estatísticos utilizados e na seção 5.5 são apresentados os resultados obtidos e a discussão dos resultados. A Figura 2 apresenta a metodologia utilizada para a validação. Cada seção a seguir, discute e detalha as etapas executadas no processo de validação, juntando-se as seções 5.4 e 5.5 na última etapa.

Figura 2 – Metodologia para a validação dos algoritmos



Fonte: Próprio autor

Como não foi possível executar os testes com experimentos reais, diferentes casos de testes foram simulados. As validações são discutidas a seguir.

5.1 VALIDAÇÃO DOS ALGORITMOS

Esta seção discute o processo de validação, definindo a quantidade de estudantes (milhares de estudantes), número de estudantes por grupo, dados considerados para os conhecimentos e as interações e o método de comparação utilizado para a validação.

5.1.1 Configuração dos dados

Como o objetivo desta pesquisa é avaliar a adoção de algoritmos genéticos para a formação de grupos na aprendizagem colaborativa, considerando um contexto massivo de dados (milhares de estudantes), em primeiro lugar foram definidas as quantidades de estudantes para as quais os algoritmos foram aplicados. As quantidades de estudantes definidas foram: 1.000, 2.000, 3.000, 4.000, 5.000, 6.000, 7.000, 8.000, 9.000 e 10.000 estudantes. Nos trabalhos relacionados, verificou-se que as quantidades de estudantes utilizadas geralmente foram de dezenas ou centenas de estudantes (ANI et al., 2010; GOGOULOU et al., 2007; MORENO; OVALLE; VICARI, 2012; WANG; LIN; SUN, 2007), sendo que estes trabalhos utilizaram dados reais em seus experimentos. Apenas um trabalho (YANNIBELLI; AMANDI, 2012) utilizou uma quantidade maior de estudantes, com milhares de estudantes, utilizando dados simulados em seus experimentos. Assim, esta pesquisa contribui com resultados para uma quantidade maior de estudantes, para além do que os trabalhos relacionados aplicaram suas soluções, utilizando dados simulados neste primeiro momento, a fim de possibilitar trabalhos futuros com dados reais.

Outra definição importante diz respeito à quantidade de estudantes por grupo. Herreid e Drake (2000) afirmam que a quantidade ideal de estudantes para a aprendizagem colaborativa fica em torno de 4-5 estudantes por grupo. MICHAELSEN; KNIGHT; FINK (2004) afirmam que para aprendizagem colaborativa os grupos geralmente têm 4 estudantes ou menos, e para as atividades de aprendizagem baseada em times (do inglês, *team-based learning*) os grupos ficam em torno de 5-7 estudantes. Como esta pesquisa efetua uma avaliação sob uma quantidade massiva de estudantes, os grupos para aprendizagem colaborativa foram definidos com 5 estudantes por grupo, considerando-se esse um número limítrofe para essa abordagem.

Além dessas definições, como a função objetivo que representa a formação de grupos busca gerar grupos heterogêneos em relação aos conhecimentos e com o maior número de interações entre os membros do grupo, foi preciso definir o conjunto de valores para os

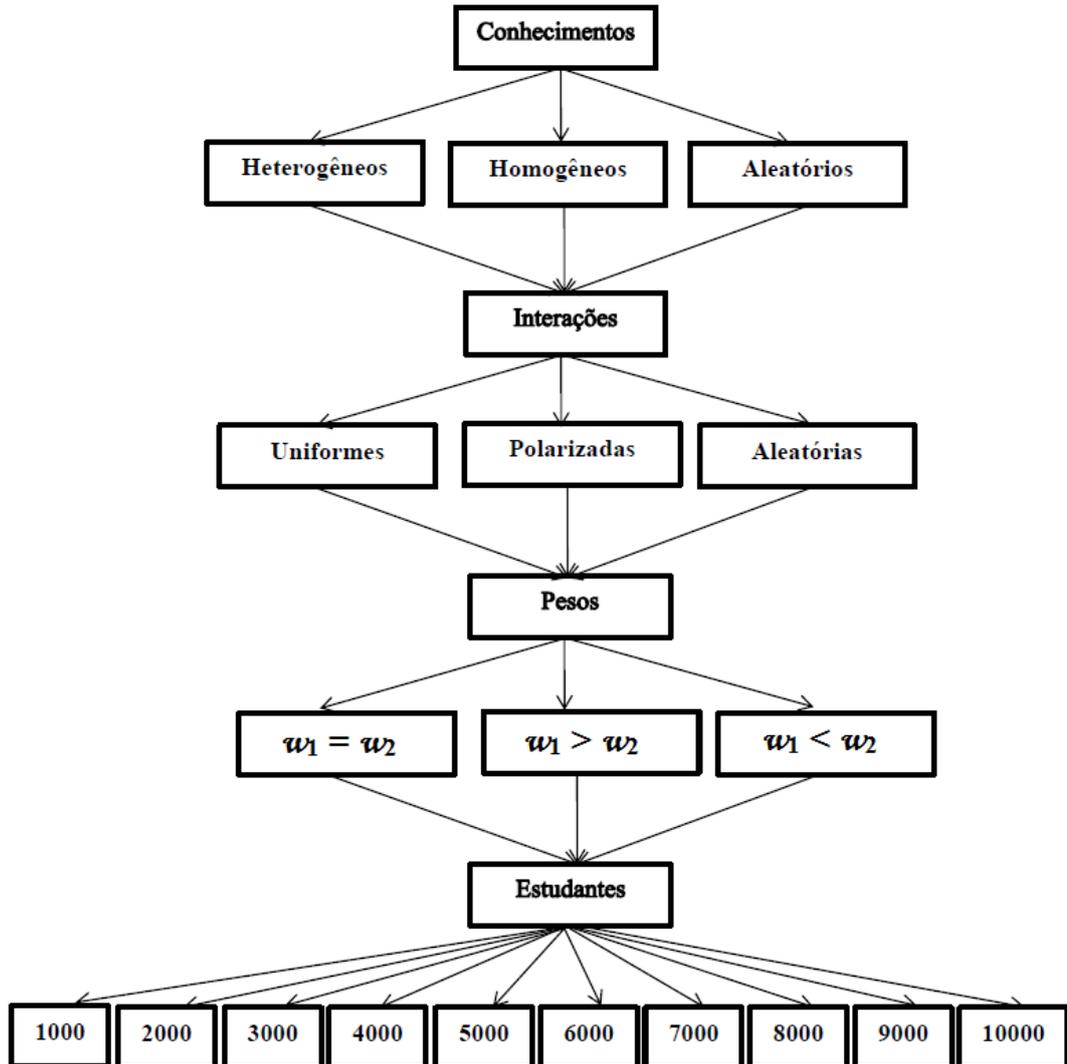
conhecimentos e para as interações dos n estudantes em cada caso de teste. Foram definidas três situações tanto para os conhecimentos como para as interações.

Para os conhecimentos, no primeiro caso, foram considerados conhecimentos **heterogêneos**, com a classe dividida em duas partes: metade dos estudantes com valor igual a 1 (um) e metade dos estudantes com valor igual a 0 (zero). Essa configuração possibilita gerar grupos ótimos em relação à heterogeneidade dos conhecimentos juntando estudantes com valores 1 (um) e 0 (zero) no mesmo grupo. No segundo caso, foram considerados conhecimentos **homogêneos**, com a classe dividida em 20% dos estudantes com valor igual a 1 (um), 20% dos estudantes com valor igual a 0 (zero) e os outros 60% (a maioria da classe) com valor igual a 0,5 (média da classe). Essa distribuição define uma classe mais homogênea e não possibilita a formação de grupos ótimos em relação à heterogeneidade dos conhecimentos, sendo que os grupos terão sempre mais estudantes na média e com o mesmo valor, nesse caso, 0,5. Assim, o desempenho do AG é avaliado para uma classe mais homogênea. No terceiro caso foram considerados valores **aleatórios** entre 0 e 10, normalizados no intervalo [0,1], para cada estudante, possibilitando a simulação de dados reais onde o pesquisador não tem apenas dados controlados.

Para as interações, no primeiro caso, foram considerados os valores máximos de interações entre todos os estudantes, por isso foram chamadas de interações **uniformes**. Essa configuração possibilita a geração de grupos ótimos em relação às interações, juntando no mesmo grupo estudantes que interagiram o máximo possível entre si. No segundo caso, foram consideradas interações **polarizadas** entre os estudantes, dividindo-se a classe em duas partes: a primeira metade interagiu apenas entre si, assim como a outra metade interagiu apenas entre si. Essa configuração possibilita uma classe mais dividida em relação às interações e não permite a formação de grupos ótimos em relação às interações. Assim, o algoritmo considera a melhor solução, juntando estudantes heterogêneos em relação aos conhecimentos, que tenham interagido entre si, sempre que possível. No terceiro caso, foram considerados valores **aleatórios** entre 0 e 100 para as interações, normalizados no intervalo [0,1], para cada par de estudante, possibilitando a simulação de dados reais onde o pesquisador não tem apenas dados controlados.

Além disso, esta pesquisa prevê a aplicação de pesos para os conhecimentos e as interações, assim, mais 3 (três) casos são possíveis: 1) pesos iguais; 2) ω_1 maior que ω_2 e 3) ω_1 menor que ω_2 . Desta forma, com a combinação desses dados estabelecidos, foram testados os algoritmos desenvolvidos. A Figura 39 resume graficamente as definições feitas.

Figura 39 – Definições dos casos de testes



Fonte: Próprio autor

São 3 (três) casos para os conhecimentos e 3 (três) casos para as interações, gerando 9 (nove) casos de testes. Desta forma, são 9 (nove) casos dos dados para cada caso possível dos pesos, gerando 27 (9x3) casos finais de testes. Como esses casos foram validados para 10 diferentes quantidades de estudantes (de mil até dez mil), foram 270 casos de testes validados.

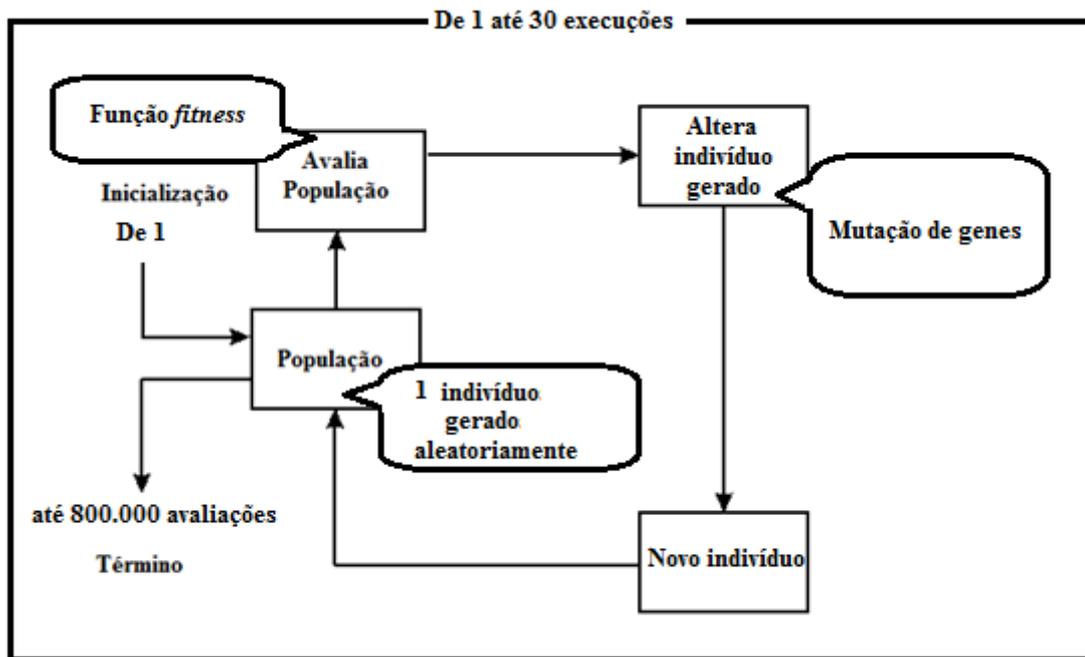
5.1.2 Método de Comparação

Para validar o AG desenvolvido, seu desempenho foi comparado com o método randômico, considerado, no âmbito desta pesquisa, um método base de comparação.

O algoritmo randômico distribui os estudantes aleatoriamente, gerando uma solução com os grupos conforme a quantidade definida de estudantes por grupo. A partir desse primeiro indivíduo é feita a avaliação (*fitness*), e o processo se repete, gerando novos

indivíduos através da mutação de dois genes aleatórios, gerando uma nova permutação dos estudantes, que é novamente avaliada, sucessivamente até o número de avaliações definido. Neste caso, 800 mil avaliações conforme feito para o AG (8.000 iterações para uma população de 100 indivíduos), conforme mostra a Figura 40. Desta forma, o AG e o algoritmo randômico são comparáveis entre si.

Figura 40 – Esquema do algoritmo randômico



Fonte: Próprio autor

O AG é um processo estocástico, e por isso não é possível considerar o resultado de apenas uma execução do algoritmo. Assim, o algoritmo foi executado 30 vezes para cada um dos casos de testes definidos, permitindo fazer uma média do desempenho do algoritmo. Esse número foi definido a fim de não sobrecarregar a execução devido ao grande número de estudantes já definidos (de mil até dez mil) e ao número de avaliações proposto (oitocentas mil). Desta forma, o algoritmo randômico também foi executado 30 vezes, utilizando-se a média do desempenho do algoritmo para a comparação com o AG. A média do *fitness*, a média do percentual de grupos heterogêneos e a média do tempo de processamento de cada uma das 30 execuções foram computadas. A partir desses dados, foram feitas as comparações entre o **AG** e o método **randômico**.

O *fitness* é calculado pela função de avaliação do algoritmo; o percentual de grupos heterogêneos é calculado considerando os grupos que alcançaram 100% da heterogeneidade possível, no caso de conhecimentos heterogêneos e pelo menos 70% (0.7) da heterogeneidade máxima possível no caso de conhecimentos homogêneos e/ou aleatórios e o tempo de

processamento é calculado considerando o início e o fim da execução de cada processamento, em segundos.

5.2 TESTES-PILOTO

A fim de verificar a correção dos algoritmos, validar a função objetivo definida (*fitness*) e identificar potenciais problemas com antecedência, testes-piloto foram realizados. Esses testes consideraram uma quantidade menor de estudantes, em grupos de 3-5 estudantes, para verificar também grupos com diferentes quantidades de estudantes.

Os testes foram realizados em um computador HP com processador AMD Phenom II X4 de 2.8Ghz, com 4 Gb de memória RAM, utilizando sistema operacional Ubuntu 14.04 LTS. Os algoritmos foram desenvolvidos em linguagem C.

Nesses testes, identificou-se, por exemplo, a necessidade:

- do balanceamento entre os grupos (inter-homogeneidade, fórmula (4));
- de se utilizar a média das distâncias do grupo na inter-homogeneidade, devido às diferentes quantidades de estudantes por grupo;
- da normalização dos dados (fórmulas (5), (6) e (7)), devido aos valores negativos gerados em alguns testes.

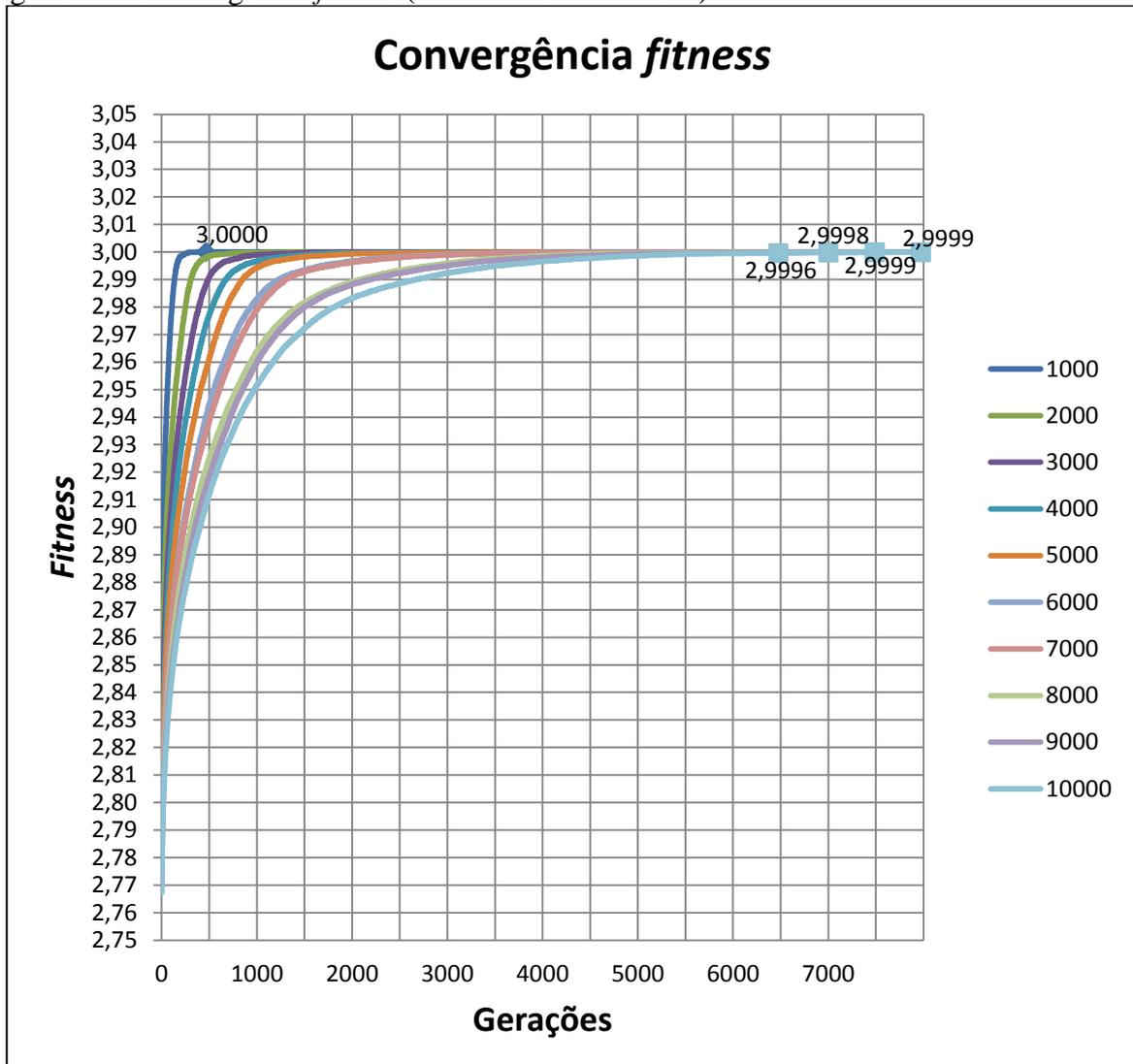
Desta forma, a função objetivo proposta inicialmente, que continha apenas as fórmulas (2) e (3), foi alterada para sua forma final, conforme descrito na seção 4.1.2, incluindo a fórmula (4).

Nos testes-piloto foi definido também o percentual de 70% para os grupos heterogêneos, no caso de conhecimentos não heterogêneos (homogêneos e/ou aleatórios), pois nesses casos não havia como se alcançar 100% da heterogeneidade. Assim, foi preciso definir um valor a ser alcançado para o qual o grupo ainda fosse considerado heterogêneo. Esse valor foi definido tomando como base a média utilizada na Universidade, já que a heterogeneidade refere-se aos conhecimentos dos estudantes. Desta forma, ao alcançar pelo menos 70% da distância máxima possível para o grupo, considerou-se o grupo heterogêneo, conforme descrito na seção 5.1.2.

Ainda nos testes-piloto foi definida a quantidade de gerações (iterações) para o AG observando-se a convergência do valor do *fitness*. O valor de 8.000 gerações foi escolhido, pois com este valor a função de avaliação alcançou o valor esperado para o *fitness*, para os casos com maior número de estudantes, pelo menos nos melhores resultados dos casos de heterogeneidade máxima dos conhecimentos e o máximo das interações possíveis. A Figura 41 mostra o gráfico da convergência dos resultados, de 1.000 a 10.000 estudantes, sendo que,

nesse caso, foi testado com uma quantidade massiva para avaliar o sucesso dos resultados. Essa simulação foi feita, pois quanto maior o número de estudantes, maior a dificuldade do algoritmo alcançar os resultados esperados. Observa-se na Figura 41 que para 1.000 estudantes o *fitness* alcança o valor igual a 3 (máximo possível para este caso) com facilidade (aproximadamente 500 gerações). No entanto, à medida que o número de estudantes aumenta, mais iterações são necessárias para que o *fitness* alcance valor igual a 3. No caso de 10.000 estudantes, o algoritmo estabiliza com *fitness* 2,9999 aproximadamente em 7.500 gerações. Considerando-se que este é o melhor caso dos dados (máximo de heterogeneidade e máximo de interações), o que facilita os resultados do algoritmo, um número maior de iterações foi pensado para atender também as situações de maior dificuldade. Assim, para garantir uma quantidade de gerações que permita ao algoritmo alcançar o resultado mais próximo possível do esperado, o número de gerações (iteraões) foi definido em 8.000 gerações.

Figura 41 – Convergência *fitness* (melhor caso dos dados)



Fonte: Próprio autor

O teste-piloto foi um processo iterativo, repetido até que se obtivessem resultados considerados satisfatórios para a formação do grupo. Consideraram-se nos resultados os dados definidos conforme a seção 5.1.1, a quantidade de estudantes por grupo e os resultados alcançados pelo AG em termos de *fitness*, percentual de grupos heterogêneos e tempo de processamento. Assim, a partir desses resultados preliminares, foram executados os testes com as quantidades massivas de estudantes, descritos a seguir.

5.3 TESTES MASSIVOS

A fim de se chegar ao objetivo geral desta pesquisa, que é avaliar a adoção de algoritmos genéticos para a formação de grupos na aprendizagem colaborativa, considerando um contexto massivo de dados (milhares de estudantes), testes com as quantidades massivas foram realizados, conforme definido na Figura 39, seção 5.1.1, tanto para o AG como para o algoritmo randômico. Esses testes foram executados considerando-se 5 estudantes por grupo e foram realizados entre maio e junho de 2015.

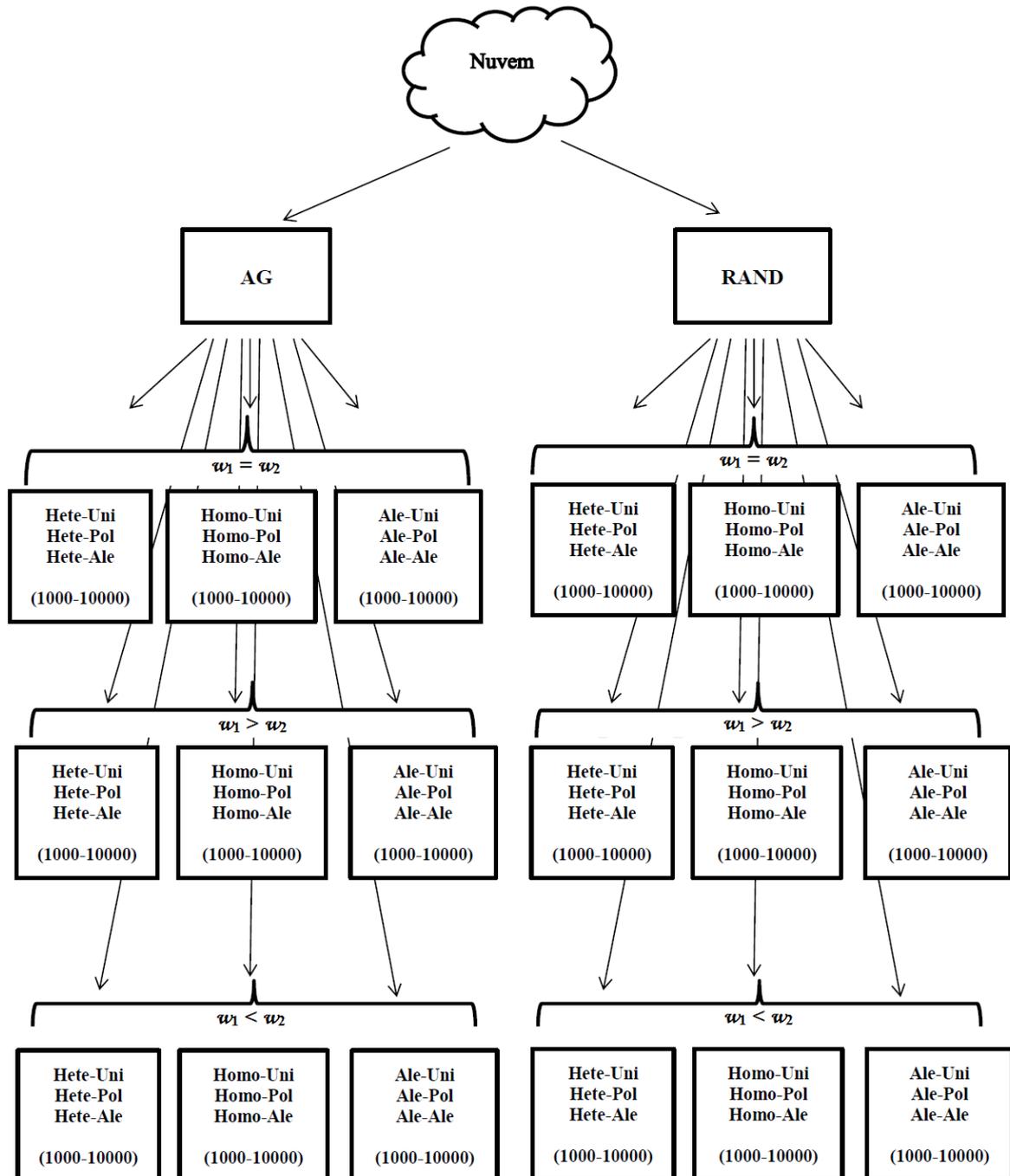
Devido à quantidade de casos de testes (270 casos) resultante da combinação da configuração de dados definida, os testes massivos foram direcionados a uma estrutura de armazenamento mais robusta. Para esse caso, então, o armazenamento e a capacidade de processamento providos pela nuvem da UDESC foram utilizados. Esse ambiente é composto, entre outros, por um servidor de 2.10 Ghz em que distintas Máquinas Virtuais (*Virtual Machines*) executando exclusivamente os algoritmos foram providas. Portanto, foram criadas 18 VMs (*Virtual Machines*), com 1 vCPU, 1 Gb da memória RAM, utilizando sistema operacional Ubuntu 14.04 LTS. Como os algoritmos não são aplicações paralelas e a quantidade de memória necessária não excede a 1 Gb, essa configuração equivale à configuração utilizada nos testes-piloto. Assim, 9 VMs foram utilizadas para os testes do AG e 9 VMs foram utilizadas para os testes do algoritmo randômico.

A Figura 42 mostra a distribuição dos testes nas VMs. Cada máquina executou a combinação de 1 (um) caso do conhecimento para os 3 (três) casos da interação, para um peso determinado, executando o mesmo teste de 1.000 até 10.000 estudantes. Desta forma, cada máquina executou 30 casos de testes (1 conhec * 3 interações * 10 quantidade de estudantes), e todos os testes foram disparados simultaneamente. Como são 9 (nove) VMs, executando 30 casos de testes cada uma, os 270 casos de testes foram executados, tanto para o AG como para o algoritmo randômico.

Cada um dos 270 casos de testes foi executado 30 vezes para possibilitar a coleta de dados. A partir dos dados, a média do *fitness*, a média do percentual de grupos heterogêneos e

a média do tempo de processamento foram computadas, para as devidas comparações entre o AG e o método randômico.

Figura 42 – Distribuição dos testes nas VMs



Fonte: Próprio autor

Análises estatísticas foram definidas para validar os resultados obtidos. Estas análises referem-se à comparação dos resultados do AG *versus* os resultados do algoritmo randômico,

através de testes de hipóteses e determinaram a validade da hipótese nula, considerando-se o nível de significância.

Para os testes estatísticos utilizou-se o programa estatístico R^4 , versão 3.2.0, através da ferramenta *RStudio*. Os métodos estatísticos utilizados são discutidos a seguir na seção 5.4 e os resultados são apresentados na seção 5.5.

5.4 MÉTODOS ESTATÍSTICOS UTILIZADOS

Métodos estatísticos paramétricos pressupõem uma distribuição normal (ou aproximadamente normal) dos dados da amostra. No entanto, para experimentos computacionais com algoritmos evolutivos, caso desta pesquisa, os dados tendem a não ter uma distribuição normal. Por isso, métodos estatísticos não paramétricos são sugeridos (LAHOZ-BELTRA; PERALES-GRAVAN, 2010; DERRAC et al., 2011).

Testes de hipóteses são testes onde uma hipótese nula (H_0) é contrastada com uma hipótese alternativa (H_1), para determinar se a hipótese nula é verdadeira, com um nível de significância α . Esse α refere-se ao erro Tipo 1 (também chamado “falso positivo”), que é a probabilidade de se rejeitar a hipótese nula quando ela é verdadeira (e nesse caso, não deveria ser rejeitada). Existe ainda o erro Tipo 2 (β , também chamado “falso negativo”), que é a probabilidade de não se rejeitar a hipótese nula quando ela é falsa (e nesse caso, deveria ser rejeitada) (MONTGOMERY; RUNGER, 2012). Os testes de significância retornam um *p*-valor (*p-value*). O *p*-valor é a probabilidade de se obter uma estatística-teste igual ou mais extrema quanto a que foi observado na amostra, assumindo que H_0 é verdadeira. Essa é uma informação útil, pois o *p*-valor fornece informações se o teste de hipótese é significativo ou não, e também indica o quão significativo é o resultado. Quanto menor for o *p*-valor, mais forte a evidência contra H_0 (DERRAC et al., 2011). Logo, se o *p*-valor for menor que o α estabelecido, significa que a hipótese nula (H_0) deve ser rejeitada (não é verdadeira).

5.4.1 Verificação da normalidade dos dados

O primeiro teste estatístico utilizado foi o teste de *Shapiro-Wilk*, que determina se a amostra é proveniente de uma distribuição normal ou não. O teste de *Shapiro-Wilk* é um teste de hipóteses, e por isso necessita do α e das hipóteses H_0 e H_1 .

Para esta pesquisa, adotou-se $\alpha = 0.05$ e o teste de *Shapiro-Wilk* avalia se a amostra informada provém de uma distribuição normal ou não, considerando as hipóteses:

- H_0 : Dados provêm de uma distribuição normal;

⁴ Disponível em <http://www.r-project.org/>

- H_1 : Dados não provêm de uma distribuição normal.

O teste de *Shapiro-Wilk* foi aplicado para algumas amostras coletadas, a fim de avaliar a distribuição dos dados. Houve casos em que algumas amostras foram consideradas normais, isto é, o *p-valor* não foi menor que o nível de significância ($\alpha = 0.05$) estabelecido, de forma que a hipótese nula (H_0) não pôde ser rejeitada. Entretanto, como em alguns casos a hipótese nula (H_0) foi rejeitada, optou-se por aplicar testes não paramétricos para todos os testes.

Segundo Weber e Sawilowsky (2009) e Devore (2012) os testes não paramétricos têm desempenho quase tão bom quanto seus concorrentes paramétricos quando a distribuição é normal; e têm melhor desempenho quando a distribuição não é normal. Assim, apesar de algumas amostras serem normais, aplicaram-se os testes não paramétricos sugeridos, conforme Derrac et al. (2011).

5.4.2 Testes de Hipóteses

A análise feita em relação aos resultados obtidos pelo AG *versus* os resultados obtidos pelo algoritmo randômico, a fim de se validar a hipótese nula estabelecida, utilizou o método *Wilcoxon rank sum test* sugerido por Derrac et al. (2011) para a comparação entre dois algoritmos. O *Wilcoxon rank sum test* é um teste de hipóteses, e por isso necessita do α e das hipóteses H_0 e H_1 .

Para esta pesquisa, adotou-se $\alpha = 0.05$, isto é, considera-se a possibilidade de 5% de erro. No caso desta pesquisa, em que as amostras são independentes, o *Wilcoxon rank sum test* considera as seguintes hipóteses:

- H_0 : As amostras (resultados) são equivalentes (as distribuições não diferem entre si);
- H_1 : As amostras (resultados) não são equivalentes (as distribuições diferem entre si).

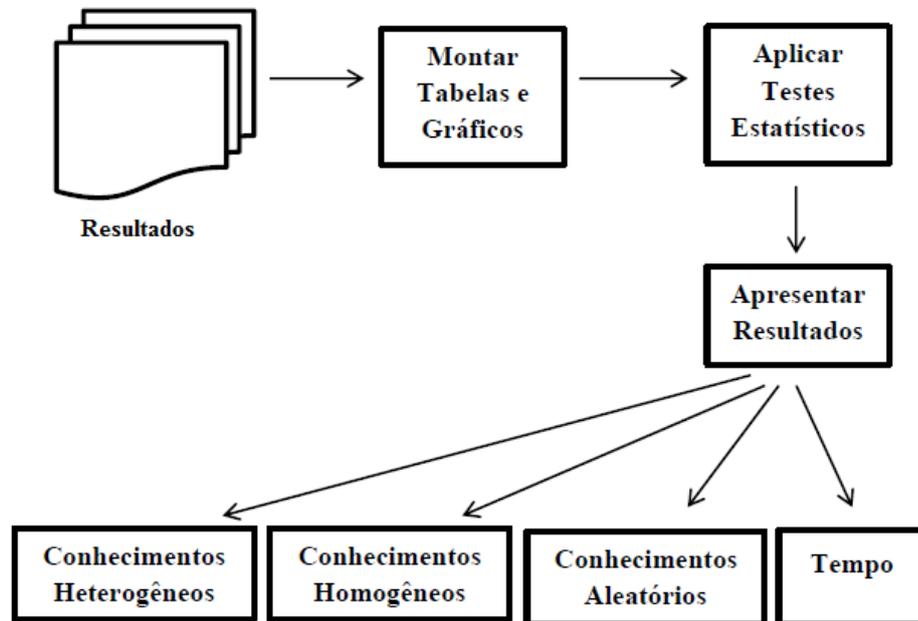
A utilização do teste está descrita a seguir, nos resultados obtidos, onde são apresentados os p-valores para os casos de testes onde se julgou necessária a aplicação do *Wilcoxon test* para validar a significância dos resultados obtidos.

5.5 RESULTADOS OBTIDOS E DISCUSSÃO

Uma vez processados todos os testes, os dados foram coletados e seguiu-se o processo de montagem das tabelas e gráficos, aplicação dos testes estatísticos e apresentação dos resultados, conforme detalhado na Figura 43. Os resultados obtidos são apresentados e discutidos nas seções a seguir, onde são apresentados os gráficos dos resultados e as tabelas

dos testes estatísticos. As tabelas principais e os gráficos dos casos que não são discutidos nas seções a seguir estão no Apêndice B (numerados internamente).

Figura 43 – Processo de análise estatística e apresentação dos resultados



Fonte: Próprio autor

Os resultados obtidos consideram a média das 30 execuções de cada teste. Os gráficos mostram de mil até dez mil estudantes no eixo x , com uma marca “*Best*” após cada quantidade, que corresponde ao melhor resultado alcançado pelo algoritmo em cada caso. O eixo y apresenta o valor do *fitness* e o percentual de grupos heterogêneos (em relação ao total de grupos), isto é, os grupos que alcançaram a heterogeneidade máxima possível. Os resultados são apresentados no mesmo gráfico, pois o percentual de grupos heterogêneos (percentual) é sempre no máximo igual a 1 (100%) e o valor do *fitness* é sempre maior do que 1.

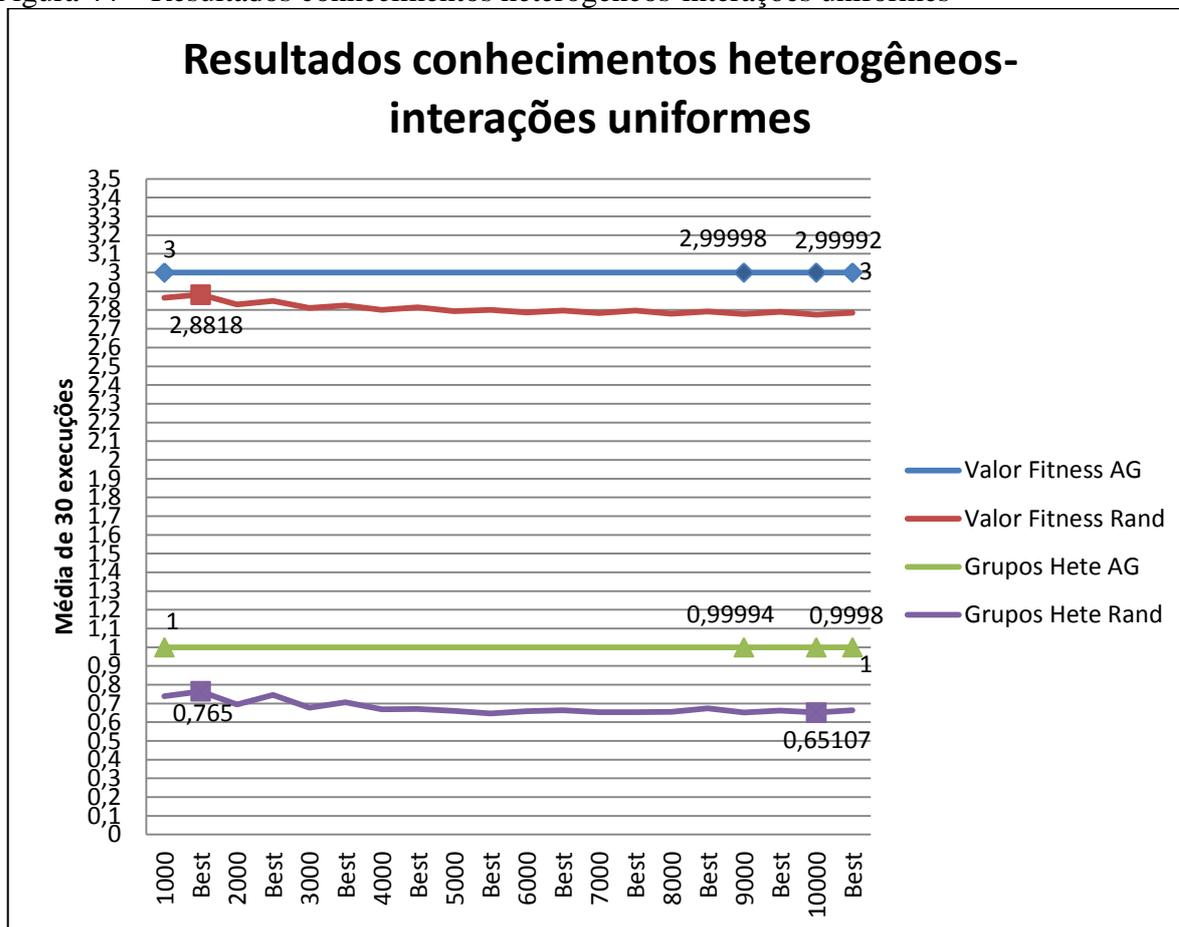
Na seção 5.5.1 são apresentados os resultados para conhecimentos heterogêneos, na seção 5.5.2 são apresentados os resultados para conhecimentos homogêneos, na seção 5.5.3 são apresentados os resultados para conhecimentos aleatórios e na seção 5.5.4 são apresentados os resultados quanto ao tempo de processamento dos algoritmos. Como para cada caso dos conhecimentos há 3 (três) possibilidades de interações e 3 (três) possibilidades de pesos, as seções a seguir apresentam destacados em negrito os resultados para: interações **uniformes**, interações **polarizadas** e interações **aleatórias**, ambas considerando todos os casos dos pesos.

5.5.1 Conhecimentos heterogêneos

Os testes com conhecimentos heterogêneos visaram permitir a formação de grupos ótimos em relação à heterogeneidade, de forma que os mesmos eram formados contendo estudantes com conhecimentos 0 (zero) e 1 (um), alcançando a máxima distância possível para a heterogeneidade. Como todos os grupos tinham o mesmo tamanho e todos alcançaram a máxima distância possível, a inter-homogeneidade dos grupos foi igual a 0 (zero), isto é, não houve diferença entre eles. Assim, o *fitness* calculado alcançou valor máximo, igual a 1 (um), para a heterogeneidade (fórmula 2) e também para a inter-homogeneidade (fórmula 4), somando 2. Desta forma, o *fitness* para conhecimentos heterogêneos é sempre maior que 2, variando conforme as interações e os pesos aplicados.

No caso das interações **uniformes**, onde todos os estudantes interagiram entre si o máximo possível, as interações (fórmula 3) alcançaram também valor máximo igual a 1 (um). Esse é o melhor caso dos dados, onde o *fitness* alcança valor igual a 3 (soma das três medidas), para pesos igual a 1 (um). Os resultados obtidos para esse caso de testes são apresentados na Figura 44, que é discutida a seguir (Tabela 1 – Apêndice B).

Figura 44 – Resultados conhecimentos heterogêneos-interações uniformes



Fonte: Próprio autor

Na Figura 44, observa-se que o *fitness* alcançado pelo AG foi igual a 3 para 8 (oito) dos 10 (dez) casos de testes (de mil até oito mil estudantes). Os casos de 9 mil e 10 mil estudantes estão com os valores marcados, respectivamente 2,99998 e 2,99992, para a média das 30 execuções. Em seus melhores resultados (*Best*) o AG chegou a 3 em todos os casos. O *fitness* alcançado pelo algoritmo randômico foi menor que o alcançado pelo AG em todos os casos (2,8818 no melhor resultado para mil estudantes). O percentual de grupos heterogêneos mostra que o AG alocou os grupos da melhor forma possível, sendo que, acompanhando o *fitness*, em 8 dos 10 casos 100% dos grupos alcançaram o máximo de heterogeneidade. Para os casos de 9 mil e 10 mil estudantes, alcançaram respectivamente 0,99994 e 0,99980, para a média das 30 execuções. Em seus melhores resultados (*Best*) o AG chegou a 1 (100%) em todos os casos. O algoritmo randômico alcançou um máximo de 76,5% (0,765) de grupos com o máximo de heterogeneidade, para o melhor resultado com mil estudantes (*Best*). As linhas no gráfico mostram que a diferença entre os resultados alcançados pelo AG e os resultados alcançados pelo algoritmo randômico, aumenta à medida que o número de estudantes cresce, tanto para o *fitness* quanto para o percentual de grupos heterogêneos. Aproximadamente uma diferença de 35% de grupos a mais (AG - 0,9998 x 0,65107 - Rand), que foram gerados com o máximo de heterogeneidade, no caso de dez mil estudantes; isso significa aproximadamente 700 grupos heterogêneos a mais gerados pelo AG do que o resultado alcançado pelo algoritmo randômico. Desta forma, observa-se que os resultados do AG foram melhores que os resultados do algoritmo randômico, em todos os casos.

Para verificar a diferença entre os resultados, as Figuras 45(a) e 45(b) e 46(a) e 46(b) mostram o gráfico *box-plot* dos resultados para $n = 1000$ e $n = 10000$ estudantes, respectivamente, a menor e a maior diferença.

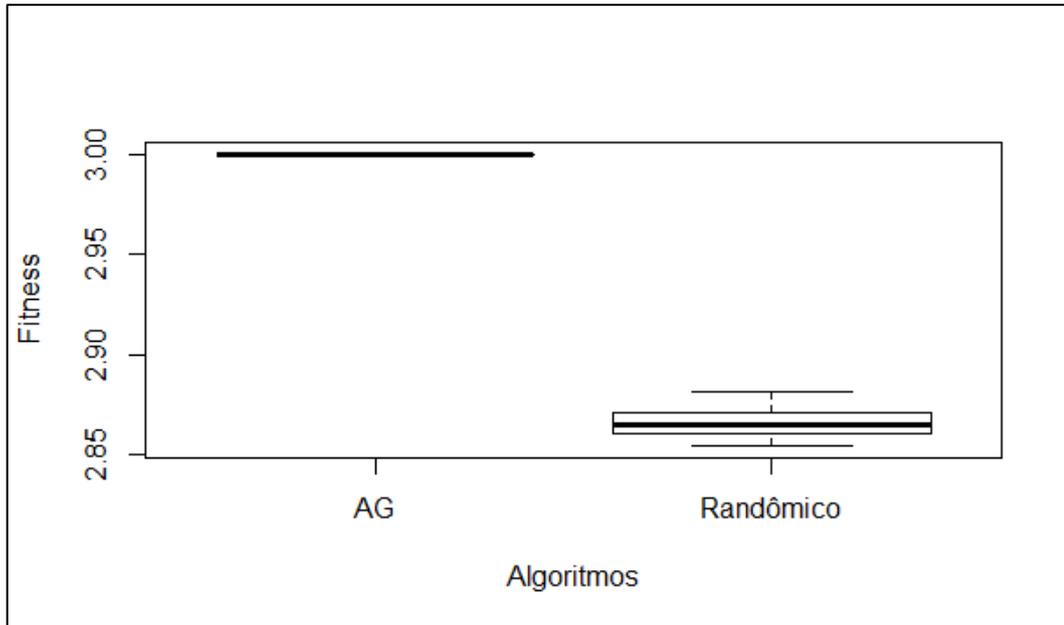
Nas Figuras 45(a) e 45(b), observa-se que, mesmo quando a diferença do *fitness* e do percentual de grupos heterogêneos é menor (caso de mil estudantes), o resultado alcançado pelo AG foi melhor que o resultado alcançado pelo algoritmo randômico, pois não há intersecção nos resultados obtidos.

Da mesma forma, ocorre para dez mil estudantes, Figuras 46(a) e 46(b), e assim conclui-se que os resultados alcançados pelo AG foram melhores que os resultados alcançados pelo algoritmo randômico.

Nesses casos, pelo gráfico *box-plot*, é possível verificar que não há intersecção dos resultados, significando que os resultados são diferentes entre si. Assim, não foi aplicado o *Wilcoxon test*.

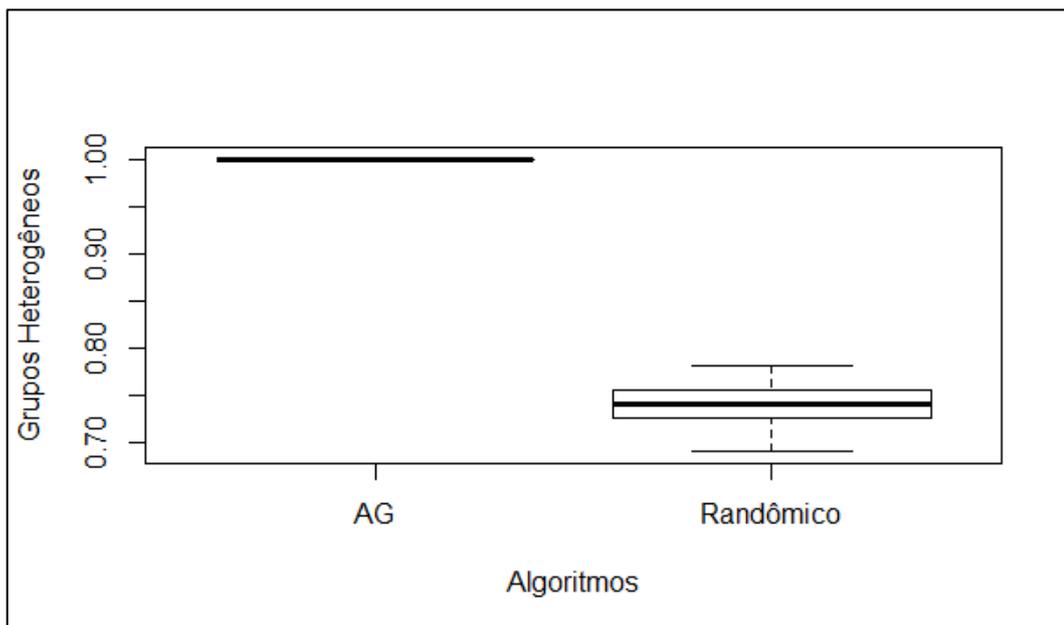
Figuras 45(a) e 45(b) com resultados do *fitness* e do percentual de grupos heterogêneos, respectivamente, para 1000 estudantes, caso em que ocorreram as menores diferenças.

Figura 45(a) – *Box-plot Fitness* para conhecimentos heterogêneos-interações uniformes, $n = 1000$



Fonte: Próprio autor

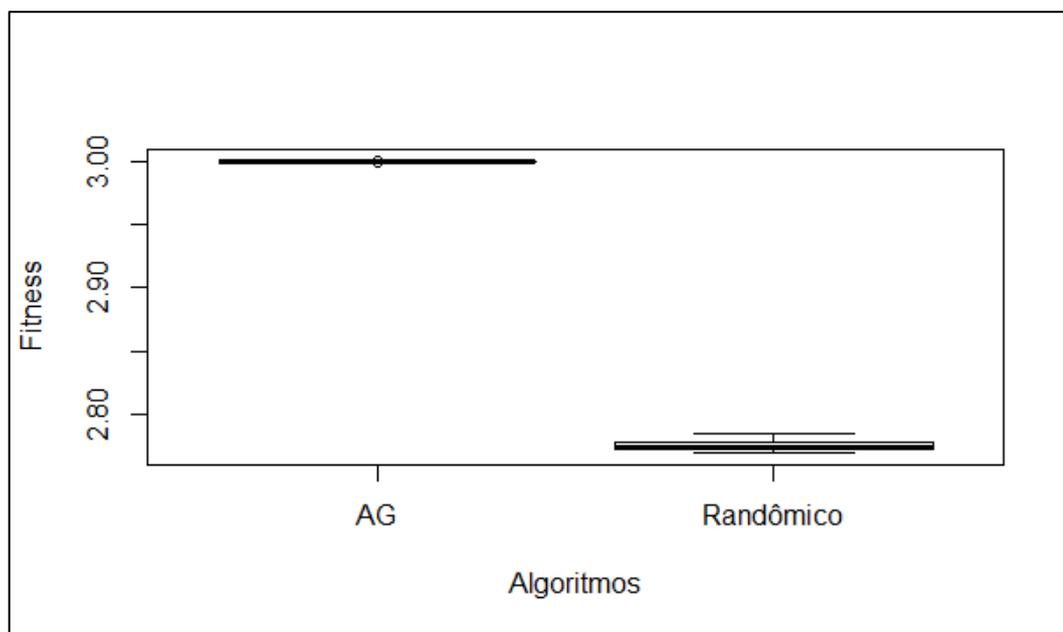
Figura 45(b) – *Box-plot Grupos Heterogêneos* para conhecimentos heterogêneos-interações uniformes, $n = 1000$



Fonte: Próprio autor

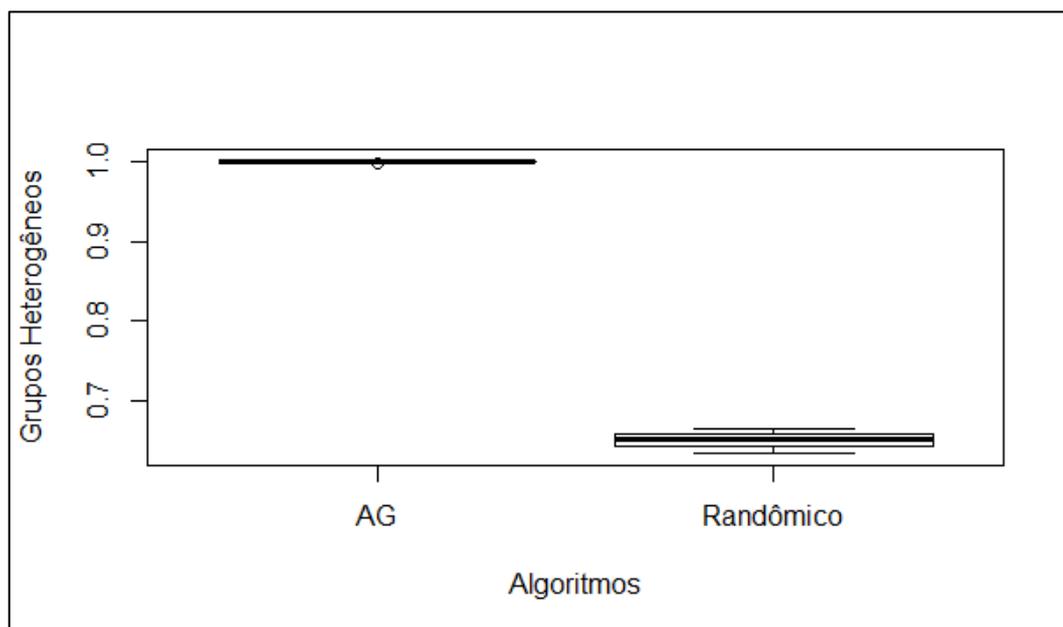
Figuras 46(a) e 46(b) com resultados do *fitness* e do percentual de grupos heterogêneos, respectivamente, para 10000 estudantes, caso em que ocorreram as maiores diferenças.

Figura 46(a) – *Box-plot Fitness* para conhecimentos heterogêneos-interações uniformes, $n = 10000$



Fonte: Próprio autor

Figura 46(b) – *Box-plot Grupos Heterogêneos* para conhecimentos heterogêneos-interações uniformes, $n = 10000$

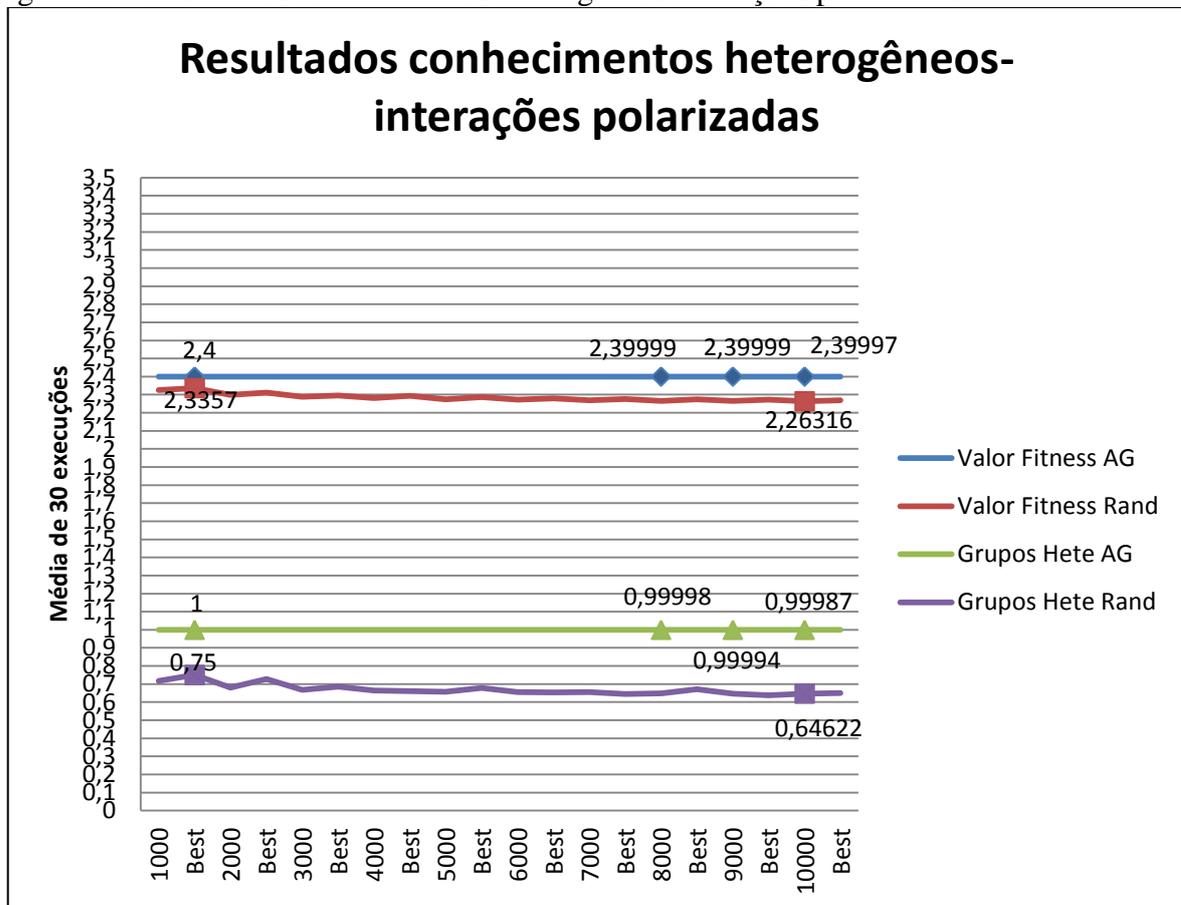


Fonte: Próprio autor

Nos casos com peso maior para os conhecimentos (ω_1 maior que ω_2) e com peso maior para as interações (ω_2 maior que ω_1), os resultados foram semelhantes aos descritos anteriormente. No entanto, como o peso considerado foi $\omega_1 = 2$ (conhecimentos) ou $\omega_2 = 2$ (interações), o *fitness* foi igual a 4,0. Observou-se que os valores do *fitness* para o AG variaram apenas em função do peso aplicado: de 3,0 para 4,0. Isso ocorreu porque, os valores desses casos de testes são valores ótimos em relação aos conhecimentos e as interações, de forma que tanto os conhecimentos quanto as interações alcançam valor máximo igual a 1 em ambos os casos. Assim, estes casos não são discutidos (Tabelas 2 e 3 - Apêndice B).

Para as interações **polarizadas**, como nesse caso nem todos os estudantes interagiram entre si, não foi possível alcançar o máximo das interações, de forma que o *fitness* máximo alcançado foi 2,4. No entanto, o percentual de grupos heterogêneos alcançado foi, pelo menos aproximadamente, 100% em todos os casos, o que indica que o AG chegou aos melhores resultados. Os resultados obtidos para esse caso de testes são apresentados na Figura 47, que é discutida a seguir (Tabela 4 – Apêndice B).

Figura 47 – Resultados conhecimentos heterogêneos-interações polarizadas



Fonte: Próprio autor

Na Figura 47, observa-se que o *fitness* alcançado pelo AG foi igual a 2,4 para 7 (sete) dos 10 (dez) casos (de mil até sete mil estudantes), e que mesmo para os casos em que a média do *fitness* não foi igual a 2,4 (oito mil até dez mil estudantes), o AG chegou a 2,4 em seus melhores resultados (*Best*). O *fitness* alcançado pelo algoritmo randômico foi menor que o alcançado pelo AG em todos os casos (2,3357 no melhor resultado para mil estudantes). O percentual de grupos heterogêneos mostra que o AG alocou os grupos da melhor forma possível, sendo que, acompanhando o *fitness*, em 7 (sete) dos 10 (dez) casos 100% dos grupos alcançaram o máximo de heterogeneidade. O algoritmo randômico alcançou um máximo de 75% (0,75) de grupos com o máximo de heterogeneidade, para o melhor resultado com mil estudantes (*Best*). As linhas no gráfico mostram que a diferença entre os resultados alcançados pelo AG e os resultados alcançados pelo algoritmo randômico, aumenta à medida que o número de estudantes cresce, conforme já verificado nos casos anteriores. As diferenças no percentual de grupos heterogêneos são semelhantes às dos casos anteriores; no entanto, as diferenças entre os valores de *fitness* alcançados pelo AG e pelo algoritmo randômico são menores que as diferenças alcançadas pelos casos anteriores, devido ao máximo valor de *fitness* possível, nesse caso, ser 2,4.

Para avaliar a significância dessas diferenças, o *Wilcoxon test* foi aplicado tanto para o *fitness* quanto para o percentual de grupos heterogêneos. Na Tabela 3, são apresentados os resultados do teste para $n = 1000$ e $n = 10000$ estudantes, respectivamente, a menor e a maior diferença.

Tabela 3 – *Wilcoxon test* conhecimentos heterogêneos-interações polarizadas

	#Estudantes	<i>Wilcoxon test</i> <i>p-valor</i>
Fitness	1000	1.199e-12
	10000	7.665e-12
Grupos Heterogêneos	1000	1.157e-12
	10000	7.665e-12

Fonte: Próprio autor

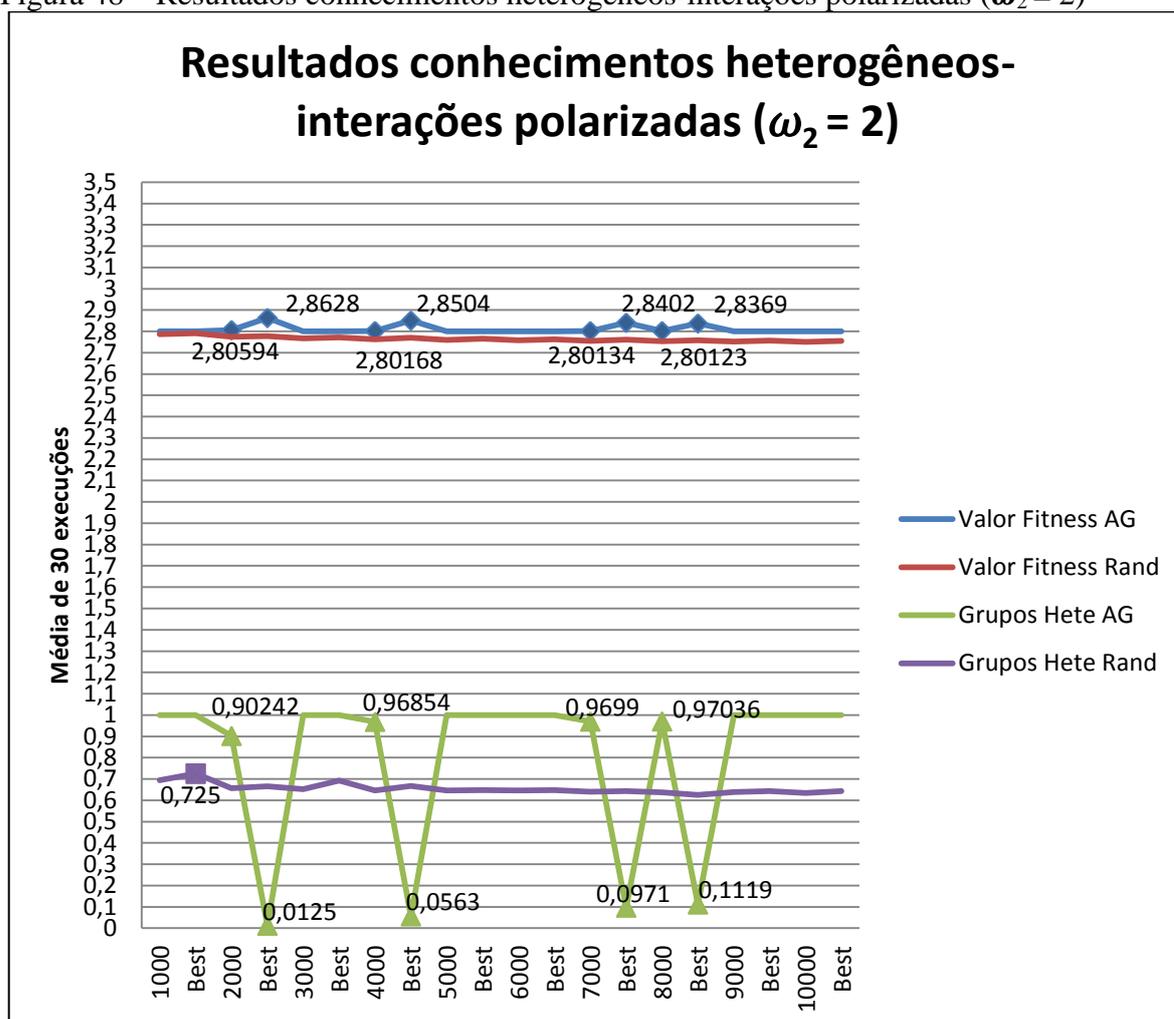
Em todos os casos, a hipótese nula foi rejeitada, com *p-valor* muito menor que o nível de significância estabelecido ($p\text{-valor} \ll \alpha = 0.05$). Assim, conclui-se que, apesar de as diferenças nos valores do *fitness* serem menores que as diferenças ocorridas nos casos anteriores, elas são significativas. E as diferenças ocorridas no percentual de grupos heterogêneos também o são. Desta forma, os resultados obtidos pelo AG foram melhores que os resultados obtidos pelo algoritmo randômico.

Nos casos com peso maior para os conhecimentos (ω_1 maior que ω_2), os resultados foram semelhantes aos descritos anteriormente (pesos iguais), sendo que os valores do *fitness*

para o AG variaram apenas em função do peso aplicado: de 2,4 para 3,4. Assim, esse caso não é discutido (Tabela 5 – Apêndice B).

Nos casos com peso maior para as interações (ω_2 maior que ω_1), o *fitness* foi igual a 2,8. Nesse caso, a Figura 48 (Tabela 6 – Apêndice B) mostra o gráfico dos resultados. Como as interações têm peso maior, houve casos em que o melhor *fitness* (*Best*) foi maior que a média, mas tem percentual de grupos heterogêneos menor (bem baixo). Isto ocorreu porque ao priorizar as interações através do peso, em alguns casos juntou no mesmo grupo estudantes que interagiram, mesmo que os conhecimentos fossem iguais. Entretanto, aconteceu em apenas quatro dos dez casos testados (dois mil, quatro mil, sete mil e oito mil estudantes), e em no máximo três, das trinta execuções para cada caso. Considerando a média das execuções, todos os casos chegaram, pelo menos, a 90% (0,90242) de heterogeneidade (caso de dois mil estudantes). Portanto, o algoritmo priorizou a formação correta, considerando o balanceamento dos grupos e a heterogeneidade, juntamente com as interações.

Figura 48 – Resultados conhecimentos heterogêneos-interações polarizadas ($\omega_2 = 2$)



Fonte: Próprio autor

O algoritmo randômico alcançou um máximo de $\approx 72,5\%$ (0,725) de grupos com o máximo de heterogeneidade, para o melhor resultado com mil estudantes. Observa-se que os valores do *fitness* foram muito semelhantes entre o AG e o algoritmo randômico, gerando diferenças bem pequenas entre os valores.

Assim, foi aplicado o *Wilcoxon test* para todas as quantidades de estudantes, tanto para o *fitness* quanto para o percentual de grupos heterogêneos. Na Tabela 4, são apresentados os resultados dos testes para a menor diferença, respectivamente, para $n = 1000$ (*fitness*) e $n = 2000$ (grupos).

Tabela 4 – *Wilcoxon test* conhecimentos heterogêneos-interações polarizadas, $\omega_2 = 2$, menores diferenças ocorridas

	#Estudantes	<i>Wilcoxon test</i> <i>p-valor</i>
Fitness	1000	1.179e-12
Grupos Heterogêneos	2000	2.463e-08

Fonte: Próprio autor

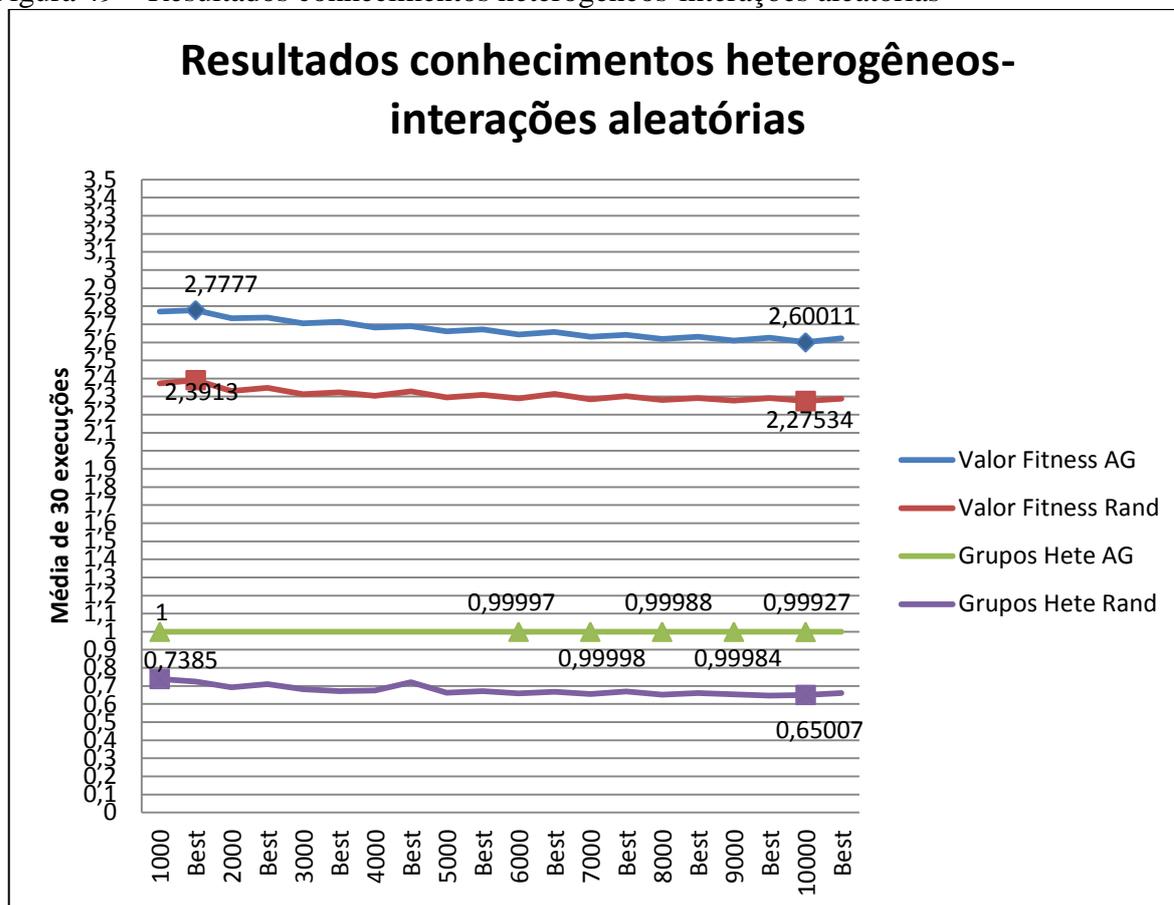
Em todos os casos, a hipótese nula foi rejeitada com *p-valor* muito menor que o nível de significância estabelecido ($p\text{-valor} \ll \alpha = 0.05$). Mesmo para a menor diferença no valor do *fitness* e do percentual de grupos heterogêneos, tem-se que a diferença é significativa, e conclui-se que os resultados do AG foram melhores que os resultados do algoritmo randômico.

Para as interações **aleatórias**, o *fitness* máximo alcançado foi $\approx 2,77$. Nesse caso, como as interações são aleatórias, não se sabe o valor “correto” do *fitness*, mas se sabe que quanto maior o *fitness*, melhor é o resultado, já que a função objetivo é uma função de maximização. O percentual de grupos heterogêneos alcançado foi, pelo menos aproximadamente, 100% em todos os casos, o que indica que o AG chegou aos melhores resultados, conforme os casos anteriores.

Na Figura 49 são apresentados os resultados obtidos para esse caso de testes (Tabela 7 – Apêndice B). Observa-se que o *fitness* alcançado pelo AG variou entre $\approx 2,60$ e $\approx 2,77$. O *fitness* alcançado pelo algoritmo randômico foi menor que o alcançado pelo AG em todos os casos. As linhas do gráfico mostram que as diferenças entre os valores de *fitness* alcançados pelo AG e pelo algoritmo randômico são maiores que as diferenças alcançadas pelos casos anteriores, e tendem a diminuir à medida que o número de estudantes cresce. Isso ocorreu devido à dificuldade desse caso de testes, com as interações aleatórias simulando uma situação real, que dificulta os resultados do algoritmo. No entanto, o percentual de grupos heterogêneos mostra que o AG alocou os grupos da melhor forma possível, sendo que em 5 (cinco) dos 10 (dez) casos, 100% dos grupos alcançaram o máximo de heterogeneidade. É um

percentual menor que o dos casos anteriores, entretanto, considerando seus melhores resultados (*Best*), o AG chegou ao máximo da heterogeneidade em todos os casos.

Figura 49 – Resultados conhecimentos heterogêneos-interações aleatórias



Fonte: Próprio autor

O algoritmo randômico alcançou um máximo de $\approx 74\%$ (0,7385) de grupos com o máximo de heterogeneidade, para o resultado com mil estudantes. As linhas do gráfico mostram que a diferença entre o resultado alcançado pelo AG e o resultado alcançado pelo algoritmo randômico, em relação ao percentual de grupos heterogêneos, aumenta à medida que o número de estudantes cresce, conforme já verificado nos casos anteriores.

O *Wilcoxon test* foi aplicado para avaliar a significância das diferenças, tanto para o *fitness* quanto para o percentual de grupos heterogêneos. Na Tabela 5, são apresentados os resultados do teste para, respectivamente, a menor e a maior diferença em cada caso.

Tabela 5 – *Wilcoxon test* conhecimentos heterogêneos-interações aleatórias

	#Estudantes	<i>Wilcoxon test</i> <i>p-valor</i>
Fitness	10000	3.018e-11
	2000	2.995e-11
Grupos Heterogêneos	1000	1.181e-12
	10000	2.418e-11

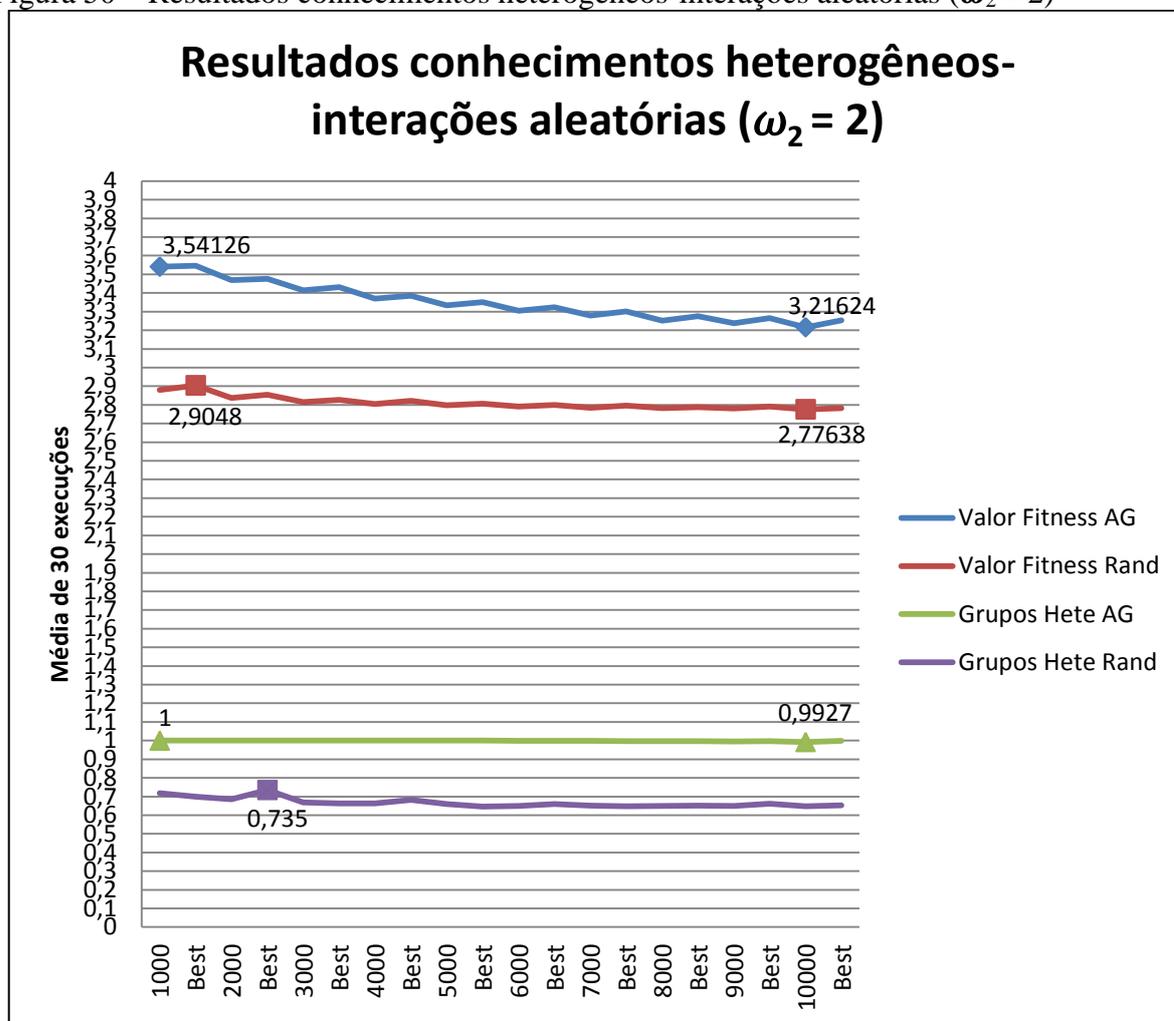
Fonte: Próprio autor

Em todos os casos, a hipótese nula foi rejeitada, com *p-valor* muito menor que o nível de significância estabelecido (*p-valor* $\ll \alpha = 0.05$). Desta forma, conclui-se que os resultados obtidos pelo AG foram melhores que os resultados obtidos pelo algoritmo randômico.

Nos casos com peso maior para os conhecimentos (ω_1 maior que ω_2), os resultados foram semelhantes aos descritos anteriormente (pesos iguais), sendo que os valores do *fitness* para o AG variaram apenas em função do peso aplicado: de $\approx 3,60$ a $\approx 3,77$. Assim, esse caso não é discutido (Tabela 8 – Apêndice B).

Nos casos com peso maior para as interações (ω_2 maior que ω_1), o *fitness* variou de $\approx 3,22$ a $\approx 3,54$. Nesse caso, a Figura 50 (Tabela 9 – Apêndice B) mostra o gráfico dos resultados.

Figura 50 – Resultados conhecimentos heterogêneos-interações aleatórias ($\omega_2 = 2$)



Fonte: Próprio autor

Na Figura 50, observa-se que o valor do *fitness* não é o máximo possível (4,0), mas o valor do *fitness* para o AG foi maior que o valor do *fitness* para o algoritmo randômico em todos os casos. Nesse caso, os valores do *fitness* obtiveram a maior diferença entre o AG e o

algoritmo randômico e, observa-se pela linha azul, que ocorreu maior variação no intervalo de valores (de $\approx 3,22$ a $\approx 3,54$). Percebe-se que o AG buscou os melhores resultados através do percentual de grupos heterogêneos, que chegou a um mínimo de $\approx 99\%$ (0,9927) de grupos com o máximo de heterogeneidade (dez mil estudantes). O algoritmo randômico alcançou um máximo de $\approx 73,5\%$ (0,735) de grupos com o máximo de heterogeneidade, para o melhor resultado com dois mil estudantes. As diferenças em relação aos grupos heterogêneos são semelhantes às ocorridas anteriormente.

Assim, foi aplicado o *Wilcoxon test* para todas as quantidades de estudantes, tanto para o *fitness* quanto para o percentual de grupos heterogêneos. Na Tabela 6, são apresentados os resultados dos testes para a menor diferença, respectivamente, para $n = 10000$ (*fitness*) e $n = 1000$ (grupos).

Tabela 6 – *Wilcoxon test* conhecimentos heterogêneos-interações aleatórias, $\omega_2 = 2$, menores diferenças ocorridas

	#Estudantes	<i>Wilcoxon test</i> <i>p-valor</i>
Fitness	10000	3.014e-11
Grupos Heterogêneos	1000	1.185e-12

Fonte: Próprio autor

Em todos os casos, a hipótese nula foi rejeitada com *p-valor* muito menor que o nível de significância estabelecido ($p\text{-valor} \ll \alpha = 0.05$). Mesmo para a menor diferença no valor do *fitness* e no percentual de grupos heterogêneos, tem-se que as diferenças são significativas, e conclui-se que os resultados do AG foram melhores que os resultados do algoritmo randômico.

Em relação à aplicação do peso para os conhecimentos, como nesses casos os conhecimentos são dados controlados e são valores ótimos em relação à heterogeneidade, conclui-se que a aplicação do peso não influenciou o comportamento do algoritmo de maneira geral, influenciando apenas nos valores alcançados em relação ao *fitness*. Esses casos não foram discutidos. No entanto, o *Wilcoxon test* foi aplicado para os casos com as menores diferenças, tanto para o *fitness* quanto para o percentual de grupos heterogêneos. Na Tabela 7, são apresentados os resultados dos testes para a menor diferença, respectivamente, para interações uniforme, interações polarizadas e interações aleatórias.

Em todos os casos, a hipótese nula foi rejeitada com *p-valor* muito menor que o nível de significância estabelecido ($p\text{-valor} \ll \alpha = 0.05$). Assim, conclui-se que os resultados alcançados pelo AG foram melhores que os resultados alcançados pelo algoritmo randômico, também para esses casos de testes, com peso maior para os conhecimentos.

Tabela 7 – *Wilcoxon test* conhecimentos heterogêneos, $\omega_1 = 2$, menores diferenças ocorridas

	Interações	#Estudantes	Wilcoxon test (p-valor)
Fitness	Uniformes	1000	1.199e-12
	Polarizadas	1000	1.195e-12
	Aleatórias	10000	3.018e-11
Grupos Heterogêneos	Uniformes	1000	1.181e-12
	Polarizadas	1000	1.161e-12
	Aleatórias	1000	1.166e-12

Fonte: Próprio autor

Em relação à aplicação do peso para as interações, conclui-se que a aplicação do peso influenciou o comportamento do algoritmo, inclusive apresentando casos com percentuais baixos de grupos heterogêneos, no caso de interações polarizadas. Entretanto, de maneira geral, o algoritmo obteve resultados médios satisfatórios ($\approx 90\%$ a 100%), considerando, além das interações, a heterogeneidade dos conhecimentos e o balanceamento dos grupos. O que, neste caso, está correto.

Conforme tudo o que foi apresentado para os testes dos conhecimentos heterogêneos, o AG alcançou o máximo de heterogeneidade para aproximadamente 100% dos casos verificados (considerando-se os melhores resultados – *Best*), com exceção dos casos de interações polarizadas e aleatórias, com peso maior para as interações, onde houve quatro casos em que os resultados ficaram entre 90% e 97% . Além disso, o AG obteve resultados melhores quando a quantidade de estudantes cresceu, variando entre $\approx 26\%$ (mil estudantes) a $\approx 35\%$ (dez mil estudantes), de grupos a mais formados com o máximo de heterogeneidade.

Assim, conclui-se que, de forma geral, o AG obteve resultados significativamente melhores que os resultados obtidos pelo algoritmo randômico, uma vez que as médias apresentadas pelo AG foram maiores que as médias do algoritmo randômico, tanto para o *fitness* quanto para o percentual de grupos heterogêneos; e os testes aplicados confirmam que as diferenças ocorridas são significativas, mesmo quando estas diferenças são pequenas.

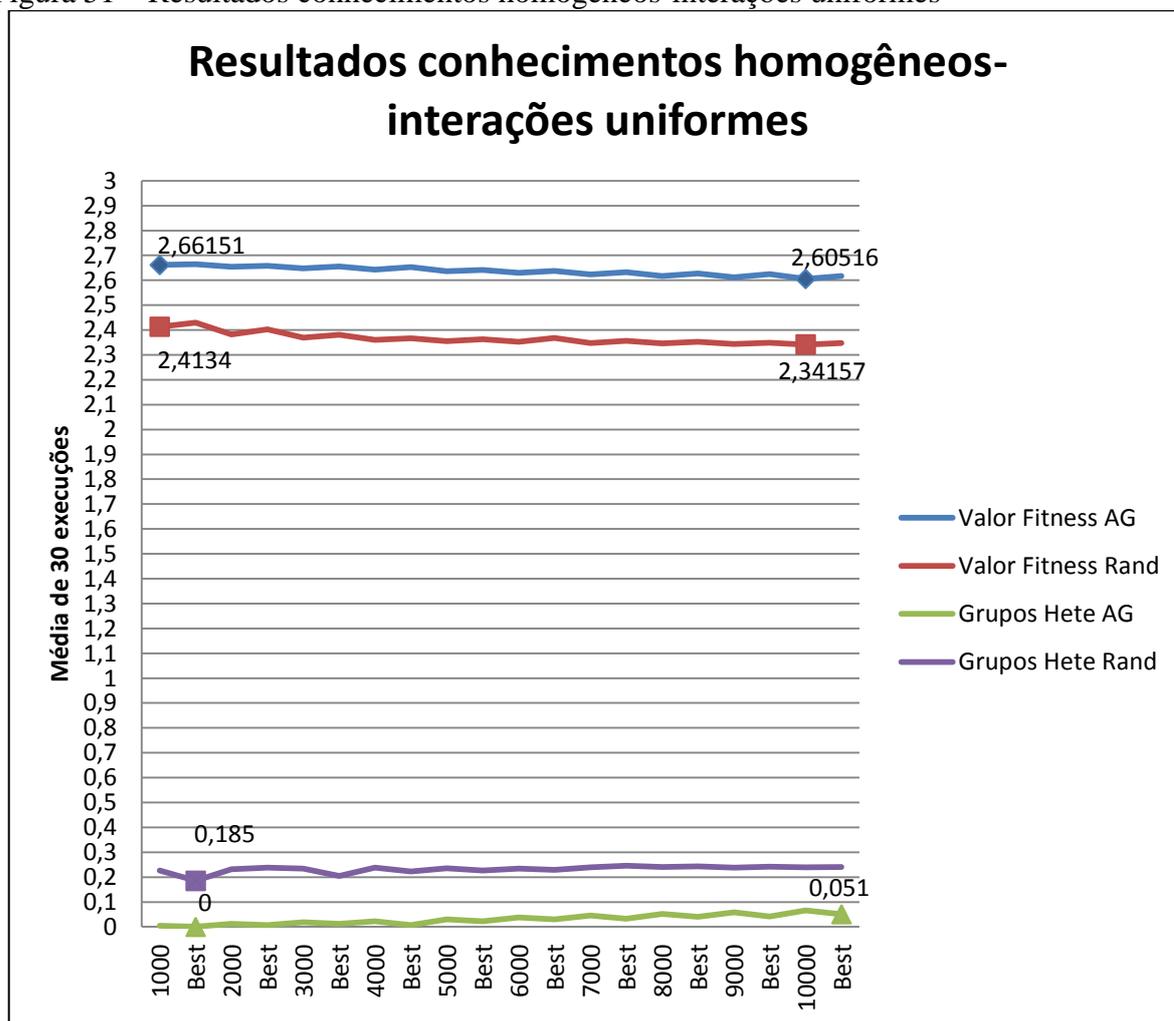
5.5.2 Conhecimentos homogêneos

Os testes com conhecimentos homogêneos visaram não permitir a formação de grupos ótimos em relação à heterogeneidade. Nesse caso, os grupos foram formados contendo estudantes com conhecimentos 0 (zero) e 1 (um), no entanto tinham mais estudantes com conhecimentos na média (0,5), pois a classe possuía 60% dos estudantes com conhecimentos na média. Como todos os grupos tinham o mesmo tamanho e todos os grupos foram formados da mesma maneira: com um estudante com conhecimento igual 1 (um), mais três estudantes com conhecimento igual a 0,5 (média), mais um estudantes com conhecimento igual a 0 (zero), a inter-homogeneidade dos grupos foi igual a 0 (zero), isto é, não houve diferença

entre os grupos; e a heterogeneidade alcançada foi menor que os 70% (0.7) estabelecidos para o cálculo do grupo como heterogêneo. Assim, nos casos dos testes com conhecimentos homogêneos, o *fitness* calculado deve ter o maior valor possível, mas o percentual de grupos heterogêneos deve ser próximo de zero, pois desta forma os grupos estão o mais balanceados possível, significando que a distribuição dos estudantes foi feita corretamente pelo algoritmo.

Na Figura 51, são apresentados os resultados para as interações **uniformes** (Tabela 10 – Apêndice B)..

Figura 51 – Resultados conhecimentos homogêneos-interações uniformes



Fonte: Próprio autor

Observa-se, na Figura 51, que o AG chegou aos resultados esperados. Isto é, o *fitness* alcançado pelo AG não chegou a 3, pois nesse caso não é possível alcançar o máximo da heterogeneidade, mas o percentual de grupos heterogêneos tendeu a 0 (zero), sendo que alcançou 0 (zero) para o melhor resultado com mil estudantes (*Best*) e alcançou 5% para o melhor resultado com dez mil estudantes. O *fitness* alcançado pelo algoritmo randômico foi menor que o alcançado pelo AG em todos os casos; e o percentual de grupos heterogêneos,

alcançado pelo algoritmo randômico, foi maior que o alcançado pelo AG em todos os casos, mostrando a situação inversa dos resultados do AG. As linhas no gráfico mostram que a diferença no valor do *fitness* fica estável ao aumentar o número de estudantes, variando de $\approx 0,25$ a $\approx 0,28$. A diferença no percentual de grupos heterogêneos ficou com valores negativos (na Tabela 10 – Apêndice B), devido ao AG ter valores mais baixos (tender a zero), o que era esperado e o melhor resultado para esses casos.

O *Wilcoxon test* foi aplicado para todas as quantidades de estudantes, tanto para o *fitness* quanto para o percentual de grupos heterogêneos. Na Tabela 8, são apresentados os resultados dos testes para, respectivamente, a menor e a maior diferença em cada caso.

Tabela 8 – *Wilcoxon test* conhecimentos homogêneos-interações uniformes

	#Estudantes	<i>Wilcoxon test</i> <i>p-valor</i>
Fitness	1000	2.857e-11
	4000	3.005e-11
Grupos Heterogêneos	10000	2.973e-11
	1000	1.896e-11

Fonte: Próprio autor

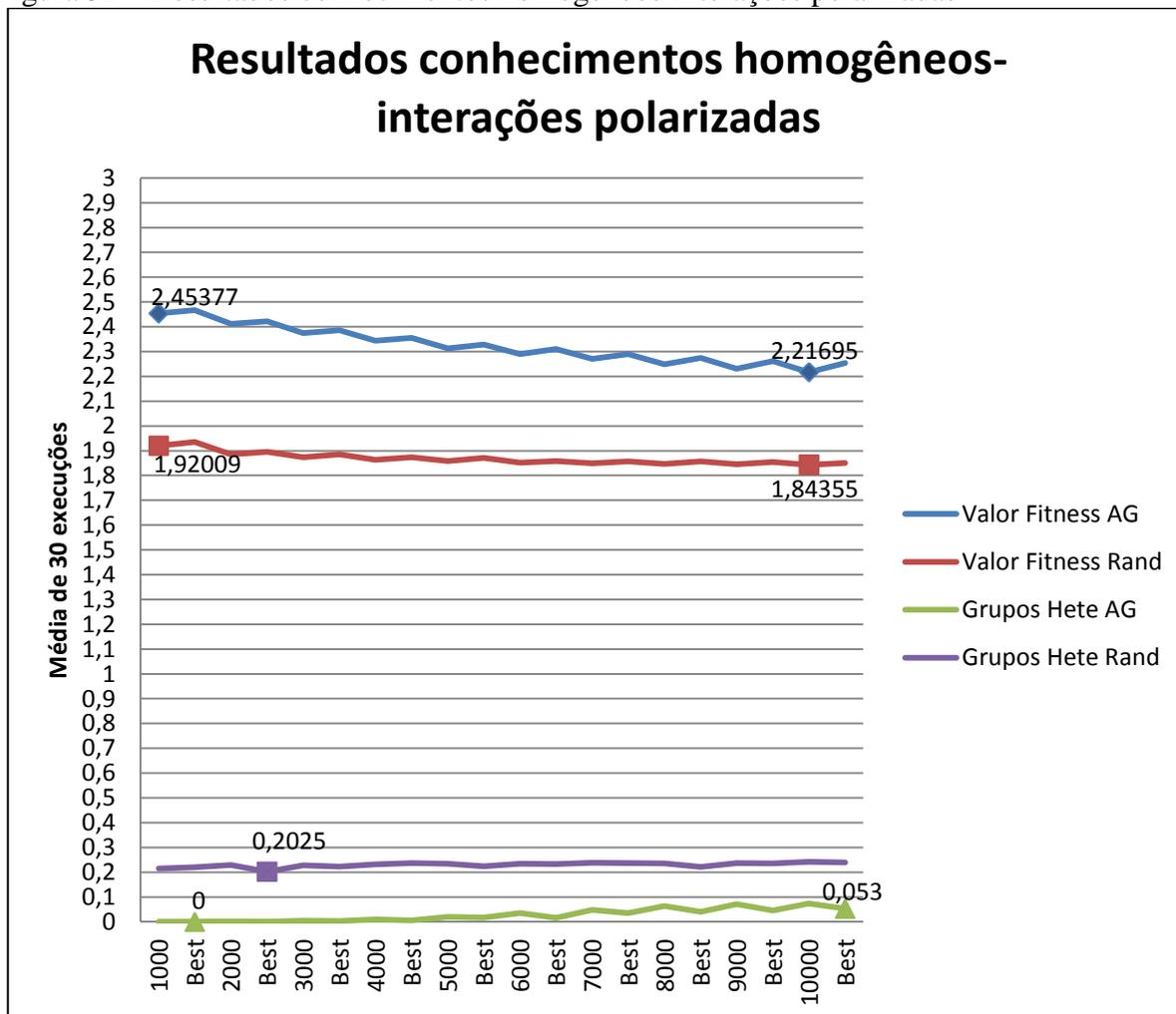
Em todos os casos, a hipótese nula foi rejeitada com *p-valor* muito menor que o nível de significância estabelecido ($p\text{-valor} \ll \alpha = 0.05$). Assim, conclui-se que os resultados obtidos pelo AG foram melhores que os resultados obtidos pelo algoritmo randômico.

Nos casos com peso maior para os conhecimentos (ω_1 maior que ω_2) e com peso maior para as interações (ω_2 maior que ω_1), os resultados obtidos foram semelhantes aos ocorridos anteriormente. Isso aconteceu porque, apesar dos conhecimentos serem homogêneos, a função *fitness* considerou o balanceamento dos grupos, de forma que os grupos balanceados obtiveram melhor *fitness*. E para as interações, os valores desses casos de testes são valores ótimos, de forma que alcançam valor máximo igual a 1 em ambos os casos. Assim, estes casos não são discutidos (Tabelas 11 e 12 - Apêndice B).

Na Figura 52 (Tabela 13 – Apêndice B) são apresentados os resultados para as interações **polarizadas** e observa-se que os resultados foram semelhantes aos descritos anteriormente para interações uniformes e pesos iguais. Isto é, o AG chegou aos resultados esperados, com o *fitness* alcançado pelo AG não chegando a 3, mas o percentual de grupos heterogêneos tendendo a 0 (zero), sendo que alcançou 0 (zero) para o melhor resultado com mil estudantes (*Best*) e alcançou 5% no melhor resultado para dez mil estudantes. O *fitness* alcançado pelo algoritmo randômico foi menor que o alcançado pelo AG em todos os casos; e o percentual de grupos heterogêneos, alcançado pelo algoritmo randômico, foi maior que o alcançado pelo AG em todos os casos, mostrando a situação inversa dos resultados do AG. As

linhas no gráfico mostram que as diferenças no valor do *fitness*, nesse caso, foram maiores. As diferenças no percentual de grupos heterogêneos foram semelhantes às ocorridas nos casos anteriores.

Figura 52 – Resultados conhecimentos homogêneos-interações polarizadas



Fonte: Próprio autor

O *Wilcoxon test* foi aplicado para todas as quantidades de estudantes, tanto para o *fitness* quanto para o percentual de grupos heterogêneos. Na Tabela 9, são apresentados os resultados dos testes para, respectivamente, a menor e a maior diferença em cada caso.

Tabela 9 – *Wilcoxon test* conhecimentos homogêneos-interações polarizadas

	#Estudantes	<i>Wilcoxon test</i> <i>p</i> -valor
Fitness	10000	3.012e-11
	1000	3.012e-11
Grupos Heterogêneos	9000	2.992e-11
	2000	1.579e-11

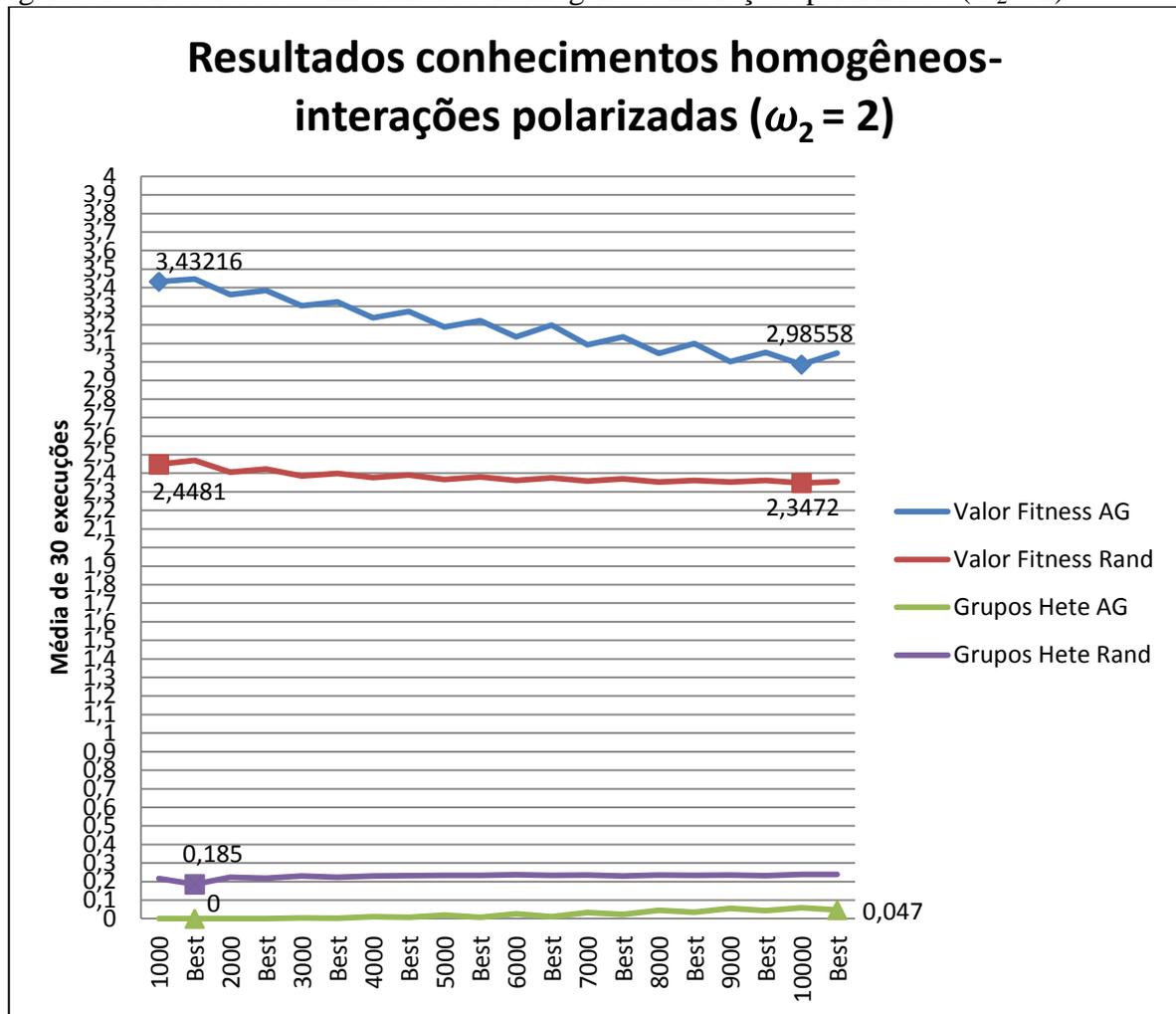
Fonte: Próprio autor

Em todos os casos, a hipótese nula foi rejeitada com *p-valor* muito menor que o nível de significância estabelecido (*p-valor* $\ll \alpha = 0.05$). Desta forma, conclui-se que os resultados obtidos pelo AG foram melhores que os resultados obtidos pelo algoritmo randômico.

Nos casos com peso maior para os conhecimentos (ω_1 maior que ω_2), os valores do *fitness* e os percentuais de grupos heterogêneos ficaram um pouco maiores (de $\approx 1\%$ a $\approx 7\%$). No entanto, no geral, os resultados foram semelhantes aos descritos anteriormente (pesos iguais). Assim, esse caso não é discutido (Tabela 14 – Apêndice B).

Nos casos com peso maior para as interações (ω_2 maior que ω_1), a Figura 53 (Tabela 15 – Apêndice B) mostra os resultados. Observa-se que os valores do *fitness* são muito diferentes entre o AG e o algoritmo randômico, gerando diferenças grandes entre os valores. O *fitness* alcançado pelo algoritmo randômico foi menor que o alcançado pelo AG em todos os casos.

Figura 53 – Resultados conhecimentos homogêneos-interações polarizadas ($\omega_2 = 2$)



Fonte: Próprio autor

O percentual de grupos heterogêneos variou de $\approx 0\%$ a $\approx 5\%$, considerando-se os melhores resultados, semelhante ao percentual obtido nos casos anteriores; e o percentual de grupos heterogêneos, alcançado pelo algoritmo randômico, foi maior que o alcançado pelo AG em todos os casos, mostrando a situação inversa dos resultados do AG. Apesar da grande diferença no valor do *fitness*, as linhas do gráfico mostram que as diferenças em relação aos grupos heterogêneos são semelhantes às ocorridas anteriormente.

Foi aplicado o *Wilcoxon test* para todas as quantidades de estudantes, tanto para o *fitness* quanto para o percentual de grupos heterogêneos. Na Tabela 10, são apresentados os resultados dos testes para, respectivamente, a menor e a maior diferença.

Tabela 10 – *Wilcoxon test* conhecimentos homogêneos-interações polarizadas ($\omega_2 = 2$)

	#Estudantes	<i>Wilcoxon test</i> <i>p-valor</i>
Fitness	10000	3.012e-11
	1000	3.01e-11
Grupos Heterogêneos	9000	2.99e-11
	3000	2.521e-11

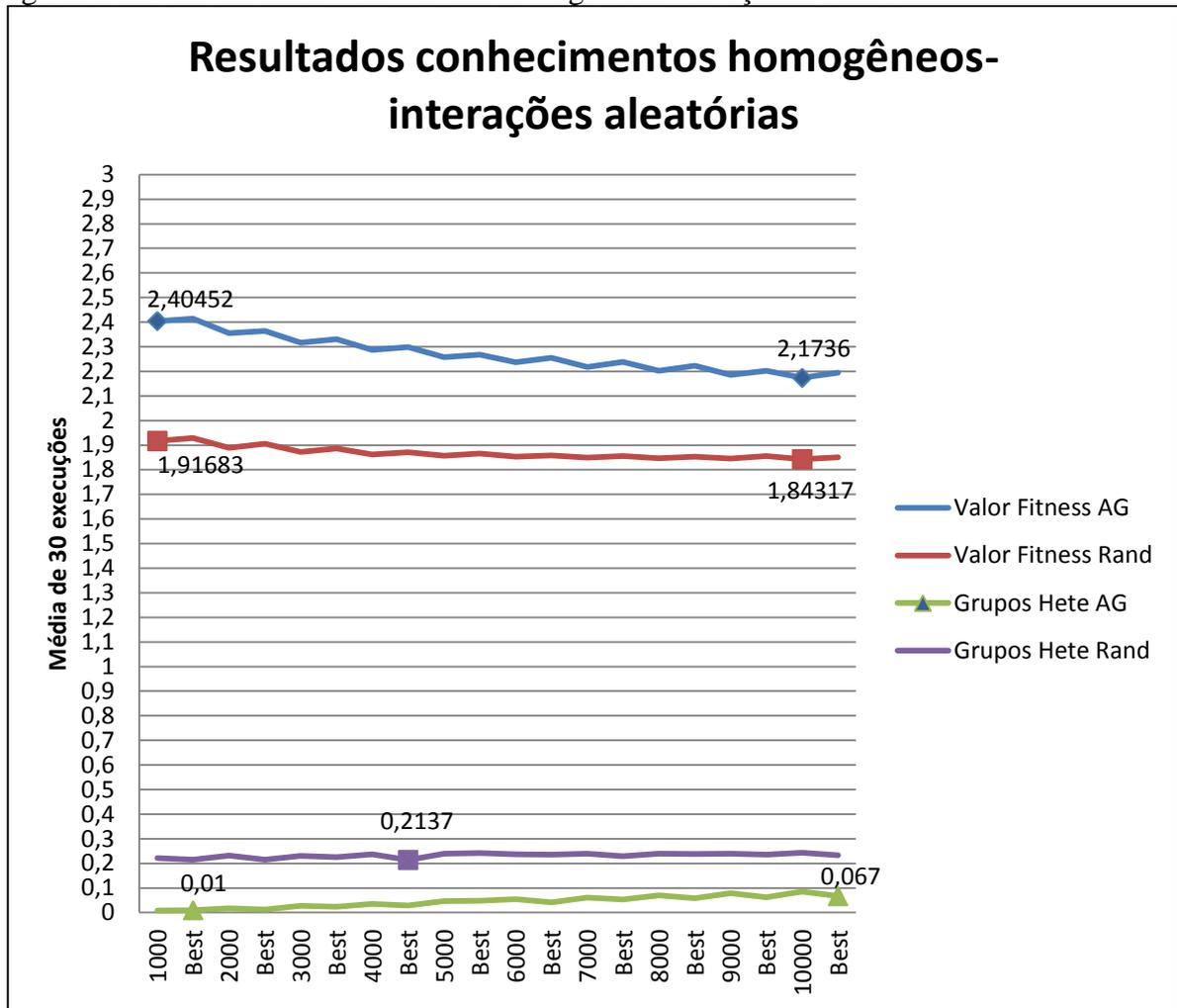
Fonte: Próprio autor

Em todos os casos, a hipótese nula foi rejeitada com *p-valor* muito menor que o nível de significância estabelecido ($p\text{-valor} \ll \alpha = 0.05$). Desta forma, conclui-se que os resultados do AG foram melhores que os resultados do algoritmo randômico.

Na Figura 54 (Tabela 16 – Apêndice B) são apresentados os resultados para as interações **aleatórias**. Observa-se, nesse caso, que o *fitness* ficou com valores menores que os casos anteriores. Entretanto, o percentual de grupos heterogêneos tendeu a 0 (zero), sendo que alcançou $\approx 1\%$ para o melhor resultado com mil estudantes e alcançou $\approx 7\%$ para o melhor resultado com dez mil estudantes, indicando que o AG alcançou bons resultados (apesar do valor do *fitness* mais baixo). O *fitness* alcançado pelo algoritmo randômico foi menor que o alcançado pelo AG em todos os casos; e o percentual de grupos heterogêneos, alcançado pelo algoritmo randômico, foi maior que o alcançado pelo AG em todos os casos, mostrando a situação inversa dos resultados do AG. E, apesar do *fitness* do AG ter apresentado valores menores, as diferenças no valor do *fitness* e as diferenças no percentual de grupos heterogêneos são semelhantes às ocorridas no caso de interações polarizadas (com pesos iguais).

Foi aplicado o *Wilcoxon test* para todas as quantidades de estudantes, tanto para o *fitness* quanto para o percentual de grupos heterogêneos. Na Tabela 11, são apresentados os resultados dos testes para, respectivamente, a menor e a maior diferença em cada caso.

Figura 54 – Resultados conhecimentos homogêneos-interações aleatórias



Fonte: Próprio autor

Tabela 11 – Wilcoxon test conhecimentos homogêneos-interações aleatórias

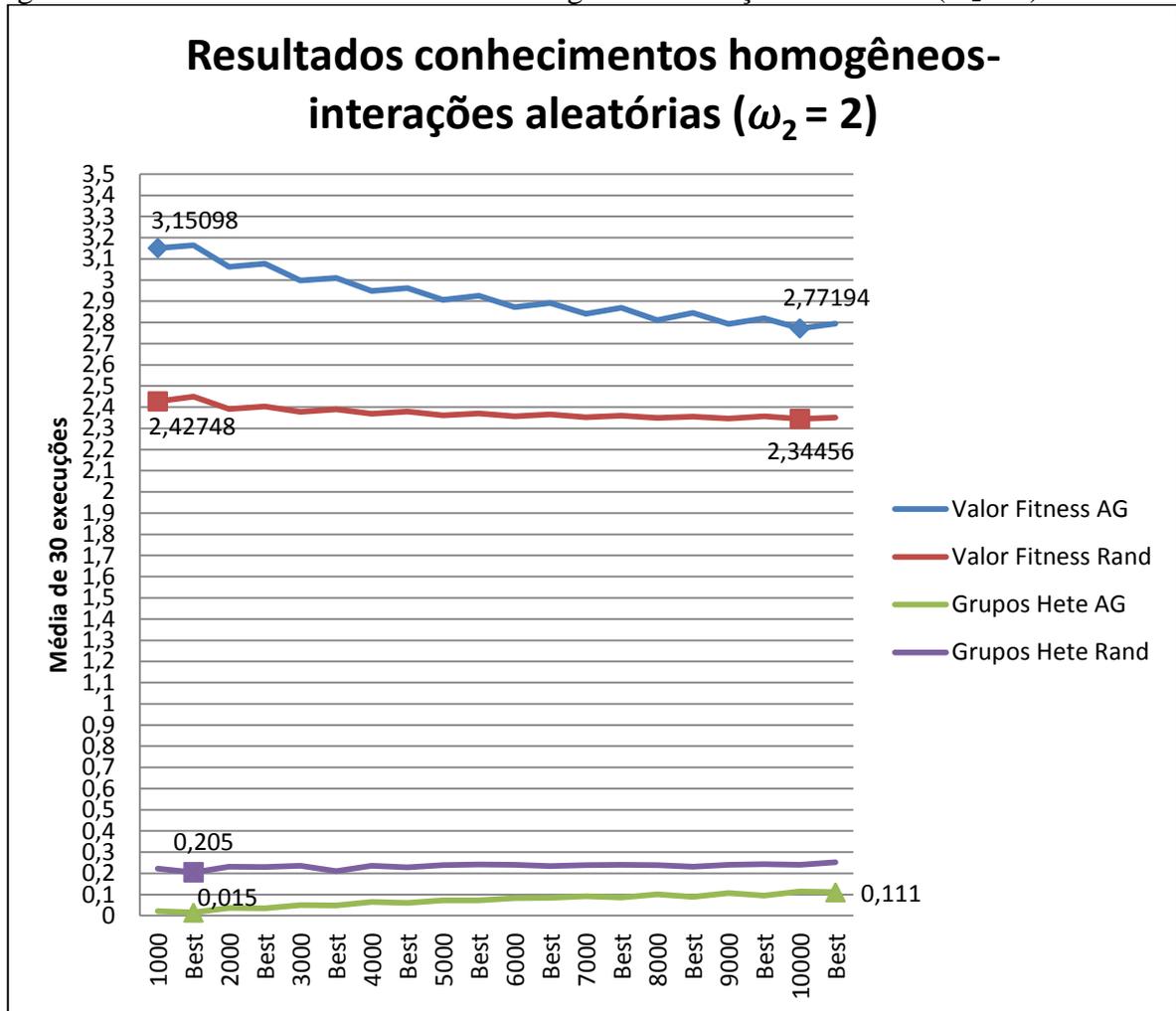
	#Estudantes	Wilcoxon test <i>p-valor</i>
Fitness	10000	3.016e-11
	1000	3.005e-11
Grupos Heterogêneos	10000	2.997e-11
	2000	2.428e-11

Fonte: Próprio autor

Em todos os casos, a hipótese nula foi rejeitada com *p-valor* muito menor que o nível de significância estabelecido ($p\text{-valor} \ll \alpha = 0.05$). Desta forma, conclui-se que os resultados obtidos pelo AG foram melhores que os resultados obtidos pelo algoritmo randômico.

Nos casos com peso maior para os conhecimentos (ω_1 maior que ω_2), os valores do *fitness* e os percentuais de grupos heterogêneos ficaram um pouco maiores (de $\approx 1\%$ a $\approx 8\%$). No entanto, no geral, os resultados foram semelhantes aos descritos anteriormente (pesos iguais). Assim, esse caso não é discutido (Tabela 17 – Apêndice B).

Nos casos com peso maior para as interações (ω_2 maior que ω_1), a Figura 55 (Tabela 18 – Apêndice B) mostra os resultados.

Figura 55 – Resultados conhecimentos homogêneos-interações aleatórias ($\omega_2 = 2$)

Fonte: Próprio autor

Nesse caso, observa-se que os valores do *fitness* foram maiores que os valores dos casos anteriores e há uma maior variação nos valores obtidos. O *fitness* alcançado pelo algoritmo randômico foi menor que o alcançado pelo AG em todos os casos. O percentual de grupos heterogêneos foi maior, mesmo considerando-se os melhores resultados, variou de $\approx 2\%$ a $\approx 11\%$. No entanto, o percentual de grupos heterogêneos alcançado pelo algoritmo randômico foi maior que o alcançado pelo AG em todos os casos. As linhas do gráfico mostram que as diferenças nos valores do *fitness* foram maiores e as diferenças nos percentuais de grupos heterogêneos foram menores que às ocorridas nos casos anteriores.

Foi aplicado o *Wilcoxon test* para todas as quantidades de estudantes, tanto para o *fitness* quanto para o percentual de grupos heterogêneos. Na Tabela 12, são apresentados os resultados dos testes para, respectivamente, a menor e a maior diferença.

Em todos os casos, a hipótese nula foi rejeitada com *p-valor* muito menor que o nível de significância estabelecido (*p-valor* $\ll \alpha = 0.05$). Mesmo para a menor diferença no valor do *fitness* e no percentual de grupos heterogêneos, tem-se que as diferenças são significativas,

e conclui-se que os resultados do AG foram melhores que os resultados do algoritmo randômico.

Tabela 12 – *Wilcoxon test* conhecimentos homogêneos-interações aleatórias ($\omega_2 = 2$)

	#Estudantes	<i>Wilcoxon test</i> <i>p-valor</i>
Fitness	10000	3.012e-11
	1000	3.016e-11
Grupos Heterogêneos	10000	2.997e-11
	1000	2.591e-11

Fonte: Próprio autor

Em relação à aplicação de pesos para os conhecimentos, assim como para conhecimentos heterogêneos, conclui-se que a aplicação do peso não influenciou o comportamento do algoritmo de maneira geral, pois os resultados foram semelhantes aos do casos anteriores com pesos iguais. Isso ocorreu devido à função *fitness* considerar o balanceamento dos grupos, em conjunto com a heterogeneidade e as interações. Assim, como nesse caso os valores dos conhecimentos estavam distribuídos de forma a gerar grupos igualmente balanceados, os algoritmos geraram os grupos semelhantes aos grupos gerados nos casos de pesos iguais. Esses casos não foram discutidos. No entanto, o *Wilcoxon test* foi aplicado para todas as quantidades de estudantes, tanto para o *fitness* quanto para o percentual de grupos heterogêneos. Na Tabela 13, são apresentados os resultados dos testes para a menor diferença, respectivamente, para interações uniformes, interações polarizadas e interações aleatórias.

Tabela 13 – *Wilcoxon test* conhecimentos homogêneos, $\omega_1 = 2$, menores diferenças ocorridas

	Interações	#Estudantes	<i>Wilcoxon test (p-valor)</i>
Fitness	Uniformes	1000	2.836e-11
	Polarizadas	10000	3.007e-11
	Aleatórias	10000	3.01e-11
Grupos Heterogêneos	Uniformes	10000	2.986e-11
	Polarizadas	10000	2.982e-11
	Aleatórias	10000	2.99e-11

Fonte: Próprio autor

Em todos os casos, a hipótese nula foi rejeitada com *p-valor* muito menor que o nível de significância estabelecido ($p\text{-valor} \ll \alpha = 0.05$). Assim, conclui-se que os resultados alcançados pelo AG foram melhores que os resultados alcançados pelo algoritmo randômico, também para conhecimentos homogêneos e peso maior para os conhecimentos.

Em relação à aplicação de pesos para as interações, assim como para os conhecimentos heterogêneos, conclui-se que a aplicação do peso influenciou o comportamento do algoritmo, inclusive apresentando casos com percentuais mais altos de grupos heterogêneos, no caso de interações aleatórias (de $\approx 2\%$ a $\approx 11\%$, e neste caso deveriam

ser mais próximos de zero). Entretanto, de maneira geral, o algoritmo obteve bons resultados em relação ao algoritmo randômico, pois seus resultados para os percentuais de grupos heterogêneos foram sempre menores que o do algoritmo randômico; e seus resultados para o *fitness* foram sempre maiores que os do algoritmo randômico. O que, neste caso, está correto.

Conforme tudo o que foi apresentado para os testes dos conhecimentos homogêneos, o AG alcançou heterogeneidade baixa, variando entre $\approx 0\%$ a $\approx 11\%$ (considerando-se os melhores resultados – *Best*), sendo o esperado para esse caso, já que os conhecimentos são homogêneos, e os grupos devem ser balanceados. Desta forma, os grupos formados foram homogêneos, isto é, não alcançaram o percentual de 70% (0.7) de heterogeneidade estabelecido como meta para considerar o grupo heterogêneo. Assim, conclui-se que, de forma geral, o AG obteve resultados significativamente melhores que os resultados obtidos pelo algoritmo randômico, uma vez que as médias apresentadas pelo AG foram maiores que as médias do algoritmo randômico para o *fitness*; e para o percentual de grupos heterogêneos as médias do AG foram menores que as médias do algoritmo randômico, conforme o esperado para esse caso.

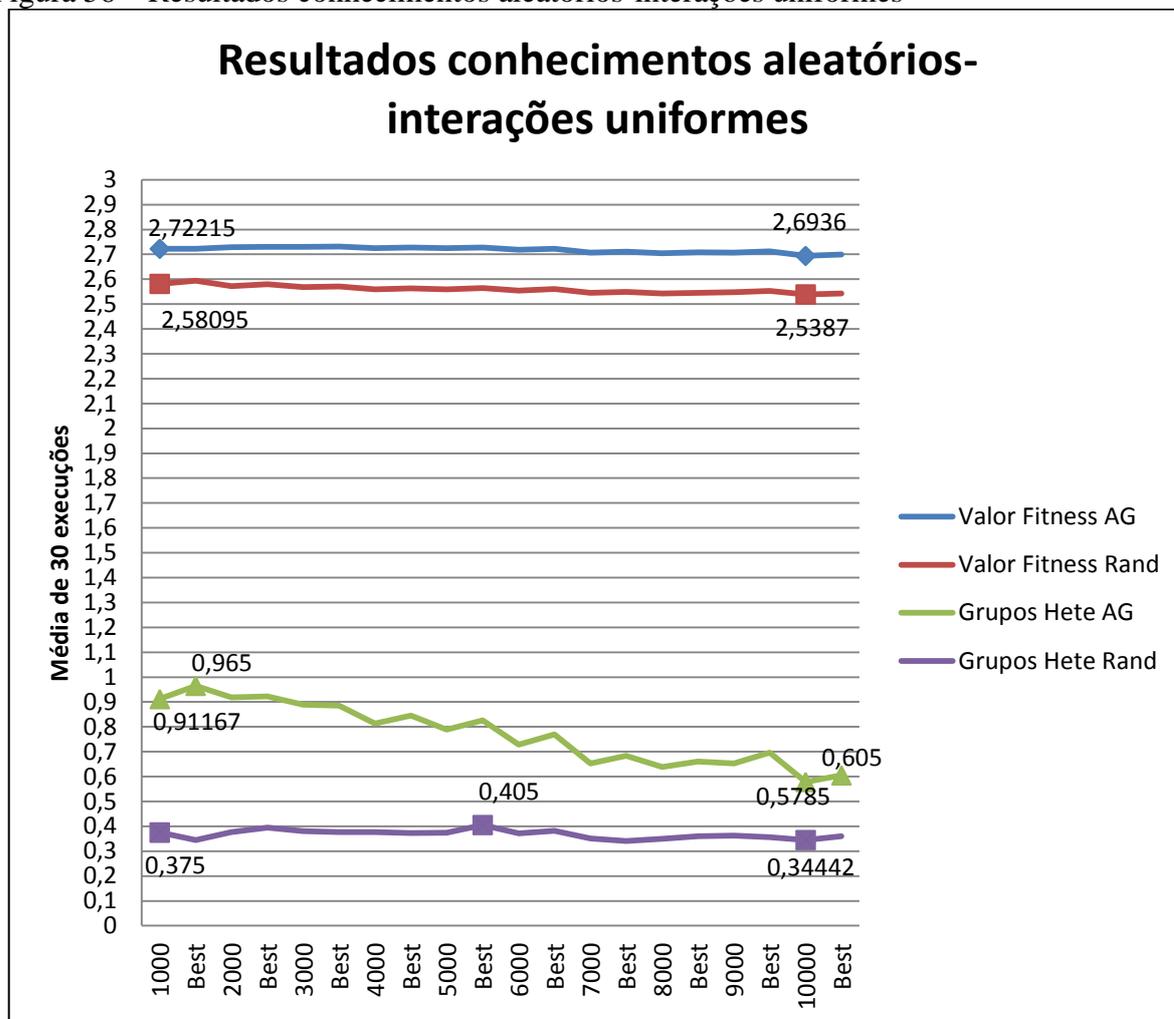
5.5.3 Conhecimentos aleatórios

Os testes com conhecimentos aleatórios tinham como objetivo simular uma situação em que o pesquisador não tivesse controle sobre os dados, especialmente no caso em que as interações também fossem aleatórias. Nesses casos, o valor do *fitness* deve ser o maior possível, mesmo sabendo que este valor não alcança o máximo de 3 (para o caso de pesos iguais) ou 4 (para os casos com pesos diferentes). Da mesma forma, o percentual de grupos heterogêneos não deve alcançar o máximo possível, mas deve alcançar um valor maior que zero, pois nesse caso a classe é uma classe heterogênea, uniformemente distribuída, com valores aleatórios no intervalo [0,1].

Na Figura 56 (Tabela 19 – Apêndice B) são apresentados os resultados para as interações **uniformes**. Observa-se que o valor do *fitness* variou de $\approx 2,69$ a $\approx 2,72$. Como o melhor caso das interações é precisamente o das interações uniformes, onde todos interagem entre si o máximo de vezes possível, sabe-se que o *fitness* máximo para esses casos será este (para esta configuração de parâmetros e pesos iguais), e considera-se que está próximo de 3 (o máximo possível). O *fitness* alcançado pelo algoritmo randômico foi menor que o alcançado pelo AG em todos os casos. A coluna de grupos heterogêneos mostra que o AG alcançou percentuais melhores para quantidades menores de estudantes, chegando a $\approx 97\%$ (0,965) no melhor caso (*Best*) com mil estudantes. Para dez mil estudantes o AG chegou a $\approx 61\%$ (0,605),

mostrando que o caso de conhecimentos aleatórios dificulta o trabalho do algoritmo. No entanto, os resultados do AG foram melhores que os do algoritmo randômico, que chegou a um máximo de $\approx 41\%$ (0,405) de grupos heterogêneos (para o melhor resultado com cinco mil estudantes).

Figura 56 – Resultados conhecimentos aleatórios-interações uniformes



Fonte: Próprio autor

Mesmo para o pior resultado, $\approx 58\%$ (0,5785 - dez mil estudantes), o AG alcançou $\approx 23\%$ a mais de grupos heterogêneos do que o algoritmo randômico. As linhas do gráfico mostram que as diferenças no valor do *fitness* são pequenas e parecidas, variando de $\approx 0,14$ a $\approx 0,17$. As diferenças no percentual de grupos heterogêneos são maiores, variando de $\approx 0,23$ a $\approx 0,54$.

Para avaliar a significância das diferenças foi aplicado o *Wilcoxon test* para todas as quantidades de estudantes, tanto para o *fitness* quanto para o percentual de grupos heterogêneos. Na Tabela 14, são apresentados os resultados dos testes para, respectivamente, a menor e a maior diferença.

Tabela 14 – *Wilcoxon test* conhecimentos aleatórios-interações uniformes

	#Estudantes	<i>Wilcoxon test</i> <i>p-valor</i>
Fitness	1000	2.806e-11
	4000	2.96e-11
Grupos Heterogêneos	10000	2.958e-11
	2000	2.809e-11

Fonte: Próprio autor

Em todos os casos, a hipótese nula foi rejeitada com *p-valor* muito menor que o nível de significância estabelecido ($p\text{-valor} \ll \alpha = 0.05$). Assim, conclui-se que os resultados do AG foram melhores que os resultados do algoritmo randômico.

Nos casos com peso maior para os conhecimentos (ω_1 maior que ω_2) e com peso maior para as interações (ω_2 maior que ω_1), os resultados obtidos foram semelhantes aos ocorridos anteriormente. Os valores dos *fitness* foram um pouco maiores devido à aplicação dos pesos. No entanto, os percentuais de grupos heterogêneos variaram de $\approx 58\%$ a $\approx 92\%$. Assim, estes casos não são discutidos (Tabelas 20 e 21 - Apêndice B).

Na Figura 57 (Tabela 22 – Apêndice B) são apresentados os resultados para as interações **polarizadas**. Nesse caso, como as interações são polarizadas, e não é possível alcançar o máximo para as interações, observa-se que o valor do *fitness* ficou menor. O percentual de grupos heterogêneos também ficou menor que o caso anterior. No entanto, o *fitness* alcançado pelo algoritmo randômico foi menor que o alcançado pelo AG em todos os casos; assim como o percentual de grupos heterogêneos. Mesmo para o pior resultado, $\approx 47\%$ (dez mil estudantes), o AG alcançou $\approx 13\%$ a mais de grupos heterogêneos do que o algoritmo randômico. As linhas do gráfico mostram que as diferenças no valor do *fitness*, ao contrário do caso anterior, foram maiores. As diferenças no percentual de grupos heterogêneos foram menores que o caso anterior.

O *Wilcoxon test* foi aplicado para avaliar a significância destas diferenças, tanto para o *fitness* quanto para o percentual de grupos heterogêneos. Na Tabela 15, são apresentados os resultados do teste para, respectivamente, a menor e a maior diferença.

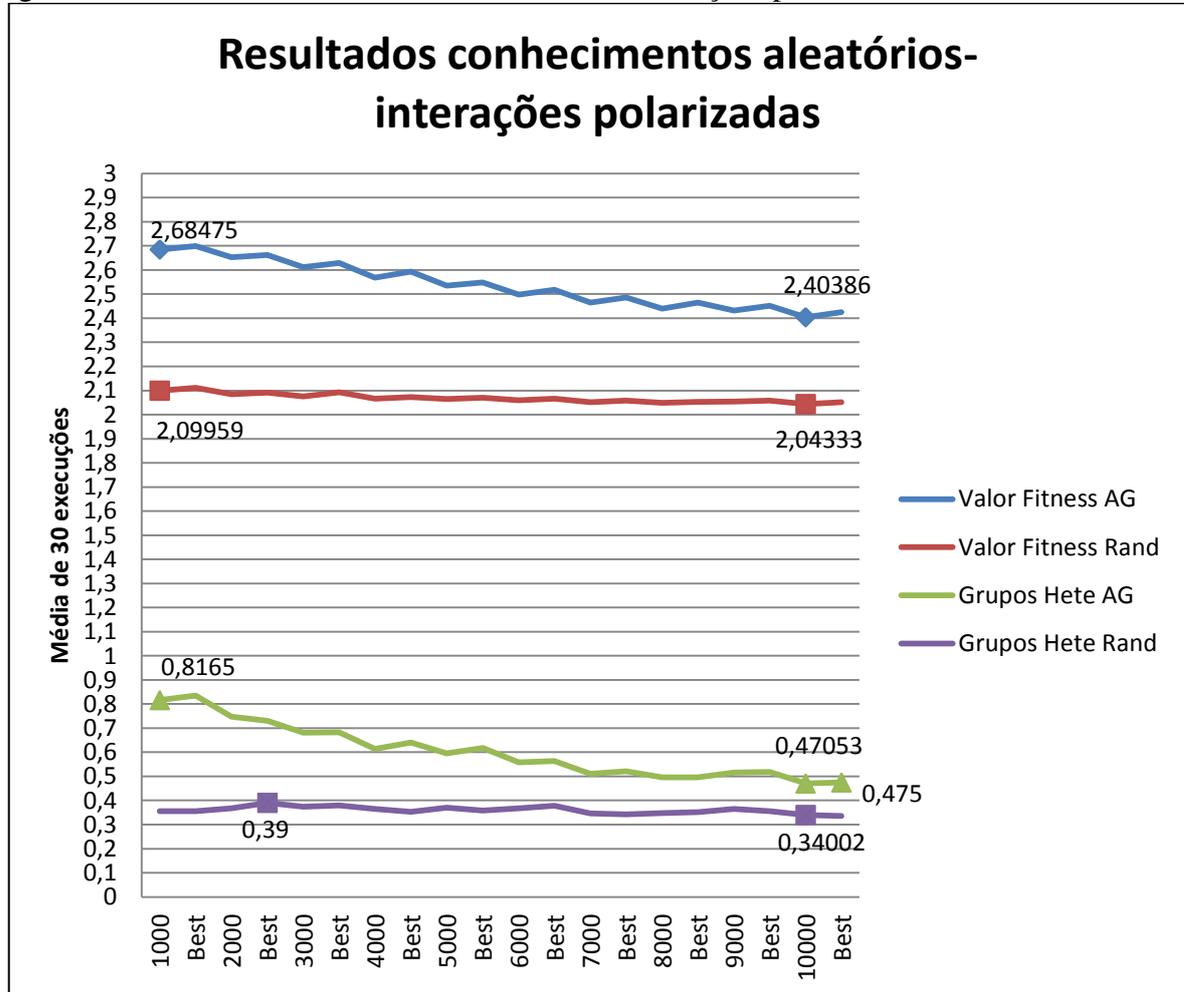
Tabela 15 – *Wilcoxon test* conhecimentos aleatórios-interações polarizadas

	#Estudantes	<i>Wilcoxon test</i> <i>p-valor</i>
Fitness	10000	3.014e-11
	1000	3.014e-11
Grupos Heterogêneos	10000	2.986e-11
	1000	2.874e-11

Fonte: Próprio autor

Em todos os casos, a hipótese nula foi rejeitada, com *p-valor* muito menor que o nível de significância estabelecido (*p-valor* $\ll \alpha = 0.05$). Assim, conclui-se que os resultados do AG foram melhores que os resultados do algoritmo randômico.

Figura 57 – Resultados conhecimentos aleatórios-interações polarizadas



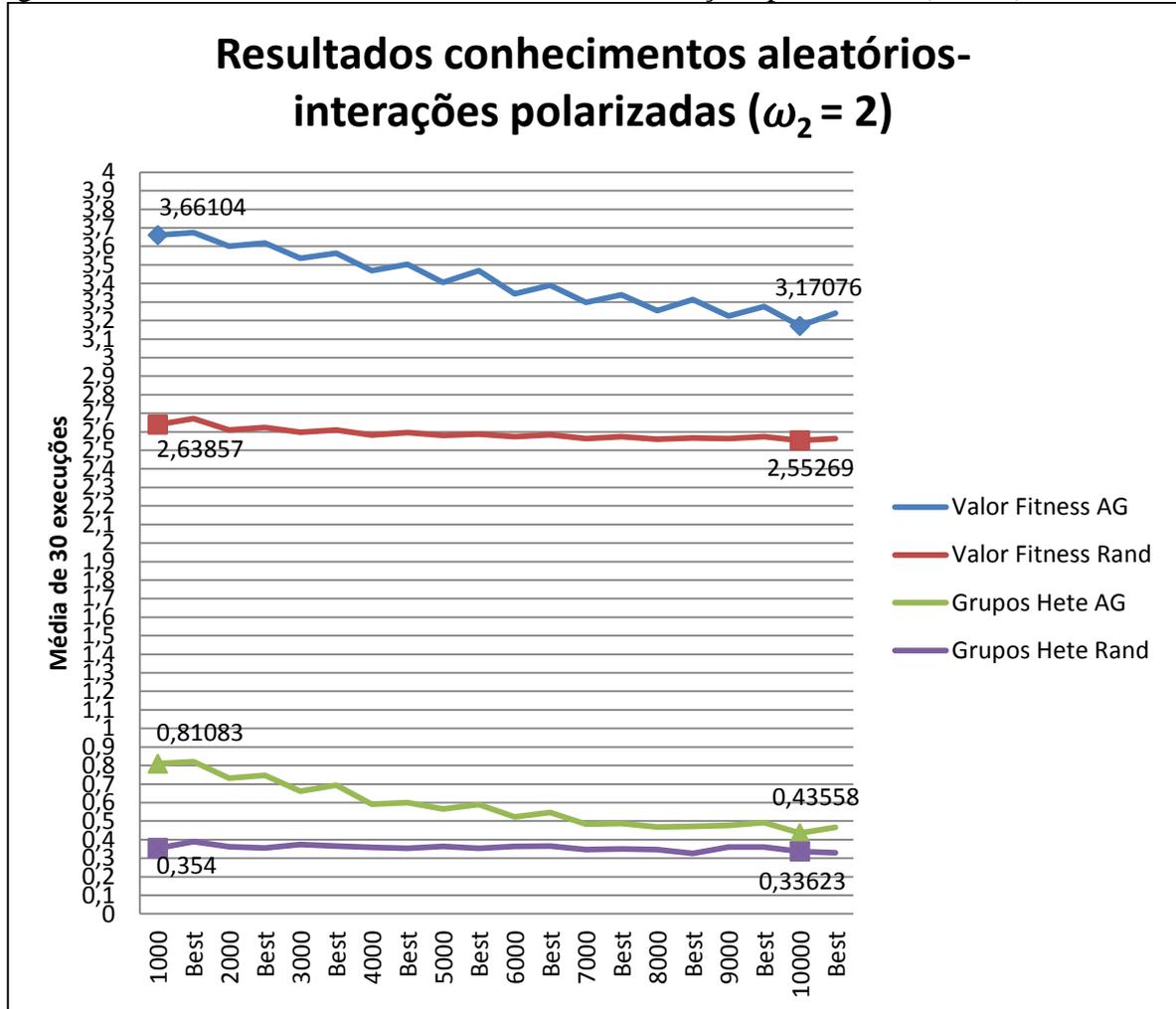
Fonte: Próprio autor

Nos casos com peso maior para os conhecimentos (ω_1 maior que ω_2), os valores do *fitness* ficaram maiores devido ao peso aplicado. No entanto, os percentuais de grupos heterogêneos foram semelhantes aos ocorridos anteriormente. Assim, esse caso não é discutido (Tabela 23 – Apêndice B).

Nos casos com peso maior para as interações (ω_2 maior que ω_1), a Figura 58 (Tabela 24 – Apêndice B) mostra os resultados. Nesse caso, observa-se que os valores do *fitness* são muito diferentes entre o AG e o algoritmo randômico, gerando diferenças grandes entre os valores. O *fitness* alcançado pelo algoritmo randômico foi menor que o alcançado pelo AG em todos os casos. Os percentuais de grupos heterogêneos variaram de $\approx 44\%$ a $\approx 81\%$, e foram menores que os percentuais obtidos nos casos anteriores. No entanto, o percentual de grupos heterogêneos alcançado pelo algoritmo randômico foi menor que o alcançado pelo AG em

todos os casos. Apesar da diferença grande no valor do *fitness*, as diferenças em relação aos grupos heterogêneos são semelhantes às ocorridas anteriormente.

Figura 58 – Resultados conhecimentos aleatórios-interações polarizadas ($\omega_2 = 2$)



Fonte: Próprio autor

Foi aplicado o *Wilcoxon test* para todas as quantidades de estudantes, tanto para o *fitness* quanto para o percentual de grupos heterogêneos. Na Tabela 16, são apresentados os resultados dos testes para, respectivamente, a menor e a maior diferença.

Tabela 16 – *Wilcoxon test* conhecimentos aleatórios-interações polarizadas ($\omega_2 = 2$)

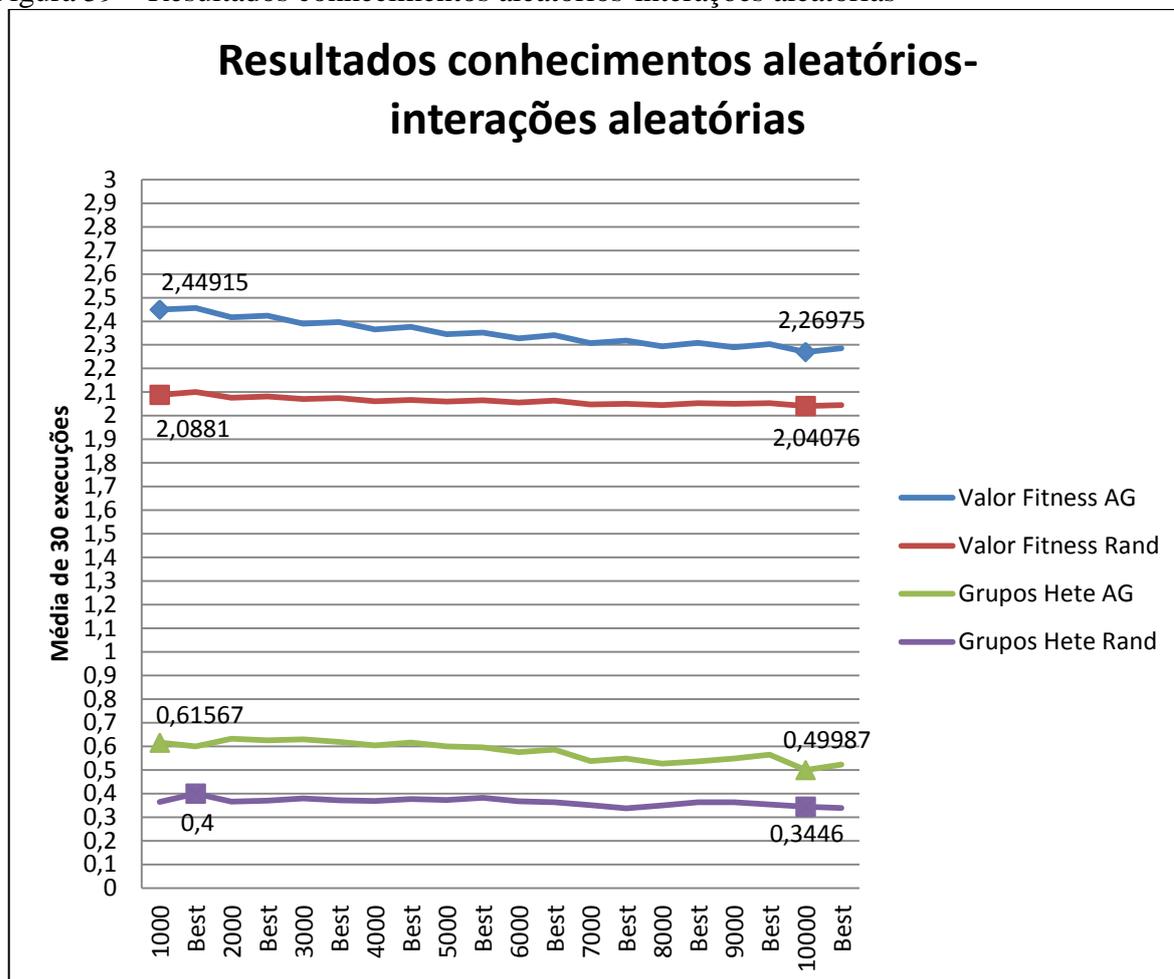
	#Estudantes	<i>Wilcoxon test</i> <i>p-valor</i>
Fitness	10000	3.016e-11
	1000	3.014e-11
Grupos Heterogêneos	10000	2.997e-11
	1000	2.848e-11

Fonte: Próprio autor

Em todos os casos, a hipótese nula foi rejeitada com *p-valor* muito menor que o nível de significância estabelecido ($p\text{-valor} \ll \alpha = 0.05$). Desta forma, conclui-se que os resultados do AG foram melhores que os resultados do algoritmo randômico.

Na Figura 59 (Tabela 25 – Apêndice B,) são apresentados os resultados para as interações **aleatórias**.

Figura 59 – Resultados conhecimentos aleatórios-interações aleatórias



Fonte: Próprio autor

Nesse caso, observa-se que o valor do *fitness* ficou ainda menor. O percentual de grupos heterogêneos também ficou menor que os casos anteriores. No entanto, o *fitness* alcançado pelo algoritmo randômico é menor que o alcançado pelo AG em todos os casos; assim como o percentual de grupos heterogêneos.

Observa-se na Figura 59 que, mesmo para o pior resultado, $\approx 50\%$ (0,49987 - dez mil estudantes), o AG alcançou $\approx 15\%$ a mais de grupos heterogêneos do que o algoritmo randômico. As linhas do gráfico mostram que as diferenças no valor do *fitness*, ao contrário do caso anterior, foram menores. As diferenças no percentual de grupos heterogêneos também foram menores.

Assim, o *Wilcoxon test* foi aplicado para avaliar a significância destas diferenças, tanto para o *fitness* quanto para o percentual de grupos heterogêneos. Na Tabela 17, são apresentados os resultados do teste para, respectivamente, a menor e a maior diferença.

Tabela 17 – *Wilcoxon test* conhecimentos aleatórios-interações aleatórias

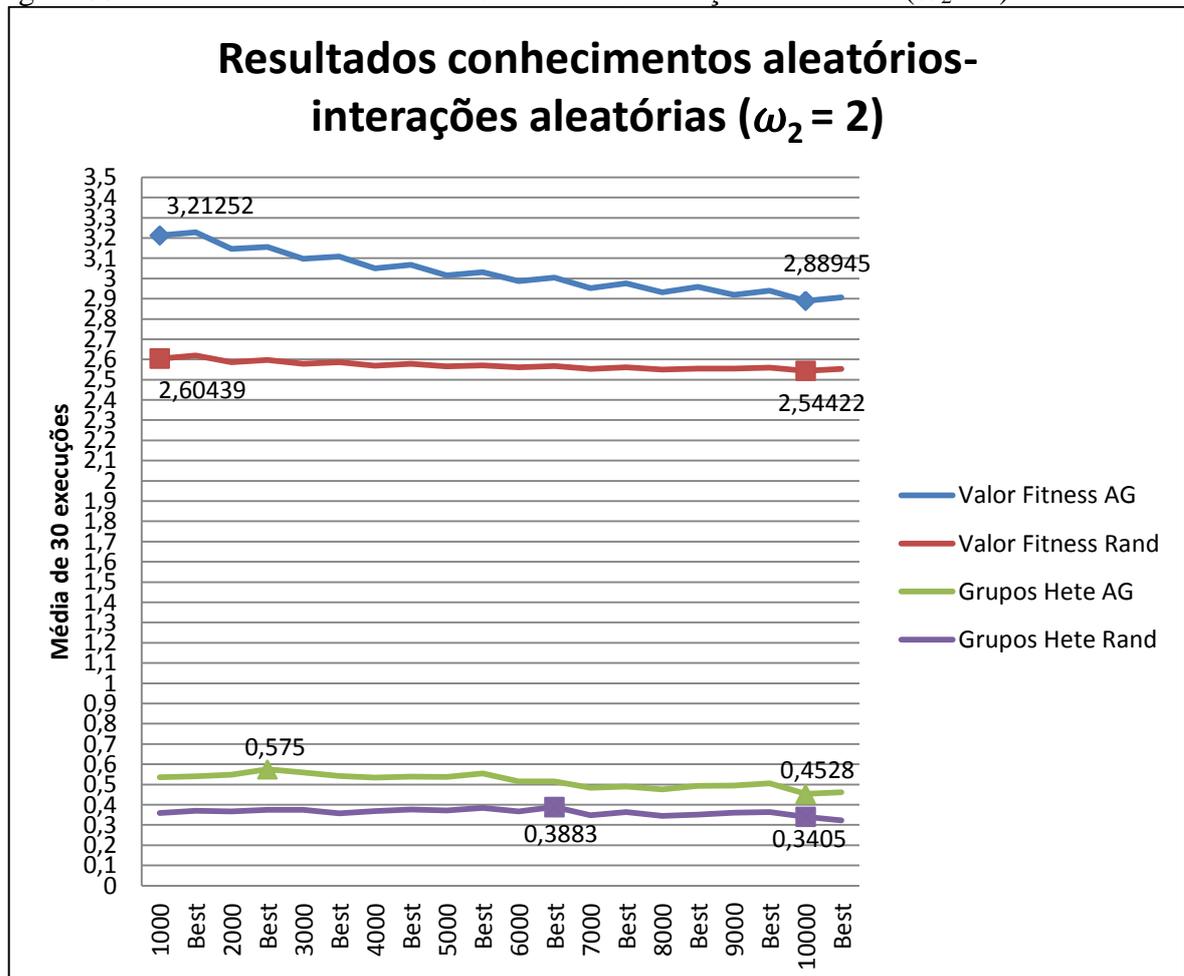
	#Estudantes	<i>Wilcoxon test</i> <i>p-valor</i>
Fitness	10000	3.001e-11
	1000	3.001e-11
Grupos Heterogêneos	10000	2.995e-11
	2000	2.965e-11

Fonte: Próprio autor

Em todos os casos, a hipótese nula foi rejeitada, com *p-valor* muito menor que o nível de significância estabelecido ($p\text{-valor} \ll \alpha = 0.05$). Desta forma, conclui-se que os resultados do AG foram melhores que os resultados do algoritmo randômico.

Nos casos com peso maior para os conhecimentos (ω_1 maior que ω_2), os valores do *fitness* e os percentuais de grupos heterogêneos ficaram um pouco maiores. No entanto, no geral, os resultados foram semelhantes aos descritos anteriormente (pesos iguais). Assim, esse caso não é discutido (Tabela 26 – Apêndice B).

Nos casos com peso maior para as interações (ω_2 maior que ω_1), a Figura 60 (Tabela 27 – Apêndice B) mostra os resultados.

Figura 60 – Resultados conhecimentos aleatórios-interações aleatórias ($\omega_2 = 2$)

Fonte: Próprio autor

Nesse caso, observa-se que os valores do *fitness* foram maiores, devido ao peso. O *fitness* alcançado pelo algoritmo randômico foi menor que o alcançado pelo AG em todos os casos. Os percentuais de grupos heterogêneos variaram de $\approx 45\%$ a $\approx 58\%$, e foram ainda menores que os percentuais obtidos nos casos anteriores. No entanto, o percentual de grupos heterogêneos alcançado pelo algoritmo randômico foi menor que o alcançado pelo AG em todos os casos. As linhas do gráfico mostram que as diferenças nos valores do *fitness* são maiores que o caso anterior. As diferenças no percentual de grupos heterogêneos são menores. Observa-se, nesse caso, que mesmo chegando a resultados piores que os casos anteriores, os resultados do AG foram melhores que os resultados do algoritmo randômico.

Assim, foi aplicado o *Wilcoxon test* para todas as quantidades de estudantes, tanto para o *fitness* quanto para o percentual de grupos heterogêneos. Na Tabela 18, são apresentados os resultados dos testes para, respectivamente, a menor e a maior diferença.

Tabela 18 – *Wilcoxon test* conhecimentos aleatórios-interações aleatórias ($\omega_2 = 2$)

	#Estudantes	<i>Wilcoxon test</i> <i>p-valor</i>
Fitness	10000	2.986e-11
	1000	3.01e-11
Grupos Heterogêneos	10000	2.999e-11
	3000	2.962e-11

Fonte: Próprio autor

Em todos os casos, a hipótese nula foi rejeitada com *p-valor* muito menor que o nível de significância estabelecido ($p\text{-valor} \ll \alpha = 0.05$). Mesmo para a menor diferença no valor do *fitness* e no percentual de grupos heterogêneos, tem-se que as diferenças são significativas, e conclui-se que os resultados do AG foram melhores que os resultados do algoritmo randômico.

Em relação à aplicação de pesos para os conhecimentos, assim como para conhecimentos heterogêneos e homogêneos, conclui-se que a aplicação do peso não influenciou o comportamento do algoritmo de maneira geral, pois os resultados foram semelhantes aos dos casos anteriores com pesos iguais. Esses casos não foram discutidos. No entanto, foi aplicado o *Wilcoxon test* para avaliar a significância de todas as quantidades de estudantes, tanto para o *fitness* quanto para o percentual de grupos heterogêneos. Na Tabela 19, são apresentados os resultados dos testes para a menor diferença, respectivamente, para interações uniforme, interações polarizadas e interações aleatórias.

Em todos os casos, a hipótese nula foi rejeitada com *p-valor* muito menor que o nível de significância estabelecido ($p\text{-valor} \ll \alpha = 0.05$). Assim, conclui-se que os resultados alcançados pelo AG foram melhores que os resultados alcançados pelo algoritmo randômico,

também para esses casos de testes, com conhecimentos aleatórios e peso maior para os conhecimentos.

Tabela 19 – *Wilcoxon test* conhecimentos aleatórios, $\omega_1 = 2$, menores diferenças ocorridas

	Interações	#Estudantes	<i>Wilcoxon test</i> (<i>p</i> -valor)
Fitness	Uniformes	1000	2.619e-11
	Polarizadas	10000	2.997e-11
	Aleatórias	10000	3.001e-11
Grupos Heterogêneos	Uniformes	10000	2.988e-11
	Polarizadas	10000	2.967e-11
	Aleatórias	10000	2.993e-11

Fonte: Próprio autor

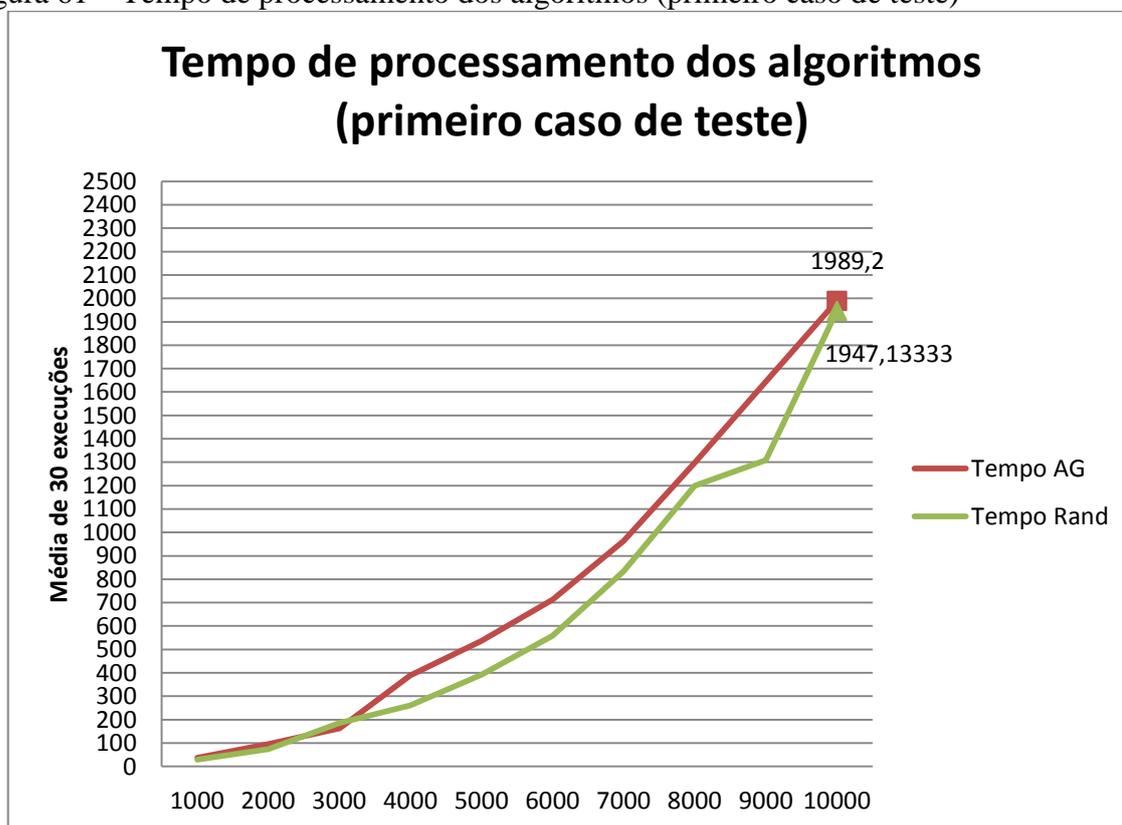
Em relação à aplicação de pesos para as interações, assim como para os conhecimentos heterogêneos e homogêneos, conclui-se que a aplicação do peso influenciou o comportamento do algoritmo, especialmente com interações aleatórias, em que apresentou percentuais de grupos heterogêneos mais baixos (de $\approx 45\%$ a $\approx 58\%$). Entretanto, de maneira geral, o algoritmo obteve bons resultados em relação ao algoritmo randômico, pois seus resultados, tanto para o *fitness* quanto para os percentuais de grupos heterogêneos, foram sempre maiores que os do algoritmo randômico. O que, neste caso, está correto.

Conforme tudo o que foi apresentado para os testes dos conhecimentos aleatórios, o AG obteve resultados bem variados, pois os dados foram simulados para que o pesquisador tivesse menos controle sobre estes. Ainda assim, o AG alcançou um percentual de $\approx 92\%$ de grupos heterogêneos, nos melhores resultados de interações uniformes, ficando com percentual acima de 70% em 60% dos casos (mil a seis mil estudantes). Nos casos de interações polarizadas, o AG alcançou $\approx 82\%$ nos melhores resultados, no entanto ficou com percentual acima de 70% em apenas 20% dos casos (mil e dois mil estudantes). Além disso, o pior resultado ocorreu para interações polarizadas, com peso maior para as interações, com dez mil estudantes, onde o AG alcançou apenas $\approx 44\%$ de grupos heterogêneos. No entanto, mesmo para o pior resultado, o AG gerou $\approx 10\%$ a mais de grupos heterogêneos do que o algoritmo randômico. Para as interações aleatórias, os percentuais de grupos heterogêneos variaram de $\approx 45\%$ a $\approx 68\%$, ficando abaixo de 70% em todos os casos. No entanto, mesmo para o pior resultado, $\approx 45\%$, com peso maior para as interações, o AG alcançou $\approx 11\%$ a mais de grupos heterogêneos do que o algoritmo randômico. Assim, conclui-se que, de forma geral, o AG obteve resultados significativamente melhores que os resultados obtidos pelo algoritmo randômico, uma vez que as médias apresentadas pelo AG foram maiores que as médias do algoritmo randômico, tanto para o *fitness* quanto para o percentual de grupos heterogêneos.

5.5.4 Tempo de Processamento

Em relação ao tempo de processamento dos algoritmos, o esperado era que o algoritmo randômico fosse mais rápido que o AG. O AG efetua mais operações (como *crossover* e mutação, por exemplo) que o algoritmo randômico, e por isso mesmo consome mais tempo de processamento. Essa expectativa se confirmou em aproximadamente 90% dos casos; em aproximadamente 10% dos casos (29 dos 270 casos) o AG foi mais rápido que o algoritmo randômico. Na Figura 61 (Tabela 28 – Apêndice B), é exemplificado um caso ocorrido, para o primeiro caso de testes, com conhecimentos heterogêneos, interações uniformes e pesos iguais. O gráfico mostra a evolução do tempo (em segundos) para cada quantidade de estudantes (de mil até dez mil).

Figura 61 – Tempo de processamento dos algoritmos (primeiro caso de teste)



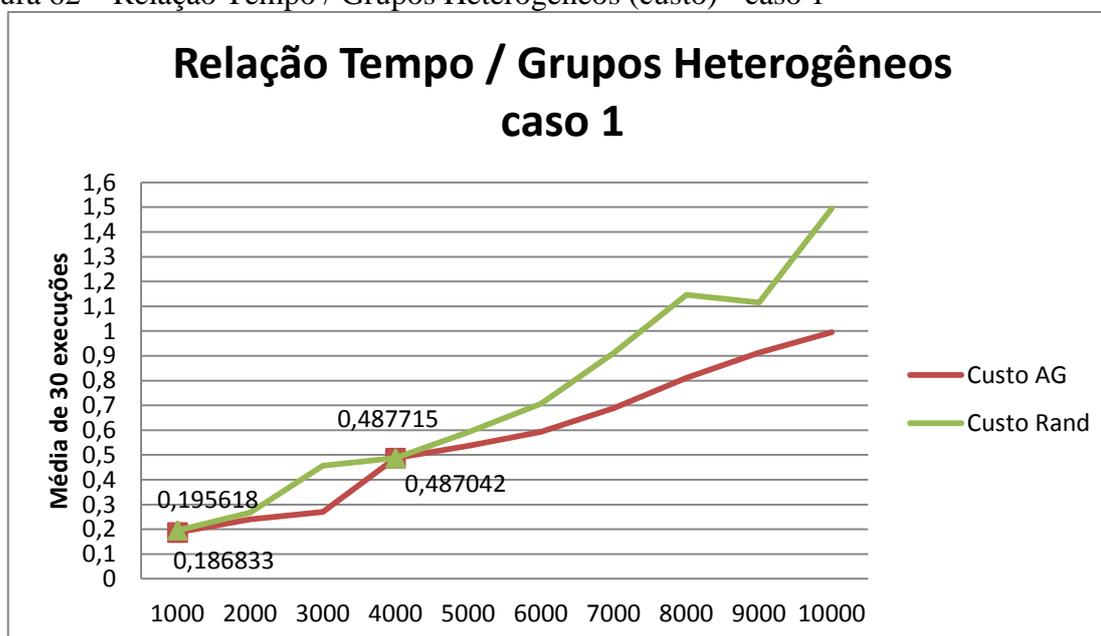
Fonte: Próprio autor

Na Figura 61, observa-se que o tempo de processamento aumenta à medida que o número de estudantes aumenta, pois isso requer um custo computacional maior devido a maior quantidade de grupos para formar e avaliar (*fitness*). O caso com três mil estudantes exemplifica a situação em que o AG foi mais rápido que o algoritmo randômico. Ainda na Figura 61, observa-se que em alguns casos o tempo do AG e do algoritmo randômico foram bem semelhantes como, por exemplo, no caso de dez mil estudantes.

Por isso a análise pura e simples em relação ao tempo de processamento não é adequada, pois já se sabe que, em geral, o AG leva mais tempo que o algoritmo randômico. Assim, a análise foi feita em relação ao custo/benefício, isto é, a relação entre tempo de processamento dividido pela quantidade de grupos heterogêneos formados. Essa relação permitiu avaliar que o AG, apesar de levar mais tempo, formou proporcionalmente mais grupos heterogêneos do que o algoritmo randômico. Exceto nos casos dos conhecimentos homogêneos, onde o percentual de grupos heterogêneos deve tender a zero (significando melhor resultado), nesse caso, essa relação foi feita em relação à quantidade de grupos balanceados, isto é, o percentual considerado foi 1 (um) menos o percentual de grupos heterogêneos gerados. O objetivo era minimizar essa relação, isto é, quanto menor o tempo gasto para formar os grupos, melhor. Assim, essa relação foi a medida da eficiência do algoritmo, denominada a seguir apenas como custo.

Nas figuras a seguir, são apresentados os resultados de alguns casos de testes, considerando a média das 30 execuções. Os gráficos apresentam o custo para cada quantidade de estudantes (de mil até dez mil).

Figura 62 – Relação Tempo / Grupos Heterogêneos (custo) - caso 1



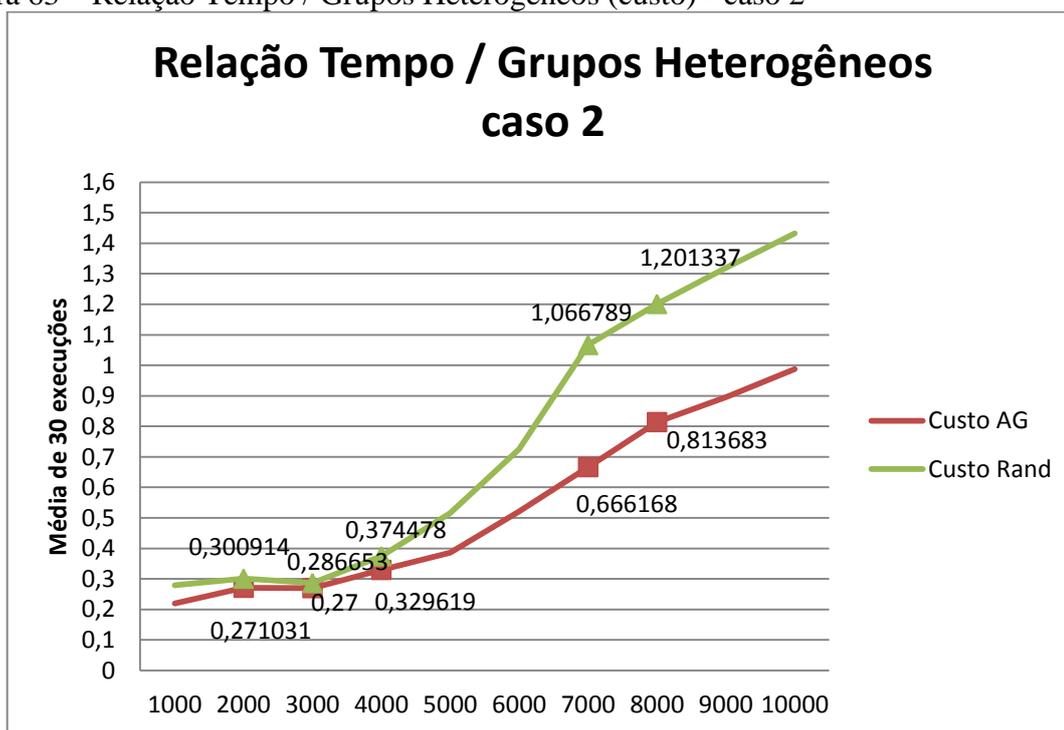
Fonte: Próprio autor

Na Figura 62 (Tabela 29 – Apêndice B), é apresentado o primeiro caso, para **conhecimentos heterogêneos, interações uniformes e pesos iguais**. Esse caso representa os conhecimentos heterogêneos, pois os resultados alcançados pelo AG e os resultados alcançados pelo algoritmo randômico foram semelhantes em relação às diferenças ocorridas para os percentuais de grupos heterogêneos, conforme discutido na seção 5.5.1. Na Figura 62,

observa-se que o custo do AG é geralmente menor que o custo do algoritmo randômico. Para o pior resultado (quatro mil estudantes), o AG obteve pelo menos um resultado igual ao do algoritmo randômico em termos de custo ($\approx 0,487$). No entanto, considerando-se esse caso, o AG gerou $\approx 33\%$ de grupos heterogêneos a mais que o algoritmo randômico, o que significa aproximadamente 264 grupos heterogêneos a mais gerados pelo AG. Para o segundo pior resultado (mil estudantes), apesar do custo próximo, o custo do algoritmo randômico é $\approx 5\%$ maior que o custo do AG ($AG - 0,186833 \times 0,195618 - Rand$); e o AG gerou $\approx 26\%$ de grupos heterogêneos a mais, o que significa aproximadamente 52 grupos heterogêneos a mais gerados pelo AG.

Para os casos de conhecimentos heterogêneos com peso maior para as interações, quando as interações eram polarizadas e/ou aleatórias, os resultados apresentados na seção 5.5.1 foram diferentes dos demais. Por isso, nas Figuras 63 e 64 (Tabelas 30 e 31 – Apêndice B) são apresentados esses casos.

Figura 63 – Relação Tempo / Grupos Heterogêneos (custo) - caso 2



Fonte: Próprio autor

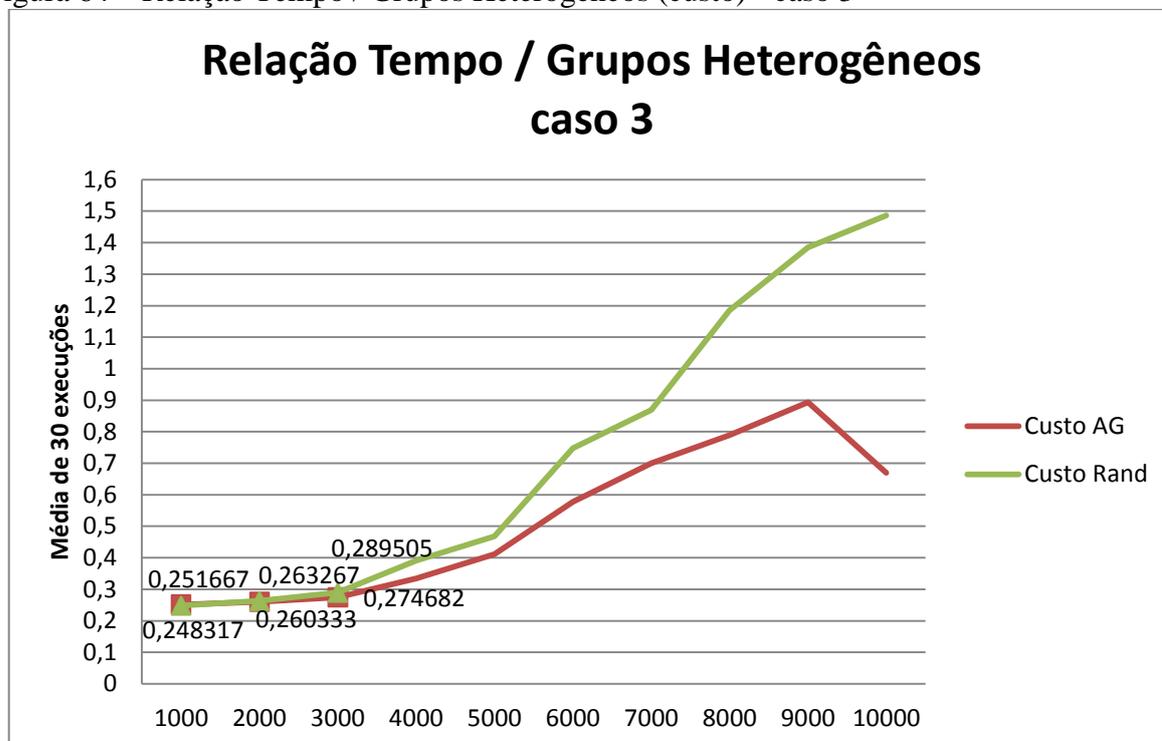
No caso da Figura 63 (caso 2 - interações polarizadas) observa-se que o custo do AG foi sempre menor que o do algoritmo randômico. Mesmo quando o AG obteve um percentual de grupos heterogêneos menor, entre $\approx 90\%$ a $\approx 97\%$ para os quatro casos em que os percentuais foram menores que 99,9% (dois mil, quatro mil, sete mil e oito mil estudantes), o AG obteve melhores resultados para o custo. Isto é, apesar de ter formado menos grupos

heterogêneos que os demais resultados, a relação em função do tempo ainda é melhor do que a relação custo/benefício do algoritmo randômico.

Ainda na Figura 63, para o pior resultado (com três mil estudantes), onde ocorre a menor diferença, apesar do custo ter valor próximo, o custo do algoritmo randômico é $\approx 6\%$ maior que o custo do AG (AG - $0,27 \times 0,286653$ – Rand); e o AG gerou $\approx 35\%$ de grupos heterogêneos a mais, o que significa aproximadamente 210 grupos heterogêneos a mais gerados pelo AG. Para os casos em que o AG apresentou menor percentual de grupos heterogêneos, para dois mil estudantes, o custo do algoritmo randômico é $\approx 11\%$ maior que o custo do AG; e o AG gerou $\approx 25\%$ de grupos heterogêneos a mais, o que significa aproximadamente 100 grupos heterogêneos a mais gerados pelo AG. Para quatro mil estudantes, o custo do algoritmo randômico é $\approx 13\%$ maior que o custo do AG; e o AG gerou $\approx 32\%$ de grupos heterogêneos a mais, o que significa aproximadamente 256 grupos heterogêneos a mais gerados pelo AG. Para os casos de sete mil e oito mil estudantes, o custo do algoritmo randômico é, respectivamente, $\approx 60\%$ e $\approx 47\%$ maior que o custo do AG.

Na Figura 64 (caso 3), são apresentados os dados para interações aleatórias, que são discutidos a seguir.

Figura 64 – Relação Tempo / Grupos Heterogêneos (custo) - caso 3



Fonte: Próprio autor

No caso da Figura 64, observa-se que para o caso de mil estudantes o AG obteve um resultado pior em relação ao custo (AG – $0,251667 \times 0,248317$ – Rand). Entretanto, é um

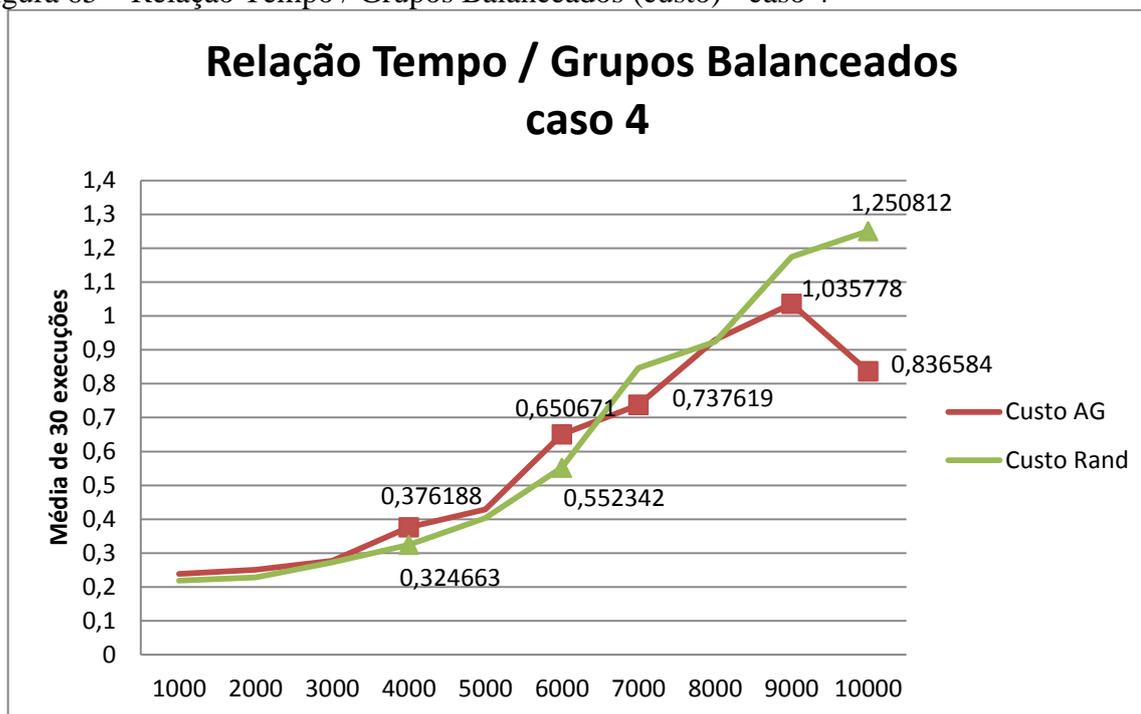
valor próximo ao do algoritmo randômico. Considerando-se esse caso, o AG gerou $\approx 28\%$ de grupos heterogêneos a mais que o algoritmo randômico, o que significa aproximadamente 56 grupos heterogêneos a mais gerados pelo AG, praticamente ao mesmo custo do algoritmo randômico. Nos demais casos o AG obteve melhores resultados para o custo. Mesmo quando essa diferença é pequena, no caso de dois mil estudantes, o custo do algoritmo randômico é $\approx 1\%$ maior que o custo do AG ($AG - 0,260333 \times 0,263267 - Rand$); e o AG gerou $\approx 31\%$ de grupos heterogêneos a mais que o algoritmo randômico, o que significa aproximadamente 62 grupos heterogêneos a mais gerados pelo AG. Para o segundo pior resultado, caso de três mil estudantes, o custo do algoritmo randômico é $\approx 5\%$ maior que o custo do AG ($AG - 0,274682 \times 0,289505 - Rand$); e o AG gerou $\approx 33\%$ de grupos heterogêneos a mais que o algoritmo randômico, o que significa aproximadamente 198 grupos heterogêneos a mais gerados pelo AG.

Nas análises em relação aos conhecimentos heterogêneos, o AG obteve um custo melhor que o algoritmo randômico em $\approx 97\%$ dos casos. Em $\approx 3\%$ dos casos o AG obteve um custo igual ao do algoritmo randômico. No entanto, o percentual de grupos heterogêneos alcançado pelo AG foi sempre maior que o percentual de grupos heterogêneos alcançado pelo algoritmo randômico, com uma diferença mínima de $\approx 25\%$ de grupos heterogêneos a mais.

No caso de **conhecimentos homogêneos**, os resultados apresentados na seção 5.5.2 foram bem semelhantes entre si. Destacam-se nesse caso, os resultados para interações aleatórias, com peso maior para as interações, onde houve a menor diferença entre os percentuais de grupos heterogêneos, de $\approx 13\%$ (dez mil estudantes). Na Figura 65 (caso 4, Tabela 32 – Apêndice B), são apresentados os resultados da relação custo/benefício para esse caso, que é em relação aos grupos balanceados.

Observa-se, na Figura 65, que o custo do AG é maior que o custo do algoritmo randômico na maioria dos casos. O pior caso ocorre para seis mil estudantes, onde a diferença é a maior. Nesse caso, o custo do AG é $\approx 18\%$ maior que o custo do algoritmo randômico ($AG - 0,650671 \times 0,552342 - Rand$); e o AG gerou apenas $\approx 15\%$ de grupos balanceados a mais que o algoritmo randômico. Para quatro mil estudantes, onde a diferença é a segunda maior, o custo do AG é $\approx 16\%$ maior que o custo do algoritmo randômico ($AG - 0,376188 \times 0,324663 - Rand$). No entanto, o AG gerou $\approx 17\%$ de grupos balanceados a mais que o algoritmo randômico. Nos demais casos, apesar do custo do AG ser maior que o custo do algoritmo randômico, a diferença percentual do custo do AG (em relação ao custo do algoritmo randômico) não é superior à diferença percentual de grupos balanceados gerados a mais pelo AG (em relação ao que foi gerado pelo algoritmo randômico).

Figura 65 – Relação Tempo / Grupos Balanceados (custo) - caso 4



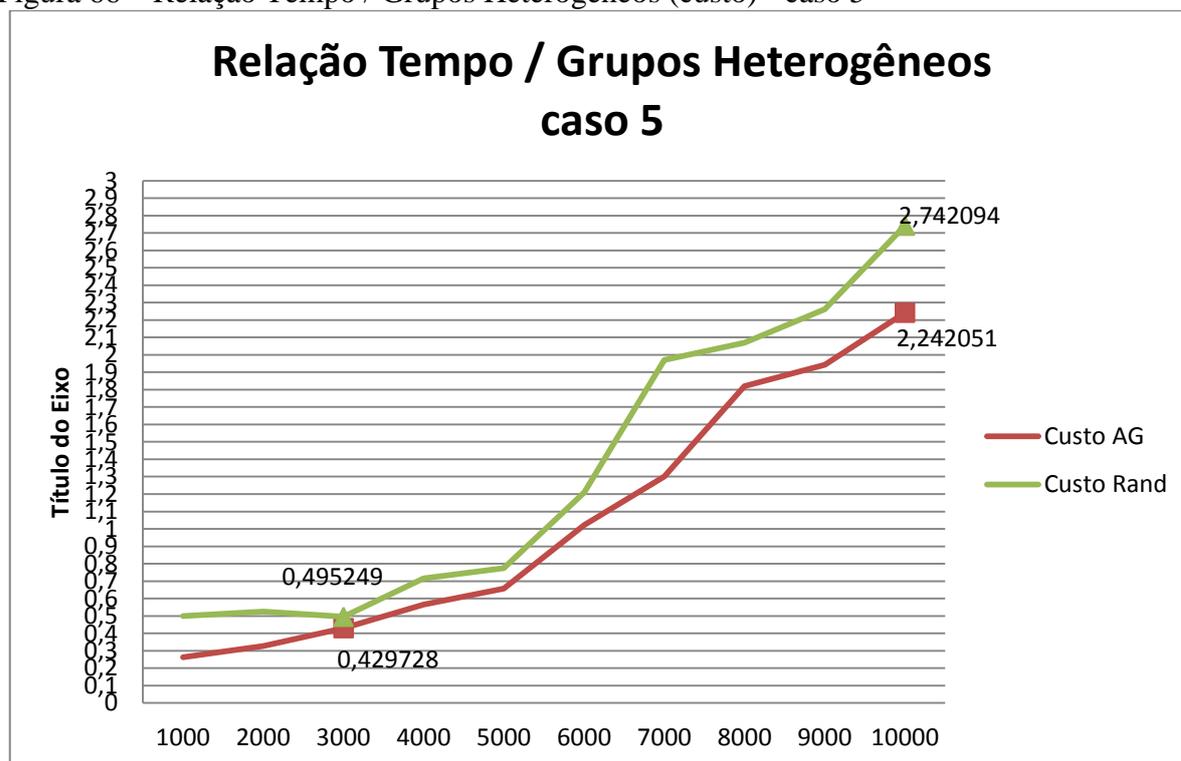
Fonte: Próprio autor

Ainda na Figura 65, em três casos (sete mil, nove mil e dez mil estudantes) o AG obteve melhor resultado em relação ao custo do que o algoritmo randômico. Analisando o caso mencionado, de dez mil estudantes, em que ocorreu o pior percentual de grupos heterogêneos ($\approx 11\%$ - e nesse caso deveria tender a zero) e a menor diferença entre o resultado do AG e do algoritmo randômico, observa-se que o custo do algoritmo randômico é $\approx 50\%$ maior que o custo do AG ($AG = 0,836584 \times 1,250812 - Rand$). Além disso, o AG gerou $\approx 13\%$ de grupos balanceados a mais que o algoritmo randômico, o que significa aproximadamente 260 grupos balanceados a mais gerados pelo AG. Observa-se ainda nesse caso que este é um daqueles casos em que o AG foi mais rápido que o algoritmo randômico (dez mil estudantes), levando menos tempo para gerar mais grupos balanceados. Por isso a linha no gráfico apresenta uma queda mais acentuada..

Nas análises em relação aos conhecimentos homogêneos, o AG obteve um custo melhor que o algoritmo randômico em $\approx 63\%$ dos casos. Em $\approx 37\%$ dos casos, o AG obteve um custo pior do que o custo do algoritmo randômico. Desses, em $\approx 94\%$ o percentual de grupos balanceados gerados a mais pelo AG superou o percentual do custo do AG sobre o algoritmo randômico. Em dois desses casos, o percentual de grupos balanceados gerados a mais pelo AG não superou o percentual do custo do AG sobre o algoritmo randômico, o que representa $\approx 2\%$ do total de casos (2 dos 90 casos).

No caso de **conhecimentos aleatórios**, os resultados apresentados na seção 5.5.3 foram bem variados. Destacam-se nesses casos, os resultados para interações polarizadas e interações aleatórias, com peso maior para as interações, onde houve os piores resultados do AG para o percentual de grupos heterogêneos, respectivamente, $\approx 44\%$ e $\approx 45\%$. Também com as menores diferenças ocorridas entre o AG e o algoritmo randômico, respectivamente, $\approx 10\%$ e $\approx 11\%$. Na Figura 66 (caso 5, Tabela 33 – Apêndice B), são apresentados os resultados para interações polarizadas.

Figura 66 – Relação Tempo / Grupos Heterogêneos (custo) - caso 5

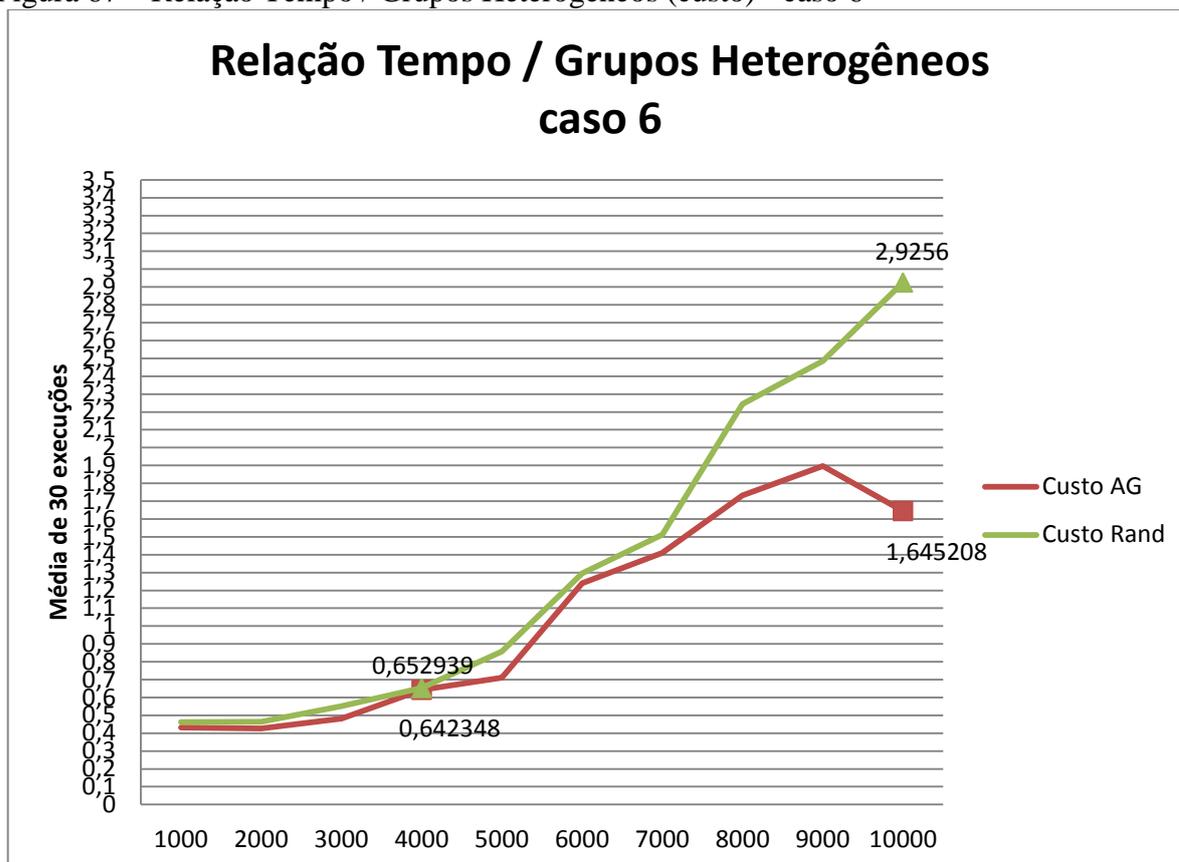


Fonte: Próprio autor

Observa-se, na Figura 66, que em todos os casos o AG obteve melhor resultado para o custo. O pior caso ocorre para três mil estudantes, onde a diferença é menor. Nesse caso, o custo do algoritmo randômico é $\approx 15\%$ maior que o custo do AG ($AG - 0,429728 \times 0,495249 - Rand$); e o AG gerou $\approx 29\%$ de grupos heterogêneos a mais que o algoritmo randômico, o que significa aproximadamente 174 grupos heterogêneos a mais gerados pelo AG. No caso com dez mil estudantes, onde ocorreu o menor percentual de grupos heterogêneos ($\approx 44\%$) e a menor diferença entre os resultados do AG e do algoritmo randômico o custo do algoritmo randômico é $\approx 22\%$ maior que o custo do AG ($AG - 2,242051 \times 2,742094 - Rand$); e o AG gerou $\approx 10\%$ de grupos heterogêneos a mais que o algoritmo randômico, o que significa aproximadamente 200 grupos heterogêneos a mais gerados pelo AG.

Na Figura 67 (caso 6, Tabela 34 – Apêndice B), é apresentado o caso com interações aleatórias. Observa-se que em todos os casos o AG obteve melhor resultado para o custo. O pior caso ocorre para quatro mil estudantes, onde a diferença é menor. Nesse caso, o custo do algoritmo randômico é $\approx 2\%$ maior que o custo do AG ($AG - 0,642348 \times 0,652939 - Rand$); e o AG gerou $\approx 17\%$ de grupos heterogêneos a mais que o algoritmo randômico, o que significa aproximadamente 135 grupos heterogêneos a mais gerados pelo AG.

Figura 67 – Relação Tempo / Grupos Heterogêneos (custo) - caso 6



Fonte: Próprio autor

Ainda na Figura 67, no caso com dez mil estudantes, onde ocorreu o menor percentual de grupos heterogêneos ($\approx 45\%$) e a menor diferença entre os resultados do AG e do algoritmo randômico, o custo do algoritmo randômico é $\approx 78\%$ maior que o custo do AG ($AG - 1,645208 \times 2,9256 - Rand$); e o AG gerou $\approx 11\%$ de grupos heterogêneos a mais que o algoritmo randômico, o que significa aproximadamente 220 grupos heterogêneos a mais gerados pelo AG.

Nas análises em relação aos conhecimentos aleatórios, o AG obteve um custo melhor que o algoritmo randômico em 100% dos casos. Em todos os casos, o percentual de grupos heterogêneos alcançado pelo AG foi maior que o percentual de grupos heterogêneos

alcançado pelo algoritmo randômico, com uma diferença mínima de $\approx 10\%$ de grupos heterogêneos a mais.

Foi aplicado o *Wilcoxon test* para a relação custo/benefício a fim de avaliar a significância das diferenças ocorridas entre os custos do AG e os custos do algoritmo randômico. Na Tabela 20, são apresentados os resultados dos testes para, respectivamente, conhecimentos heterogêneos, conhecimentos homogêneos e conhecimentos aleatórios.

Tabela 20 – *Wilcoxon test* para a relação custo/benefício (tempo/grupos)

Conhecimentos	<i>Wilcoxon test</i> <i>p-valor</i>
Heterogêneos	0.004281
Homogêneos	0.379
Aleatórios	0.0001426

Fonte: Próprio autor

Nos casos de conhecimentos heterogêneos e conhecimentos aleatórios, a hipótese nula foi rejeitada com *p-valor* muito menor que o nível de significância estabelecido ($p\text{-valor} \ll \alpha = 0.05$). Assim, para esses casos, conclui-se que o AG foi mais eficiente que o algoritmo randômico, gerando mais grupos heterogêneos a um custo menor ou igual ao custo do algoritmo randômico.

No caso de conhecimentos homogêneos, a hipótese nula é verdadeira, com *p-valor* maior que o nível de significância estabelecido ($p\text{-valor} > \alpha = 0.05$). Assim, para esse caso, conclui-se que não há diferença significativa entre os custos do AG e do algoritmo randômico.

5.6 CONSIDERAÇÕES DO CAPÍTULO

Nesse capítulo foram apresentadas as validações dos algoritmos desenvolvidos, tanto do AG como do algoritmo randômico, como método base de comparação. A validação propôs uma série com 270 casos de testes, divididos em três seções em relação aos conhecimentos: heterogêneos, homogêneos e aleatórios. Foram apresentados os resultados para esses casos de testes, que foram analisados para avaliar a significância estatística dos dados.

Para o *fitness* e para o percentual de grupos heterogêneos, em todos os casos a hipótese nula foi rejeitada com *p-valor* muito menor que o nível de significância estabelecido ($p\text{-valor} \ll \alpha = 0.05$); e assim, conclui-se que os resultados do AG foram significativamente melhores que os resultados do algoritmo randômico.

A eficácia está geralmente associada à capacidade do programa de alcançar os objetivos estabelecidos, neste caso, a formação de grupos heterogêneos na aprendizagem colaborativa, considerando uma quantidade massiva de estudantes. Como o AG obteve resultados melhores em relação ao *fitness* e ao percentual de grupos heterogêneos em todos os

casos, conclui-se que o **AG é um recurso eficaz para a formação de grupos heterogêneos na aprendizagem colaborativa considerando um contexto massivo de dados (milhares de estudantes)**.

De forma geral, a eficiência está associada aos recursos necessários para o programa atingir os objetivos, neste caso a relação custo/benefício (em função do tempo). Nos testes em relação ao tempo de processamento, foi feita a análise da relação custo/benefício, considerando o tempo gasto para formar os grupos em relação à quantidade de grupos formados adequadamente. O AG obteve melhores resultados que o algoritmo randômico, conclui-se, assim, que o AG foi mais eficiente que o algoritmo randômico, gerando mais grupos heterogêneos que o algoritmo randômico a um custo menor ou no mínimo igual a esse, para os casos de conhecimentos heterogêneos e/ou aleatórios. No caso de conhecimentos homogêneos, não houve diferença significativa nos resultados alcançados pelo AG. Assim, para esse caso, mais testes são necessários para que novas análises possam ser feitas.

Conclui-se que o **AG é mais eficiente** que o algoritmo randômico **para conhecimentos heterogêneos e aleatórios. Para conhecimentos homogêneos apresenta a mesma eficiência**, no entanto, é mais eficaz, pois gera mais grupos balanceados.

Em relação aos pesos aplicados, observou-se que a aplicação do peso para os conhecimentos não influenciou o comportamento do AG, pois seus resultados foram semelhantes aos resultados para pesos iguais. Mais testes são necessários para que novas análises possam ser feitas. Já para o caso das interações, a aplicação do peso influenciou os resultados, conforme discutido. Assim, em um caso real, se o professor desejar que as interações entre os estudantes tenham uma influência maior na formação dos grupos, isso poderá ser feito utilizando-se o peso maior para as interações.

6 CONCLUSÃO

A colaboração tem um papel importante para a construção do conhecimento. No entanto, a colaboração por si só não produz resultados na aprendizagem, ela depende da efetiva interação dos envolvidos em um grupo. Por isso, a formação de grupos é importante para a aprendizagem colaborativa. A formação de grupos parece uma tarefa simples, mas torna-se complexa, à medida que cresce o número de estudantes e/ou as características definidas para o agrupamento. A explosão combinatória relativa ao número de estudantes torna o problema da formação de grupos um problema de otimização combinatória, que é *NP-hard*, isto é, um problema que não tem solução ótima em tempo polinomial. Junto a isso, um novo paradigma se apresenta no âmbito educacional, o dos *Massive Open Online Courses* (MOOCs), em que uma quantidade massiva de estudantes pode se inscrever para cursos abertos, disponíveis na Internet.

Por isso, nesta pesquisa, o problema tratado foi como formar grupos para aprendizagem colaborativa, considerando uma quantidade massiva de estudantes (milhares de estudantes). No desenvolvimento dos estudos, observou-se que o uso de algoritmos genéticos tem provado ser eficiente na resolução de problemas complexos de otimização. Assim, o objetivo geral desta pesquisa foi avaliar a adoção de algoritmos genéticos para a formação de grupos na aprendizagem colaborativa, considerando um contexto massivo de dados (milhares de estudantes).

Da investigação sobre a formação de grupos para aprendizagem colaborativa destacam-se os resultados obtidos no mapeamento sistemático, adotados nesta pesquisa para a formação de grupos de *forma automática*, utilizando a *abordagem selecionada* e como *critério* grupos **heterogêneos**, utilizando **algoritmos genéticos (AG)**. A fim de simplificar a formação de grupos, posto que já se considerou uma quantidade massiva de estudantes, esta pesquisa utilizou os **conhecimentos** e as **interações** dos estudantes como características para a formação de grupos. Como parte do problema, considerou-se que a heterogeneidade e as interações nos grupos deviam ser maximizadas, com o objetivo de gerar grupos **heterogêneos** em relação aos **conhecimentos** e com o **maior número de interações** possível entre os membros do grupo, gerando **ainda grupos balanceados** entre si (**inter-homogêneos**).

Dos trabalhos relacionados a esta pesquisa, seis deles passaram pela revisão sistemática, e foram descritos em detalhes no capítulo 3. Todos os trabalhos contribuíram direta ou indiretamente para as definições adotadas nesta pesquisa como, por exemplo, o uso da distância Euclidiana como medida de similaridade (GOGOULOU et al., 2007) ou as visões

de grupos balanceados (ANI et al., 2010; MORENO; OVALLE; VICARI, 2012; WANG, LIN; SUN, 2007). O trabalho de Yannibelli e Amandi (2012), em especial, destacou-se em relação às definições adotadas para o AG. As autoras utilizaram, para a representação dos indivíduos no AG, a representação inteira por permutação. Essa representação foi considerada a mais adequada ao problema (que é uma permutação dos estudantes em grupos). Além disso, as autoras validaram o AG proposto com uma quantidade maior de estudantes (milhares) e obtiveram bons resultados em comparação com um método base de comparação (randômico). Assim, para o AG definido nesta pesquisa, a representação adotada foi a inteira por **permutação**; o que influenciou na escolha de operadores genéticos, como o **crossover de ordem** para a recombinação e o **swap mutation** para a mutação, compatíveis com essa representação e que também foram utilizados por Yannibelli e Amandi (2012). Ainda nos trabalhos relacionados e no mapeamento sistemático realizado, observou-se uma variedade de características adotadas para a formação dos grupos. Yannibelli e Amandi (2012) utilizaram papéis preferidos pelos estudantes, por exemplo. Já Wang, Lin e Sun (2007) utilizaram variáveis psicológicas, como estilo de aprendizagem. Alguns trabalhos utilizaram, inclusive, várias características (GOGOULOU et al., 2007; MORENO; OVALLE; VICARI, 2012). No entanto, estilo de aprendizagem e papéis não são características facilmente encontradas em ambientes virtuais de aprendizagem como o Moodle ou na plataforma *EdX*, uma plataforma MOOC *open source*, por exemplo. Assim, esta pesquisa verificou a disponibilidade das características utilizadas, sendo que **conhecimentos** e **interações** são dados geralmente disponíveis nos ambientes virtuais de aprendizagem. Além disso, destaca-se o fato de considerar as interações, que têm um papel importante na aprendizagem colaborativa.

Dois algoritmos foram desenvolvidos para as devidas comparações: **AG** e **randômico** (método base de comparação). Os algoritmos foram desenvolvidos na linguagem C e testados na nuvem da UDESC com quantidades massivas de estudantes (milhares de estudantes).

Para avaliar os algoritmos desenvolvidos, 270 casos de testes com dados simulados foram executados. Os resultados foram testados estatisticamente e mostraram que o AG foi mais eficaz que o algoritmo randômico em todos os casos, com diferenças significativas. Em relação à eficiência dos algoritmos, o **AG foi mais eficiente** que o algoritmo randômico **para conhecimentos heterogêneos e aleatórios**; **para conhecimentos homogêneos**, foi eficaz, mas apresentou a **mesma eficiência que o algoritmo randômico**. Assim, para conhecimentos homogêneos, mais testes são necessários para que novas análises possam ser feitas.

Esta dissertação contribui com respostas iniciais para a formação de grupos na aprendizagem colaborativa num contexto massivo de estudantes, de forma a permitir o

desenvolvimento de trabalhos futuros nesta temática. Além dos resultados apresentados, dos quais se pode apreender que **é eficaz utilizar algoritmos genéticos para a formação de grupos na aprendizagem colaborativa, considerando-se milhares de estudantes**, a dissertação contribui com o **modelo matemático** gerado para a função de avaliação (*fitness*), que considera as características **conhecimentos e interações** na formação do grupo, além de gerar grupos balanceados entre si (**inter-homogêneos**). Considerando-se que essas características estão geralmente disponíveis nos ambientes virtuais de aprendizagem, e a importância das interações para a aprendizagem colaborativa, este modelo pode ser adotado por diferentes pesquisadores interessados em formação de grupos para a aprendizagem colaborativa.

Trabalhos foram publicados, decorrentes da realização das pesquisas e entendimentos iniciais até delimitar o objetivo final da pesquisa, além do mapeamento sistemático descrito no capítulo 2. Esses trabalhos estão citados no Apêndice A.

Como trabalhos futuros, várias possibilidades se apresentam, visto que esta é uma contribuição inicial na temática (contexto massivo), entre elas:

- Realizar mais experimentos utilizando diferentes configurações dos parâmetros (que foram escolhidos empiricamente) e analisar se alguma combinação de parâmetros apresenta melhores resultados. Essa configuração dos parâmetros pode ser feita apenas escolhendo-se empiricamente outros parâmetros como, por exemplo, utilizar 90% para o *crossover* e/ou 5% para a mutação, ou pode ser feita automaticamente pelo algoritmo (controle automático de parâmetros), o que poderia trazer algum benefício para o resultado final do algoritmo (ANDRÉ; PARPINELLI, 2014);
- Utilizar outro método para a seleção de pais e outros operadores genéticos (*crossover* e mutação), e analisar se diferentes métodos de seleção e/ou outros operadores genéticos apresentam outros resultados. As definições adotadas nessa pesquisa basearam-se especialmente em Yannibelli e Amandi (2012), que apresentaram bons resultados para uma quantidade maior de estudantes, já na casa dos milhares. No entanto, a comparação de resultados com diferentes operadores genéticos pode contribuir com escolhas mais adequadas para o problema em questão como, por exemplo, se utilizar o ciclo *crossover* tem-se melhores resultados que o *crossover* de ordem na formação de grupos em um contexto com milhares de estudantes?;

- Paralelizar (multi-thread) os algoritmos, e avaliar os impactos na eficiência (tempo de processamento). Essa alteração pode contribuir para que a eficiência do AG para conhecimentos homogêneos seja melhor, por exemplo. Se o algoritmo for paralelo, poderá se beneficiar dos recursos disponíveis para realizar mais tarefas em menos tempo;
- Permitir diferentes critérios para a formação de grupos (homogêneos, por exemplo) e avaliar os resultados. Alguns trabalhos apontados pelo mapeamento utilizavam grupos homogêneos, inclusive o trabalho revisado de Gogoulou et al. (2007), permite a formação de grupos homogêneos também. Permitir diferentes critérios para a formação de grupos pode contribuir para uma investigação mais profunda sobre o impacto de se utilizar diferentes formações de grupos (heterogêneos x homogêneos) na aprendizagem colaborativa, por exemplo;
- Aplicar a formação de grupos numa situação real de aprendizagem e avaliar os desdobramentos das diferentes formações de grupo (heterogêneos x randômico ou heterogêneos x homogêneos) nos resultados da aprendizagem e na satisfação dos estudantes. Implantar a plataforma *EdX*, por exemplo, e aplicar a formação de grupos do AG em um curso real, permitiria avaliar os resultados na aprendizagem e na satisfação dos estudantes (conforme feito por Moreno, Ovalle e Vicari (2012) e Wang, Lin e Sun (2007)). Essas comparações poderiam contribuir com resultados interessantes num contexto massivo de estudantes, já que MOOCs ainda é um assunto novo e carente de pesquisas.

Assim, os resultados são promissores e muitas possibilidades se abrem ante esse primeiro passo de contribuição na pesquisa sobre formação de grupos na aprendizagem colaborativa para um contexto massivo de estudantes.

REFERÊNCIAS

- ABNAR, S.; OROOJI, F.; TAGHIYAREH, F. An evolutionary algorithm for forming mixed groups of learners in web based collaborative learning environments. In: **Technology Enhanced Education (ICTEE), 2012 IEEE International Conference on**. IEEE, p. 1-6, 2012.
- ALAVI, M. Computer-mediated collaborative learning: An empirical evaluation. In: **MIS quarterly**, p. 159-174, 1994.
- ALVES, I. F.; KEMCZINSKI, A.; HOUNSELL, M. da S.; GASPARINI, I. Diagnóstico da Avaliação de Aprendizagem em Ambientes E-learning. In: **XIX Simpósio Brasileiro De Informática Na Educação (Sbie), Fortaleza. Tecnologia E Educação Para Todos. Fortaleza: Sociedade Brasileira de Computação**, 2008.
- ANDRÉ, L.; PARPINELLI, R. S. Tutorial Sobre o Uso de Técnicas para Controle de Parâmetros em Algoritmos de Inteligência de Enxame e Computação Evolutiva. In: **Revista de Informática Teórica e Aplicada**, v. 21, n. 2, p. 83-128, 2014.
- ANI, Z. C.; YASIN, A.; HUSIN, M. Z.; HAMID, Z. A. A method for group formation using genetic algorithm. In: **International Journal on Computer Science and Engineering**, v. 2, n. 9, p. 3060-3064, 2010.
- APARICIO, M.; BACAO, F. E-learning concept trends. In: **Proceedings of the 2013 International Conference on Information Systems and Design of Communication**. ACM, p. 81-86, 2013.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS (ABNT). **NBR ISO/IEC 9126-1: Engenharia de Software – Qualidade de Produto. Parte 1 – Modelo de Qualidade**, 2003.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS (ABNT). **NBR ISO/IEC 9126-4: Engenharia de Software – Qualidade de Produto. Parte 4 – Qualidade em uso**, 2004.
- BARKLEY, E.; CROSS, K. P; MAJOR, C. H. **Collaborative Learning Techniques: A Practical Guide to Promoting Learning in Groups**. San Francisco, CL: Jossey Bass, 2005.
- BEKELE, R. **Computer-Assisted Learner Group Formation Based on Personality Traits**. PhD thesis, Universität Hamburg, 2005.
- BELBIN, R. M. **Management teams: Why they succeed or fail**. Oxford: Butterworth-Heinemann, 1981.
- BELBIN, R. M. **Team roles at work**. Oxford: Butterworth-Heinemann, 1993.
- BENBUNAN-FICH; R., HILTZ; S. R.; TUROFF, M. A comparative content analysis of face-to-face vs. asynchronous group decision making. In: **Decision Support Systems**, v. 34, n. 4, p.457–469, 2003.
- BROPHY, D. Understanding, measuring, and enhancing collective creative problem-solving efforts. In: **Creative Research Journal**, v. 3, p.199–299, 1998.

CASTELLS, M.; GERHARDT, K. B. **A sociedade em rede**. São Paulo: Paz e Terra, 2000.

CASTILLO, L. M.; CAZARINI, E. W. Conceitos importantes na gestão do conhecimento e a Taxonomia do conhecimento. In: **Simpósio de Administração da Produção, Logística e Operações Internacionais SIMPOI**, 2010.

CASTRO, A.; MENEZES, C. Aprendizagem Colaborativa com Suporte Computacional. In: PIMENTEL, M.; FUKS, H. (orgs). **Sistemas Colaborativos**. Rio de Janeiro: Elsevier, 2011. p. 135-153.

CHAN, K. W. Issues of heterogeneous grouping for engaged learning. In: **APER A Conference 2006**, 2006.

CITADIN, J. R.; KEMCZINSKI, A.; DE MATOS, A. V. Formação de Grupos para Aprendizagem Colaborativa: Um mapeamento sistemático da literatura. In: **Nuevas Ideas en Informática Educativa TISE 2014**, p.46-54, 2014.

CORMEN, T. H.; LEISERSON, C. E.; RIVEST, R. L.; STEIN, C. **Introduction to Algorithms**, 2 ed. MIT Press. Cambridge, MA, 2011.

CRUZ, W. M.; ISOTANI, S. Group Formation Algorithms in Collaborative Learning Contexts: A Systematic Mapping of the Literature. In: **Collaboration and Technology**. Springer International Publishing, p. 199-214, 2014.

DANIEL, J.S. Making Sense of MOOCs: Musings in a Maze of Myth, Paradox and Possibility. In: **Journal of Interactive Media in Education (JIME)**, 2012.

DARWIN, C. **On the Origin of the Species by Means of Natural Selection: Or, The Preservation of Favoured Races in the Struggle for Life**. John Murray, 1859.

DE PAIVA, I. A.; PERNAMBUCO, M. M. C. A. **Educação e realidade: interdisciplinar**. Universidade Federal do Rio Grande do Norte, 2005.

DERRAC, J; GARCIA, S; MOLINA, D.; HERRERA, F. A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. **Swarm and Evolutionary Computation**, v. 1, n. 1, p. 3-18, 2011.

DEVORE, J.L. Probabilidade e estatística: para engenharia e ciências. São Paulo: Cengage Learning, 2012.

DILLENBOURG, P. What do you mean by collaborative learning? In: **Collaborative-learning: Cognitive and computational approaches**. Elsevier, p.1-19, 1999.

DILLENBOURG, P. Over-scripting CSCL: The risks of blending collaborative learning with instructional design. In: **Three worlds of CSCL. Can we support CSCL?**. Heerlen: Open University Nederland, p.61-91, 2002.

DILLENBOURG, P.; JÄRVELÄ, S.; FISCHER, F. The evolution of research on computer-supported collaborative learning. In: **Technology-enhanced learning**. Springer Netherlands, 2009.

DOWNES, S. Learning networks and connective knowledge. In: **Instructional Technology Forum**, paper 92, 2006.

EIBEN, A. E.; SMITH, J. E. **Introduction to evolutionary computing**. 2. ed. Springer, 2007.

ENGELBRECHT, A. P. **Computational intelligence: an introduction**. John Wiley & Sons, 2 ed., 2007.

FJERMESTAD, J. An analysis of communication mode in group support systems research. In: **Decision Support Systems**, v. 37, n. 2, p.239–263, 2004.

GIL, A. C. **Como elaborar projetos de pesquisa**. São Paulo: Atlas, 1991.

GOGOULOU, A.; GOULI, E.; BOAS, G.; LIAKOU, E.; GRIGORIADOU, M. Forming homogeneous, heterogeneous and mixed groups of learners. In: P. Brusilovsky, M. Grigoriadou, K. Papanikolaou (Eds.): **Proceedings of Workshop on Personalisation in E-Learning Environments at Individual and Group Level**, 11th International Conference on User Modeling, p.33-40, 2007.

GRAF, S.; BEKELE, R. Forming Heterogeneous Groups for Intelligent Collaborative Learning Systems with Ant Colony Optimization. In: **Proceedings of Intelligent Tutoring Systems**, p. 217 – 226, 2006.

HERREID, C.F.; DRAKE, K. **Team Learning: Cooperative Learning in the Science Classroom**. University at Buffalo, 2000.

HOLLAND, J. H. **Adaptation in Natural and Artificial Systems**. The University of Michigan Press, 1975.

ISOTANI, S. **An Ontological Engineering Approach to Computer- Supported Collaborative Learning: From Theory to Practice**. PhD. Thesis in Information Engineering. Osaka University, 2009.

ISOTANI, S; MIZOGUCHI, R. Theory-Driven Group Formation through Ontologies. In: **Intelligent Tutoring Systems**. Springer, p. 646-655, 2008.

JANIS, I. L. Groupthink. In: **Psychology today**, v. 5, n. 6, p. 43-46, 1971.

JOHNSON, D. W.; JOHNSON, R. **Cooperation and competition: Theory and research**. Edina, MN: Interaction, 1989.

JOHNSON, D. W.; JOHNSON, R. T.; SMITH, K. A. Cooperative learning returns to college: What evidence is there that it works? In: **Change**, v. 30, n. 4, p.26–35, 1998.

KEMCZINSKI, A. **Método de Avaliação para Ambientes E-Learning**. PPEP – UFSC, Florianópolis-SC. Tese de Doutorado em Engenharia de Produção, 2005.

KIRSCHNER, P. A. Using integrated electronic environments for collaborative teaching/learning. In: **Learning and Instruction**, v. 10(Suppl. 1), p.1–9, 2001.

LAHOZ-BELTRA, R.I; PERALES-GRAVAN, C. A survey of nonparametric tests for the statistical analysis of evolutionary computational experiments. In: **International Journal “Information Theories and Applications”**, v. 17, n. 1, p. 49, 2010.

LINDEN, R. Técnicas de agrupamento. In: **Revista de Sistemas de Informação da FSMA**, n. 4, p. 18-36, 2009.

LINDEN, R. **Algoritmos Genéticos**. Rio de Janeiro: Editora Ciência Moderna, 3. ed., 2012.
LITTO, F.M.; FORMIGA, M.M.M. (orgs.) **Educação a distância: o estado da arte**. São Paulo: Pearson Education do Brasil, 2009.

MCAULEY, A.; STEWART, B.; SIEMENS, G.; CORMIER, D. The MOOC Model for Digital Practice. 2010. Disponível em https://oerknowledgecloud.org/sites/oerknowledgecloud.org/files/MOOC_Final_0.pdf.

MELIÁN, B.; PÉREZ, J. A. M.; VEGA, J. M. M. Metaheuristics: A Global View. In: **Revista Iberoamericana de Inteligencia Artificial**. Asociación Española de Inteligencia Artificial, v. 2, n. 19, 2003.

MICHAELSEN, L. K.; KNIGHT, A. B.; FINK, L. D. (Ed.). **Team-based learning: A transformative use of small groups**. Greenwood Publishing Group, 2004.

MINICUCCI, A. **Dinâmica de Grupo: teorias e sistemas**. São Paulo: Atlas, 1997.

MONTGOMERY, D. C.; RUNGER, G. C. **Estatística aplicada e probabilidade para engenheiros**. Rio de Janeiro: LTC, 5 ed., 2012.

MORENO, J.; OVALLE, D. A.; VICARI, R. M. A genetic algorithm approach for group formation in collaborative learning considering multiple student characteristics. In: **Computers & Education**, v. 58, n.1, p.560-569, 2012.

OUNNAS, A. **Enhancing the automation of forming groups for education with semantics**. PhD Thesis. University of Southampton, 2010.

OUNNAS, A.; DAVIS, H. C.; MILLARD, D. E. Towards Semantic Group Formation. In: **Proceedings of The 7th IEEE International Conference on Advanced Learning Technologies (ICALT 2007)**. IEEE Computer Society Press, Los Alamitos, pp.825-827, 2007.

OWENS, D. A.; MANNIX, E. A.; NEALE, M. A. Strategic formation of groups: Issues in task-performance and team member selection. In: D. H. Gruenfel (editor), **Research on managing groups and teams: Composition**, v. 1, p.149–165, 1998.

PANITZ, T. A definition of collaborative versus cooperative learning. *Deliberations* [Online], 1996. Disponível em <http://www.londonmet.ac.uk/deliberations/collaborative-learning/panitz-paper.cfm>

PETERSEN, K.; FELDT, R.; MUJTABA, S.; MATTSSON, M. Systematic mapping studies in software engineering. In: **12th International Conference on Evaluation and Assessment in Software Engineering**, v. 17, n. 1, 2008.

POPOV, V.; NOROOZI, O.; BARRETT, J. B.; BIEMANS, H. J.; TEASLEY, S. D.; SLOF, B.; MULDER, M. Perceptions and experiences of, and outcomes for, university students in culturally diversified dyads in a computer-supported collaborative learning environment. In: **Computers in Human Behavior**, v. 32, p. 186-200, 2014.

RESTA, P.; LAFERRIÈRE, T. Technology in support of collaborative learning. In: **Education and Psychology Review**, v. 19, p.65-83, 2007.

ROSHELLE, J.; TEASLEY, S. The construction of shared knowledge in collaborative problem solving. In C. E. O'Malley (Ed.), **Computer supported collaborative learning**. Berlin Heidelberg New York: Springer, p. 69-197, 1995.

STAHL, G.; KOSCHMANN, T.; SUTHERS, D. Computer-supported collaborative learning: An historical perspective. **Cambridge handbook of the learning science**. Cambridge: Cambridge University Press, p.409-426, 2006.

SIEMENS, G. Connectivism: A Learning Theory for the Digital Age. In: **International Journal of Instructional Technology and Distance Learning**, v. 2, n.1, 2004.

SILVA, E. L. **Metodologia da pesquisa e elaboração de dissertação**. 3 ed. Florianópolis, 2001.

SILVA, K do N.; JOYE, C.R. Medidas de Avaliação de Aprendizagem. In. ALMEIDA, M.G.; FREITAS, M. C. D. (orgs). **Desafios Permanentes: projeto político pedagógico, gestão escolar, métricas no contexto das TICs (A escola do Século XXI; v.4)**. Rio de Janeiro: Brasport, 2015.

STEGMANN, K.; WECKER, C.; WEINBERGER, A.; FISCHER, F. Collaborative argumentation and cognitive elaboration in a computer-supported collaborative learning environment. In: **Instructional Science**, v. 40, n. 2, p. 297-323, 2012.

STERNBERG, R. J. Thinking styles: theory and assessment at the interface between intelligence and personality. In: R. J. Sterberg & P. Ruzgis (Eds.), **Personality and intelligence**. New York: Cambridge University Press, p. 169-187, 1994.

TOSCANI, L. V.; VELOSO, P. A. **Complexidade de algoritmos**. Porto Alegre: Sagra-Luzzato, 2009.

ZINS, C. Conceptual approaches for defining data, information, and knowledge. In: **Journal of the American society for information science and technology**, v. 58, n. 4, p. 479-493, 2007.

WANG, D.; LIN, S. S. J.; SUN, C. DIANA: A computer-supported heterogeneous grouping system for teachers to conduct successful small learning groups. In: **Computers in Human Behavior**, v. 23, n. 4, p. 1997-2010, 2007.

WANG, F.; CARLEY, K. M.; ZENG, D.; MAO, W. Social Computing: From Social Informatics to Social Intelligence. In: **Intelligent Systems**, IEEE, v. 22, n. 2, p.79-83, 2007.

WEBB, N. M. Predicting learning from student interaction: Defining the interaction variable. In: **Educational Psychologist**, v. 18, p.33-41, 1983.

WEBB, N. M. Testing a theoretical model of student interaction and learning in small groups. In: Hertz-Lazarowitz, R.; Miller, N. (editors) **Interaction in Cooperative Groups: The theoretical anatomy of group learning**, p. 102-119. Cambridge University Press, 1992.

WEBER, M.; SAWIŁOWSKY, S.S. Comparative Power Of The Independent t, Permutation t, and Wilcoxon Tests. In: **Journal of Modern Applied Statistical Methods**, v. 8, 2009.

WEISE, T. **Global optimization algorithms-theory and application**. Self-Published, 2009.

WESSNER, M.; PFISTER, H. R. Group formation in computer-supported collaborative learning. In: **Proceedings of the 2001 international ACM SIGGROUP conference on supporting group work**. ACM, p.24-31, 2001.

WOLPERT, D. H.; MACREADY, W. G. No free lunch theorems for optimization. In: **Evolutionary Computation**, IEEE Transactions, v.1, n. 1, p.67-82, 1997.

YANNIBELLI, V.; AMANDI, A. A deterministic crowding evolutionary algorithm to form learning teams in a collaborative learning context. In: **Expert Systems with Applications**, v. 39, n. 10, p. 8584-8592, 2012.

YUAN, L.; POWELL, S. MOOCs and Open Education: Implications for Higher Education. In: **JISC CETIS**, 2013.

APÊNDICE A – TRABALHOS PUBLICADOS

CITADIN, J. R.; KEMCZINSKI, A.; DE MATOS, A. V. Análisis de Herramientas de Colaboración con MOOCs. In: **Anais III Workshop Internacional sobre criação de MOOC com anotações multimídias**, 2014.

CITADIN, J. R.; KEMCZINSKI, A.; DE MATOS, A. V. Colaboração em Massive Open Online Courses (MOOCs). In: **Anais do Computer on the Beach**, p. 233-242, 2014.

CITADIN, J. R.; KEMCZINSKI, A.; DE MATOS, A. V. Formação de Grupos para Aprendizagem Colaborativa: Um mapeamento sistemático da literatura. In: **Nuevas Ideas en Informática Educativa TISE 2014**, p. 46-54, 2014.

DE ARAÚJO, L. P.; MANTAU, M. J.; CITADIN, J. R.; BERKENBROCK, C. D. M.; KEMCZINSKI, A.; BERKENBROCK, G. R.; MATTOS, M. M. Heuristic evaluation for mobile groupware: Evaluating two Audience Response Systems. In: **Journal of Applied Computing Research**, v. 3, n. 2, p. 64-77, 2014.

PESSINI, A.; CITADIN, J.; KEMCZINSKI, A.; GASPARINI, I. Avaliação da Acessibilidade das Funções de Privacidade do Facebook com Pessoas com Deficiência Visual. In: **Revista Latino-Americana de Inovação e Engenharia de Produção**, v. 1, n.1, 2013.

APÊNDICE B – TABELAS DOS RESULTADOS

As tabelas a seguir apresentam os resultados obtidos, considerando-se a média das 30 execuções. As linhas correspondem à quantidade de estudantes, de mil até dez mil, com uma linha abaixo de cada quantidade apresentando o melhor resultado alcançado pelos algoritmos (linha *Best*). As colunas apresentam o valor do *fitness* com o respectivo desvio padrão e o percentual de grupos heterogêneos (em relação ao total de grupos), isto é, os grupos que alcançaram a heterogeneidade máxima possível, com o respectivo desvio padrão. Há ainda duas colunas com as diferenças entre o resultado do AG e o resultado do algoritmo randômico, para o valor do *fitness* e para o percentual de grupos heterogêneos.

B.1 RESULTADOS OBTIDOS PARA CONHECIMENTOS HETEROGÊNEOS-INTERAÇÕES UNIFORMES

Tabela 1 – Resultados conhecimentos heterogêneos-interações uniformes

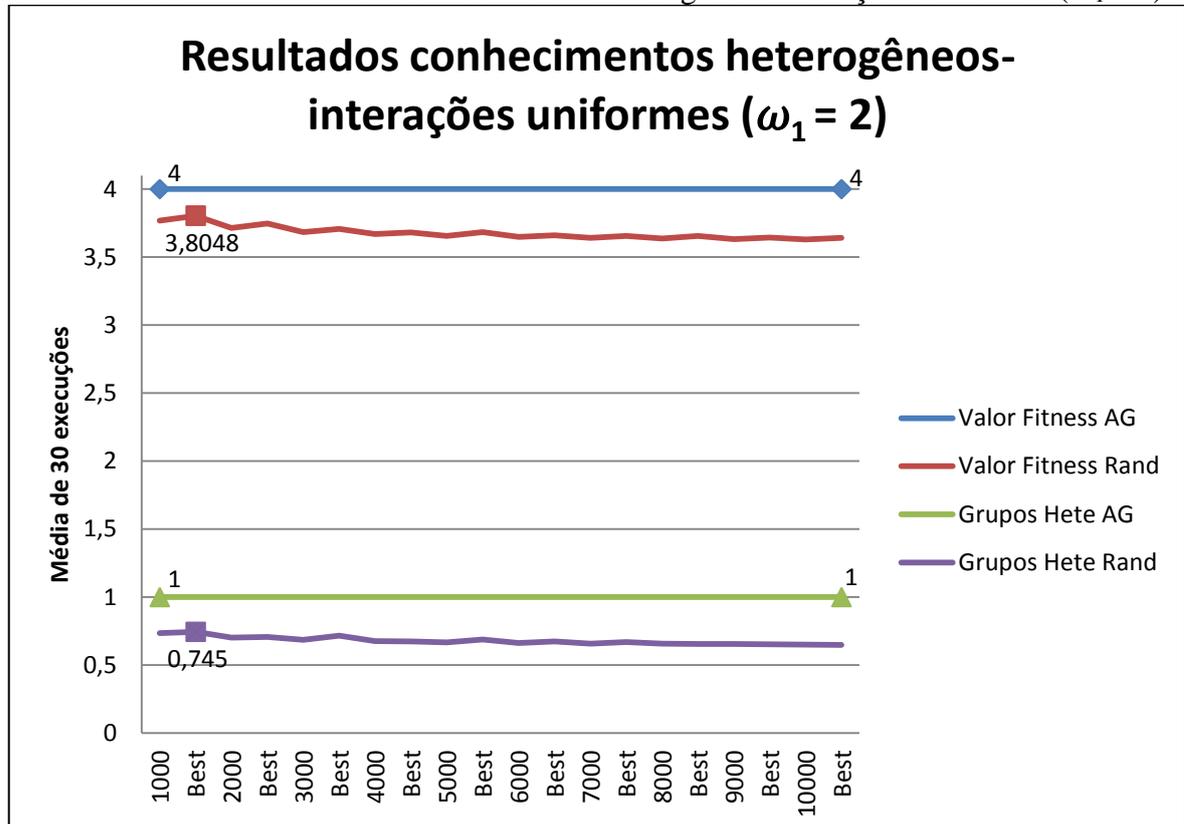
Média de 30 execuções										
#Est	AG				Randômico				Diferença	
	Valor <i>Fitness</i>	+/-	Grupos Hete	+/-	Valor <i>Fitness</i>	+/-	Grupos Hete	+/-	Valor <i>Fitness</i>	Grupos Hete
1000	3,00000	0,00	1,00000	0,00	2,86574	0,01	0,73783	0,02	0,13426	0,26217
<i>Best</i>	3,0000		1,0000		2,8818		0,7650			
2000	3,00000	0,00	1,00000	0,00	2,83041	0,01	0,69450	0,02	0,16959	0,30550
<i>Best</i>	3,0000		1,0000		2,8485		0,7450			
3000	3,00000	0,00	1,00000	0,00	2,81151	0,00	0,67800	0,01	0,18849	0,32200
<i>Best</i>	3,0000		1,0000		2,8239		0,7067			
4000	3,00000	0,00	1,00000	0,00	2,80066	0,01	0,66842	0,01	0,19934	0,33158
<i>Best</i>	3,0000		1,0000		2,8150		0,6700			
5000	3,00000	0,00	1,00000	0,00	2,79408	0,00	0,66083	0,01	0,20592	0,33917
<i>Best</i>	3,0000		1,0000		2,8009		0,6470			
6000	3,00000	0,00	1,00000	0,00	2,78772	0,00	0,65880	0,01	0,21228	0,34120
<i>Best</i>	3,0000		1,0000		2,7977		0,6642			
7000	3,00000	0,00	1,00000	0,00	2,78296	0,00	0,65405	0,01	0,21704	0,34595
<i>Best</i>	3,0000		1,0000		2,7967		0,6529			
8000	3,00000	0,00	1,00000	0,00	2,78039	0,00	0,65450	0,01	0,21961	0,34550
<i>Best</i>	3,0000		1,0000		2,7920		0,6737			
9000	2,99998	0,00	0,99994	0,00	2,77809	0,00	0,65168	0,01	0,22189	0,34826
<i>Best</i>	3,0000		1,0000		2,7909		0,6617			
10000	2,99992	0,00	0,99980	0,00	2,77481	0,00	0,65107	0,01	0,22511	0,34873
<i>Best</i>	3,0000		1,0000		2,7851		0,6630			

Fonte: Próprio autor

Tabela 2 – Resultados conhecimentos heterogêneos-interações uniformes ($\omega_1 = 2$)

#Est	AG				Randômico				Diferença	
	Valor Fitness	+/-	Grupos Hete	+/-	Valor Fitness	+/-	Grupos Hete	+/-	Valor Fitness	Grupos Hete
1000	4,00000	0,00	1,00000	0,00	3,76816	0,01	0,73567	0,02	0,23184	0,26433
<i>Best</i>	4,0000		1,0000		3,8048		0,7450			
2000	4,00000	0,00	1,00000	0,00	3,71432	0,01	0,70317	0,02	0,28568	0,29683
<i>Best</i>	4,0000		1,0000		3,7478		0,7075			
3000	4,00000	0,00	1,00000	0,00	3,68439	0,01	0,68522	0,02	0,31561	0,31478
<i>Best</i>	4,0000		1,0000		3,7085		0,715			
4000	4,00000	0,00	1,00000	0,00	3,66903	0,01	0,67633	0,01	0,33097	0,32367
<i>Best</i>	4,0000		1,0000		3,6824		0,675			
5000	4,00000	0,00	1,00000	0,00	3,65591	0,01	0,66643	0,01	0,34409	0,33357
<i>Best</i>	4,0000		1,0000		3,6834		0,689			
6000	4,00000	0,00	1,00000	0,00	3,64955	0,01	0,66214	0,01	0,35045	0,33786
<i>Best</i>	4,0000		1,0000		3,6613		0,6733			
7000	4,00000	0,00	1,00000	0,00	3,64160	0,01	0,65850	0,01	0,35840	0,34150
<i>Best</i>	4,0000		1,0000		3,6565		0,67			
8000	3,99993	0,00	0,99992	0,00	3,63702	0,01	0,65648	0,01	0,36291	0,34344
<i>Best</i>	4,0000		1,0000		3,6565		0,6562			
9000	3,99996	0,00	0,99994	0,00	3,63164	0,01	0,65407	0,01	0,36832	0,34587
<i>Best</i>	4,0000		1,0000		3,645		0,6533			
10000	3,99983	0,00	0,99978	0,00	3,62877	0,01	0,65012	0,01	0,37106	0,34967
<i>Best</i>	4,0000		1,0000		3,6427		0,649			

Fonte: Próprio autor

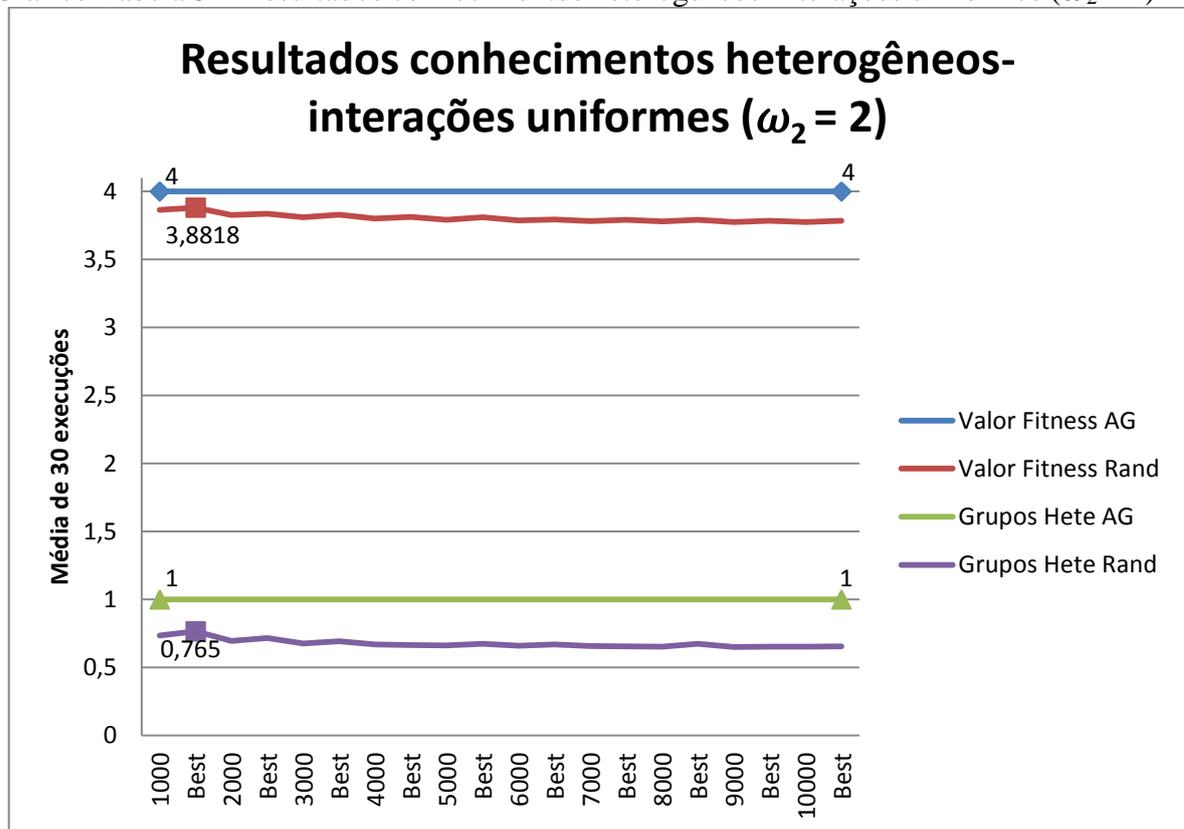
Gráfico Tabela 2 – Resultados conhecimentos heterogêneos-interações uniformes ($\omega_1 = 2$)

Fonte: Próprio autor

Tabela 3 – Resultados conhecimentos heterogêneos-interações uniformes ($\omega_2 = 2$)

Média de 30 execuções										
#Est	AG				Randômico				Diferença	
	Valor Fitness	+/-	Grupos Hete	+/-	Valor Fitness	+/-	Grupos Hete	+/-	Valor Fitness	Grupos Hete
1000	4,00000	0,00	1,00000	0,00	3,86583	0,01	0,73600	0,03	0,13417	0,26400
<i>Best</i>	4,0000		1,0000		3,8818		0,765			
2000	4,00000	0,00	1,00000	0,00	3,82807	0,00	0,69567	0,02	0,17193	0,30433
<i>Best</i>	4,0000		1,0000		3,8358		0,715			
3000	4,00000	0,00	1,00000	0,00	3,81112	0,01	0,67662	0,01	0,18888	0,32338
<i>Best</i>	4,0000		1,0000		3,8289		0,6917			
4000	4,00000	0,00	1,00000	0,00	3,80093	0,01	0,66936	0,02	0,19907	0,33064
<i>Best</i>	4,0000		1,0000		3,8125		0,6637			
5000	4,00000	0,00	1,00000	0,00	3,79273	0,01	0,66287	0,01	0,20727	0,33713
<i>Best</i>	4,0000		1,0000		3,8116		0,673			
6000	4,00000	0,00	1,00000	0,00	3,78702	0,00	0,65939	0,01	0,21298	0,34061
<i>Best</i>	4,0000		1,0000		3,7944		0,67			
7000	4,00000	0,00	1,00000	0,00	3,78320	0,00	0,65683	0,01	0,21680	0,34317
<i>Best</i>	4,0000		1,0000		3,7912		0,6543			
8000	4,00000	0,00	1,00000	0,00	3,78011	0,00	0,65223	0,01	0,21989	0,34777
<i>Best</i>	4,0000		1,0000		3,7922		0,6744			
9000	3,99994	0,00	0,99982	0,00	3,77659	0,00	0,65034	0,01	0,22334	0,34948
<i>Best</i>	4,0000		1,0000		3,7848		0,6539			
10000	3,99989	0,00	0,99973	0,00	3,77557	0,00	0,65222	0,01	0,22432	0,34752
<i>Best</i>	4,0000		1,0000		3,7842		0,6545			

Fonte: Próprio autor

Gráfico Tabela 3 – Resultados conhecimentos heterogêneos-interações uniformes ($\omega_2 = 2$)

Fonte: Próprio autor

B.2 RESULTADOS OBTIDOS PARA CONHECIMENTOS HETEROGÊNEOS- INTERAÇÕES POLARIZADAS

Tabela 4 – Resultados conhecimentos heterogêneos-interações polarizadas

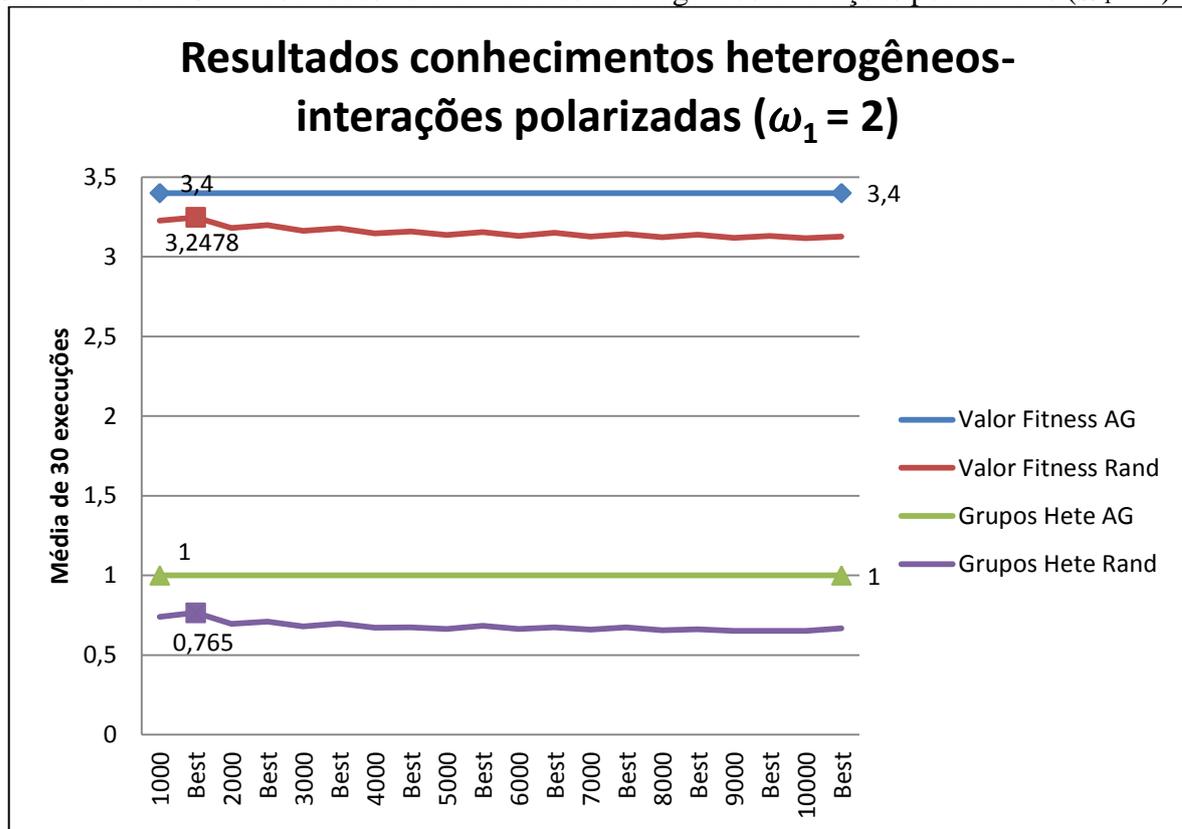
Média de 30 execuções										
#Est	AG				Randômico				Diferença	
	Valor <i>Fitness</i>	+/-	Grupos Hete	+/-	Valor <i>Fitness</i>	+/-	Grupos Hete	+/-	Valor <i>Fitness</i>	Grupos Hete
1000	2,40000	0,00	1,00000	0,00	2,32487	0,00	0,71800	0,02	0,07513	0,28200
<i>Best</i>	2,4000		1,0000		2,3357		0,75			
2000	2,40000	0,00	1,00000	0,00	2,29970	0,00	0,67992	0,02	0,10030	0,32008
<i>Best</i>	2,4000		1,0000		2,312		0,7275			
3000	2,40000	0,00	1,00000	0,00	2,28849	0,00	0,66700	0,02	0,11151	0,33300
<i>Best</i>	2,4000		1,0000		2,2963		0,685			
4000	2,40000	0,00	1,00000	0,00	2,28192	0,00	0,66383	0,01	0,11808	0,33617
<i>Best</i>	2,4000		1,0000		2,2946		0,6613			
5000	2,40000	0,00	1,00000	0,00	2,27478	0,00	0,65677	0,01	0,12522	0,34323
<i>Best</i>	2,4000		1,0000		2,2873		0,679			
6000	2,40000	0,00	1,00000	0,00	2,27184	0,00	0,65467	0,01	0,12816	0,34533
<i>Best</i>	2,4000		1,0000		2,2799		0,6533			
7000	2,40000	0,00	1,00000	0,00	2,26983	0,00	0,65495	0,01	0,13017	0,34505
<i>Best</i>	2,4000		1,0000		2,2765		0,6443			
8000	2,39999	0,00	0,99998	0,00	2,26583	0,00	0,64880	0,01	0,13416	0,35118
<i>Best</i>	2,4000		1,0000		2,2736		0,6706			
9000	2,39999	0,00	0,99994	0,00	2,26578	0,00	0,64643	0,01	0,13421	0,35351
<i>Best</i>	2,4000		1,0000		2,2732		0,6372			
10000	2,39997	0,00	0,99987	0,00	2,26316	0,00	0,64622	0,01	0,13682	0,35365
<i>Best</i>	2,40000	0,00	1,00000	0,00	2,2691		0,6495			

Fonte: Próprio autor

Tabela 5 – Resultados conhecimentos heterogêneos-interações polarizadas ($\omega_1 = 2$)

Média de 30 execuções										
#Est	AG				Randômico				Diferença	
	Valor <i>Fitness</i>	+/-	Grupos Hete	+/-	Valor <i>Fitness</i>	+/-	Grupos Hete	+/-	Valor <i>Fitness</i>	Grupos Hete
1000	3,40000	0,00	1,00000	0,00	3,22596	0,01	0,73933	0,02	0,17404	0,26067
<i>Best</i>	3,4000		1,0000		3,2478		0,765			
2000	3,40000	0,00	1,00000	0,00	3,18145	0,01	0,69642	0,02	0,21855	0,30358
<i>Best</i>	3,4000		1,0000		3,1982		0,71			
3000	3,40000	0,00	1,00000	0,00	3,16229	0,01	0,67867	0,02	0,23771	0,32133
<i>Best</i>	3,4000		1,0000		3,1787		0,6983			
4000	3,40000	0,00	1,00000	0,00	3,14643	0,01	0,67074	0,01	0,25357	0,32926
<i>Best</i>	3,4000		1,0000		3,158		0,6725			
5000	3,40000	0,00	1,00000	0,00	3,13704	0,01	0,66410	0,01	0,26296	0,33590
<i>Best</i>	3,4000		1,0000		3,1544		0,684			
6000	3,40000	0,00	1,00000	0,00	3,13066	0,01	0,66391	0,01	0,26934	0,33609
<i>Best</i>	3,4000		1,0000		3,1502		0,6733			
7000	3,40000	0,00	1,00000	0,00	3,12746	0,01	0,65898	0,01	0,27254	0,34102
<i>Best</i>	3,4000		1,0000		3,1427		0,6729			
8000	3,39999	0,00	0,99998	0,00	3,12237	0,01	0,65552	0,01	0,27761	0,34446
<i>Best</i>	3,4000		1,0000		3,1392		0,6606			
9000	3,39998	0,00	0,99996	0,00	3,11917	0,00	0,65232	0,01	0,28081	0,34764
<i>Best</i>	3,4000		1,0000		3,1301		0,6517			
10000	3,39991	0,00	0,99985	0,00	3,11576	0,00	0,65127	0,01	0,28415	0,34858
<i>Best</i>	3,4000		1,0000		3,1269		0,6665			

Fonte: Próprio autor

Gráfico Tabela 5 – Resultados conhecimentos heterogêneos-interações polarizadas ($\omega_1 = 2$)

Fonte: Próprio autor

Tabela 6 – Resultados conhecimentos heterogêneos-interações polarizadas ($\omega_2 = 2$)

Média de 30 execuções

#Est	AG				Randômico				Diferença	
	Valor Fitness	+/-	Grupos Hete	+/-	Valor Fitness	+/-	Grupos Hete	+/-	Valor Fitness	Grupos Hete
1000	2,80000	0,00	1,00000	0,00	2,78630	0,00	0,69367	0,04	0,01370	0,30633
Best	2,8000		1,0000		2,7915		0,725			
2000	2,80594	0,02	0,90242	0,29	2,77414	0,00	0,65633	0,02	0,03180	0,24608
Best	2,8628		0,0125		2,778		0,665			
3000	2,80000	0,00	1,00000	0,00	2,76792	0,00	0,65255	0,02	0,03208	0,34745
Best	2,8000		1,0000		2,7724		0,6933			
4000	2,80168	0,01	0,96854	0,17	2,76324	0,00	0,64601	0,02	0,03844	0,32253
Best	2,8504		0,0563		2,7703		0,6675			
5000	2,80000	0,00	1,00000	0,00	2,75974	0,00	0,64653	0,01	0,04026	0,35347
Best	2,8000		1,0000		2,7653		0,648			
6000	2,80000	0,00	1,00000	0,00	2,75756	0,00	0,64616	0,01	0,04244	0,35384
Best	2,8000		1,0000		2,7621		0,6483			
7000	2,80134	0,01	0,96990	0,16	2,75583	0,00	0,63970	0,01	0,04551	0,33020
Best	2,8402		0,0971		2,7605		0,6436			
8000	2,80123	0,01	0,97036	0,16	2,75386	0,00	0,63660	0,01	0,04737	0,33376
Best	2,8369		0,1119		2,7579		0,625			
9000	2,80000	0,00	0,99996	0,00	2,75268	0,00	0,63896	0,01	0,04732	0,36100
Best	2,8000		1,0000		2,7562		0,6433			
10000	2,80000	0,00	0,99973	0,00	2,75134	0,00	0,63463	0,01	0,04866	0,36510
Best	2,8000		1,0000		2,7551		0,6435			

Fonte: Próprio autor

B.3 RESULTADOS OBTIDOS PARA CONHECIMENTOS HETEROGÊNEOS- INTERAÇÕES ALEATÓRIAS

Tabela 7 – Resultados conhecimentos heterogêneos-interações aleatórias

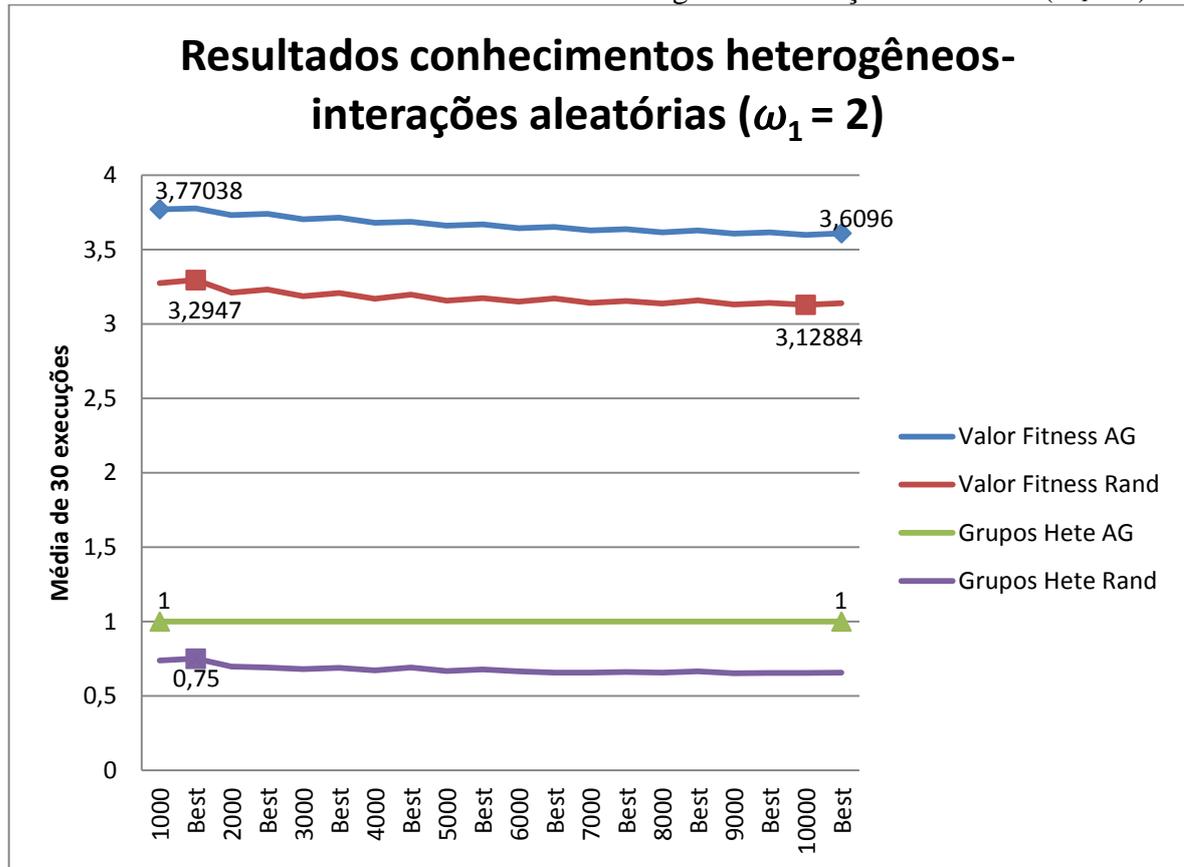
Média de 30 execuções										
#Est	AG				Randômico				Diferença	
	Valor <i>Fitness</i>	+/-	Grupos Hete	+/-	Valor <i>Fitness</i>	+/-	Grupos Hete	+/-	Valor <i>Fitness</i>	Grupos Hete
1000	2,77111	0,00	1,00000	0,00	2,37270	0,01	0,73850	0,02	0,39841	0,26150
<i>Best</i>	2,7777		1,0000		2,3913		0,725			
2000	2,73313	0,00	1,00000	0,00	2,33157	0,01	0,69333	0,02	0,40156	0,30667
<i>Best</i>	2,7377		1,0000		2,3486		0,71			
3000	2,70466	0,00	1,00000	0,00	2,31240	0,01	0,68178	0,02	0,39225	0,31822
<i>Best</i>	2,7144		1,0000		2,3235		0,6717			
4000	2,68216	0,00	1,00000	0,00	2,30450	0,01	0,67440	0,01	0,37765	0,32560
<i>Best</i>	2,6901		1,0000		2,3291		0,72			
5000	2,66142	0,01	1,00000	0,00	2,29501	0,01	0,66320	0,01	0,36641	0,33680
<i>Best</i>	2,6715		1,0000		2,309		0,671			
6000	2,64394	0,01	0,99997	0,00	2,28943	0,01	0,65964	0,01	0,35451	0,34033
<i>Best</i>	2,6575		1,0000		2,3145		0,6683			
7000	2,63161	0,01	0,99998	0,00	2,28475	0,01	0,65563	0,01	0,34686	0,34435
<i>Best</i>	2,6419		1,0000		2,3026		0,6693			
8000	2,61936	0,01	0,99988	0,00	2,28182	0,00	0,65262	0,01	0,33754	0,34726
<i>Best</i>	2,6311		1,0000		2,2926		0,66			
9000	2,60983	0,01	0,99984	0,00	2,27812	0,00	0,65290	0,01	0,33171	0,34694
<i>Best</i>	2,6251		1,0000		2,2919		0,6472			
10000	2,60011	0,01	0,99927	0,00	2,27534	0,00	0,65007	0,01	0,32477	0,34920
<i>Best</i>	2,6226		1,0000		2,2879		0,6615			

Fonte: Próprio autor

Tabela 8 – Resultados conhecimentos heterogêneos-interações aleatórias ($\omega_1 = 2$)

Média de 30 execuções										
#Est	AG				Randômico				Diferença	
	Valor <i>Fitness</i>	+/-	Grupos Hete	+/-	Valor <i>Fitness</i>	+/-	Grupos Hete	+/-	Valor <i>Fitness</i>	Grupos Hete
1000	3,77038	0,00	1,00000	0,00	3,27349	0,01	0,73817	0,02	0,49689	0,26183
<i>Best</i>	3,7762		1,0000		3,2947		0,75			
2000	3,73205	0,00	1,00000	0,00	3,20887	0,01	0,69658	0,01	0,52317	0,30342
<i>Best</i>	3,7396		1,0000		3,231		0,69			
3000	3,70370	0,00	1,00000	0,00	3,18681	0,01	0,67973	0,01	0,51689	0,32027
<i>Best</i>	3,7134		1,0000		3,2068		0,6883			
4000	3,67938	0,00	1,00000	0,00	3,16987	0,01	0,67141	0,01	0,50951	0,32859
<i>Best</i>	3,6863		1,0000		3,1978		0,69			
5000	3,66024	0,00	1,00000	0,00	3,15711	0,01	0,66813	0,01	0,50314	0,33187
<i>Best</i>	3,6682		1,0000		3,174		0,678			
6000	3,64237	0,01	1,00000	0,00	3,15063	0,01	0,66414	0,01	0,49174	0,33586
<i>Best</i>	3,6526		1,0000		3,1706		0,6567			
7000	3,62885	0,01	1,00000	0,00	3,14150	0,01	0,65652	0,01	0,48735	0,34348
<i>Best</i>	3,6371		1,0000		3,154		0,6607			
8000	3,61635	0,01	0,99998	0,00	3,13791	0,01	0,65609	0,01	0,47844	0,34389
<i>Best</i>	3,6293		1,0000		3,1581		0,6644			
9000	3,60583	0,01	0,99981	0,00	3,13072	0,01	0,65270	0,01	0,47511	0,34711
<i>Best</i>	3,6162		1,0000		3,1411		0,655			
10000	3,59754	0,01	0,99967	0,00	3,12884	0,01	0,65457	0,01	0,46870	0,34510
<i>Best</i>	3,6096		1,0000		3,1392		0,6555			

Fonte: Próprio autor

Gráfico Tabela 8 – Resultados conhecimentos heterogêneos-interações aleatórias ($\omega_1 = 2$)

Fonte: Próprio autor

Tabela 9 – Resultados conhecimentos heterogêneos-interações aleatórias ($\omega_2 = 2$)

#Est	AG				Randômico				Diferença	
	Valor Fitness	+/-	Grupos Hete	+/-	Valor Fitness	+/-	Grupos Hete	+/-	Valor Fitness	Grupos Hete
1000	3,54126	0,00	1,00000	0,00	2,88013	0,01	0,71817	0,03	0,66113	0,28183
Best	3,5467		1,0000		2,9048		0,7			
2000	3,46881	0,01	1,00000	0,00	2,83685	0,01	0,68625	0,02	0,63196	0,31375
Best	3,4756		1,0000		2,854		0,735			
3000	3,41442	0,01	0,99994	0,00	2,81573	0,01	0,66877	0,01	0,59870	0,33118
Best	3,4315		1,0000		2,8276		0,6633			
4000	3,36942	0,01	0,99984	0,00	2,80537	0,01	0,66320	0,01	0,56405	0,33664
Best	3,3848		1,0000		2,822		0,6825			
5000	3,33341	0,01	0,99973	0,00	2,79775	0,00	0,66023	0,01	0,53566	0,33950
Best	3,3501		1,0000		2,8073		0,646			
6000	3,30415	0,01	0,99895	0,00	2,79026	0,00	0,65028	0,01	0,51389	0,34867
Best	3,3226		0,9983		2,7989		0,6592			
7000	3,27950	0,01	0,99861	0,00	2,78506	0,00	0,65162	0,01	0,49444	0,34699
Best	3,3018		0,9979		2,797		0,6471			
8000	3,25236	0,01	0,99647	0,00	2,78252	0,00	0,64914	0,01	0,46984	0,34733
Best	3,2759		0,9975		2,7878		0,6512			
9000	3,23712	0,01	0,99563	0,00	2,78103	0,00	0,64913	0,01	0,45609	0,34651
Best	3,2647		0,9978		2,7909		0,6611			
10000	3,21624	0,02	0,99270	0,00	2,77638	0,00	0,64735	0,01	0,43986	0,34535
Best	3,2526		0,998		2,7823		0,653			

Fonte: Próprio autor

B.4 RESULTADOS OBTIDOS PARA CONHECIMENTOS HOMOGÊNEOS- INTERAÇÕES UNIFORMES

Tabela 10 – Resultados conhecimentos homogêneos-interações uniformes

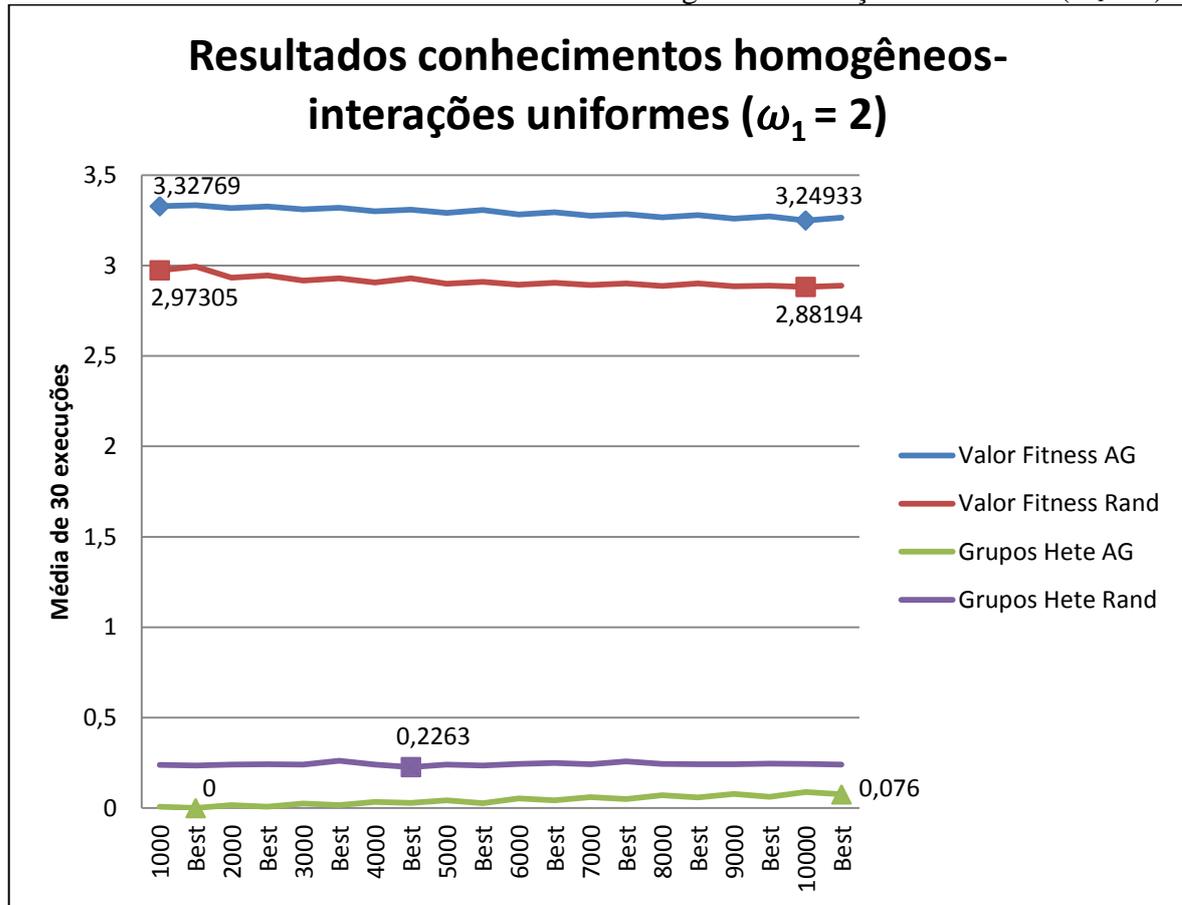
Média de 30 execuções										
#Est	AG				Randômico				Diferença	
	Valor Fitness	+/-	Grupos Hete	+/-	Valor Fitness	+/-	Grupos Hete	+/-	Valor Fitness	Grupos Hete
1000	2,66151	0,00	0,00417	0,00	2,41340	0,00	0,22617	0,02	0,24811	-0,22200
<i>Best</i>	2,6644		0		2,4302		0,185			
2000	2,65384	0,00	0,01183	0,00	2,38240	0,01	0,23167	0,01	0,27144	-0,21983
<i>Best</i>	2,6585		0,0075		2,4021		0,2375			
3000	2,64783	0,00	0,01839	0,00	2,36967	0,00	0,23445	0,01	0,27816	-0,21606
<i>Best</i>	2,655		0,0117		2,3806		0,205			
4000	2,64239	0,00	0,02263	0,00	2,36061	0,00	0,23870	0,01	0,28178	-0,21607
<i>Best</i>	2,6531		0,0075		2,3671		0,2225			
5000	2,63624	0,00	0,03047	0,00	2,35569	0,00	0,23613	0,01	0,28055	-0,20567
<i>Best</i>	2,6414		0,023		2,3633		0,227			
6000	2,62967	0,00	0,03797	0,00	2,35248	0,00	0,23461	0,01	0,27719	-0,19664
<i>Best</i>	2,6373		0,03		2,3677		0,2292			
7000	2,62351	0,00	0,04538	0,01	2,34732	0,00	0,23958	0,01	0,27619	-0,19421
<i>Best</i>	2,6322		0,0321		2,3561		0,2457			
8000	2,61771	0,00	0,05189	0,01	2,34572	0,00	0,24052	0,01	0,27199	-0,18862
<i>Best</i>	2,6268		0,04		2,3527		0,2437			
9000	2,61208	0,01	0,05882	0,01	2,34326	0,00	0,23774	0,01	0,26882	-0,17892
<i>Best</i>	2,6249		0,0417		2,3486		0,2422			
10000	2,60516	0,01	0,06645	0,01	2,34157	0,00	0,23892	0,01	0,26360	-0,17247
<i>Best</i>	2,6167		0,051		2,3472		0,241			

Fonte: Próprio autor

Tabela 11 – Resultados conhecimentos homogêneos-interações uniformes ($\omega_1 = 2$)

Média de 30 execuções										
#Est	AG				Randômico				Diferença	
	Valor Fitness	+/-	Grupos Hete	+/-	Valor Fitness	+/-	Grupos Hete	+/-	Valor Fitness	Grupos Hete
1000	3,32769	0,00	0,00667	0,00	2,97305	0,01	0,23900	0,02	0,35464	-0,23233
<i>Best</i>	3,3333		0		2,9953		0,235			
2000	3,31762	0,00	0,01667	0,00	2,93367	0,01	0,24058	0,01	0,38395	-0,22392
<i>Best</i>	3,3265		0,0075		2,9456		0,2425			
3000	3,31036	0,00	0,02440	0,01	2,91730	0,01	0,24023	0,01	0,39306	-0,21583
<i>Best</i>	3,3192		0,0167		2,9296		0,2617			
4000	3,29985	0,00	0,03470	0,00	2,90686	0,01	0,24132	0,01	0,39299	-0,20662
<i>Best</i>	3,3082		0,0288		2,9288		0,2263			
5000	3,29214	0,01	0,04340	0,01	2,89887	0,00	0,24103	0,01	0,39327	-0,19763
<i>Best</i>	3,3079		0,027		2,9093		0,236			
6000	3,28277	0,00	0,05347	0,01	2,89447	0,00	0,24403	0,01	0,38830	-0,19056
<i>Best</i>	3,2942		0,0433		2,9048		0,25			
7000	3,27608	0,00	0,05981	0,01	2,89165	0,00	0,24307	0,01	0,38443	-0,18326
<i>Best</i>	3,2843		0,05		2,9007		0,2586			
8000	3,26573	0,01	0,07135	0,01	2,88774	0,00	0,24472	0,01	0,37800	-0,17337
<i>Best</i>	3,2782		0,0594		2,9016		0,2425			
9000	3,25879	0,01	0,07902	0,01	2,88450	0,00	0,24286	0,01	0,37429	-0,16384
<i>Best</i>	3,2725		0,0628		2,8893		0,2456			
10000	3,24933	0,01	0,08868	0,01	2,88194	0,00	0,24400	0,01	0,36739	-0,15532
<i>Best</i>	3,2641		0,076		2,8885		0,24			

Fonte: Próprio autor

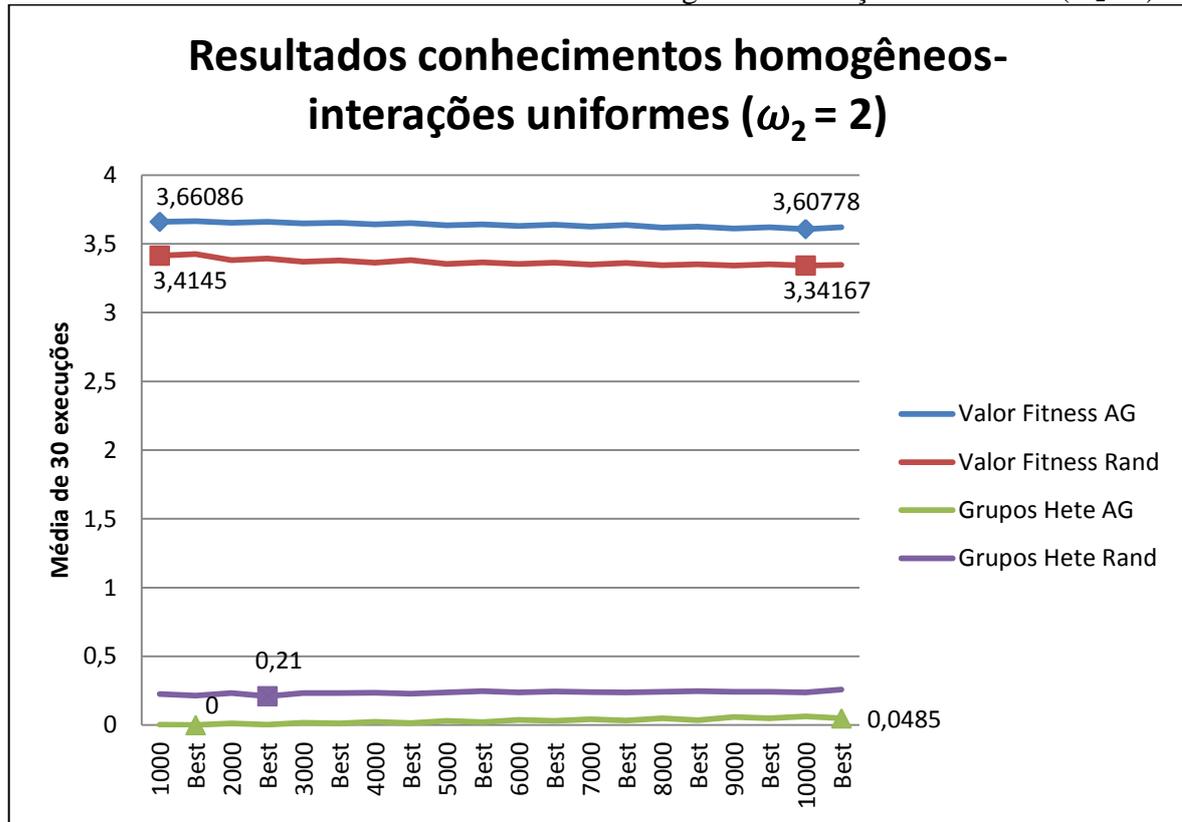
Gráfico Tabela 11 – Resultados conhecimentos homogêneos-interações uniformes ($\omega_1 = 2$)

Fonte: Próprio autor

Tabela 12 – Resultados conhecimentos homogêneos-interações uniformes ($\omega_2 = 2$)

#Est	Média de 30 execuções									
	AG				Randômico				Diferença	
	Valor Fitness	+/-	Grupos Hete	+/-	Valor Fitness	+/-	Grupos Hete	+/-	Valor Fitness	Grupos Hete
1000	3,66086	0,00	0,00400	0,00	3,41450	0,01	0,22633	0,02	0,24636	-0,22233
<i>Best</i>	3,6644		0		3,4257		0,215			
2000	3,65364	0,00	0,01192	0,00	3,38270	0,00	0,23383	0,02	0,27093	-0,22192
<i>Best</i>	3,6603		0,0025		3,3931		0,21			
3000	3,64838	0,00	0,01673	0,00	3,37042	0,00	0,23206	0,01	0,27795	-0,21533
<i>Best</i>	3,6531		0,0117		3,3794		0,2333			
4000	3,64225	0,00	0,02420	0,00	3,36245	0,01	0,23437	0,01	0,27980	-0,21017
<i>Best</i>	3,6506		0,015		3,3828		0,2288			
5000	3,63500	0,00	0,03153	0,01	3,35489	0,00	0,23690	0,01	0,28011	-0,20537
<i>Best</i>	3,6417		0,022		3,3653		0,247			
6000	3,62983	0,00	0,03788	0,00	3,35283	0,00	0,23686	0,01	0,27699	-0,19898
<i>Best</i>	3,6394		0,03		3,3627		0,245			
7000	3,62552	0,00	0,04320	0,00	3,34947	0,00	0,24041	0,01	0,27605	-0,19721
<i>Best</i>	3,6376		0,0329		3,3603		0,2379			
8000	3,61888	0,01	0,05008	0,01	3,34558	0,00	0,24137	0,01	0,27330	-0,19129
<i>Best</i>	3,6255		0,0362		3,3527		0,2462			
9000	3,61212	0,01	0,05920	0,01	3,34332	0,00	0,24154	0,01	0,26880	-0,18234
<i>Best</i>	3,6215		0,0494		3,3513		0,2428			
10000	3,60778	0,01	0,06383	0,01	3,34167	0,00	0,23833	0,01	0,26611	-0,17450
<i>Best</i>	3,6219		0,0485		3,3472		0,2585			

Fonte: Próprio autor

Gráfico Tabela 12 – Resultados conhecimentos homogêneos-interações uniformes ($\omega_2=2$)

Fonte: Próprio autor

B.5 RESULTADOS OBTIDOS PARA CONHECIMENTOS HOMOGÊNEOS- INTERAÇÕES POLARIZADAS

Tabela 13 – Resultados conhecimentos homogêneos-interações polarizadas

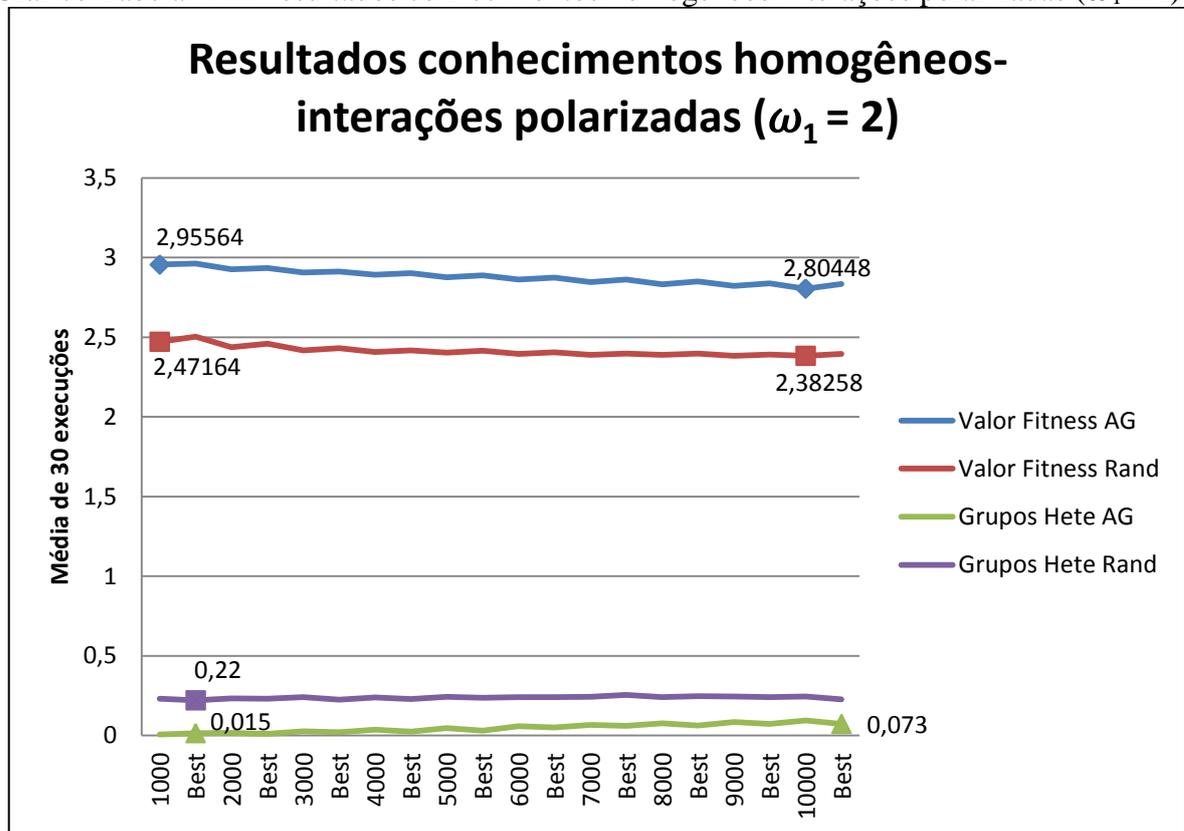
#Est	Média de 30 execuções									
	AG				Randômico				Diferença	
	Valor Fitness	+/-	Grupos Hete	+/-	Valor Fitness	+/-	Grupos Hete	+/-	Valor Fitness	Grupos Hete
1000	2,45377	0,01	0,00017	0,00	1,92009	0,01	0,21550	0,02	0,53369	-0,21533
Best	2,4666		0		1,9352		0,22			
2000	2,41210	0,00	0,00133	0,00	1,88547	0,00	0,22875	0,02	0,52663	-0,22742
Best	2,4217		0		1,8962		0,2025			
3000	2,37453	0,00	0,00467	0,00	1,87332	0,00	0,22783	0,01	0,50121	-0,22316
Best	2,386		0,0033		1,8853		0,2233			
4000	2,34428	0,01	0,00996	0,00	1,86304	0,00	0,23171	0,01	0,48124	-0,22175
Best	2,3549		0,0063		1,8741		0,2362			
5000	2,31340	0,01	0,01970	0,01	1,85792	0,00	0,23457	0,01	0,45548	-0,21487
Best	2,328		0,017		1,8707		0,224			
6000	2,28987	0,01	0,03473	0,01	1,85196	0,00	0,23456	0,01	0,43791	-0,19984
Best	2,3101		0,0158		1,8586		0,2325			
7000	2,27089	0,01	0,04782	0,01	1,84953	0,00	0,23838	0,01	0,42136	-0,19056
Best	2,2897		0,035		1,8572		0,2371			
8000	2,24830	0,02	0,06325	0,01	1,84736	0,00	0,23506	0,01	0,40094	-0,17181
Best	2,2742		0,04		1,857		0,2213			
9000	2,23130	0,02	0,07131	0,01	1,84503	0,00	0,23713	0,01	0,38627	-0,16582
Best	2,2618		0,045		1,8548		0,2356			
10000	2,21695	0,01	0,07425	0,01	1,84355	0,00	0,24133	0,01	0,37340	-0,16708
Best	2,2542		0,053		1,8508		0,24			

Fonte: Próprio autor

Tabela 14 – Resultados conhecimentos homogêneos-interações polarizadas ($\omega_1 = 2$)

Média de 30 execuções										
#Est	AG				Randômico				Diferença	
	Valor Fitness	+/-	Grupos Hete	+/-	Valor Fitness	+/-	Grupos Hete	+/-	Valor Fitness	Grupos Hete
1000	2,95564	0,00	0,00550	0,00	2,47164	0,01	0,23083	0,03	0,48400	-0,22533
<i>Best</i>	2,9617		0,015		2,503		0,22			
2000	2,92569	0,00	0,01525	0,00	2,43659	0,01	0,23167	0,01	0,48910	-0,21642
<i>Best</i>	2,9348		0,01		2,4593		0,23			
3000	2,90598	0,00	0,02695	0,00	2,41716	0,01	0,23988	0,01	0,48882	-0,21293
<i>Best</i>	2,913		0,02		2,4306		0,225			
4000	2,89224	0,00	0,03676	0,01	2,40684	0,00	0,23921	0,01	0,48540	-0,20245
<i>Best</i>	2,9021		0,025		2,418		0,2288			
5000	2,87605	0,01	0,04633	0,01	2,40295	0,01	0,24260	0,01	0,47310	-0,19627
<i>Best</i>	2,8891		0,031		2,4156		0,236			
6000	2,86162	0,01	0,05799	0,01	2,39445	0,01	0,24061	0,01	0,46717	-0,18262
<i>Best</i>	2,8736		0,0492		2,4042		0,2408			
7000	2,84686	0,01	0,06694	0,01	2,39008	0,00	0,24188	0,01	0,45678	-0,17494
<i>Best</i>	2,8614		0,0607		2,3964		0,255			
8000	2,83268	0,01	0,07692	0,01	2,38821	0,00	0,24081	0,01	0,44447	-0,16389
<i>Best</i>	2,8499		0,0625		2,3974		0,2475			
9000	2,82216	0,01	0,08341	0,01	2,38360	0,00	0,24386	0,01	0,43856	-0,16045
<i>Best</i>	2,8382		0,0722		2,3911		0,2406			
10000	2,80448	0,01	0,09362	0,01	2,38258	0,00	0,24385	0,01	0,42190	-0,15023
<i>Best</i>	2,8343		0,073		2,3945		0,2275			

Fonte: Próprio autor

Gráfico Tabela 14 – Resultados conhecimentos homogêneos-interações polarizadas ($\omega_1 = 2$)

Fonte: Próprio autor

Tabela 15 – Resultados conhecimentos homogêneos-interações polarizadas ($\omega_2 = 2$)

Média de 30 execuções										
#Est	AG				Randômico				Diferença	
	Valor <i>Fitness</i>	+/-	Grupos Hete	+/-	Valor <i>Fitness</i>	+/-	Grupos Hete	+/-	Valor <i>Fitness</i>	Grupos Hete
1000	3,43216	0,01	0,00017	0,00	2,44810	0,01	0,21567	0,02	0,98406	-0,21550
<i>Best</i>	3,4459		0		2,4689		0,185			
2000	3,36337	0,01	0,00075	0,00	2,40570	0,01	0,22325	0,02	0,95767	-0,22250
<i>Best</i>	3,3848		0		2,4222		0,2175			
3000	3,30204	0,01	0,00417	0,00	2,38524	0,01	0,23061	0,01	0,91681	-0,22644
<i>Best</i>	3,3241		0,0017		2,3995		0,2233			
4000	3,23791	0,01	0,01121	0,00	2,37603	0,01	0,23017	0,01	0,86188	-0,21896
<i>Best</i>	3,2726		0,0075		2,3898		0,2325			
5000	3,18778	0,02	0,01923	0,01	2,36712	0,00	0,23343	0,01	0,82066	-0,21420
<i>Best</i>	3,2226		0,008		2,3802		0,233			
6000	3,13440	0,02	0,02573	0,01	2,36123	0,00	0,23711	0,01	0,77317	-0,21138
<i>Best</i>	3,1979		0,0117		2,3745		0,2342			
7000	3,09199	0,03	0,03337	0,01	2,35825	0,01	0,23488	0,01	0,73374	-0,20151
<i>Best</i>	3,1349		0,0236		2,3697		0,23			
8000	3,04654	0,03	0,04507	0,01	2,35274	0,00	0,23540	0,01	0,69380	-0,19033
<i>Best</i>	3,0988		0,0356		2,3614		0,2338			
9000	3,00107	0,03	0,05526	0,01	2,35225	0,00	0,23556	0,01	0,64882	-0,18030
<i>Best</i>	3,0517		0,0444		2,3621		0,2322			
10000	2,98558	0,03	0,05827	0,01	2,34720	0,00	0,23862	0,01	0,63838	-0,18035
<i>Best</i>	3,0484		0,047		2,3543		0,238			

Fonte: Próprio autor

B.6 RESULTADOS OBTIDOS PARA CONHECIMENTOS HOMOGÊNEOS-INTERAÇÕES ALEATÓRIAS

Tabela 16 – Resultados conhecimentos homogêneos-interações aleatórias

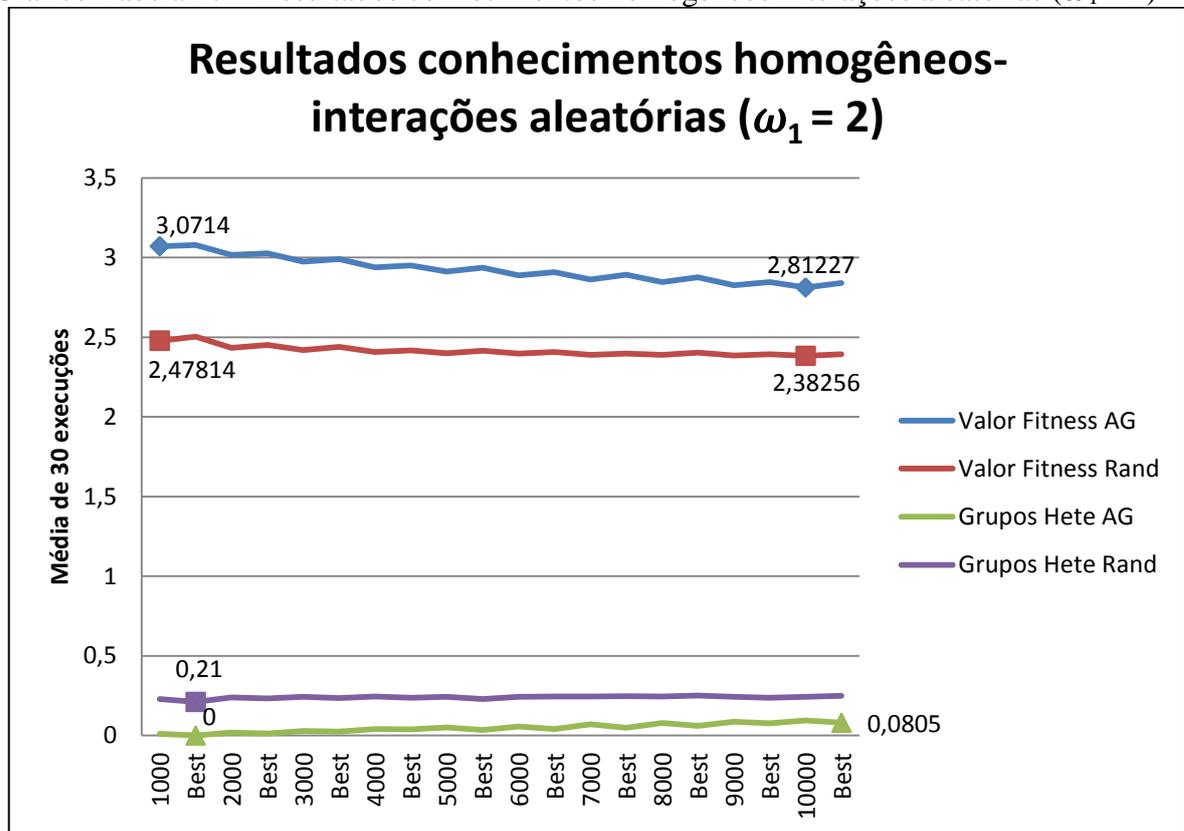
Média de 30 execuções										
#Est	AG				Randômico				Diferença	
	Valor <i>Fitness</i>	+/-	Grupos Hete	+/-	Valor <i>Fitness</i>	+/-	Grupos Hete	+/-	Valor <i>Fitness</i>	Grupos Hete
1000	2,40452	0,00	0,00883	0,00	1,91683	0,01	0,22117	0,02	0,48769	-0,21233
<i>Best</i>	2,4138		0,01		1,9285		0,215			
2000	2,35529	0,00	0,01783	0,00	1,88868	0,01	0,23200	0,01	0,46662	-0,21417
<i>Best</i>	2,3644		0,0125		1,9059		0,215			
3000	2,31633	0,01	0,02723	0,00	1,87250	0,01	0,23050	0,01	0,44383	-0,20327
<i>Best</i>	2,3307		0,0233		1,8868		0,225			
4000	2,28711	0,01	0,03495	0,00	1,86215	0,00	0,23650	0,01	0,42496	-0,20154
<i>Best</i>	2,2988		0,0288		1,8717		0,2137			
5000	2,25786	0,01	0,04620	0,01	1,85658	0,00	0,23963	0,01	0,40127	-0,19343
<i>Best</i>	2,2682		0,048		1,8656		0,242			
6000	2,23660	0,01	0,05414	0,01	1,85263	0,00	0,23722	0,01	0,38398	-0,18309
<i>Best</i>	2,2547		0,0417		1,8578		0,2358			
7000	2,21816	0,01	0,06123	0,01	1,84921	0,00	0,23921	0,01	0,36895	-0,17798
<i>Best</i>	2,2385		0,0536		1,8561		0,2293			
8000	2,20263	0,01	0,06943	0,00	1,84692	0,00	0,23935	0,01	0,35571	-0,16992
<i>Best</i>	2,2229		0,0587		1,8538		0,2381			
9000	2,18534	0,01	0,07850	0,01	1,84530	0,00	0,24001	0,01	0,34003	-0,16151
<i>Best</i>	2,203		0,0622		1,8552		0,2361			
10000	2,17360	0,01	0,08472	0,01	1,84317	0,00	0,24278	0,01	0,33043	-0,15807
<i>Best</i>	2,1949		0,067		1,851		0,233			

Fonte: Próprio autor

Tabela 17 – Resultados conhecimentos homogêneos-interações aleatórias ($\omega_1 = 2$)

Média de 30 execuções											
#Est	AG				Randômico				Diferença		
	Valor <i>Fitness</i>	+/-	Grupos Hete	+/-	Valor <i>Fitness</i>	+/-	Grupos Hete	+/-	Valor <i>Fitness</i>	Grupos Hete	
1000	3,07140	0,00	0,00917	0,00	2,47814	0,01	0,22900	0,02	0,59326	-0,21983	
<i>Best</i>	3,0785		0		2,5027		0,21				
2000	3,01626	0,00	0,01825	0,00	2,43419	0,01	0,23917	0,02	0,58208	-0,22092	
<i>Best</i>	3,0267		0,0125		2,4513		0,2325				
3000	2,97442	0,01	0,02827	0,00	2,41871	0,01	0,24355	0,01	0,55571	-0,21528	
<i>Best</i>	2,9905		0,025		2,4391		0,235				
4000	2,93771	0,01	0,04095	0,01	2,40775	0,00	0,24432	0,01	0,52996	-0,20338	
<i>Best</i>	2,9507		0,0387		2,4175		0,2375				
5000	2,91256	0,01	0,04950	0,01	2,39931	0,01	0,24267	0,01	0,51324	-0,19317	
<i>Best</i>	2,936		0,034		2,4157		0,228				
6000	2,88898	0,01	0,05678	0,01	2,39655	0,00	0,24295	0,01	0,49244	-0,18616	
<i>Best</i>	2,9088		0,0408		2,4079		0,2442				
7000	2,86242	0,01	0,06922	0,01	2,38989	0,00	0,24511	0,01	0,47254	-0,17589	
<i>Best</i>	2,8916		0,0479		2,3981		0,2471				
8000	2,84577	0,01	0,07723	0,01	2,38947	0,00	0,24530	0,00	0,45630	-0,16807	
<i>Best</i>	2,8761		0,0594		2,4024		0,2512				
9000	2,82659	0,01	0,08532	0,01	2,38468	0,00	0,24305	0,01	0,44191	-0,15774	
<i>Best</i>	2,8466		0,0761		2,3928		0,2367				
10000	2,81227	0,01	0,09412	0,01	2,38256	0,00	0,24317	0,01	0,42971	-0,14905	
<i>Best</i>	2,84		0,0805		2,393		0,2495				

Fonte: Próprio autor

Gráfico Tabela 17 – Resultados conhecimentos homogêneos-interações aleatórias ($\omega_1 = 2$)

Fonte: Próprio autor

Tabela 18 – Resultados conhecimentos homogêneos-interações aleatórias ($\omega_2 = 2$)

Média de 30 execuções										
#Est	AG				Randômico				Diferença	
	Valor <i>Fitness</i>	+/-	Grupos Hete	+/-	Valor <i>Fitness</i>	+/-	Grupos Hete	+/-	Valor <i>Fitness</i>	Grupos Hete
1000	3,15098	0,01	0,02117	0,01	2,42748	0,01	0,22267	0,02	0,72350	-0,20150
<i>Best</i>	3,1647		0,015		2,449		0,205			
2000	3,06254	0,01	0,03633	0,01	2,39095	0,01	0,23100	0,01	0,67159	-0,19467
<i>Best</i>	3,0776		0,035		2,4039		0,23			
3000	2,99730	0,01	0,05056	0,00	2,37771	0,00	0,23523	0,01	0,61959	-0,18466
<i>Best</i>	3,0094		0,0483		2,3896		0,21			
4000	2,94793	0,01	0,06463	0,00	2,36812	0,01	0,23587	0,01	0,57981	-0,17124
<i>Best</i>	2,9626		0,06		2,3798		0,2288			
5000	2,90710	0,01	0,07327	0,01	2,36170	0,00	0,23850	0,01	0,54539	-0,16523
<i>Best</i>	2,9253		0,072		2,3699		0,242			
6000	2,87177	0,01	0,08261	0,01	2,35647	0,00	0,24106	0,01	0,51530	-0,15845
<i>Best</i>	2,891		0,085		2,3655		0,2342			
7000	2,84080	0,02	0,09203	0,01	2,35220	0,00	0,23937	0,01	0,48860	-0,14734
<i>Best</i>	2,8692		0,0857		2,3596		0,2407			
8000	2,81112	0,02	0,10076	0,01	2,34989	0,00	0,23905	0,01	0,46123	-0,13829
<i>Best</i>	2,8456		0,0894		2,3556		0,2319			
9000	2,79207	0,02	0,10752	0,01	2,34646	0,00	0,24047	0,01	0,44560	-0,13294
<i>Best</i>	2,8192		0,095		2,3572		0,2433			
10000	2,77194	0,01	0,11463	0,01	2,34456	0,00	0,24100	0,01	0,42737	-0,12637
<i>Best</i>	2,7942		0,111		2,3514		0,252			

Fonte: Próprio autor

B.7 RESULTADOS OBTIDOS PARA CONHECIMENTOS ALEATÓRIOS-INTERAÇÕES UNIFORMES

Tabela 19 – Resultados conhecimentos aleatórios-interações uniformes

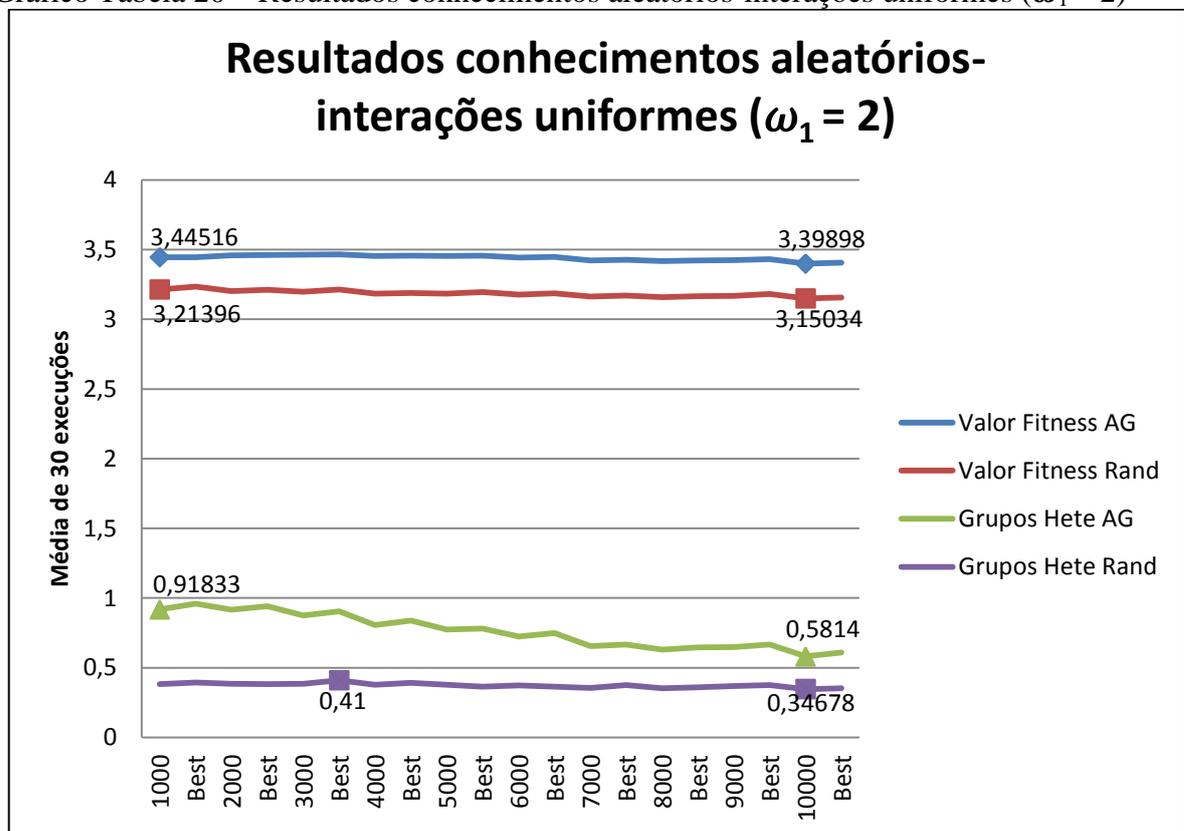
Média de 30 execuções										
#Est	AG				Randômico				Diferença	
	Valor <i>Fitness</i>	+/-	Grupos Hete	+/-	Valor <i>Fitness</i>	+/-	Grupos Hete	+/-	Valor <i>Fitness</i>	Grupos Hete
1000	2,72215	0,00	0,91167	0,02	2,58095	0,00	0,37500	0,02	0,14120	0,53667
<i>Best</i>	2,7225		0,965		2,5946		0,345			
2000	2,72880	0,00	0,91792	0,01	2,57202	0,00	0,37633	0,02	0,15679	0,54158
<i>Best</i>	2,7295		0,9225		2,5804		0,395			
3000	2,73047	0,00	0,88862	0,01	2,56805	0,00	0,38111	0,01	0,16242	0,50751
<i>Best</i>	2,7315		0,885		2,5705		0,3767			
4000	2,72511	0,00	0,81345	0,02	2,55920	0,00	0,37645	0,01	0,16591	0,43700
<i>Best</i>	2,7272		0,845		2,5635		0,3725			
5000	2,72440	0,00	0,78910	0,02	2,55881	0,00	0,37417	0,01	0,16559	0,41493
<i>Best</i>	2,7271		0,826		2,5649		0,405			
6000	2,71793	0,00	0,72885	0,02	2,55379	0,00	0,37147	0,01	0,16414	0,35738
<i>Best</i>	2,7224		0,7692		2,5601		0,3825			
7000	2,70694	0,00	0,65272	0,02	2,54583	0,00	0,35114	0,01	0,16111	0,30157
<i>Best</i>	2,7105		0,6836		2,5494		0,3414			
8000	2,70435	0,00	0,63876	0,02	2,54306	0,00	0,34993	0,01	0,16129	0,28883
<i>Best</i>	2,7081		0,66		2,5453		0,3606			
9000	2,70692	0,00	0,65230	0,02	2,54824	0,00	0,36261	0,01	0,15868	0,28969
<i>Best</i>	2,7125		0,6961		2,5524		0,3567			
10000	2,69360	0,00	0,57850	0,02	2,53870	0,00	0,34442	0,01	0,15490	0,23408
<i>Best</i>	2,6998		0,605		2,5427		0,36			

Fonte: Próprio autor

Tabela 20 – Resultados conhecimentos aleatórios-interações uniformes ($\omega_1 = 2$)

Média de 30 execuções										
#Est	AG				Randômico				Diferença	
	Valor Fitness	+/-	Grupos Hete	+/-	Valor Fitness	+/-	Grupos Hete	+/-	Valor Fitness	Grupos Hete
1000	3,44516	0,00	0,91833	0,02	3,21396	0,01	0,38283	0,03	0,23120	0,53550
<i>Best</i>	3,4454		0,96		3,2344		0,395			
2000	3,45950	0,00	0,91633	0,02	3,20288	0,00	0,38542	0,01	0,25662	0,53092
<i>Best</i>	3,4603		0,9425		3,2119		0,3825			
3000	3,46390	0,00	0,87494	0,01	3,19729	0,00	0,38516	0,01	0,26661	0,48978
<i>Best</i>	3,4652		0,905		3,2127		0,41			
4000	3,45447	0,00	0,80546	0,01	3,18381	0,00	0,37733	0,01	0,27066	0,42813
<i>Best</i>	3,4565		0,8387		3,189		0,3925			
5000	3,45450	0,00	0,77440	0,01	3,18414	0,00	0,37697	0,01	0,27035	0,39743
<i>Best</i>	3,4575		0,782		3,1945		0,365			
6000	3,44360	0,00	0,72333	0,02	3,17638	0,00	0,37311	0,01	0,26722	0,35022
<i>Best</i>	3,4474		0,75		3,1857		0,3642			
7000	3,42275	0,00	0,65521	0,01	3,16286	0,00	0,35562	0,01	0,25988	0,29959
<i>Best</i>	3,4278		0,6664		3,1698		0,3757			
8000	3,41714	0,00	0,63090	0,01	3,15862	0,00	0,35257	0,01	0,25852	0,27833
<i>Best</i>	3,423		0,6469		3,1665		0,3606			
9000	3,42522	0,00	0,64769	0,01	3,16866	0,00	0,36964	0,01	0,25656	0,27805
<i>Best</i>	3,4317		0,6656		3,1808		0,3767			
10000	3,39898	0,01	0,58140	0,01	3,15034	0,00	0,34678	0,01	0,24863	0,23462
<i>Best</i>	3,407		0,61		3,1568		0,353			

Fonte: Próprio autor

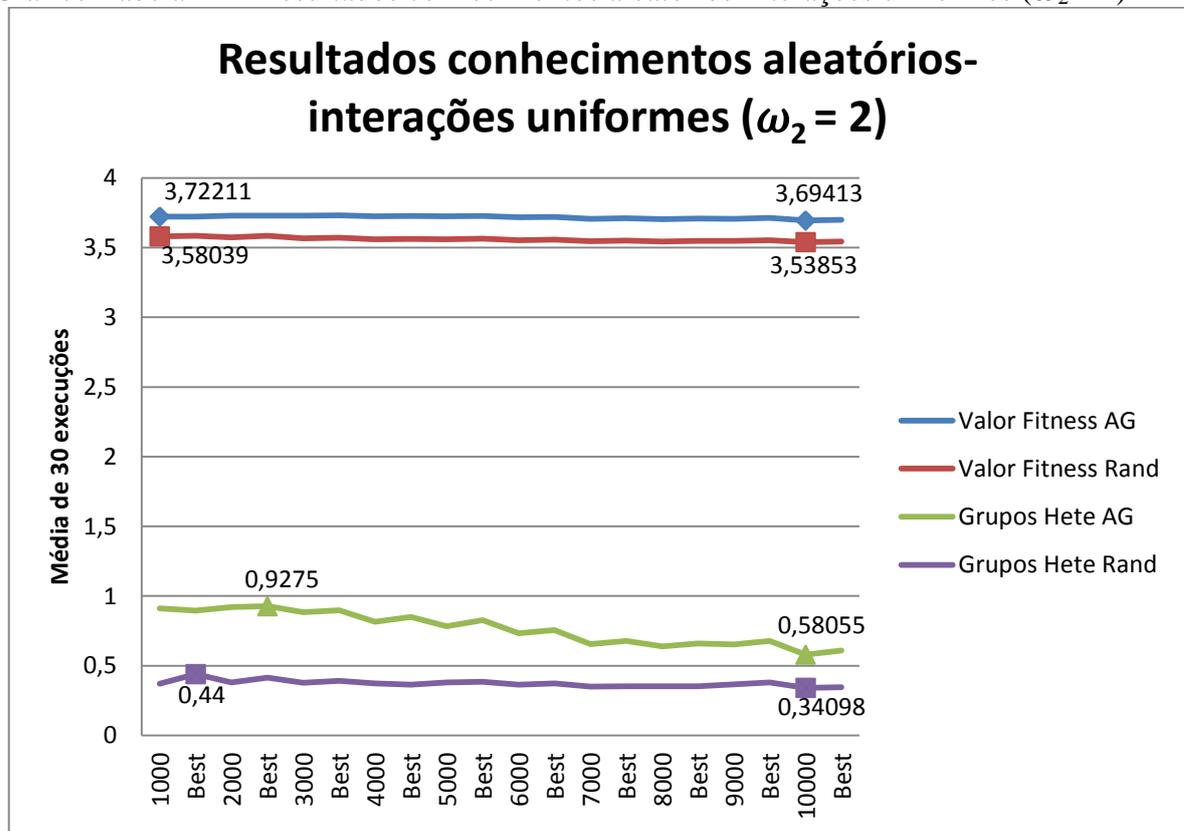
Gráfico Tabela 20 – Resultados conhecimentos aleatórios-interações uniformes ($\omega_1 = 2$)

Fonte: Próprio autor

Tabela 21 – Resultados conhecimentos aleatórios-interações uniformes ($\omega_2 = 2$)

Média de 30 execuções										
#Est	AG				Randômico				Diferença	
	Valor Fitness	+/-	Grupos Hete	+/-	Valor Fitness	+/-	Grupos Hete	+/-	Valor Fitness	Grupos Hete
1000	3,72211	0,00	0,91283	0,02	3,58039	0,00	0,37183	0,03	0,14172	0,54100
<i>Best</i>	3,7225		0,895		3,5841		0,44			
2000	3,72879	0,00	0,92100	0,01	3,57276	0,00	0,38075	0,02	0,15603	0,54025
<i>Best</i>	3,7294		0,9275		3,5854		0,415			
3000	3,73031	0,00	0,88377	0,02	3,56674	0,00	0,37711	0,01	0,16357	0,50666
<i>Best</i>	3,7316		0,8983		3,5719		0,3917			
4000	3,72504	0,00	0,81617	0,01	3,55865	0,00	0,37280	0,01	0,16639	0,44337
<i>Best</i>	3,7265		0,8488		3,5627		0,365			
5000	3,72417	0,00	0,78403	0,02	3,55882	0,00	0,38077	0,01	0,16535	0,40327
<i>Best</i>	3,7263		0,827		3,5636		0,386			
6000	3,71849	0,00	0,73306	0,02	3,55365	0,00	0,36367	0,01	0,16485	0,36940
<i>Best</i>	3,721		0,7567		3,558		0,3742			
7000	3,70697	0,00	0,65464	0,02	3,54572	0,00	0,35067	0,01	0,16125	0,30397
<i>Best</i>	3,7112		0,6771		3,5503		0,3521			
8000	3,70450	0,00	0,63878	0,02	3,54375	0,00	0,35378	0,01	0,16074	0,28500
<i>Best</i>	3,7078		0,66		3,5486		0,3519			
9000	3,70734	0,00	0,65205	0,02	3,54823	0,00	0,36599	0,01	0,15911	0,28606
<i>Best</i>	3,7124		0,6783		3,5529		0,38			
10000	3,69413	0,00	0,58055	0,01	3,53853	0,00	0,34098	0,01	0,15560	0,23957
<i>Best</i>	3,6992		0,609		3,5426		0,3465			

Fonte: Próprio autor

Gráfico Tabela 21 – Resultados conhecimentos aleatórios-interações uniformes ($\omega_2 = 2$)

Fonte: Próprio autor

B.8 RESULTADOS OBTIDOS PARA CONHECIMENTOS ALEATÓRIOS- INTERAÇÕES POLARIZADAS

Tabela 22 – Resultados conhecimentos aleatórios-interações polarizadas

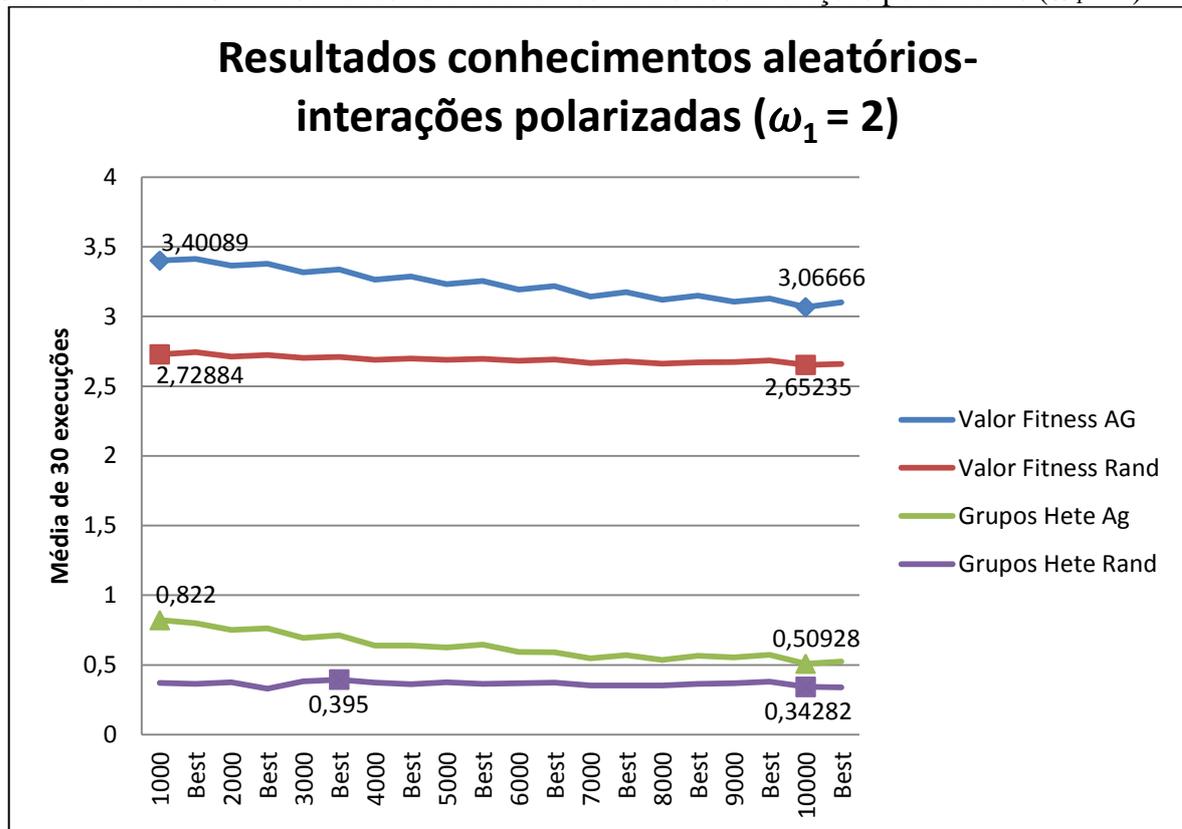
Média de 30 execuções										
#Est	AG				Randômico				Diferença	
	Valor Fitness	+/-	Grupos Hete	+/-	Valor Fitness	+/-	Grupos Hete	+/-	Valor Fitness	Grupos Hete
1000	2,68475	0,01	0,81650	0,02	2,09959	0,01	0,35600	0,02	0,58516	0,46050
<i>Best</i>	2,6986		0,835		2,1113		0,355			
2000	2,65234	0,01	0,74792	0,02	2,08540	0,00	0,36800	0,02	0,56694	0,37992
<i>Best</i>	2,6617		0,73		2,0915		0,39			
3000	2,61173	0,01	0,68184	0,02	2,07581	0,00	0,37411	0,01	0,53592	0,30772
<i>Best</i>	2,6286		0,6833		2,0922		0,38			
4000	2,56852	0,01	0,61325	0,02	2,06646	0,00	0,36551	0,01	0,50206	0,24774
<i>Best</i>	2,5931		0,64		2,0727		0,3525			
5000	2,53468	0,01	0,59567	0,01	2,06463	0,00	0,36987	0,01	0,47005	0,22580
<i>Best</i>	2,5479		0,618		2,0696		0,358			
6000	2,49797	0,01	0,55764	0,01	2,05946	0,00	0,36819	0,01	0,43851	0,18945
<i>Best</i>	2,5182		0,5633		2,0669		0,3783			
7000	2,46496	0,01	0,51028	0,01	2,05185	0,00	0,34700	0,01	0,41311	0,16328
<i>Best</i>	2,486		0,5214		2,0586		0,3421			
8000	2,43982	0,02	0,49604	0,01	2,04857	0,00	0,34763	0,01	0,39126	0,14841
<i>Best</i>	2,465		0,4963		2,0532		0,3519			
9000	2,43208	0,01	0,51565	0,01	2,05402	0,00	0,36489	0,01	0,37806	0,15076
<i>Best</i>	2,4522		0,5183		2,0586		0,3556			
10000	2,40386	0,01	0,47053	0,01	2,04333	0,00	0,34002	0,01	0,36053	0,13052
<i>Best</i>	2,4249		0,475		2,0517		0,3355			

Fonte: Próprio autor

Tabela 23 – Resultados conhecimentos aleatórios-interações polarizadas ($\omega_1 = 2$)

Média de 30 execuções										
#Est	AG				Randômico				Diferença	
	Valor Fitness	+/-	Grupos Hete	+/-	Valor Fitness	+/-	Grupos Hete	+/-	Valor Fitness	Grupos Hete
1000	3,40089	0,01	0,82200	0,02	2,72884	0,01	0,37000	0,02	0,67204	0,45200
<i>Best</i>	3,4142		0,8		2,7452		0,365			
2000	3,36397	0,01	0,75125	0,02	2,71235	0,01	0,37642	0,02	0,65162	0,37483
<i>Best</i>	3,3783		0,7625		2,7232		0,33			
3000	3,31710	0,01	0,69321	0,02	2,70397	0,00	0,38306	0,01	0,61313	0,31015
<i>Best</i>	3,3368		0,7133		2,7101		0,395			
4000	3,26432	0,01	0,63887	0,01	2,69005	0,00	0,37418	0,01	0,57427	0,26469
<i>Best</i>	3,2876		0,64		2,6982		0,3613			
5000	3,23179	0,01	0,62590	0,01	2,68901	0,00	0,37673	0,01	0,54278	0,24917
<i>Best</i>	3,2549		0,645		2,697		0,364			
6000	3,19280	0,01	0,59439	0,01	2,68124	0,00	0,36950	0,01	0,51156	0,22489
<i>Best</i>	3,2185		0,5908		2,6921		0,3733			
7000	3,14313	0,02	0,54831	0,01	2,66644	0,00	0,35187	0,01	0,47669	0,19644
<i>Best</i>	3,1755		0,5707		2,6779		0,3521			
8000	3,12093	0,02	0,53636	0,01	2,66270	0,00	0,35176	0,01	0,45823	0,18460
<i>Best</i>	3,1493		0,565		2,6708		0,3638			
9000	3,10726	0,02	0,55536	0,01	2,67212	0,00	0,36907	0,01	0,43514	0,18629
<i>Best</i>	3,13		0,5717		2,6841		0,38			
10000	3,06666	0,02	0,50928	0,01	2,65235	0,00	0,34282	0,01	0,41431	0,16647
<i>Best</i>	3,1015		0,524		2,6591		0,3395			

Fonte: Próprio autor

Gráfico Tabela 23 – Resultados conhecimentos aleatórios-interações polarizadas ($\omega_1 = 2$)

Fonte: Próprio autor

Tabela 24 – Resultados conhecimentos aleatórios-interações polarizadas ($\omega_2 = 2$)

Média de 30 execuções

#Est	AG				Randômico				Diferença	
	Valor Fitness	+/-	Grupos Hete	+/-	Valor Fitness	+/-	Grupos Hete	+/-	Valor Fitness	Grupos Hete
1000	3,66104	0,01	0,81083	0,02	2,63857	0,01	0,35400	0,03	1,02247	0,45683
<i>Best</i>	3,6744		0,82		2,6719		0,39			
2000	3,60063	0,01	0,73108	0,02	2,60924	0,01	0,36233	0,02	0,99139	0,36875
<i>Best</i>	3,6174		0,7475		2,6228		0,355			
3000	3,53543	0,02	0,66205	0,02	2,59751	0,00	0,37411	0,02	0,93792	0,28794
<i>Best</i>	3,5632		0,6933		2,6097		0,365			
4000	3,46852	0,02	0,59216	0,02	2,58303	0,00	0,35939	0,01	0,88549	0,23277
<i>Best</i>	3,5035		0,6		2,5952		0,3538			
5000	3,40566	0,02	0,56570	0,01	2,57967	0,00	0,36377	0,01	0,82599	0,20193
<i>Best</i>	3,4689		0,59		2,5868		0,353			
6000	3,34475	0,02	0,52322	0,01	2,57316	0,00	0,36288	0,01	0,77159	0,16033
<i>Best</i>	3,3905		0,5475		2,5846		0,365			
7000	3,29813	0,02	0,48404	0,01	2,56377	0,00	0,34702	0,01	0,73436	0,13702
<i>Best</i>	3,3387		0,4871		2,5728		0,3493			
8000	3,25316	0,03	0,46893	0,01	2,55922	0,00	0,34705	0,01	0,69394	0,12188
<i>Best</i>	3,3125		0,4713		2,566		0,3262			
9000	3,22440	0,03	0,47641	0,01	2,56405	0,00	0,36085	0,01	0,66035	0,11557
<i>Best</i>	3,2752		0,4917		2,5728		0,36			
10000	3,17076	0,03	0,43558	0,01	2,55269	0,00	0,33623	0,01	0,61807	0,09935
<i>Best</i>	3,2403		0,4665		2,5642		0,33			

Fonte: Próprio autor

B.9 RESULTADOS OBTIDOS PARA CONHECIMENTOS ALEATÓRIOS- INTERAÇÕES ALEATÓRIAS

Tabela 25 – Resultados conhecimentos aleatórios-interações aleatórias

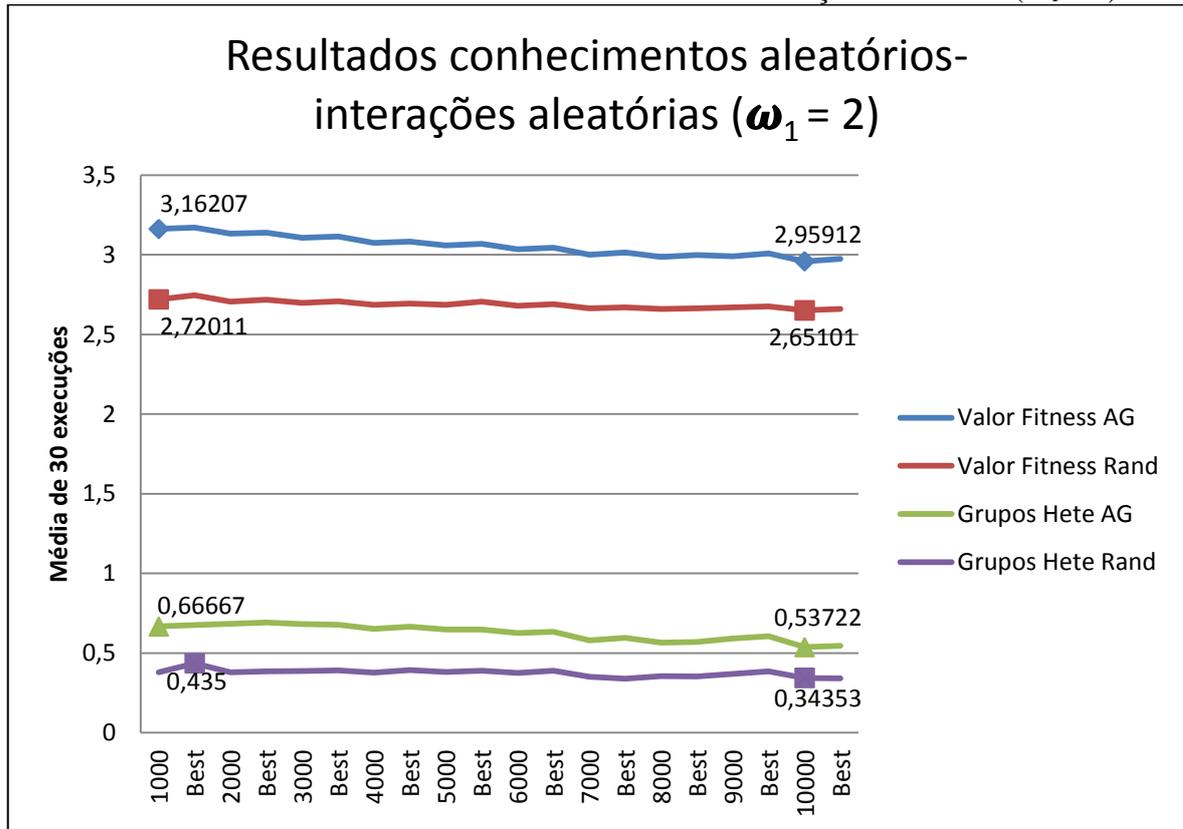
Média de 30 execuções										
#Est	AG				Randômico				Diferença	
	Valor <i>Fitness</i>	+/-	Grupos Hete	+/-	Valor <i>Fitness</i>	+/-	Grupos Hete	+/-	Valor <i>Fitness</i>	Grupos Hete
1000	2,44915	0,00	0,61567	0,02	2,08810	0,00	0,36467	0,02	0,36105	0,25100
<i>Best</i>	2,4557		0,6		2,1003		0,4			
2000	2,41683	0,00	0,63167	0,02	2,07579	0,00	0,36600	0,02	0,34104	0,26567
<i>Best</i>	2,4235		0,625		2,0814		0,37			
3000	2,39043	0,00	0,62967	0,01	2,07099	0,00	0,37961	0,01	0,31944	0,25006
<i>Best</i>	2,3967		0,6183		2,0754		0,3717			
4000	2,36506	0,01	0,60359	0,01	2,06183	0,00	0,36843	0,01	0,30323	0,23516
<i>Best</i>	2,3757		0,6162		2,0672		0,3762			
5000	2,34492	0,00	0,59920	0,01	2,06048	0,00	0,37307	0,01	0,28444	0,22613
<i>Best</i>	2,3516		0,596		2,0657		0,382			
6000	2,32744	0,01	0,57527	0,01	2,05615	0,00	0,36667	0,01	0,27129	0,20861
<i>Best</i>	2,3409		0,5867		2,0639		0,3633			
7000	2,30704	0,01	0,53683	0,01	2,04766	0,00	0,35101	0,01	0,25938	0,18582
<i>Best</i>	2,3184		0,5479		2,0503		0,3371			
8000	2,29337	0,01	0,52619	0,01	2,04575	0,00	0,34969	0,01	0,24762	0,17650
<i>Best</i>	2,3088		0,5363		2,0527		0,3625			
9000	2,28996	0,01	0,54810	0,01	2,04995	0,00	0,36326	0,01	0,24001	0,18484
<i>Best</i>	2,3031		0,5644		2,0531		0,3539			
10000	2,26975	0,01	0,49987	0,01	2,04076	0,00	0,34460	0,01	0,22899	0,15527
<i>Best</i>	2,2856		0,523		2,0445		0,339			

Fonte: Próprio autor

Tabela 26 – Resultados conhecimentos aleatórios-interações aleatórias ($\omega_1 = 2$)

Média de 30 execuções										
#Est	AG				Randômico				Diferença	
	Valor <i>Fitness</i>	+/-	Grupos Hete	+/-	Valor <i>Fitness</i>	+/-	Grupos Hete	+/-	Valor <i>Fitness</i>	Grupos Hete
1000	3,16207	0,00	0,66667	0,02	2,72011	0,01	0,37850	0,02	0,44196	0,28817
<i>Best</i>	3,1709		0,675		2,7465		0,435			
2000	3,13195	0,00	0,68275	0,01	2,70649	0,01	0,37942	0,02	0,42546	0,30333
<i>Best</i>	3,139		0,6925		2,7179		0,385			
3000	3,10620	0,00	0,68134	0,01	2,69863	0,00	0,38600	0,01	0,40757	0,29534
<i>Best</i>	3,1138		0,6783		2,7081		0,39			
4000	3,07474	0,00	0,65133	0,01	2,68617	0,00	0,37608	0,01	0,38857	0,27525
<i>Best</i>	3,0831		0,665		2,6939		0,3937			
5000	3,05840	0,01	0,64770	0,01	2,68650	0,00	0,38177	0,01	0,37189	0,26593
<i>Best</i>	3,0691		0,647		2,705		0,389			
6000	3,03425	0,01	0,62489	0,01	2,67908	0,00	0,37408	0,01	0,35517	0,25081
<i>Best</i>	3,0452		0,6333		2,6901		0,3892			
7000	3,00119	0,01	0,57920	0,01	2,66380	0,00	0,35164	0,01	0,33739	0,22755
<i>Best</i>	3,0143		0,5957		2,6699		0,3379			
8000	2,98671	0,01	0,56531	0,01	2,65989	0,00	0,35524	0,01	0,32682	0,21007
<i>Best</i>	2,9976		0,5687		2,6645		0,3538			
9000	2,98978	0,01	0,59180	0,01	2,67026	0,00	0,36884	0,01	0,31952	0,22297
<i>Best</i>	3,0077		0,6056		2,6755		0,3839			
10000	2,95912	0,01	0,53722	0,01	2,65101	0,00	0,34353	0,01	0,30811	0,19368
<i>Best</i>	2,974		0,5445		2,6591		0,3405			

Fonte: Próprio autor

Gráfico Tabela 26 – Resultados conhecimentos aleatórios-interações aleatórias ($\omega_1 = 2$)

Fonte: Próprio autor

Tabela 27 – Resultados conhecimentos aleatórios-interações aleatórias ($\omega_2 = 2$)

#Est	AG				Randômico				Diferença	
	Valor Fitness	+/-	Grupos Hete	+/-	Valor Fitness	+/-	Grupos Hete	+/-	Valor Fitness	Grupos Hete
1000	3,21252	0,01	0,53600	0,03	2,60439	0,01	0,35817	0,02	0,60812	0,17783
Best	3,2277		0,54		2,619		0,37			
2000	3,14587	0,00	0,54900	0,02	2,58656	0,00	0,36725	0,02	0,55931	0,18175
Best	3,1558		0,575		2,5981		0,375			
3000	3,09700	0,01	0,55994	0,01	2,57798	0,00	0,37394	0,02	0,51902	0,18600
Best	3,1089		0,5417		2,5856		0,3567			
4000	3,05003	0,01	0,53476	0,01	2,56851	0,00	0,36776	0,01	0,48151	0,16700
Best	3,0669		0,5387		2,5785		0,3762			
5000	3,01527	0,01	0,53727	0,01	2,56585	0,00	0,37190	0,01	0,44942	0,16537
Best	3,0313		0,555		2,5706		0,384			
6000	2,98765	0,01	0,51463	0,01	2,56134	0,00	0,36637	0,01	0,42631	0,14827
Best	3,0036		0,515		2,5673		0,3883			
7000	2,95212	0,01	0,48329	0,01	2,55283	0,00	0,34762	0,01	0,39928	0,13567
Best	2,9761		0,49		2,5608		0,3643			
8000	2,93190	0,01	0,47546	0,01	2,55046	0,00	0,34463	0,01	0,38144	0,13083
Best	2,9586		0,4931		2,5541		0,3513			
9000	2,91940	0,01	0,49519	0,01	2,55445	0,00	0,35989	0,01	0,36496	0,13529
Best	2,9404		0,5056		2,5597		0,3633			
10000	2,88945	0,01	0,45280	0,01	2,54422	0,00	0,34050	0,01	0,34523	0,11230
Best	2,9062		0,4615		2,5533		0,323			

Fonte: Próprio autor

B.10 RESULTADOS OBTIDOS EM RELAÇÃO AO TEMPO

Tabela 28 – Tempo de processamento dos algoritmos (primeiro caso de teste)

Média de 30 execuções							
#Estudantes	AG			Randômico			Diferença
	Grupos Hete	Tempo (s)	+/-	Grupos Hete	Tempo (s)	+/-	
1000	1,00000	37,36667	8,24816	0,73783	28,86667	5,14933	8,50000
2000	1,00000	96,13333	13,20539	0,69450	74,16667	4,55400	21,96667
3000	1,00000	162,10000	24,08367	0,67800	185,76667	23,77209	-23,66667
4000	1,00000	389,63333	106,82930	0,66842	260,80000	52,86423	128,83333
5000	1,00000	536,60000	189,55889	0,66083	391,63333	146,20659	144,96667
6000	1,00000	712,13333	330,40599	0,65880	558,80000	194,55940	153,33333
7000	1,00000	964,00000	439,95371	0,65405	834,33333	354,75732	129,66667
8000	1,00000	1298,20000	680,16490	0,65450	1199,80000	527,78268	98,40000
9000	0,99994	1644,20000	784,65315	0,65168	1308,06667	600,84790	336,13333
10000	0,99980	1989,20000	921,08959	0,65107	1947,13333	728,07526	42,06667

Fonte: Próprio autor

Nas tabelas a seguir, as linhas correspondem à quantidade de estudantes, de mil até dez mil. As colunas apresentam o percentual de grupos heterogêneos e o tempo de processamento. Há ainda as colunas com a relação do tempo de processamento dividido pelos grupos heterogêneos para cada algoritmo (custo) e a coluna com a diferença entre o resultado do AG e o resultado do algoritmo randômico, para esta relação.

Tabela 29 – Relação Tempo / Grupos Heterogêneos (custo) - caso 1

Média de 30 execuções							
#Estudantes	AG		Randômico		Custo		Diferença
	Grupos Hete	Tempo (s)	Grupos Hete	Tempo (s)	AG	Rand	
1000	1,00000	37,36667	0,73783	28,86667	0,186833	0,195618	-0,00878
2000	1,00000	96,13333	0,69450	74,16667	0,240333	0,266979	-0,02665
3000	1,00000	162,10000	0,67800	185,76667	0,270167	0,456654	-0,18649
4000	1,00000	389,63333	0,66842	260,80000	0,487042	0,487715	-0,00067
5000	1,00000	536,60000	0,66083	391,63333	0,5366	0,592636	-0,05604
6000	1,00000	712,13333	0,65880	558,80000	0,593444	0,706837	-0,11339
7000	1,00000	964,00000	0,65405	834,33333	0,688571	0,911168	-0,2226
8000	1,00000	1298,20000	0,65450	1199,80000	0,811375	1,145722	-0,33435
9000	0,99994	1644,20000	0,65168	1308,06667	0,913499	1,115129	-0,20163
10000	0,99980	1989,20000	0,65107	1947,13333	0,994799	1,495341	-0,50054

Fonte: Próprio autor

Tabela 30 – Relação Tempo / Grupos Heterogêneos (custo) - caso 2

Média de 30 execuções							
#Estudantes	AG		Randômico		Custo		Diferença
	Grupos Hete	Tempo (s)	Grupos Hete	Tempo (s)	AG	Rand	
1000	1,00000	43,90000	0,69367	38,73333	0,2195	0,279193	-0,05969
2000	0,90242	97,83333	0,65633	79,00000	0,271031	0,300914	-0,02988
3000	1,00000	162,00000	0,65255	112,23333	0,27	0,286653	-0,01665
4000	0,96854	255,40000	0,64601	193,53333	0,329619	0,374478	-0,04486
5000	1,00000	386,23333	0,64653	333,30000	0,386233	0,515519	-0,12929
6000	1,00000	625,30000	0,64616	562,36667	0,521083	0,725264	-0,20418
7000	0,96990	904,56667	0,63970	955,40000	0,666168	1,066789	-0,40062
8000	0,97036	1263,30000	0,63660	1223,63333	0,813683	1,201337	-0,38765
9000	0,99996	1611,43333	0,63896	1517,06667	0,895277	1,319048	-0,42377
10000	0,99973	1973,80000	0,63463	1818,23333	0,987163	1,432507	-0,44534

Fonte: Próprio autor

Tabela 31 – Relação Tempo / Grupos Heterogêneos (custo) - caso 3

Média de 30 execuções							
#Estudantes	AG		Randômico		Custo		Diferença
	Grupos Hete	Tempo (s)	Grupos Hete	Tempo (s)	AG	Rand	
1000	1,00000	50,33333	0,71817	35,66667	0,251667	0,248317	0,003349
2000	1,00000	104,13333	0,68625	72,26667	0,260333	0,263267	-0,00293
3000	0,99994	164,80000	0,66877	116,16667	0,274682	0,289505	-0,01482
4000	0,99984	267,73333	0,66320	207,70000	0,33472	0,391471	-0,05675
5000	0,99973	411,23333	0,66023	308,83333	0,411343	0,467764	-0,05642
6000	0,99895	692,13333	0,65028	583,80000	0,577386	0,748143	-0,17076
7000	0,99861	978,33333	0,65162	793,10000	0,69978	0,869372	-0,16959
8000	0,99647	1258,23333	0,64914	1231,56667	0,789182	1,185761	-0,39658
9000	0,99563	1601,90000	0,64913	1618,73333	0,893848	1,385394	-0,49155
10000	0,99270	1329,96667	0,64735	1924,26667	0,669873	1,486265	-0,81639

Fonte: Próprio autor

Tabela 32 – Relação Tempo / Grupos Balanceados (custo) - caso 4

Média de 30 execuções							
#Estudantes	AG		Randômico		Custo		Diferença
	Grupos Hete	Tempo (s)	Grupos Hete	Tempo (s)	AG	Rand	
1000	0,02117	46,66667	0,22267	34,03333	0,238379	0,218911	0,019468
2000	0,03633	96,83333	0,23100	70,16667	0,251211	0,22811	0,023101
3000	0,05056	157,96667	0,23523	125,16667	0,277299	0,272775	0,004524
4000	0,06463	281,50000	0,23587	198,46667	0,376188	0,324663	0,051525
5000	0,07327	397,23333	0,23850	307,56667	0,428638	0,403896	0,024742
6000	0,08261	716,30000	0,24106	503,03333	0,650671	0,552342	0,098329
7000	0,09203	937,63333	0,23937	901,06667	0,737619	0,846162	-0,10854
8000	0,10076	1337,53333	0,23905	1124,60000	0,929628	0,923677	0,005951
9000	0,10752	1663,93333	0,24047	1605,20000	1,035778	1,174113	-0,13834
10000	0,11463	1481,36667	0,24100	1898,73333	0,836584	1,250812	-0,41423

Fonte: Próprio autor

Tabela 33 – Relação Tempo / Grupos Heterogêneos (custo) - caso 5

Média de 30 execuções							
#Estudantes	AG		Randômico		Custo		Diferença
	Grupos Hete	Tempo (s)	Grupos Hete	Tempo (s)	AG	Rand	
1000	0,81083	42,56667	0,35400	35,33333	0,262487	0,499058	-0,23657
2000	0,73108	95,50000	0,36233	76,16667	0,32657	0,525529	-0,19896
3000	0,66205	170,70000	0,37411	111,16667	0,429728	0,495249	-0,06552
4000	0,59216	267,76667	0,35939	205,53333	0,565233	0,714869	-0,14964
5000	0,56570	371,86667	0,36377	281,43333	0,657357	0,773664	-0,11631
6000	0,52322	641,46667	0,36288	526,60000	1,021671	1,209296	-0,18762
7000	0,48404	881,76667	0,34702	957,00000	1,301192	1,969814	-0,66862
8000	0,46893	1364,93333	0,34705	1149,56667	1,819213	2,070267	-0,25105
9000	0,47641	1666,06667	0,36085	1469,46667	1,942835	2,262375	-0,31954
10000	0,43558	1953,20000	0,33623	1843,96667	2,242051	2,742094	-0,50004

Fonte: Próprio autor

Tabela 34 – Relação Tempo / Grupos Heterogêneos (custo) - caso 6

Média de 30 execuções							
#Estudantes	AG		Randômico		Custo		Diferença
	Grupos Hete	Tempo (s)	Grupos Hete	Tempo (s)	AG	Rand	
1000	0,53600	46,33333	0,35817	33,13333	0,432214	0,462541	-0,03033
2000	0,54900	93,60000	0,36725	68,10000	0,42623	0,463581	-0,03735
3000	0,55994	161,63333	0,37394	123,96667	0,481106	0,552525	-0,07142
4000	0,53476	274,80000	0,36776	192,10000	0,642348	0,652939	-0,01059
5000	0,53727	382,36667	0,37190	318,66667	0,711689	0,856861	-0,14517
6000	0,51463	766,10000	0,36637	569,70000	1,240527	1,295833	-0,05531
7000	0,48329	954,76667	0,34762	736,20000	1,411112	1,512736	-0,10162
8000	0,47546	1317,33333	0,34463	1236,80000	1,731644	2,242985	-0,51134
9000	0,49519	1690,90000	0,35989	1609,96667	1,89704	2,485253	-0,58821
10000	0,45280	1489,90000	0,34050	1992,33333	1,645208	2,9256	-1,28039

Fonte: Próprio autor