

UNIVERSIDADE FEDERAL DO AMAZONAS
INSTITUTO DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

ILMARA MONTEVERDE MARTINS RAMOS

**RECOMENDAÇÃO PARA FORMAÇÃO DE GRUPOS PARA ATIVIDADES
COLABORATIVAS UTILIZANDO A CARACTERIZAÇÃO DOS APRENDIZES
BASEADA EM TRILHAS DE APRENDIZAGEM**

Manaus

2017

ILMARA MONTEVERDE MARTINS RAMOS

**RECOMENDAÇÃO PARA FORMAÇÃO DE GRUPOS PARA ATIVIDADES
COLABORATIVAS UTILIZANDO A CARACTERIZAÇÃO DOS APRENDIZES
BASEADA EM TRILHAS DE APRENDIZAGEM**

Dissertação apresentada ao Programa de Pós-Graduação em Informática do Instituto de Computação da Universidade Federal do Amazonas como requisito para obtenção do título de Mestre em Informática, na área de concentração em Inteligência Artificial.

Este trabalho foi desenvolvido com o apoio do Governo do Estado do Amazonas, por meio Fundação de Amparo à Pesquisa do Estado do Amazonas, com a concessão de bolsa de estudo.

Orientadora: Prof^ª. Dra. Elaine Harada Teixeira de Oliveira.

Coorientador: Prof. Dr. Bruno Freitas Gadelha.

Manaus

2017

Ficha Catalográfica

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

R175r Ramos, Ilmara Monteverde Martins
Recomendação para Formação de Grupos para Atividades Colaborativas utilizando a caracterização dos Aprendizes baseada em Trilhas de Aprendizagem / Ilmara Monteverde Martins Ramos. 2017
113 f.: il. color; 31 cm.

Orientador: Elaine Harada Teixeira de Oliveira
Orientador: Bruno Freitas Gadelha
Dissertação (Mestrado em Informática) - Universidade Federal do Amazonas.

1. Ambiente Virtual de Aprendizagem. 2. Recomendação de Grupos. 3. Framework Conceitual. 4. Trilhas de Aprendizagem. I. Oliveira, Elaine Harada Teixeira de II. Universidade Federal do Amazonas III. Título



PODER EXECUTIVO
MINISTÉRIO DA EDUCAÇÃO
INSTITUTO DE COMPUTAÇÃO

PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA



UFAM

FOLHA DE APROVAÇÃO

"Recomendação para formação de grupos para atividades colaborativas utilizando a caracterização dos aprendizes baseada em trilhas de aprendizagem"

ILMARA MONTEVERDE MARTINS RAMOS

Dissertação de Mestrado defendida e aprovada pela banca examinadora constituída pelos Professores:

Profa. Elaine Harada Teixeira de Oliveira - PRESIDENTE

Prof. José Francisco de Magalhães Netto - MEMBRO INTERNO

Profa. Kátia Cilene Neles da Silva - MEMBRO EXTERNO

Manaus, 20 de Outubro de 2017

DEDICATÓRIA

Aos meus pais por todo amor e incentivo, por sempre acreditarem em mim, ajudar naquilo que podiam e por não medirem esforços para que eu chegasse até essa etapa da minha vida.

À minha irmã pelos momentos de diversão, compartilhamento de ideias e emoções.

Ao meu marido por todo amor, apoio e compreensão. Que apesar dos sacrifícios que passamos, esteve sempre incentivando a minha decisão.

Ao meu filho Luiz Antônio, o primeiro coração batendo eternamente dentro de mim. Dedico essa conquista ao meu bebê, que é a maior de todas as minhas vitórias.

AGRADECIMENTOS

Agradeço em primeiro lugar a Deus e a toda minha família. Em especial, aos meus pais, que sempre me incentivaram e confiaram que a realização deste trabalho seria possível, à minha irmã pelo apoio de sempre e a todos meus familiares, por todo o carinho e incentivo. Meu agradecimento ao meu marido David Ramos, pelo apoio em todos os momentos, pela compreensão, pelo carinho e por acreditar em mim. Meus agradecimentos a todos os professores do Instituto de Computação - ICOMP da UFAM pelos muitos ensinamentos, parceria e incentivo. Em especial, à minha orientadora, Elaine Harada, agradeço por todos seus ensinamentos, sua paciência, seus conselhos e incentivos, que foram essenciais para a conclusão deste trabalho. E também, ao professor Bruno Gadelha por seus ensinamentos, conselhos e incentivos, tão importantes para a conclusão deste trabalho. Aos amigos que fiz durante o mestrado, em especial aos amigos de longa jornada Priscilla Nascimento, Hellen Sobrinho, Oziel Antunes, Isomar Lima, Euler Vieira, Marcel Rios e Dhanielly Lima, que me incentivaram, cada um de sua maneira, para completar essa jornada e a todos do grupo GSI que me ajudaram direta ou indiretamente para conclusão desse trabalho. Obrigada pela parceria, apoio e incentivo.

“Se tu o desejas, podes voar, só tens de confiar muito em ti.”

Steve Jobs

“Força, Foco e Fé.”

Ilmar de Souza Martins. Meu pai.

RESUMO

Grupos são estruturas sociais básicas, tanto dentro quanto fora do mundo acadêmico, eles se formam e se modificam em vários modos para múltiplos propósitos. Enquanto alunos, formar grupos fora de sala de aula aparenta ser mais fácil, por sua vez, em ambiente educativo formar grupos pode ser um processo complexo. Contudo, para que a aprendizagem colaborativa seja bem-sucedida, é importante formar grupos que possam ser efetivos e eficientes em realizar os objetivos da tarefa. Nesse sentido, busca-se melhorar as interações dos discentes principalmente em atividades em grupo. O trabalho em grupo é um recurso bastante usado pelo docente, com o intuito de incentivar a interação dos alunos nas atividades colaborativas. Há nos AVAs uma deficiência em dar suporte à criação de grupos. As Trilhas de Aprendizagem (TAs) podem ser recursos para sugerir grupos de aprendizes. O uso de TAs é promissora porque mostra os caminhos percorridos pelos alunos nos AVAs, o que possibilita sugerir grupos baseados nessas interações, pois realizar grupos na modalidade EAD é realmente uma tarefa desafiadora para o docente. No geral, nessa modalidade, os alunos só possuem 20% do contato presencial entre si e com o docente, dificultando estabelecer critérios para realizar os agrupamentos. Diante deste contexto, nesta pesquisa é proposta um *framework* conceitual para formação de grupos em atividades colaborativas, por meio dos dados extraídos dos grafos das TAs, para auxiliar o docente no processo de ensino-aprendizagem. Para verificar a exequibilidade do *framework* conceitual proposto foi desenvolvida a ferramenta M-CLUSTER que analisa os atributos descritos a seguir e sugere a formação de grupos. O mecanismo dá ênfase à formação dos grupos aplicando o algoritmo *K-Means*, que é utilizado com três métricas de similaridade, que são as distâncias: Euclidiana, *Manhattan* e Cosseno, usando os atributos (média de acesso de vértices, quantidade, dispersão e variâncias das arestas padrão, avanço e retorno e id do aluno) obtidos por meio dos dados extraídos das TAs. Foram realizadas reuniões com os docentes (especialistas) para validar os resultados. No estudo de caso, o M-CLUSTER usou os atributos e os classificou com o *K-Means*, obtendo três resultados de agrupamentos, um para cada métrica. O docente escolhe dentre as sugestões geradas e disponibiliza uma delas aos estudantes para eles escolherem seus parceiros dentro do cluster, assim formando os grupos. Esses grupos sugeridos pela ferramenta foram validados e visualizados pelo docente, a partir de duas representações, uma descritiva e outra visual por meio de gráficos de bolhas. Para validar os grupos sugeridos, foram realizadas duas atividades na disciplina, na primeira atividade, os estudantes escolheram seus parceiros de grupo e na segunda, os grupos foram formados de acordo com as sugestões da ferramenta. De acordo com os resultados obtidos do estudo de caso mostra que, a ferramenta obteve resultados satisfatórios onde 75% dos alunos igualaram ou melhoraram suas notas individuais em relação às alcançadas na primeira atividade.

Palavras-chave: Ambiente Virtual de Aprendizagem, Recomendação de grupos, Framework Conceitual, Trilhas de Aprendizagem.

ABSTRACT

Groups are a basic social structure, and as inside as outside of the academic world they form and change themselves in various ways for multiple purposes. While students form groups easily out of class, forming groups on a course can be an unnatural process. However, for collaborative learning to succeed, it is important to form groups that can be effective and efficient in accomplishing the objectives of the task. In this sense, it is sought to improve the interactions of students mainly in group activities. The group work is a resource widely used by the teacher, with the intention of encouraging students' interaction in collaborative activities. There is a lack of support for group creation in Virtual Learning Environments (VLEs). In this sense, Learning Path (LPs) can be resources to suggest groups of learners. The use of LPs is promising because it shows the paths taken by students in VLEs, which makes it possible to suggest groups based on these interactions, since to perform groups in the EAD modality is really a challenging task for the teacher. In general, in this modality, students only have 20% of face-to-face contact with each other and with the teacher, making it difficult to establish criteria for grouping. In this context, this research proposes a conceptual framework for the formation of groups in collaborative activities, through data extracted from the LPs graphs, to assist the teacher in the teaching-learning process. To verify the feasibility of the proposed conceptual framework, the M-CLUSTER tool was developed that analyzes the attributes described below and suggests the formation of groups. The mechanism emphasizes the formation of groups by applying the *K-Means* algorithm, which is used with three similarity metrics, which are the distances: Euclidean, *Manhattan* and Cosine, using attributes (vertex access, quantity, dispersion and variances of the standard edges, forward and return, and student id) obtained through the data extracted from the LPs. Meetings were held with the teachers (specialists) to validate the results. In the case study, M-CLUSTER used the attributes and classified them with *K-Means*, obtaining three clustering results, one for each metric. The teacher chooses among the suggestions generated and makes one available to the students so they can choose their partners within the cluster, thus forming the groups. These groups suggested by the tool were validated and visualized by the teacher, from two representations, one descriptive and one visual through bubble charts. To validate the suggested groups, two activities were carried out in the discipline, in the first activity, the students chose their group partners and in the second, the groups were formed according to the suggestions of the tool. According to the results obtained from the case study shows that the tool obtained satisfactory results where 75% of students matched or improved their individual scores in relation to those achieved in the first activity.

Keywords: Virtual Learning Environments, Group Recommendation, Conceptual Framework, Learning Path.

LISTA DE FIGURAS

Figura 1: Expressão da similaridade da Distância Euclidiana.....	31
Figura 2: Expressão da similaridade da Distância de <i>Manhattan</i>	32
Figura 3: Expressão da Distância do Cosseno.....	33
Figura 4: Filtragem Híbrida.....	36
Figura 5: Grafo da Trilha de Aprendizagem	37
Figura 6: Tipos de Grafo.	38
Figura 7: Artigos retornados pela String.	42
Figura 8: Resultado do Primeiro Filtro.....	43
Figura 9: Artigos Aceitos e Rejeitados.....	43
Figura 10: Resultado do Segundo Filtro.....	44
Figura 11: Fases para Formação dos Grupos.....	46
Figura 12: Comparação de Avaliação entre Métodos.	48
Figura 13: Visão Geral do <i>Framework</i>	54
Figura 14: Exemplo da impressão das Trilhas de Aprendizagem dos Discentes	56
Figura 15: Dados coletados dos TAs de forma manual.....	58
Figura 16: Agrupamento de Alunos usando Eclipse e WEKA	61
Figura 17: Arquivo arff refinado.	64
Figura 18: Arquivo arff com média de acesso dos vértices.....	65
Figura 19: Arquivo arff da turma de Matemática Discreta 2016/1.	66

Figura 20: Arquitetura do Trabalho.....	68
Figura 21: Bloco M-CLUSTER	72
Figura 22: Camadas do M-CLUSTER	72
Figura 23: Parte da tela do M-CLUSTER.	74
Figura 24: Exemplo de Visualização da Representação Descritiva	76
Figura 25: Exemplo de Visualização do Gráfico de Bolhas.....	77
Figura 26: Resultado da Distância do Cosseno	79
Figura 27: Resultado da Distância de <i>Manhattan</i>	79
Figura 28: Resultados da Distância Euclidiana	80
Figura 29: Notas dos alunos de Matemática Discreta.	81

LISTA DE TABELAS

Tabela 1: Resultados dos atributos <i>K-Means</i> com Distância Euclidiana	61
Tabela 2: Resultados dos atributos versão um com todas as distâncias turma 2015/1.....	95
Tabela 3: Resultados dos atributos versão dois.	95
Tabela 4: Resultados dos atributos versão três.....	96
Tabela 5: Resultados dos atributos versão quatro.	96
Tabela 6: Resultados dos atributos versão cinco.	97
Tabela 7: Resultados dos atributos versão seis.....	97
Tabela 8: Resultados dos atributos versão sete.	98
Tabela 9: Resultados dos atributos versão oito.	98
Tabela 10: Resultado dos agrupamentos da turma 2016/1.....	99

LISTA DE QUADROS

Quadro 1: Diferenças das abordagens de formação de grupos.....	18
Quadro 2: Termo de busca.....	41
Quadro 3: Critérios de Inclusão e Exclusão	41
Quadro 4: Comparação entre os trabalhos similares e a pesquisa.	49

LISTA DE ABREVIATURAS E SIGLAS

API	<i>Application Programming Interface</i>
AVA	Ambiente Virtual de Aprendizagem
CE	Critérios de Exclusão
CI	Critérios de Inclusão
CMS	<i>Content Management System</i>
D3	<i>Data-Driven Documents</i>
EaD	Educação a Distância
EFC	C-Means Fuzzy Evolutivo
GNU	<i>General Public License</i>
HCM	<i>Hybrid Clustering Method</i>
HTML	<i>Hypertext Markup Language</i>
I	Intervenção
IC	<i>Instituto de Computação</i>
IDE	<i>Integrated Development Environment</i>
IEEE	<i>Institute of Electrical and Electronics Engineers</i>
ITS	<i>Intelligent Tutoring System</i>
LaPES	Laboratório de Pesquisa em Engenharia de Software
LMS	<i>Learning Management System</i>
LCMS	<i>Learning Content Management System</i>
JSON	<i>JavaScript Object Notation</i>
KNN	<i>K -Nearest Neighbor</i>
MOODLE	<i>Modular Object-Oriented Dynamic Learning Environment</i>

M-CLUSTER	<i>Moodle Cluster</i>
MySQL	<i>My Structured Query Language</i>
NIED	Núcleo de Informática Aplicada à Educação
P	População
PHP	<i>Hypertext Preprocessor</i>
RSL	Revisão Sistemática da Literatura
SQL	<i>Structured Query Language</i>
SR	Sistema de Recomendação
SGBD	Sistema de Gerenciamento de Banco de Dados
TA	Trilhas de Aprendizagem
TARGAVA	<i>Framework</i> conceitual baseado em Trilhas de Aprendizagem para Recomendação de Grupos de aluno em um AVA
TIC	Tecnologias de Informação e Comunicação
VLE	<i>Virtual Learning Environment</i>
WEKA	<i>Waikato Environment for Knowledge Analysis</i>

SUMÁRIO

1. INTRODUÇÃO	16
1.1 DEFINIÇÃO DO PROBLEMA	17
1.2 MOTIVAÇÃO E JUSTIFICATIVA	19
1.3 OBJETIVOS	21
1.4 ORGANIZAÇÃO DO TRABALHO	21
2 REFERENCIAL TEÓRICO	22
2.1 AMBIENTES VIRTUAIS DE APRENDIZAGEM.....	22
2.2 TRILHAS DE APRENDIZAGEM	24
2.3 FORMAÇÃO DE GRUPOS	25
2.4 ALGORITMO DE CLUSTERIZAÇÃO	29
2.5 MEDIDAS DE DISTÂNCIA	31
2.6 SISTEMA DE RECOMENDAÇÃO	33
2.7 GRAFOS	36
3 TRABALHOS RELACIONADOS	39
3.1 REVISÃO SISTEMÁTICA DE LITERATURA	39
3.2 TRABALHOS RELACIONADOS SOBRE RECOMENDAÇÃO DE GRUPOS EM AMBIENTES VIRTUAIS DE APRENDIZAGEM.....	45
3.3 COMPARAÇÃO ENTRE TRABALHOS SIMILARES	49
4 PROPOSTA DE TRABALHO	51
4.1 METODOLOGIA DE PESQUISA	51
4.2 <i>FRAMEWORK</i> CONCEITUAL DE RECOMENDAÇÃO PARA FORMAÇÃO DE GRUPOS DE APRENDIZES EM AVA	52
4.3 ARQUITETURA DE TRABALHO.....	68
4.4 FERRAMENTAS TECNOLÓGICAS	69
5 O DESENVOLVIMENTO DA FERRAMENTA M-CLUSTER.....	71
5.1 O BLOCO M-CLUSTER	71
5.2 LÓGICA DE CAMADAS DO M-CLUSTER	72
5.3 IMPLEMENTAÇÃO DO NÚCLEO DO M-CLUSTER	74
5.4 IMPLEMENTAÇÃO DA REPRESENTAÇÃO DESCRITIVA	75
5.5 IMPLEMENTAÇÃO DA REPRESENTAÇÃO VISUAL	76

6 VALIDAÇÃO INICIAL E RESULTADOS DA PROPOSTA.....	78
7 CONSIDERAÇÕES FINAIS.....	82
7.1 CONTRIBUIÇÕES	83
7.2 TRABALHO FUTUROS	84
REFERÊNCIAS	86
APÊNDICE A – PUBLICAÇÕES PRODUZIDAS NO MESTRADO	92
APÊNDICE B – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO	93
APÊNDICE C – RESULTADOS DAS ANÁLISES DOS ATRIBUTOS	95
APÊNDICE D – ARTIGO	100

1. INTRODUÇÃO

O uso de recursos tecnológicos é bastante disseminado na área da educação, a partir do olhar sobre a utilização tecnológica dentro do contexto escolar, ao qual demonstra contribuições significativas ao ensino-aprendizagem. No entanto, para a sua inserção é necessário verificar novas práticas e estratégias, levando em consideração seu uso de forma eficaz ao ambiente educacional.

Nessa perspectiva, o desenvolvimento de recursos tecnológicos vem como um dos fatores que promovem a interação entre pessoas, transformando a informação em grande escala em elemento primordial às sociedades modernas. Nessa nova conjuntura social, começa a existir um investimento expressivo em tecnologias telemáticas de alta velocidade e, conseqüentemente, a uma nova formatação da educação em prol do aprendizado (MORAN, 2013).

Segundo, Prensky (2001) os estilos tradicionais não são adequados aos aprendizes digitalmente alfabetizados, pois estes já cresceram habituados com o uso da tecnologia e, portanto, necessitam de uma forma diferenciada de ensino. Essa nova formatação deve acompanhar novos hábitos em sociedade, por exemplo: o ato de ler, que não apenas se faz com um livro hoje em dia, como também, a partir de diversos dispositivos móveis e diversas formas de compartilhamento.

Diante deste cenário, o ambiente escolar pode fazer o uso dos vastos recursos tecnológicos para informatizar o processo educacional em quaisquer modalidades de ensino. Porém, observa-se, em alguns casos, que a utilização destes recursos - nesse contexto - possui certa resistência, afinal, a proposta requer modificações nas práticas pedagógicas dos docentes, pois o uso dos recursos pode fazer parte da praxe do professor, portanto, essa transição da utilização dos meios tecnológicos é bastante complexa. (ALTOÉ; FUGIMOTO, 2009).

A modalidade da Educação a Distância (EaD), torna-se relevante nesse cenário, principalmente pelo surgimento da internet que possibilitou à EaD realizar mudanças significativas na educação. Logo, Moran (2013) afirma que a EaD é uma modalidade que vem crescendo ininterruptamente, em todas as áreas, utilizando-se da evolução tecnológica e moldes diferenciados de aprendizagem, principalmente usando Ambientes Virtuais de Aprendizagem (AVAs).

Dessa forma, os AVAs funcionam como salas de aula, fazendo com que o enfoque seja nas pesquisas em tempo real, em atividades individuais e em grupo, mudando as metodologias tradicionais para aprendizagem colaborativa e personalizada.

Assim, Lima *et al.* (2014) relatam que os Ambientes Virtuais de Aprendizagem (AVAs) não possuem ferramentas suficientes para ajudar o docente a detectar as relações sociais entre os discentes, impedindo assim a identificação de grupos. Além disso, existe nos AVAs uma deficiência em dar suporte à criação de grupos. Para ajudar a suprir essa deficiência, utilizou-se as Trilhas de Aprendizagem (TAs) como recurso para sugerir ao docente grupos de aprendizes para realizar atividades que requerem colaboração.

No contexto desta dissertação, o termo TA será utilizado assim como nos dizeres de Ramos *et al.* (2016b), para designar a trajetória percorrida pelo aluno durante a sua interação com o AVA, aqui representada na forma de grafo.

Portanto de acordo com o contexto acima é relevante conhecer o que o aluno faz no AVA, a fim de tentar criar grupos de alunos com comportamentos parecidos. Essa informação sobre o comportamento dos alunos pode ser dado por meio da análise das TAs e o resultado dessa análise possibilita sugerir grupos baseados nessas interações. Essa possibilidade de sugestão de agrupamentos é de suma importância para o docente, pois realizar grupos na modalidade EAD, torna-se uma tarefa desafiadora. No geral, nessa modalidade, os alunos só possuem 20% do contato presencial entre si e com o docente, tornando difícil o trabalho de estabelecer critérios para realizar agrupamentos.

Assim, a presente pesquisa se norteou pelos seguintes questionamentos: Como se constroem grupos homogêneos para as atividades colaborativas de aprendizagem? Como é possível formá-los? As trilhas de aprendizagem (TA) podem contribuir nessa formação?

1.1 DEFINIÇÃO DO PROBLEMA

Para aprimorar os AVAs é de fundamental importância conhecer o aluno. Para isso, é necessário ter as informações sobre os perfis dos usuários que descrevam as suas preferências e interesses para poder ocorrer um processo de recomendação baseado em tais perfis (CAZELLA; NUNES; REATEGUI, 2010).

À medida que o processo de coleta das preferências dos interesses vai se aperfeiçoando, auxilia na melhora no processo de recomendação, principalmente em atividades que requerem colaboração. Das propostas de recurso didático do professor, o trabalho de grupo se apresenta como atividade bastante solicitada dentro das plataformas

educacionais, por incentivar a participação ativa dos alunos no processo de aprendizagem e buscar a construção coletiva do conhecimento.

Segundo Ounnas, Davis e Millard (2007), existem três tipos de abordagens para a formação de grupos, que são:

1. Grupos selecionados aleatoriamente: são propostos pelo professor que atribui estudantes a grupos de forma aleatória, sem nenhum critério específico definido. Este tipo de abordagem geralmente é utilizado para formar grupos informais e temporários;
2. Grupos de autosseleção: são grupos onde os alunos têm permissão para escolher o grupo ao qual querem pertencer e negociar com quem querem trabalhar;
3. Grupos selecionados pelo professor: são iniciados pelo professor, permitindo-os criar ou direcionar a formação de grupos, considerando os interesses compartilhados e os laços sociais.

Como visto, no Quadro 1, para grupos aleatórios não são estabelecidos critérios de agrupamento. Em grupos autosselecionados, os critérios são estabelecidos pelos próprios alunos, não pelo professor. Por sua vez, na última abordagem, é o professor que decide a composição dos grupos com base em informações que estão mais evidentes em turmas presenciais, ou seja, o grupo de estudantes envolve a distribuição simultânea de todos os alunos da turma sobre n grupos, onde n é igual ao número de alunos sobre o tamanho do grupo ideal; ou escolher alguns alunos de toda a turma para formar um grupo.

Quadro 1: Diferenças das abordagens de formação de grupos

Fonte:(Ounnas; Davis;Millard, 2007)

Abordagens de Formação de grupos	Iniciador	Identificação dos membros	Negociação da formação
Grupos Selecionados Aleatoriamente	Professor	Aleatoriamente	Nenhuma
Grupos de autosseleção	Alunos	Identificando potenciais pares	Negociar com quem querem trabalhar
Grupos Selecionados pelo professor	Professor	Formar um grupo	
		Identificando potenciais pares	Consulta potencial do aluno, não precisa de negociação
		Grupo com todos os Alunos	

		Distribui os Alunos nos grupos	Assegura a equidade da formação + maximiza o benefício de cada aluno
--	--	--------------------------------	--

O último caso acontece quando a colaboração só é necessária para um número de alunos da turma, como usar exemplos de alunos de toda a população da turma ou selecionar os melhores alunos para um desafio específico. Assim, o primeiro caso diz respeito a distribuição dos alunos uniformemente para construir grupos equilibrados em termos de restrições de formação, considerando ao mesmo tempo os benefícios máximos dos alunos de participar dos grupos.

Desta forma, percebe-se o quão desafiador é para o professor realizar a formação de grupos dentro de sala de aula tanto na modalidade presencial quanto na EaD. Ajudá-lo nessa formação se faz necessário, principalmente quando é imprescindível estabelecer grupos para realização de atividades colaborativas dentro do AVA, onde no geral, na modalidade à distância, os alunos só possuem 20% do contato presencial entre si e com o docente, dificultando estabelecer critérios para realizar agrupamentos. Nesse contexto é de suma importância que os docentes tenham ferramentas que possam auxiliar nessa tarefa.

A problemática abordada por esse estudo consiste em permitir ao professor que faz uso do AVA, uma possibilidade de ferramenta de trabalho que possa auxiliá-lo no processo de formação de grupos discentes, a partir da utilização de um *framework* conceitual.

À medida que essas inquietações foram postas em forma de projeto e desenvolvidas na pesquisa de mestrado, obtivemos um *framework* conceitual para recomendação de grupos para atividades colaborativas, que utiliza a caracterização dos aprendizes baseada em trilhas de aprendizagem no AVA. O uso das TAs foi promissor, por permitir analisar as interações dos alunos e compará-las por meio das métricas de similaridade, o que possibilitou o surgimento de grupos baseados nessas interações.

1.2 MOTIVAÇÃO E JUSTIFICATIVA

Com o avanço das tecnologias e a ampliação do acesso à internet, novas modalidades educacionais tem atuado como recurso na difusão do conhecimento. Com esse propósito, o ensino a distância (EaD) é uma modalidade que preza pela inserção das práticas de otimização dos ensinamentos. As construções dessas novas concepções de aprendizagem

colaborativa visam à utilização das novas tecnologias que impõem a possibilidade do aprender de forma não convencional.

A EaD seria essa inegável realidade da expansão do conhecimento cuja modalidade é a mais crescente no Brasil nos últimos anos (ABED, 2015).

O número de matrículas de bacharelado, licenciatura e cursos superiores tecnológicos a distância aumentou significativamente, cuja soma está mais de 1 milhão, de acordo com último Censo de Educação Superior, divulgado em 2013 (INEP, 2013). Fato que a EaD mostra como vantagens a flexibilidade, economia de tempo, comodidade, bem como a utilização dos mais diversos recursos de comunicação a partir da rede mundial de computadores (LAVRADOR, 2015).

Conforme o contexto descrito anteriormente, justifica-se a importância em se debruçar sobre os moldes diferenciados de aprendizagem.

Visto que, com a inserção da tecnologia, pode-se realizar otimização do ensino tentando melhorar as práticas convencionais. No caso da Educação, as tecnologias, sem dúvidas permitem expandir o conceito de aula, espaço, tempo, interação, interatividade, estabelecendo novas conexões entre o estar junto fisicamente e a distância (MORAN, 2013).

Nesse sentido, os Ambientes Virtuais de Aprendizagem (AVA) funcionam como as salas de aula do modelo tradicional, mas com a diferenciação nos recursos de fluxo, espaço e tempo utilizando tecnologia.

Essa descoberta de novas dimensões de significação, estabelecem pontes entre a reflexão e a ação, entre a experiência e a conceituação, entre teoria e prática, entre atividades individuais e em grupo *online*, participando da discussão entre as metodologias tradicionais e aprendizagem colaborativa e personalizada.

Essa busca perpassa por mecanismo que possam ajudar no processo de ensino-aprendizagem entre discentes e docentes. Na EaD, os AVAs relacionam a formação de grupos para atividades em grupo *online*, uma metodologia adotada pelo professor como uma opção para aprendizagem dos alunos no ambiente educacional digital.

Assim, a relevância desta pesquisa está em dar suporte ao professor no processo de formação de grupos visando melhorar o processo de ensino-aprendizagem por meio do uso das TAs

A abordagem proposta nesta dissertação pretende se aliar aos demais esforços como as pesquisas realizadas por (DARADOUMIS *et al*, 2002; JAGADISH, 2014) no que diz respeito à formação de grupos de aprendizes em ambientes virtuais de aprendizagem. Nesse

sentido, propomos um *framework* conceitual para recomendar grupos para atividades colaborativas em um AVA de forma automática.

1.3 OBJETIVOS

Considerando o exposto, a presente pesquisa tem como objetivo geral desenvolver e validar um *framework* conceitual para recomendação de grupos para atividades colaborativas, utilizando a caracterização dos aprendizes baseada em trilhas de aprendizagem (TAs) em um AVA. Como objetivos específicos, o estudo percorre pelo processo de:

- Identificar os dados mais relevantes das TAs para serem usados como atributos para formação dos grupos;
- Desenvolver um mecanismo para recomendação de grupos de aprendizes de acordo com o *framework* estabelecido para os agrupamentos;
- Validar o *framework*, por meio do M-CLUSTER, em turmas reais buscando identificar a contribuição da abordagem proposta para a atividade docente.

1.4 ORGANIZAÇÃO DO TRABALHO

A sequência deste trabalho está organizada da seguinte forma: no Capítulo 2 é apresentado o referencial teórico com os conceitos que norteiam esta pesquisa; o Capítulo 3 apresenta o levantamento bibliográfico através de uma Revisão Sistemática da Literatura (RSL) juntamente com os trabalhos relacionados; o Capítulo 4 descreve a proposta do projeto, apresentando a estrutura do *framework* conceitual, a metodologia utilizada e as ferramentas usadas para compor o *framework*; no Capítulo 5 descreve a aplicação M-CLUSTER que serviu de ferramenta para análise e validação da formação dos grupos obtidos a partir do *framework* conceitual e; o Capítulo 6 discorre sobre os experimentos realizados para validação do *framework* conceitual por meio do *plugin* M-CLUSTER, que foi utilizado para sugerir grupos de alunos em um ambiente real no AVA de uma turma em andamento. O Capítulo 7 apresenta as considerações finais.

2 REFERENCIAL TEÓRICO

Neste capítulo, serão abordadas as bases teóricas que darão suporte à pesquisa, buscando auxiliar a compreensão das tecnologias utilizadas para subsidiar a sua construção. Serão apresentados os princípios teóricos sobre: Ambientes Virtuais de Aprendizagem com relação ao seu contexto nos cursos ofertados na modalidade a distância e apresentar algumas das plataformas utilizadas na EaD, como: TelEduc, E-Proinfo e Moodle; as Trilhas de Aprendizagem na conjuntura dos AVAs; a formação de grupos com relação ao conceitos de grupo, tipos de grupo, tamanho, abordagens para sua formação; o algoritmo de clusterização *K-Means* conceito e funcionamento; Medidas de Distância no que concerne as distâncias Euclidiana, *Manhattan* e Cosseno; Sistemas de Recomendação com relação ao conceito, coleta de informações do usuário e técnicas de recomendação e, por fim, grafos compreendendo noções básicas e tipos de grafos.

2.1 AMBIENTES VIRTUAIS DE APRENDIZAGEM

Na EaD são utilizadas diversas plataformas de ensino, as quais permitem alusão às salas de aula presenciais e possibilitam a aprendizagem aos alunos que delas se utilizam. Essas plataformas de ensino se denominam Ambientes Virtuais de Aprendizagem (AVAs). Segundo Almeida, Vieira e Luciano (2001), os AVAs são ambientes que habitam o ciberespaço¹ e incluem ferramentas que favorecem a comunicação dos discentes, docentes e tutores, oferecendo recursos que facilitem a aprendizagem individual e colaborativa.

Com os AVAs, pretende-se mediar à construção do conhecimento, motivar e prover habilidades de aprendizagem autônoma nos aprendizes (MESSA, 2010), proporcionando a melhoria no processo educativo destes discentes. A modalidade à distância, permite juntamente com os AVAs, que o estudante tenha uma flexibilidade no que se refere ao momento e local ao qual este pretende realizar seus estudos.

Atualmente, os cursos à distância ofertados pelas Instituições de Ensino (IES) utilizam em sua estruturação e concepção os AVAs, sendo que, para atender às suas necessidades institucionais, em sua maioria, esses ambientes são plataformas gratuitas que permitem adequações para realizar customizações de novas funcionalidades. Essas personalizações ajudam os docentes nas suas práticas didático-pedagógicas e educacionais.

¹ É a interconexão digital entre computadores ligados em rede.

Nesse contexto, serão descritos sucintamente, a seguir, três AVAs: TelEduc, E-Proinfo e Moodle. De acordo com a pesquisa realizada por Gabardo, Quevedo e Ulbright (2010), esses ambientes de aprendizagem estão entre as oito plataformas de ensino mais pesquisadas no Brasil e usadas em instituições de ensino que ofertam cursos na modalidade à distância.

De acordo com Otsuka e Rocha (2002), o TelEduc é uma plataforma de concepção, participação e gerenciamento de cursos na modalidade à distância, que foi criado no ano de 1997, conjuntamente pelo Núcleo de Informática Aplicada à Educação (NIED) e Instituto de Computação (IC), ambos da Universidade Estadual de Campinas (UNICAMP).

O TelEduc é um ambiente de aprendizagem fundamentado em uma abordagem educacional construtivista e de aprendizagem colaborativa. Nunes e Prass (2012) afirma que a plataforma de ensino foi desenvolvida de acordo com os requisitos solicitados dos seus próprios usuários, resultando em uma ferramenta interativa e de fácil acesso. Neste âmbito, o TelEduc é uma plataforma de software livre sob a licença *General Public License* (GNU), com código aberto, sendo que suas versões não estão sendo atualizadas constantemente.

O E-Proinfo, de acordo com Marinho (2014), é uma plataforma de ensino colaborativo de aprendizagem que visa a criação e gerenciamento de cursos à distância, contribuindo com o processo de ensino-aprendizagem. O E-Proinfo foi uma iniciativa do Ministério da Educação por meio do programa educacional ProInfo, onde o ambiente de ensino veio com o intuito de complementar o projeto, com o objetivo de inserir as Tecnologias de Informação e Comunicação (TICs) nas escolas públicas do Brasil (SILVA, 2007).

Por sua vez, o Moodle, segundo Ferro (2011), é uma plataforma virtual de aprendizagem muito utilizada pelas instituições de ensino que possuem cursos na modalidade à distância. Ele permite customizações, de maneira que os docentes possam adaptá-lo de acordo com as suas práticas pedagógicas (NASCIMENTO; BRANDÃO, 2010). Um exemplo dessas personalizações pode ser dado com o ColabWeb², uma plataforma de ensino baseada no Moodle utilizada pela Universidade Federal do Amazonas em que concerne ao Instituto de Computação- IComp.

Essa flexibilidade para realizar customizações tem interessado diversas instituições de ensino para utilização do Moodle. De acordo com Santos, Botelho e Bichet (2014), um dos

² <http://colabweb.ufam.edu.br/>

AVAs mais populares hoje é o Moodle, por ser de fácil utilização, proporcionar modularidade (diversos *plugins*) e personalizações de funcionalidades.

No âmbito desse trabalho, foi utilizado a plataforma Moodle como AVA para validar o framework conceitual por meio do M-CLUSTER. A ferramenta desenvolvida foi acoplada ao ColabWeb durante a realização da disciplina de matemática discreta do curso do Instituto de Computação da Universidade Federal do Amazonas (UFAM).

2.2 TRILHAS DE APRENDIZAGEM

Os cursos ofertados na EaD tendem a utilizar os AVAs para facilitar o processo de ensino-aprendizagem. Dentro do AVA, as trilhas de aprendizagem que, segundo Tafner, Tomelin e Muller (2012), equivalem aos percursos de aprendizagem dentro das plataformas virtuais, são capazes de incentivar e ampliar capacidades pertencentes ao conhecimento, à habilidade, à atitude, à participação, à interação e à autonomia.

Enfatiza-se que a concepção das trilhas de aprendizagem é construída, por meio das interações dos aprendizes com o ambiente virtual ou em relação aos caminhos que os alunos percorrem ao utilizar a plataforma. Partindo do princípio que as trilhas de aprendizagem podem auxiliar na construção dos perfis dos aprendizes dentro de um AVA (RAMOS; OLIVEIRA; RAMOS, 2015).

Com os perfis individuais de cada aluno baseados nas trilhas de aprendizagem foi possível realizar a formação de grupos em AVAs para atividades que requerem colaboração. Sendo que a trilha de aprendizagem, conforme Ramos *et al.* (2015), pode ser considerada sob duas perspectivas sendo:

1. A trilha tem uma sucessão de conteúdos (wiki, recurso educacional, atividade, fórum, chat,...) estabelecida pelo docente, ao realizar o planejamento da sua disciplina, e em seguida é fornecida ao AVA; ou
2. A trilha é o caminho realizado pelo discente, no decorrer da sua participação com o AVA.

Nesta pesquisa, as trilhas de aprendizagem auxiliaram na base para a caracterização dos aprendizes e essa foi utilizada no framework conceitual de recomendação para formação de grupos para atividades colaborativas, possibilitando sugerir ao professor agrupamentos de alunos por meio destas trilhas.

2.3 FORMAÇÃO DE GRUPOS

Para realizar atividades colaborativas dentro dos AVAs é necessário formar grupos de alunos, para que estes realizem as atividades. Para isso, é indispensável saber o contexto de grupo que será abordado para realizar o agrupamento dos aprendizes.

Segundo Santos, Castro e Castro (2007), grupo é um pequeno conjunto de indivíduos que se reúnem por meio de uma determinada peculiaridade. Para que ocorra essa reunião se deve considerar diversas particularidades, como: abordagens, critérios, características e dados dos estudantes. Nessa situação de agrupamentos é importante verificar de que forma podemos gerar essa motivação de envolvimento.

Grupos são estruturas sociais básicas e, tanto dentro quanto fora do mundo acadêmico, eles se formam e se modificam em vários modos para múltiplos propósitos. Em um campus universitário, estudantes se reúnem em grupos por qualquer razão, de clubes e comitês a times de atletas e irmandades. Enquanto estudantes formam grupos facilmente fora de sala de aula, formar grupos em um curso pode ser um processo não natural.

Contudo, para que a aprendizagem colaborativa seja bem-sucedida, é importante formar grupos que possam ser efetivos e eficientes em realizar os objetivos da tarefa. Seguindo o descrito em Barkley, Majore Cross (2014), a seguir são apresentados os seguintes tópicos: tipos de grupo, tamanho de grupo, adesão a grupo e seleção de membros do grupo.

2.3.1 Tipos de Grupos

Grupos de aprendizagem colaborativa têm vários tamanhos e formas, trabalhando para alcançar muitos objetivos. Eles variam dependendo dos objetivos, atividades e tempo. Nesse sentido, os grupos são classificados em informal, formal ou base.

Grupos informais são formados de forma rápida e aleatória. Geralmente são formados para atividades realizadas no momento da aula. Técnicas de aprendizagem colaborativa que utilizam este tipo de grupo são: *Think-Pair-Share*, *Round Table* (Mesa Redonda), *Buzz Groups*.

Grupos formais destinam-se a objetivos mais complexos, em geral compartilhados. Busca reunir diferentes talentos e conhecimento, e maximizar o aprendizado de todos do grupo. A adesão não é aleatória, mas sim intencional e pensada. Usado por algumas técnicas de atividade colaborativa como, por exemplo: *Critical Debate* (Debate Crítico), estudo de caso, ou investigação de grupo.

Grupos base são grupos de longo prazo, que trabalham em várias tarefas, que podem durar todo o ano. O propósito deste tipo de grupo é dar suporte aos membros, como encorajamento, como, por exemplo, às comunidades. Nos cursos, um grupo base é formado para trabalhar em um projeto de longo prazo ou uma série de atividades que requerem atividades colaborativas. Atividades colaborativas como *Dialogue Journal* (diálogos), *Round Robin* (rodízio), agrupamento por afinidade, escrita colaborativa de processo, seminário.

Assim a escolha do tipo de grupo depende da finalidade do grupo, da tarefa e do tempo para sua realização.

2.3.2 Tamanho do Grupo

Grupos colaborativos geralmente variam de dois a seis estudantes. Grupos de dois estudantes trabalham melhor para a maioria dos casos, pois, em geral, são melhores para trocas rápidas e as interrupções são mínimas. Pequenos grupos trabalham melhor, pois podem maximizar o envolvimento e facilitar a reunião dos membros. O tamanho do grupo também pode ser determinado em razão do espaço físico utilizado, como por exemplo, atividades em laboratório.

Bean (1996, p. 160) afirma que cinco seria o número de participantes mais efetivo em um grupo. Ele explica que grupos de quatro indivíduos se dividem em duplas, grupos de três, dividem-se em uma dupla e um aluno. Grupos de seis podem dissolver a experiência e criar problemas logísticos.

Em resumo, os grupos devem ser pequenos o suficiente para que todos participem e confie um nos outros, mas grande o suficiente para que haja recursos e diversidade para se concluir a tarefa.

2.3.3 Adesão a Grupo

Grupos podem ser constituídos por adesão aleatória, escolha entre alunos, indicação do professor. O agrupamento pode ser baseado em interesses, habilidades, atitudes entre outras características, além de poder ser um grupo homogêneo ou heterogêneo. Em geral, as pesquisas apoiam grupos heterogêneos por expor os alunos a ideias, conhecimentos e experiências diferentes. Barkley, Major e Cross (2014), apontam que esta experiência é o maior valor educacional da aprendizagem colaborativa.

Grupos heterogêneos podem apresentar algumas desvantagens. A primeira é que opiniões diferentes podem gerar desentendimentos. Ao distribuir alunos que compõem minorias entre os grupos pode ocorrer que eles sejam marginalizados pelos demais membros, sendo ignorados ou estereotipados, impedindo seu progresso acadêmico. Caso os grupos sejam formados centrados nos alunos que se destacam, isso pode reduzir o contato entre eles, e também pode inibir uma participação mais efetiva de alunos com aproveitamento menor.

Grupos homogêneos possuem algumas vantagens, por exemplo, os estudantes compartilham características, o que facilita o entendimento entre eles. Também é mais eficiente no desenvolvimento de habilidades, por terem um nível de conhecimento similar, por exemplo, ao aprender um novo idioma. É um agrupamento mais natural, o que melhora a satisfação do trabalho em grupo. A grande desvantagem dos grupos homogêneos é que não permitem aos seus integrantes uma experiência mais rica e desafiadora.

No contexto desta pesquisa grupos homogêneos são grupos de alunos que tem TAs similares durante sua interação com o ambiente virtual. E grupos heterogêneos são grupos de alunos que tem TAs complementares possibilitando agrupar indivíduos que tenham conhecimentos adicionais.

Na ausência de parâmetros pode ocorrer de o grupo ser estabelecido de forma aleatória. O ideal é que os agrupamentos dependam dos objetivos do curso e da tarefa a ser realizada.

2.3.4 Selecionando Membros do Grupo

Os membros podem ser selecionados de três formas diferentes para criação dos grupos: aleatoriamente, selecionado pelos estudantes e estabelecido pelo professor. A seguir, apresentamos cada uma das possibilidades de agrupamento.

- **Agrupamento aleatório**

O professor forma os grupos selecionando os alunos aleatoriamente. A seleção aleatória é ideal para grupos informais e de curto tempo. Também pode ser usado para garantir que o aluno interaja com outros alunos de sua classe. Os alunos podem ser selecionados de forma livre, por exemplo, alunos sentados próximos a outros, pelas iniciais dos nomes, por números anotados em papel, etc.

- **Agrupamento por seleção do estudante**

O professor pode permitir que os estudantes escolham seus parceiros de grupo, assim eles se sentem mais confortáveis e motivados para trabalhar. Os estudantes podem formar

seus grupos livremente, também podem usar o método *team-hiring*, onde alguns alunos terão a opção de escolher quem fará parte de sua equipe. É importante ressaltar que permitir que o aluno escolha o grupo pode gerar grupos homogêneos.

- **Agrupamento estabelecido pelo professor**

Grupos formais ou grupos base, em geral, são determinados pelo professor, baseados nos interesses ou em características de cada estudante.

O agrupamento baseado em interesses é útil para atividades colaborativas com *Role-Play* e *Critical Debates*, onde o ponto de vista é usado como critério de agrupamento. Entretanto, essa exposição pode ser desconfortável para o estudante, bem como reforça a homogeneidade dos grupos.

Os estudantes também podem ser agrupados baseados em suas características, o que é conhecido como estratificação. É uma abordagem usada para criar grupos (em geral, heterogêneos), porém os mesmo dados, características, podem criar grupos homogêneos. A seguir, são apresentadas algumas técnicas onde o agrupamento é realizado pelo professor (BARKLEY; MAJOR; CROSS, 2014):

- *Show of Hands*: os estudantes levantam as mãos para responder uma série de perguntas e são agrupados conforme as respostas. Em AVA isso pode ser feito usando o chat.
- *Student Sign-Up*: Tópicos para a investigação são escritas em papéis de inscrição e os alunos são solicitados a se inscreverem em suas preferências. Em um ambiente online, o Google Docs pode ser usado.
- *Single-Statement Likert Scale Rating*: o professor apresenta uma afirmação cuja resposta deve ser dada usando a escala de *Likert*. Para formar grupos homogêneos, os estudantes que deram as mesmas respostas são reunidos. Para grupos heterogêneos, alunos com respostas diferentes formam grupos. Ferramentas que geram *quizzes* podem ser usadas neste caso.
- *Corners*: Cada canto de uma sala representa uma resposta para uma questão de múltipla escolha que o professor fará. Para grupos homogêneos, agrupar estudantes com a mesma resposta. Para grupos heterogêneos, um aluno de cada canto é adicionado ao grupo.
- *Essay*: os estudantes devem escrever uma redação com base em um tema determinado pelo professor. Com base em análise de suas respostas, por suas semelhanças ou diferenças.

- *Data Sheet*: os alunos recebem uma ficha de dados para preencher. As repostas são usadas como informação para criar os grupos.
- *Course-Base Test Score*: o professor aplica um pré-teste ou usa dados de um teste recente para agrupar os alunos com base em seus níveis de conhecimento.
- *Learning Styles*: o professor aplica um dos métodos de descoberta de estilo de aprendizagem dos estudantes e forma grupos com base nos perfis de estilo.

Para essa pesquisa, a realização da formação de grupos será por meio da coleta de dados implícita e, para isso, é necessário combinar as preferências individuais de cada usuário de forma que o sistema de recomendação para grupos possa verificar os perfis semelhantes e sugerir uma formação de grupo.

Segundo Citadin, Kemczinski e Matos (2014) os grupos podem ser gerados manualmente e automaticamente no sistema. Manualmente, quando o agrupamento é formado fora do sistema, ou seja, o professor realiza a formação do grupo de acordo com critérios estabelecidos por ele e posteriormente informa a formação ao sistema. Essa abordagem ocorre comumente em cursos presenciais. Automaticamente, quando o próprio sistema recomenda a formação do grupo de acordo com critérios estabelecidos pelo professor resultando nos membros do agrupamento.

Para realizar o agrupamento dos aprendizes foi necessário estabelecer noções de como seria a seleção dos membros do grupo. Nesta pesquisa, utilizou-se como base a seleção dos membros do grupo definido por Barkley; Major; Cross (2014), no caso, agrupamentos selecionados pelo professor com uma abordagem que considera a caracterização dos aprendizes baseada nas trilhas de aprendizagem.

Nesta pesquisa a forma de geração do agrupamento será automática, visto que desenvolveu-se um *framework* conceitual para realizar os agrupamentos e, a partir dele, uma ferramenta para sugerir as recomendações de grupos no ColabWeb.

2.4 ALGORITMO DE CLUSTERIZAÇÃO

Para realizar o agrupamento dos alunos para a recomendação se faz necessário aplicar algum algoritmo de clusterização. Nessa pesquisa foi utilizado o algoritmo *K-Means* para realizar as sugestões dos agrupamentos para o professor a partir das TAs.

2.4.1 Algoritmo *K-Means*

Segundo Diniz e Louzada (2000), o algoritmo *K-Means* é um dos algoritmos de agrupamento de aprendizagem não supervisionada mais conhecidos e simples. Proposto por J. Macqueen em 1967, é um dos mais utilizados para realização de clusterização por ser eficiente, de fácil implementação e simples (JAIN, 2008). O algoritmo *K-Means* funciona da seguinte forma:

O algoritmo inicia com a escolha dos k elementos que formaram as sementes iniciais. A seguir é apresentado o passo a passo do algoritmo *K-Means*:

1. **Inicialização dos centróides:** Os k centróides geralmente são escolhidos como as k primeiras amostras da base. Esses centróides podem ser escolhidos aleatoriamente ou por critérios pré-estabelecidos.
2. **Cálculo da distância entre as amostras da base e os centróides.** Se existem N pontos e k centróides, deve-se calcular $N \times k$ distâncias neste passo. Nesse momento exige-se mais processamentos para os cálculos.
3. **Agrupar cada ponto de acordo com as suas distâncias dos centróides.** O centróide mais próximo de cada amostra vai agregá-la, ou seja, a amostra vai pertencer ao grupo representado pelo mesmo.
4. **Cálculo dos novos centróides para cada classe.** Os valores das coordenadas dos centróides são refinados. Para cada *cluster* que possuir mais de uma amostra, o novo valor dos centróides é calculado por meio da média dos atributos de todas as amostras do *cluster*.
5. **Repetir até a convergência.** O algoritmo retorna ao passo 2 repetindo iterativamente o refinamento do cálculo das coordenadas dos centróides até que todas as amostras não mudem mais de grupo.

Segundo Vargas (2012), o algoritmo *K-Means* é um algoritmo que pode ser utilizado para diferentes domínios de problemas. Nessa pesquisa foi utilizado o algoritmo *K-Means* por ser um algoritmo bastante utilizado para realizar agrupamentos, conforme resultados obtidos da Revisão Sistemática da Literatura (RSL), por se tratar de um algoritmo eficiente para clusterização, estes descritos na seção 3 desta dissertação, ao qual permitiu, também, obter resultados satisfatórios para a pesquisa, como descrito no capítulo 7 desta dissertação.

2.5 MEDIDAS DE DISTÂNCIA

Para a realização da recomendação para formação dos grupos desta pesquisa foi usado o algoritmo *K-Means* sendo que, para calcular a distância entre as amostras, deve-se especificar uma métrica para medir a proximidade entre as mesmas. A seguir, são descritas as medidas de distâncias utilizadas por essa pesquisa.

2.5.1 Distância Euclidiana

É um conceito matemático que representa a menor proximidade entre dois pontos na geometria Euclidiana (ÁVILA, 2001). Segundo Latchi e Rocha (2005), a Distância Euclidiana ou função Euclidiana é uma métrica bastante popular usada para calcular similaridades em características contínuas.

Para calcular a distância Euclidiana pode-se utilizar a expressão a seguir:

$$d_2(x_i, x_j) = \|\sum(x_{i,k} - x_{j,k})\| \quad \text{para } k=1, \dots, d$$

Figura 1: Expressão da similaridade da Distância Euclidiana

Fonte: (LATCHI; ROCHA, 2005).

A Figura 1 acima mostra a expressão utilizada para calcular a Distância Euclidiana, onde:

d_2 = distância entre os padrões x_i, x_j no espaço de dimensão 2.

$x_{i,k}$ = componente k do padrão x_i

$x_{j,k}$ = componente k do padrão x_j

A distância Euclidiana é uma métrica de dissimilaridade ($d(i,j)$), mas a partir da dissimilaridade pode-se obter o cálculo da similaridade ($s(i,j)$) entre pontos por meio da fórmula a seguir:

$$s(i, j) = 1 - d(i, j)$$

O resultado da similaridade é obtido da diferença do valor 1 menos o valor da dissimilaridade. Nesta pesquisa foi usado o algoritmo *K-Means* juntamente com métrica da distância Euclidiana para obter um dos resultados dos agrupamentos dos aprendizes por meio dos dados das TAs.

Esta métrica foi utilizada por ser umas das métricas mais utilizadas para o cálculo do centro dos *clusters* (VARGAS, 2012). Outras métricas foram usadas com o mesmo algoritmo para o agrupamento dos aprendizes, estas serão descritas nas seções 3.5.2 e 3.5.3 a seguir.

2.5.2 Distância de *Manhattan*

Segundo Amaral (2016), é uma métrica que calcula a proximidade entre dois pontos baseada na soma das diferenças absolutas de suas coordenadas. A distância de *Manhattan* também é conhecida como *City-Block* esta pode ser definida pela expressão a seguir:

$$d(x, y) = \sum_{i=1}^n |x_i - y_i|$$

Figura 2: Expressão da similaridade da Distância de *Manhattan*

Fonte: (GOLDSCHMIT; BEZERRA; PASSOS, 2015).

A Figura 2 acima mostra a expressão utilizada para calcular a Distância Euclidiana, onde:

$d(x, y)$ = é a distância entre os padrões x_i, y_i no espaço de dimensão 1.

n = é a quantidade de componentes (coordenadas) de cada ponto.

x_i e y_i são respectivamente o valor do i -ésimo atributo para os dados x e y .

A distância de *Manhattan* é uma versão simples da distância Euclidiana. Esta distância foi utilizada juntamente com o algoritmo *K-Means*, para obter mais um resultado de recomendação para formação dos grupos de aprendizes. Tornando-se mais uma opção para o professor escolher de acordo com a sua necessidade.

2.5.3 Distância do Cosseno

A distância do cosseno ou similaridade do cosseno é uma métrica muito conhecida em sistemas de recomendação em geral (TESTA, 2013). Esta métrica calcula a relação entre dois vetores de acordo com ângulo de separação entre eles (SALAZAR, 2015). A distância do cosseno pode ser definida pela seguinte expressão na Figura 3:

$$d_{ij} = 1 - \frac{\sum_{k=1}^p x_{ik}x_{jk}}{\sqrt{\sum_{i=1}^p x_i^2} \sqrt{\sum_{j=1}^p x_j^2}}$$

Figura 3: Expressão da Distância do Cosseno.

Fonte: (SALAZAR, 2015)

A Figura 3 mostra que o resultado da similaridade é obtido da diferença do valor 1 menos o valor da distância do cosseno. Com o valor resultante pode-se verificar se a amostra é similar ou não. Se o valor for próximo de 1, mais similares são as amostras e quanto mais próximo de 0, menos semelhanças existem entre elas. Essas métricas de similaridade foram utilizadas devido às amostras (atributos) serem representadas, na implementação, por vetores e por serem métricas usadas em alguns trabalhos relacionados com essa pesquisa.

A distância do cosseno foi utilizada nessa pesquisa para validar os resultados das demais métricas usadas e, principalmente, para realizar análises futuras para agrupar alunos com características heterogêneas.

2.6 SISTEMA DE RECOMENDAÇÃO

No ambiente educacional há várias atividades que auxiliam na aprendizagem dos discentes. Existem algumas atividades que requerem colaboração, com isso, muitas vezes é necessário formar grupos de aprendizes para realizar essas atividades. Neste cenário, os Sistemas de Recomendação (SR) veem auxiliar na concepção da formação destes grupos.

De acordo com Cazella, Reategui e Machado (2009), sistemas de recomendação são sistemas que funcionam literalmente como filtros de informação, pois tendem a ajudar o usuário na pesquisa e escolha de um conteúdo baseado em seu perfil. Nessa pesquisa, o foco do sistema de recomendação estará no agrupamento de aprendizes para realizar atividades colaborativas.

Segundo Carvalho e Macedo (2014), os SR podem ser divididos em: sistemas de recomendação para indivíduos e sistemas de recomendação para grupos. Corroborando com essa ideia, Ferro *et al.* (2011), definem os sistemas de recomendação para indivíduos como sistemas que visam recomendar itens para usuários individuais. Os sistemas de recomendação para grupos são sistemas que realizam recomendações para um grupo de pessoas (BORATTO; CARTA, 2010). No que se refere ao tema desta pesquisa, sistemas de

recomendação de grupos são sistemas que visam recomendar ao professor, grupos de alunos para realização de atividades colaborativas.

2.6.1 Coleta de Informações de Usuário

Para caracterizar o perfil dos discentes em um sistema de recomendação é necessário realizar a coleta de dados para determinar as preferências dos usuários. Segundo Neto (2011), a coleta de informações é considerada sob duas perspectivas:

- Explícita: são informações que são solicitadas aos usuários por meio de técnicas de elicitación de requisitos como, por exemplo, o questionário.
- Implícita: são informações que são extraídas a partir da navegação do usuário, sem que estes tenham que fornecê-las diretamente, possibilitando inferir suas necessidades e preferências.

A coleta de dados é primordial para extrair as preferências e necessidades dos usuários, pois permite saber o interesse individual de cada membro e suas características enquanto aprendiz no auxílio da formação de grupos. Essa caracterização irá ajudar a criação tanto dos sistemas de recomendação para indivíduos quanto para grupos. Na abordagem desta pesquisa, utilizou-se coleta de dados implícita, para evitar constantes solicitações aos usuários.

2.6.2 Técnicas de Recomendação

Para a recomendação de grupos, faz-se necessário aplicar algumas das técnicas de recomendação ou a combinação delas. Segundo Cazella, Nunes e Reategui (2010), as técnicas de recomendação são utilizadas para o reconhecimento de padrões de comportamento em sistemas de recomendação. A seguir, são descritas algumas técnicas de recomendação.

2.6.2.1 Filtragem Baseada em Conteúdo

Segundo Herlocker (2000), há anos os cientistas têm direcionado seus esforços para amenizar a problemática causada pela sobrecarga de informações por meio de projetos que integram tecnologias automáticas que reconhecem e classificam as informações. Alguns *softwares* têm como finalidade gerar de maneira automatizada descrições dos conteúdos dos

itens e compará-las com os interesses dos usuários, propondo averiguar se o item é ou não relevante para cada um (BALABANOVIC; SHOHAM, 1997).

A filtragem baseada em conteúdo busca recomendar itens semelhantes já pesquisados anteriormente pelo usuário. Corroborando com esta ideia, Herlocker (2010) define a filtragem baseada em conteúdo em uma técnica que realiza uma seleção baseada na análise de conteúdo dos itens e do perfil do usuário, levando em consideração também histórico anterior.

2.6.2.2 Filtragem Colaborativa

A Filtragem Colaborativa é uma das técnicas de recomendação mais populares, consequentemente, uma das mais usadas em muitos sistemas existentes na Internet (SCHAFFER; KONSTAN; RIEDL, 2001). A técnica está baseada na análise de preferências semelhantes dentro de um grupo de pessoas. A essência desta técnica está na transferência mútua de experiências entre os usuários, que têm interesses similares e possuem “gostos” comuns por itens (CAZELLA; REATEGUI; MACHADO, 2009).

Em linhas gerais, esta técnica usa as experiências coletivas de um grupo para determinar a recomendação para um membro do grupo. A técnica de filtragem colaborativa foi desenvolvida para suprir necessidades que foram detectadas na filtragem baseada em conteúdo (HERLOCKER, 2000). Essa técnica se distingue da filtragem baseada em conteúdo principalmente por não demandar a compreensão ou reconhecimento do conteúdo dos itens (CAZELLA; NUNES; REATEGUI, 2010).

2.6.2.3 Filtragem Híbrida

A filtragem híbrida é uma técnica que busca unir o melhor das técnicas de filtragem baseada em conteúdo e colaborativa, possibilitando desenvolver um sistema que atenda às necessidades do usuário (HERLOCKER, 2000). A Figura 4, a seguir, apresenta as vantagens proporcionadas pela filtragem híbrida.

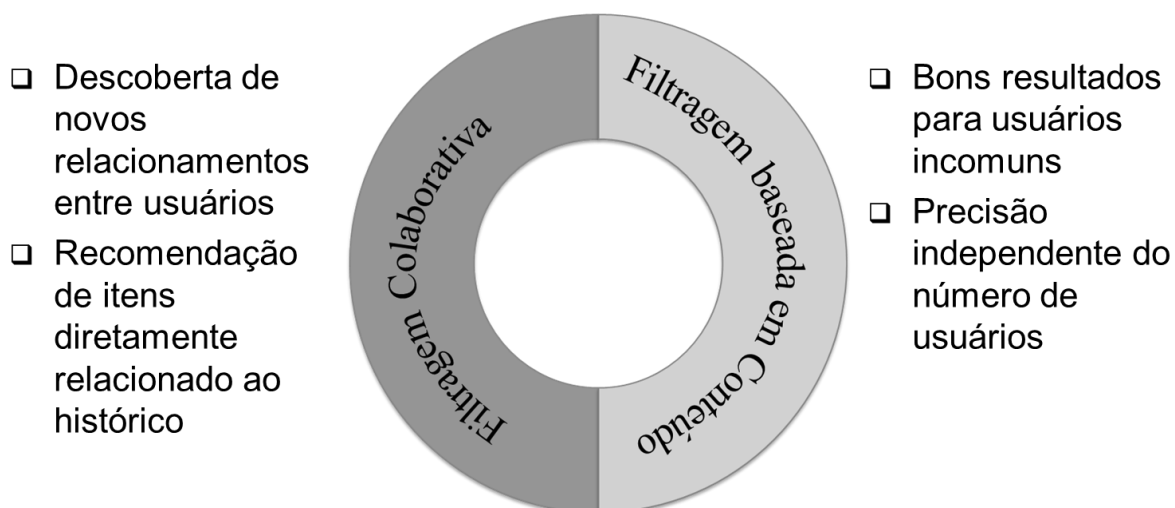


Figura 4: Filtragem Híbrida.

Fonte:(CAZELLA; NUNES; REATEGUI, 2010)

Martins *et al.* (2011) citam algumas características da filtragem híbrida que constituem-se em: descoberta de novos relacionamentos entre usuários; recomendação de itens diretamente correlatos ao histórico; bons resultados para usuários incomuns e precisão independentemente do número de usuários.

2.7 GRAFOS

Nesta seção serão abordados diversos conceitos que auxiliarão no entendimento da utilização de grafos para visualização das trilhas de aprendizagem, visto que, essas servem de base para caracterizar os aprendizes na recomendação dos grupos de alunos desta pesquisa. A seguir, serão descritos tópicos que ajudam na compreensão dos conceitos.

2.7.1 Grafo: Noções Básicas

Para entender como os grafos foram utilizados para visualização das trilhas de aprendizagem é imprescindível a compreensão do conceito de grafo. Nos relatos Goldbarg e Goldbarg (2012), grafo é uma estrutura abstrata que representa um conjunto de vértices e arestas. Os vértices ou nós são os elementos dos conjuntos que podem ser representados por pontos ou círculos. As arestas ou arcos são as relações ligadas aos vértices que podem ser representadas por traços ou setas ligando os pontos.

Nessa pesquisa, os vértices são representados por círculos e as arestas por setas ligando os círculos. Os grafos que foram utilizados nesse trabalho são os idealizados no trabalho de Ramos (2016) que tem como objetivo criar e validar um modelo de identificação e representação visual das trilhas de aprendizagem.

As trilhas são baseadas nas ações de interação dos alunos. Como resultado da pesquisa, obteve-se o LPGraph como *plugin* no Moodle para gerar um grafo individual de cada discente levando em consideração sua trilha de aprendizagem. A Figura 5 apresenta um exemplo do grafo que foi usado por esta pesquisa.

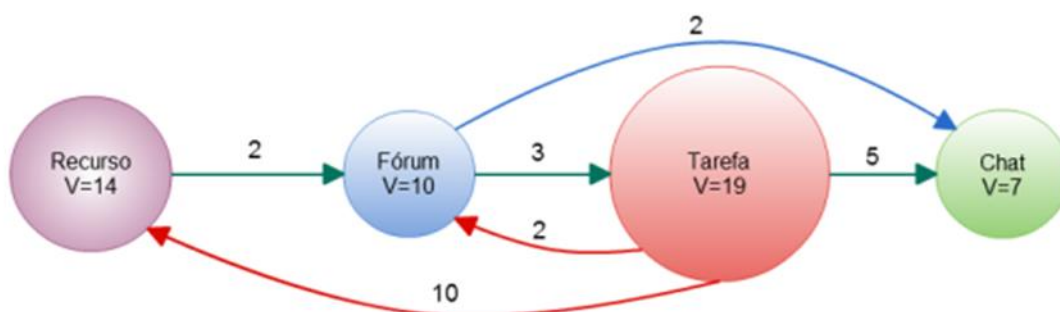


Figura 5: Grafo da Trilha de Aprendizagem

Fonte: (RAMOS, 2016)

A Figura 5 apresenta um grafo individual (por aluno) da trilha de aprendizagem gerado pelo *plugin* LPGraph. Segundo Ramos *et al.* (2016a), cada vértice representa um recurso/atividade que possui uma cor e o diâmetro de cada vértice é proporcional ao número de visualizações (V) realizadas pelo aluno. As arestas possuem 3 cores, a de cor verde é a padrão representa a estrutura de TA pretendida pelo professor no AVA, a de cor azul é a de avanço e representa o percurso que o aluno realizou a partir de um recurso/atividade para o outro diretamente. As arestas, nas cores vermelhas, representam os retornos realizados pelo aluno nos recursos/atividades anteriores. Os valores nas arestas indicam o número de vezes que o aluno percorreu de um vértice A para um vértice B e a seta indica o sentido de transição.

2.7.2 Tipos de Grafos

Para idealizar um grafo que será utilizado em uma pesquisa é necessário saber qual a problemática envolvida e o tipo de grafo que irá se adequar às necessidades do problema para

o desenvolvimento da solução. A seguir, na Figura 6, serão apresentados alguns tipos de grafos.

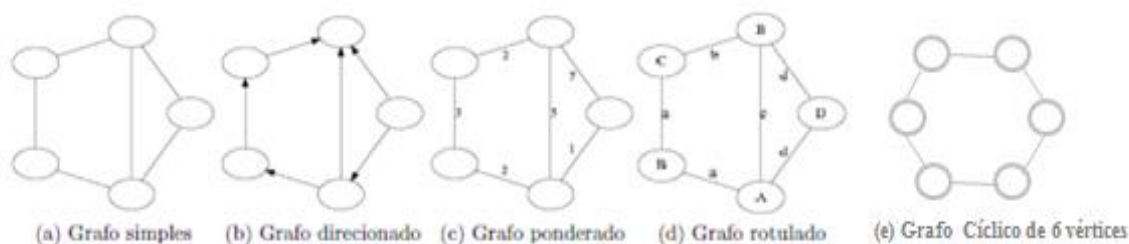


Figura 6: Tipos de Grafo.

Fonte: (TEIXEIRA, 2011)

As definições dos tipos de grafos adotadas nesta pesquisa baseiam-se nas descrições de Netto (2011) e Goldbarg e Goldbarg (2012). O grafo simples, conforme Figura 6, é um grafo que não inclui laços ou arestas paralelas. Laço é uma aresta que liga um vértice a ele mesmo. O grafo ponderado é um grafo em que as arestas ou vértices têm valores numéricos (pesos) na sua estrutura. O grafo direcionado ou orientado é quando o sentido da aresta é relevante, as arestas possuem setas e estas são denominadas de arcos. O grafo rotulado é um grafo em que os vértices e arestas possuem tipos ou rótulos e o grafo cíclico é um grafo que possui ciclos.

Todos os grafos podem ser visualizados na Figura 6. O grafo que será utilizado nesta pesquisa possui como características ser direcionado, ponderado e cíclico. Direcionado porque o sentido da aresta é importante. Ponderado, pois há arestas com valores (pesos) e estes podem ter valores diferentes e por último, pode ser cíclico, caso os alunos retornem para algum recurso que já utilizaram.

Nessa pesquisa foram analisados os grafos das TAs dos alunos. Os grafos, segundo Medeiros (2013), são um recurso heurístico utilizado para determinar níveis de semelhança. Esses níveis de semelhanças entre os grafos serão utilizados para auxiliar no agrupamento dos discentes. A partir da análise realizada, foram extraídos dados das TAs baseada em grafo para compor os atributos necessários para realizar o agrupamento dos aprendizes. Mais detalhes serão elucidados no capítulo 4.

3 TRABALHOS RELACIONADOS

Este capítulo apresentará o planejamento elaborado para realização da revisão bibliográfica nos padrões de uma Revisão Sistemática da Literatura (RSL) e os trabalhos mais similares que de alguma forma se relacionam com as características desta pesquisa.

3.1 REVISÃO SISTEMÁTICA DE LITERATURA

Após a escolha do tema de pesquisa, foi realizada uma revisão bibliográfica nos moldes de uma Revisão Sistemática da Literatura (RSL), seguindo as orientações de Kitchenham (2004).

Esta revisão constitui como base fundamental deste trabalho. Um estudo sobre recomendação de grupos em ambientes virtuais foi realizado, por meio de uma RSL, cujo resultado está registrado em Ramos e Oliveira (2015).

A revisão pode ser classificada como uma revisão *quasi*-sistemática, uma vez que não houve comparações entre publicações, bem como, uma pesquisa exploratória, preservando o mesmo formalismo e seguindo o mesmo processo da revisão sistemática (TRAVASSOS *et al.*, 2008).

Durante o processo de revisão sistemática, utilizou-se uma ferramenta especialmente para este fim, chamada de StArt³, que segundo Menezes, Vale e Cruz (2013), tem objetivo de auxiliar no Planejamento, Execução e Análise Final de uma RSL.

Esta ferramenta foi desenvolvida na Universidade Federal de São Carlos, pelo Laboratório de Pesquisa em Engenharia de Software (LaPES). A versão 2.3.4.2 foi utilizada para auxiliar esta pesquisa.

3.1.1 Objetivo e Questão da Pesquisa

O objetivo da RSL foi identificar modelos de recomendação de grupos para atividades colaborativas que fazem o uso de coleta de dados implícitas, utilizados dentro dos ambientes virtuais. A descrição do objetivo, conforme o paradigma *Goal-Question-Metric* (BASILI; CALDIERA; ROMBACH, 1994), pode ser observada abaixo:

³ http://lapes.dc.ufscar.br/tools/start_tool

Analisar publicações científicas através de um estudo baseado em revisão sistemática. **Com o propósito de** identificar modelos ou combinações de modelos de recomendação de grupos para atividades colaborativas utilizando coletas de dados implícitas utilizados na literatura voltados para Ambientes Virtuais de Aprendizagem. **Com relação ao** agrupamento de alunos, **do ponto de vista dos** pesquisadores, por meio do **contexto** acadêmico ou industrial. A RSL buscou responder a seguinte questão principal: *Como se constroem os agrupamentos para realizar atividades colaborativas de aprendizagem em Ambientes Virtuais?*

3.1.2 Métodos de Busca das Publicações

Utilizou-se as máquinas de buscas específicas, das bibliotecas selecionadas, sendo consultados os conjuntos de palavras-chaves pré-estabelecidos e apenas fontes digitais foram consideradas nesta pesquisa. As fontes digitais consultadas foram:

- IEEE⁴ *Computer Science Digital Library*;
- Scopus⁵.

A Scopus destaca-se por armazenar publicações de diversas fontes como, por exemplo: *Springer, ACM, Science Direct/Elsevier, British Computer Society*.

Além disso, Kitchenham e Chartes (2007) afirmam que a *Scopus* é a maior base de dados de indexação de resumos e citações de literatura científica. A *Scopus* é uma base multidisciplinar abrangendo as áreas de ciência, tecnologia, medicina, ciências sociais e artes e humanidades.

3.1.3 Expressão da Busca

A expressão de busca foi definida após a execução de testes auxiliados com base em artigos de controle e é apresentada no Quadro 2. Ela é composta por palavras-chave que representam a população (P) e a intervenção (I).

A composição básica da expressão de busca segue a seguinte estrutura: (P) AND (I). As palavras-chave da expressão de busca são apresentadas abaixo. Para a busca na Scopus, as aspas (") foram substituídas por chaves ({}), sendo inicializada com TITLE-ABS-KEY.

⁴ <https://www.ieee.org/index.html>

⁵ <http://www.scopus.com>

Quadro 2: Termo de busca

População
"learning management system" OR "virtual learning environment" OR "moodle" OR "distance education" OR "distance learning" OR "e-learning recommender system"
Intervenção
"clustering" OR "group recommendation" OR "group modeling" OR "group recommender systems" OR "Profiles Aggregation"

3.1.4 Critérios de Inclusão e Exclusão

Os critérios de inclusão (CI) e de exclusão (CE), listados no Quadro 3, foram preparados a partir dos artigos de controle.

Quadro 3: Critérios de Inclusão e Exclusão

Critérios de Inclusão
(CI1) Artigos que apresentem representações de modelos, técnicas, processos, algoritmos de agrupamento sobre a formação de grupos de alunos em ambientes virtuais. (CI2) Artigos que apresentem propostas de adaptação de trilhas de aprendizagem em ambientes virtuais.
Critérios de Exclusão
(CE1) Não serão selecionadas publicações em que as palavras-chave da busca não apareçam no título, resumo e/ou palavras-chaves. (CE2) Não serão selecionadas publicações em que as siglas equivalentes não signifiquem: LMS - “Learning Management System, VLE – “Virtual Learning Environment”. (CE3) Serão excluídas publicações que não tem disponibilidade de conteúdo para leitura (por exemplo, onde os trabalhos são pagos ou não disponibilizados pelas máquinas de buscas). (CE4) Artigos que apenas apresentam relatos de experiência de algum modelo, técnica ou ferramenta que não estejam relacionadas a grupo. (CE5) Artigos que apenas citam sistemas de e-Learning (LMS, VLE, etc.), mas não o abordam como foco principal. (CE6) Não serão selecionadas publicações que não relacionam formação de grupos e Ambientes Virtuais de Aprendizagem (LMS, CMS, ITS, LCMS). (CE7) Não serão selecionadas publicações que apresentam ambientes virtuais de aprendizagem, mas não descrevam formação de grupos ou recomendação de grupos. (CE8) Não serão selecionadas publicações que descrevam formação de grupos ou recomendação de grupos, mas não relacionados a ambientes virtuais de aprendizagem.

3.1.5 Extração dos Dados

Os dados extraídos durante a leitura dos artigos que foram aprovados no primeiro e segundo filtro seguiram uma lista de itens predefinidos. Todos os itens são de preenchimento obrigatório, exceto quando não informados ou quando não se aplicam à publicação. Dados

básicos da publicação: título, autores, local de publicação, ano de publicação, palavras-chave e status.

Os dados relativos às características que auxiliam na resposta da questão de pesquisa: técnicas utilizadas no trabalho; ambiente virtual utilizado; tipo de AVA; público-alvo do trabalho; abrangência do trabalho; tipo da pesquisa; resumo de cada trabalho; quais técnicas para o agrupamento foram utilizadas?; a ferramenta apresenta formação de grupos?; para a formação dos grupos, foram utilizados os *logs* dos usuários?

3.1.6 Condução da Revisão

Após a fase de planejamento, iniciou-se em novembro de 2015, a fase de condução da RSL, tendo como resultado da execução da (*string*) expressão de busca 316 artigos, sendo 195 encontrados na biblioteca da Scopus artigos e 121 na IEEE.

Estes artigos foram importados para a ferramenta StArt e após a importação foram identificados 35 artigos duplicados e 281 artigos para análise do primeiro filtro, como mostra a Figura 7.

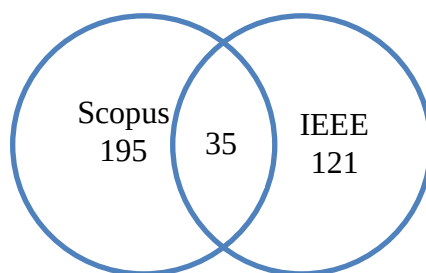


Figura 7: Artigos retornados pela String.

No primeiro filtro, foram analisados os 281 artigos por meio dos critérios de inclusão e exclusão, como mostra a Figura 8. Para isso, o título, resumo/*abstract* e as palavras-chaves de cada publicação foram analisadas, para verificar quais artigos são úteis para o contexto da pesquisa.

Os artigos, cujos conteúdos para leitura não foram obtidos diretamente da fonte de pesquisa (Scopus e IEEE), foram pesquisados via máquina de busca (*google.com.br*) e/ou os autores foram contatados via e-mail.

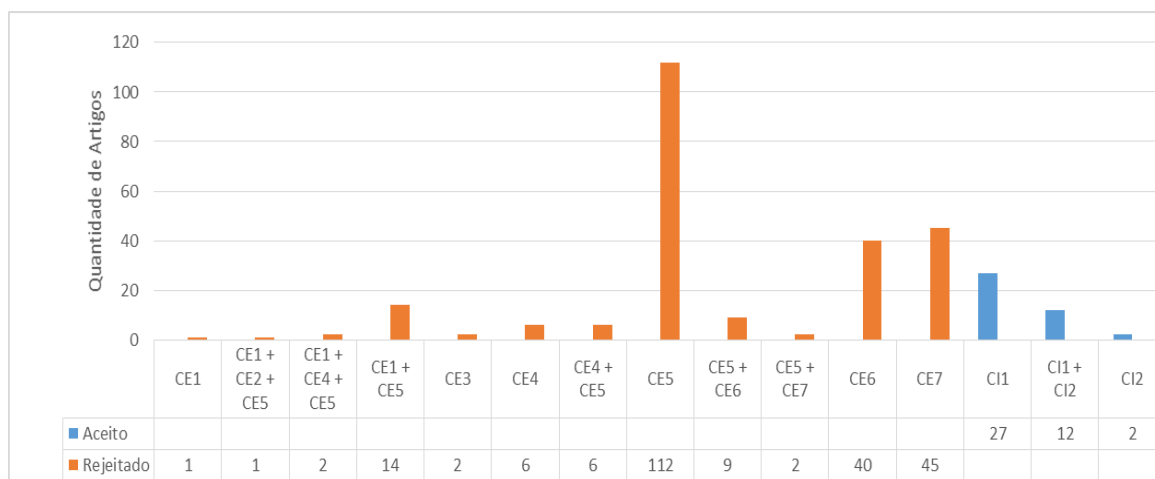


Figura 8: Resultado do Primeiro Filtro

Após execução do primeiro filtro, 41 artigos foram aceitos para execução do segundo filtro e 240 rejeitados conforme visualização da Figura 9 a seguir.

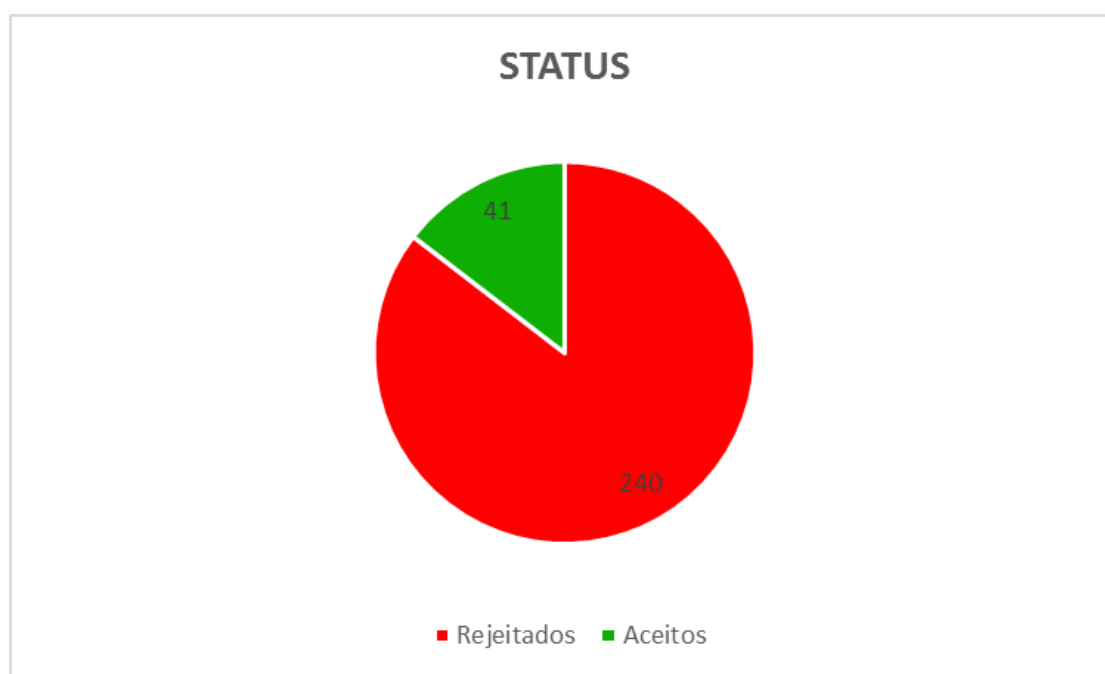


Figura 9: Artigos Aceitos e Rejeitados.

No processo de execução do segundo filtro, foram analisados os 41 artigos aceitos no primeiro filtro. A partir dessa constatação os artigos foram lidos na íntegra e analisados conforme os critérios de inclusão e exclusão.

Após a execução do segundo filtro, obteve-se 19 artigos rejeitados, 3 duplicados e 19 aceitos, conforme apresentado na Figura 10 a seguir.

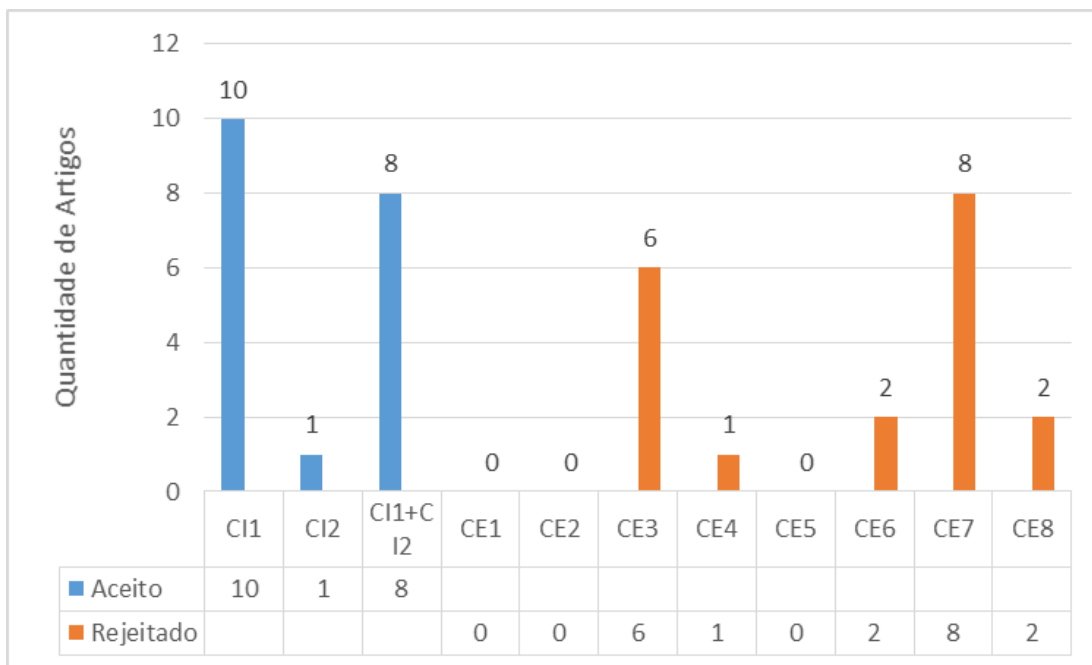


Figura 10: Resultado do Segundo Filtro.

3.1.7 Análise e Discussão dos Dados

Os trabalhos relacionados foram encontrados durante a revisão *quasi-sistemática*, resultando em um *survey* sobre formação de grupos em ambientes virtuais.

Na RSL a questão principal era saber como se constroem os agrupamentos para realizar as atividades de aprendizagem em ambientes virtuais. Ao final da RSL, foram identificadas 19 publicações pertinentes, das quais cinco estão relacionadas similarmente ao objetivo desta proposta de trabalho.

Foi percebida, na análise dos resultados da revisão, que para a formação de grupos de aprendizes, as coletas de dados para se extrair as informações sobre as preferências dos discentes, é feita 52,63% de forma explícita, ou seja, os dados são solicitados aos aprendizes por meio de alguma técnica de coleta de dados, e 47,36% de forma implícita por meio dos *logs* de acesso dos usuários.

Nota-se também que 58,33% usam em sua composição para formação de grupos de alunos o algoritmo de agrupamento *K-Means* e que 61,53% dos trabalhos que usam os modelos ou sistemas de recomendação são aplicados na plataforma Moodle.

Os atributos na coleta de dados implícita, que são utilizados para caracterizar os aprendizes, são em sua maioria atividades como quiz, fórum e chat extraídos dos *logs* de acesso do AVA.

O grande foco de pesquisa nos trabalhos foi identificar modelos ou combinações de modelos de recomendação de grupos para atividades colaborativas utilizando coletas de dados implícitas utilizados na literatura voltados para Ambientes Virtuais de Aprendizagem.

Verificou-se também que 73,68% dos trabalhos analisados foram publicados somente nos últimos 5 anos, sendo que o trabalho mais antigo foi publicado em 2002.

3.2 TRABALHOS RELACIONADOS SOBRE RECOMENDAÇÃO DE GRUPOS EM AMBIENTES VIRTUAIS DE APRENDIZAGEM

Os trabalhos a seguir apresentam pontos de vista e formas diferentes de agrupamentos de alunos em AVAs. Eles fazem parte do resultado da RSL e mostram também que muitas técnicas diferentes podem ser aplicadas. A seguir, são descritos os trabalhos mais similares ao trabalho proposto.

3.2.1 Supporting the Composition of Effective Virtual Groups for Collaborative Learning

O artigo de Daradoumis *et al.* (2002) relata a formação de grupos de alunos em ambientes virtuais de aprendizagem, como uma dinâmica no processo colaborativo.

Este defende que os estudantes devem se conhecer em um ambiente informal, para começar a desenvolver relacionamentos, avaliar e aprender uns com os outros, antes de propriamente realizarem atividades no AVA, permitindo a eficácia da equipe posteriormente.

Neste contexto, a pesquisa menciona princípios básicos do ciclo de vida do grupo. Para implementá-lo, o autor propõe um regime de quatro fases, que consiste de processos bem definidos cujo objetivo é envolver os alunos em atividades que irão levar à criação do bom funcionamento dos grupos de aprendizes. A seguir, a Figura 11 mostra a abordagem geral seguida pelo autor.

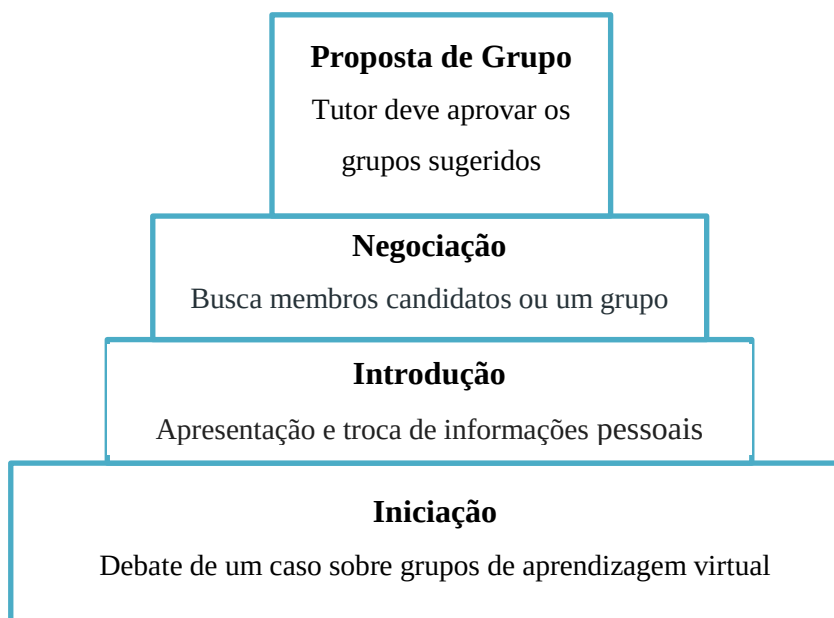


Figura 11: Fases para Formação dos Grupos.

Fonte: (DARADOUMIS *et al.*, 2002)

Um breve resumo sobre as quatro fases relatadas por Daradoumis *et al.* (2002):

- As fases de iniciação permitem que os alunos aumentem a sua compreensão sobre a noção e a função dos grupos de aprendizagem colaborativa;
- A fase de introdução incentiva os alunos a definir metas e expectativas individuais, enquanto fomenta a abertura e honestidade quanto às suas capacidades, habilidades e atitudes para a colaboração.
- Negociação: Para formar um grupo de aprendizagem que satisfaça tanto o individual e os objetivos do grupo ou para procurar um grupo aberto que se encaixa metas e necessidades pessoais melhores de cada um.
- Proposta de grupo: informa e pergunta do tutor se ele quer aprovar a formação definitiva dos grupos de aprendizagens sugeridos.

O ambiente virtual utilizado para a formação dos grupos foi o Campus Virtual, um ambiente da própria instituição de ensino. Esta abordagem permite estabelecer grupos de pessoas que se escolhem para ser parte do mesmo grupo, em razão de algum interesse compartilhado.

3.2.2 Grouping in Collaborative E-Learning Environment Based On Interaction among Students

O trabalho de Jagadish (2014) descreve a formação de grupos de aprendizes por meio da aprendizagem colaborativa utilizando a plataforma Moodle. O algoritmo *KNN* (*k-Nearest Neighbor*) foi utilizado para realizar a formação dos grupos de aprendizes.

Esta pesquisa usa o chat da plataforma para realizar as interações entre os alunos, possibilitando extrair dados sobre as preferências dos usuários por meio dos *logs* de acesso do Moodle. A coleta implícita de informações é utilizada para classificar os perfis dos discentes por meio do algoritmo *KNN*. Ainda na fase de coleta de dados, também foi usado o preenchimento de questionários para a obtenção de traços de personalidade e estilos de aprendizagem dos alunos.

3.2.3 An Evolutionary Algorithm for Forming Mixed Groups of Learners in Web Based Collaborative Learning Environments

A pesquisa de Abnar, Orooji e Taghiyareh (2012) propõe um processo interativo com base em algoritmo genético e escala de *Likert*. O algoritmo é flexível para o número e o tipo dos atributos. Para diferentes contextos, conjuntos diferentes de atributos podem ser usados para formar grupos de aprendizagem. Na verdade, o professor tem a possibilidade de escolher um conjunto diferente de atributos e classificá-las com base em seu impacto sobre a formação de grupos. Além disso, o modelo de estilo de aprendizagem de *Felder-Silverman* e grafos são utilizados também para representação.

Nesse contexto, o aluno é representado como um vetor, cada dimensão desse vetor descreve um atributo pessoal do discente na escala de *Likert*. De acordo com Abnar, Orooji e Taghiyareh (2012), são utilizados também dois tipos de grafos na proposta, que são:

- Grafo de distância: é um grafo ponderado, onde seus nós são os aprendizes e os pesos das arestas entre os nós indica a distância entre os alunos.
- Grafo binário de distância: seus nós representam os alunos e suas arestas indicam se os alunos estão perto suficientes ou não.

Essa abordagem permite que o professor possa adicionar, selecionar e classificar as características de construção do grupo, criar grupos, e avaliar a formação dos aprendizes após a realização de atividades de colaboração, permitindo também ao aluno avaliar os seus colegas durante o tempo que ficaram interagindo. O ambiente virtual utilizado foi o Moodle.

3.2.4 Learner Classification Based on Learning Behavior and Performance

O estudo proposto por Yathongchai *et al.* (2013) aborda o agrupamento de estudantes na plataforma Moodle usando as técnicas de árvore de decisão, mineração de dados e índice de *Hartigan* considerando duas etapas principais: 1) Aplicar o algoritmo de agrupamento *K-Means* para analisar o comportamento de aprendizagem de cada aluno, com base no perfil gerado por meio dos *logs* de acesso do Moodle. Como resultado, temos quatro grupos apropriados que são criados com base no perfil dos aprendizes; 2) Aplicação de um classificador de árvore de decisão para gerar o modelo de classificação do aluno, baseado nos comportamentos e desempenho dos aprendizes por meio de suas notas.

3.2.5 A New Approach in E-Learners Grouping Using Hybrid Clustering Method

A pesquisa de Montazer e Rezaei (2013) aborda a criação de um método híbrido de agrupamento que consiste na junção dos métodos *K-means* e *C-Means Fuzzy*. Este método híbrido foi avaliado comparando com outros três métodos tradicionais: *K-means*, *Fuzzy C-Means* e *C-Means Fuzzy Evolutivo*. Como pode ser visualizado na Figura 12 abaixo.

methods index	Fuzzy C- means	K-means	EFC	HCM
Gat(C)	0.91	.094	0.91	0.95
Pur(C)	0.45	0.45	0.45	0.47
PG(C)	1.36	1.39	1.36	1.42

Figura 12: Comparação de Avaliação entre Métodos.
Fonte:(MONTAZER; REZAEI, 2013)

A coleta de dados é realizada de forma implícita por meio dos *logs* de acesso da plataforma, sendo que esta não foi informada no trabalho. Podemos perceber, por meio da Figura 12, que o Método Híbrido se destacou na avaliação em relação aos demais algoritmos de agrupamento. Na seção a seguir, é apresentada a comparação entre os trabalhos elencados e o trabalho proposto.

3.3 COMPARAÇÃO ENTRE TRABALHOS SIMILARES

Os trabalhos elencados no quadro a seguir, apresentam os trabalhos similares em relação a esta pesquisa. Estes foram de fundamental importância para ajudar na composição e subsidiar olhares para esta proposta.

É possível verificar no Quadro 4, o comparativo dos trabalhos similares relacionados ao contexto desta pesquisa. É válido esclarecer que os campos com traço são dados que não puderam ser identificados nas publicações.

Quadro 4: Comparação entre os trabalhos similares e a pesquisa.

	Tipo de Coleta de Dados	Usado com Moodle	Fonte dos dados	Técnicas usadas para formar os grupos
(DARADOUMIS et al., 2002)	Explícito	Não	Questionários	Raciocínio Baseado em Casos
(JAGADISH, 2014)	Explícito	Sim	Questionário	Algoritmo <i>KNN (k-Nearest Neighbor)</i>
(ABNAR; OROOJI; TAGHIYAREH, 2012)	Implícito	Sim	<i>Logs(-)</i>	Algoritmo genético
(YATHONGCHA I et al., 2013)	Implícito	Sim	<i>Logs</i> (Frequência de login, número de tempo de download, entre outros)	<i>K-Means e C4.5</i>
(MONTAZER; REZAEI, 2013)	Implícito	-	<i>Logs (-)</i>	<i>K-Means e C-Means Fuzzy</i>
Proposta de trabalho	Implícito	Sim	Dados das Trilhas de Aprendizagem representadas por grafos	<i>K-Means</i>

No Quadro 4, cada linha apresenta as informações de cada trabalho analisado, a última linha representa a proposta de trabalho. As colunas possuem os seguintes significados:

- Tipo de coleta de dados: quais as formas de coleta de dados que podem ser extraídos as informações para a formação dos perfis dos usuários.
- Usado com Moodle: indica que o trabalho utilizou Moodle para aplicar a técnica apresentada.
- Fonte dos dados: como são extraídos os dados do banco de dados.
- Técnicas usadas para formar os grupos: quais tipos de técnicas são usados para realizar o agrupamento dos aprendizes.

Como visto no Quadro 4, os cinco trabalhos citados estão relacionados com esta proposta, sendo que as abordagens de (DARADOUMIS *et al* 2012 e JAGADISH, 2014), realizaram a formação de grupos de forma explícita, já os trabalhos de (ABNAR; OROOJI e TAGHIYAREH, 2012; YATHONGCHAI *et al.*, 2013 e MONTAZER; REZAEI, 2013) formaram os grupos de forma implícita, sendo que a maioria destas pesquisas usaram o Moodle e técnicas de mineração de dados ou raciocínio baseada em casos.

As fontes de dados usadas para caracterizar os perfis dos aprendizes nas duas primeiras pesquisas foram por meio de questionários e das três últimas por logs de acessos extraídos da plataforma Moodle.

Logo, essa pesquisa difere das apresentadas, pois não se encontrou na literatura trabalhos que usaram as TAs para a formação dos grupos dos aprendizes. Além de que, nos trabalhos acima, a técnica de mineração de dados utilizada, em sua maioria, usa uma métrica de similaridade, alcançando um único resultado.

Já nesta abordagem de pesquisa ora apresentada é utilizado o algoritmo *K-Means* com três métricas de similaridade: Euclidiana, *Manhattan* e Cosseno.

Com cada métrica o algoritmo *K-Means* classifica os atributos e sugere os clusters de alunos, obtendo-se um resultado para cada distância, permitindo com que o docente tenha três opções de sugestões para recomendar grupos aos alunos de forma automática. Mais detalhes sobre as sugestões de recomendações podem ser elucidados no Capítulo 5.

4 PROPOSTA DE TRABALHO

Neste capítulo, serão apresentadas a metodologia, a visão geral do framework, a proposta da arquitetura do trabalho e a descrição das principais ferramentas que foram utilizadas para concluir este trabalho.

4.1 METODOLOGIA DE PESQUISA

A proposta de pesquisa foi iniciada com a realização da pesquisa bibliográfica, por meio de uma Revisão Sistemática da Literatura sobre recomendação de grupos para atividades colaborativas em ambientes virtuais. Tal revisão sistemática ocorreu a partir de consultas em bases de conhecimento científico, como: Scopus e IEEE.

Para a etapa seguinte do planejamento proposto foram realizadas pesquisas sobre as temáticas de métricas de similaridades e técnicas de mineração de dados, haja vista, a obtenção de subsídios suficientes para extrair os dados dos grafos baseados nas trilhas de aprendizagem dos alunos e verificar a similaridade entre as amostras e, consequentemente, usar o *K-Means* para classificar os alunos.

O passo posterior consistiu na identificação do cenário educacional correspondente à aprendizagem em grupo a distância. Nesta fase ocorreu uma pesquisa de campo, de onde foram coletadas as informações de duas turmas fechadas da disciplina de Matemática Discreta, dos semestres de 2015/2 e 2016/1 e do curso de Ciência e Engenharia da Computação no Moodle. Os dados foram analisados e posteriormente transformados em atributos sendo organizados para a realização dos experimentos iniciais de formação de grupos. Esses dados foram colocados em um formato independente do ambiente virtual, no ambiente Eclipse com a integração da *Application Programming Interface* (API) do *Waikato Environment for Knowledge Analysis* (WEKA), usando a linguagem de programação JAVA. O objetivo dessa análise foi verificar como as trilhas de aprendizagem, baseadas em grafos, em conjunto com as métricas de similaridade e o *K-Means*, formariam o cenário de agrupamento dos aprendizes.

Em seguida, foi desenvolvido um *framework* conceitual e para verificar a sua factibilidade foi criado um *plugin*, o M-CLUSTER, que utilizou os dados produzidos pelo passo anterior e sugeriu a formação dos grupos de acordo com a análise realizada. Após, foi realizada uma nova pesquisa de campo com uma turma real em andamento da disciplina

Matemática Discreta dos cursos de Ciência e Engenharia da Computação do semestre de 2016/2 da UFAM, para validar o *framework* conceitual.

O *plugin* foi incorporado na plataforma ColabWeb (*Moodle* customizado da UFAM) onde foram analisados e validados os dados do passo anterior de acordo com o *framework*, buscando identificar a contribuição da abordagem proposta para a atividade docente. Neste passo foram aplicados questionários e reuniões foram realizadas com os docentes (especialistas) para coletar os resultados da utilização da ferramenta, bem como, verificar o desempenho dos alunos na disciplina (notas), em relação às atividades de colaboração que foram avaliadas durante o curso para validar a criação dos grupos sugeridos pela ferramenta. A partir dessas informações, o *framework* conceitual usado por esta proposta foi concebido e é apresentado na próxima seção.

4.2 FRAMEWORK CONCEITUAL DE RECOMENDAÇÃO PARA FORMAÇÃO DE GRUPOS DE APRENDIZES EM AVA

Diante da problemática apresentada sobre formação grupos em Ambiente Virtuais de Aprendizagem, a presente pesquisa se apoia nas várias iniciativas anteriores, elencadas nos trabalhos relacionados, que objetivam viabilizar agrupamentos de forma automática ou não para AVAs. Nesse contexto, propõem-se a criação de um *framework* conceitual baseado nos dados extraídos das Trilhas de Aprendizagem para Recomendação de Grupos de Aluno em um AVA (TARGAVA).

Esta seção está organizada em subseções que tratam sobre considerações iniciais acerca do *framework*, visão geral da proposta, os componentes e suas finalidades.

4.2.1 Considerações Iniciais

Antes de apresentar o *framework* conceitual proposto por esta dissertação é necessário elucidar alguns conceitos, convenções adotadas e esclarecimentos que ajudam ao leitor no melhor entendimento de nossa proposta.

Segundo, Gaertner (2005) *frameworks* são esquemas usados como uma forma simplificada de clarificar temas complexos em formas que possam ser estudadas e analisadas, e ainda podem segundo Shehabuddeen *et al.* (2000):

- Representar um problema para um escopo definido;
- Mostrar a relação entre vários elementos;

- Permitir uma visão ampla de uma determinada situação para ser capturada;
- Demonstrar opções para se resolver um determinado problema;
- Fornecer uma abordagem estruturada para lidar com um propósito específico.

Nesta dissertação, o conceito de *framework* usado será o estabelecido por Gaertner (2005). O *framework* criado por esta pesquisa pode ser reusável e portátil podendo ser aprimorado ou usado para outros propósitos com a mesma finalidade. A seguir é apresentada uma visão geral do *framework*.

4.2.2 Visão Geral do *Framework*

O *framework* conceitual proposto nesta dissertação, denominado TARGAVA tem como objetivo recomendar grupos de alunos para atividades colaborativas, utilizando a caracterização dos aprendizes baseadas nas TAs em um AVA, para auxiliar o professor no processo de ensino-aprendizagem.

Foram realizadas coletas de dados do banco de dados do AVA Moodle, para fazer análises das TAs criadas por Ramos (2016), buscando definir atributos para serem utilizados na concepção dos grupos de alunos. Esses atributos foram encontrados e validados, por meio de entrevistas com a professora especialista.

A entrevista é uma técnica de levantamento de requisitos da engenharia de software (SOMMERVILLE, 2011) e um dos critérios de validade internos, para a classificação de agrupamentos não-supervisionados (VARGAS, 2012).

Tendo seus principais componentes devidamente explicados na seção 4.2.3, a Figura 13 mostra como todas essas peças se encaixam para realizar as recomendações de grupos de alunos usando a importância da opinião da professora especialista. Este *framework* foi baseado na taxonomia de Ounnas, Davis e Millard, (2009).

O fluxo ideal de funcionamento, partindo da navegação dos alunos na plataforma e usuários já cadastrados, funciona da seguinte forma: o aluno interage com a plataforma Moodle, mais especificamente com o ColabWeb e a partir disso, são coletados os logs de arquivos e dados dos alunos.

Dessas informações são extraídos os dados para formar as trilhas de aprendizagem dos alunos. Das TAs são extraídos atributos que serão utilizados para classificar e formar os grupos de aprendizes usando o algoritmo *K-Means* em conjunto com as três métricas de similaridade e os resultados dessas classificações são exibidas ao docente/tutor da turma.

Neste passo, a importância da opinião do professor especialista está desde a concepção dos atributos para formar os agrupamentos na fase inicial de concepção do *framework*, até a escolha dos clusters formados pelo *plugin* M-CLUSTER para serem utilizados nas atividades colaborativas com os alunos.

Ao final da classificação, o professor da disciplina faz as avaliações sobre os clusters de alunos que lhes foram recomendados. Informações mais detalhadas sobre os componentes e suas funcionalidades são apresentadas na seção a seguir.

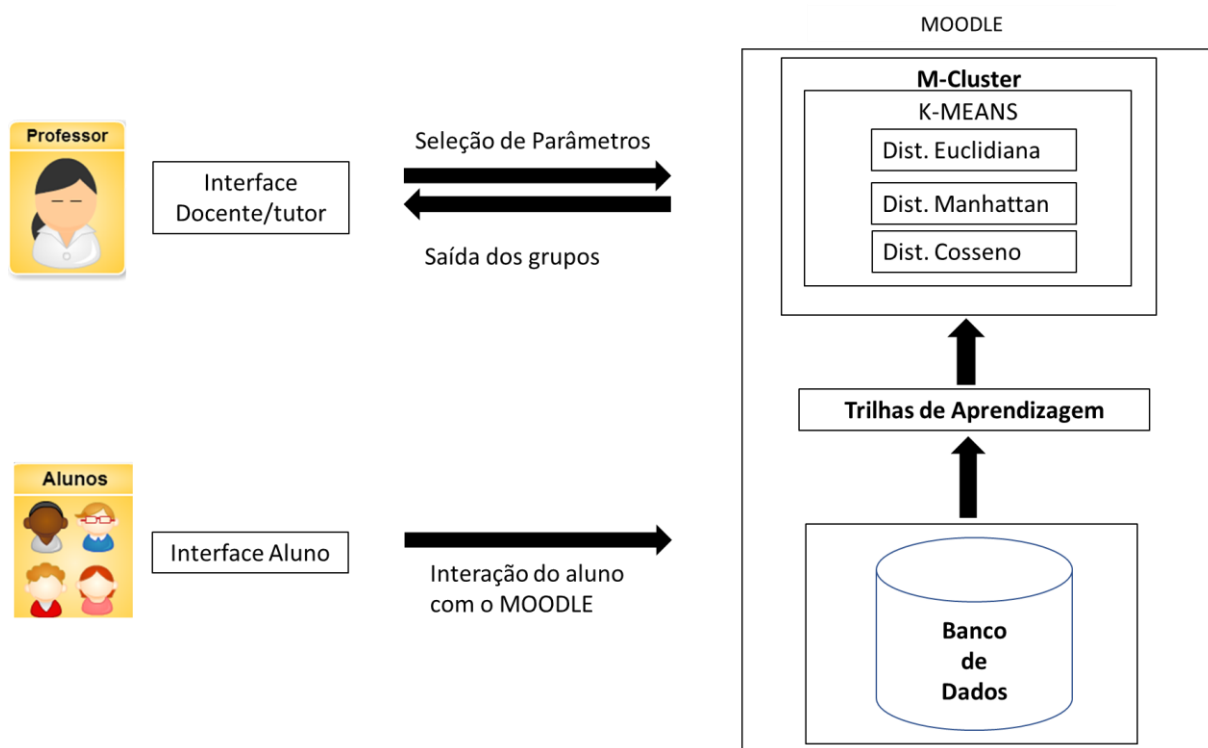


Figura 13: Visão Geral do *Framework*

O *framework* proposto por este trabalho é flexível, afinal conseguimos acrescentar no módulo principal, que corresponde ao M-CLUSTER, mais algoritmos de clusterização e consequentemente, mais métricas de similaridade, podendo potencializar ainda mais o *framework* para outras formações de grupos.

Não só homogêneas como foi o caso realizado, como também para heterogêneas e híbridas (homogêneas e heterogêneas).

Podemos perceber que a plataforma também é portátil, pois sua implementação pode ser reproduzida para outros ambientes virtuais e reutilizável para finalidades relacionadas à recomendação de grupos.

4.2.3 Componentes e suas Finalidades

4.2.3.1 A interface Aluno

Nesta interface o aluno interage com o Moodle. E a partir disso, são extraídos suas preferências e interesses de forma automática não necessitando de formulários para a coleta de informações, assim como sua interação durante o curso na plataforma virtual.

4.2.3.2 Banco de Dados

Todas essas informações são armazenadas no banco de dados do Moodle, dessas as mais relevantes para esta dissertação são as relacionadas com as interações dos alunos durante o curso no AVA. A estrutura do banco de dados utilizada foi a definida por Ramos (2016).

Nesta estrutura foram criadas duas tabelas virtuais, uma para os registros dos logs e outra para os recursos e atividades. Nas tabelas de logs do Moodle são armazenados os registros sobre os acessos a cada módulo no AVA.

As principais informações são: o *timestamp* (data/hora), o id do curso, o id do usuário, o nome do módulo, o id da instância do módulo acessado.

Todas essas informações estão na tabela *view* chamada *view_paths*, utilizada para identificar tudo o que o aprendiz acessou. Existem outras como a *course_modules*, que mantém a identificação única dos recursos/atividades chamada de *cmid*.

Assim, da tabela *course_modules* em conjunto com a principal tabela de cada módulo é possível coletar os dados de todas as instâncias necessárias para a criação do grafo.

Para integrar todas as informações relacionadas aos logs e recursos/atividade foi criada a tabela *view_instances*.

Com essa estrutura se facilitou a coleta dos dados no Moodle, pois este possui mais de 250 tabelas e com essa simplificação possibilitou a geração das TAs baseadas em grafo. Essas que serviram de base para o desenvolvimento desta pesquisa, os detalhes sobre as TAs são apresentados a seguir.

4.2.3.3 Trilhas de Aprendizagem

Os grafos que foram utilizados nesse trabalho são os idealizados no trabalho de Ramos (2016), estes representam as TAs dos alunos. Onde os vértices são representados por

círculos coloridos e as arestas por setas ligando os pontos. Segundo Ramos *et al.* (2016a), cada vértice representa um recurso/atividade que possui uma cor e o diâmetro de cada vértice é proporcional ao número de interações realizadas pelo aluno.

As arestas possuem 3 cores, a de cor verde é a padrão representa a estrutura de TA definida pelo professor no AVA, a de cor azul é a de avanço representa o percurso que o aluno realizou de um recurso/atividade para o outro mais à frente do que a estabelecida pelo professor, por fim, as arestas nas cores vermelhas representam os retornos realizados pelos alunos nos recursos/atividades.

Os valores nas arestas indicam o número de vezes que o aluno percorreu de um vértice A para um vértice B e a seta indica o sentido de transição.

Para geração dos grafos das TAs foi utilizado o *plugin* LPGraph para criar os grafos individuais de cada discente, das turmas analisadas. Todos os grafos individuais por aluno das turmas foram impressos, conforme pode ser visualizado na Figura 14.

Estas impressões foram realizadas, pois a visualização do grafo individual da ferramenta era limitada para gerar a visualização de todos os módulos da disciplina de Matemática Discreta de uma só vez na tela.

Os grafos impressos apresentavam 8 módulos da disciplina, juntamente com os caminhos percorridos por cada aluno durante a sua interação com o AVA.

A partir das impressões dos grafos foi realizada uma análise para investigar os possíveis atributos das TAs que seriam relevantes para a concepção desta dissertação.

Na análise, buscou-se perceber as interações similares existentes entre os discentes, ao realizarem seus estudos no curso e consequentemente encontrar possíveis atributos para formação dos grupos de alunos. Primeiro, a análise foi realizada de forma manual e posteriormente de forma automática.

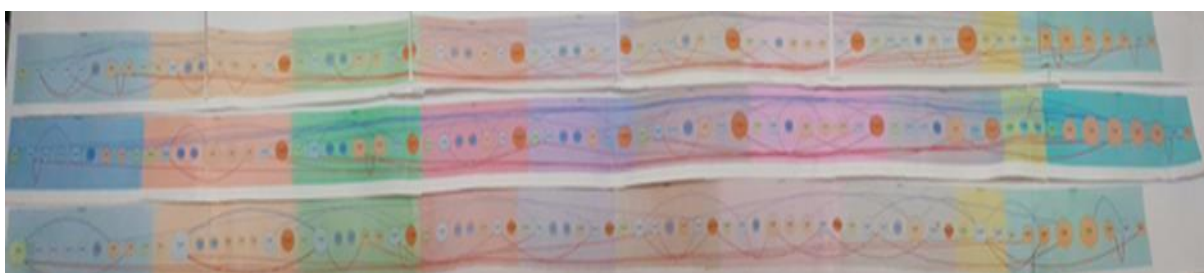


Figura 14: Exemplo da impressão das Trilhas de Aprendizagem dos Discentes

A forma manual se iniciou com a análise de todos os 44 grafos da turma de 2015/1 e todos os 40 da turma de 2016/1. Em nível de comparação entre os grafos se verificou durante o processo que havia similaridade ou a proximidade da quantidade de acesso aos recursos/atividade entre os alunos.

Outra percepção foi que alguns alunos tinham suas TAs similares em relação às arestas de avanço e retorno, sendo detectadas em algumas partes durante o processo de realização da disciplina. Apontadas as percepções iniciais de investigação da possibilidade desses dados serem relevantes para a recomendação de grupos, estas necessitavam ser validadas de forma concreta. Realizou-se uma entrevista com a professora especialista que ministrou as aulas para essas turmas, para buscar sua percepção de vivência durante a disciplina com os alunos.

Expostas as informações preliminares sobre a análise dos grafos das TAs, a professora especialista solicitou que investigasse mais a fundo os dados percebidos, obtendo como parâmetro principal 4 alunos da turma de 2015/1.

A justificativa para o foco nesses alunos é que pela sua convivência durante as aulas, percebeu que eles tinham perfis de aprendizagem parecidos, visto que, já se conheciam, realizavam trabalhos em conjunto e haviam estudado anteriormente em outra unidade de ensino juntos.

Partindo dessa premissa foram feitas análises mais detalhadas em relação aos dados preliminares identificados, conforme pode ser visto na Figura 15.

1	Quantidade de acesso Por Recurso/Atividade					Dispersão de valores das arestas de avanço (destino-origem)										d-dividiu p-passou direto n-não tem									
2																									
3	Alunos	ARQUIVO	URL	FORUM	TAREFA	QUESTIONARIO																			
4	ALUNO 1	24	5	11	203	60	0-14	1	2	3	4	5	6	7	8	9	10-13	11	12	13	14	15			
5							14	0	0	0	0	0	0	0	0	0	3	0	0	0	0	1			
6	ALUNO 2	1991	209	94	4.242	1987	0-9	1	2	3-10	4-6	5-32	6	7	8	9	10-13	11	12	13	14	15			
7							9	0	0	7	2	27	0	0	0	0	1	0	0	0	0	1			
8	ALUNO 3	26	2	12	683	50	0-13	1	2-10	3	4	5	6-45	7	8	9	10-13	11	12	13	14	15			
9							13	0	8	0	0	0	39	0	0	0	3	0	0	0	0	1			
10	ALUNO 4	18	2	10	131	36	0-10	1	2	3	4	5	6	7-72	8	9	10-13	11	12	13-19	14	15			
11							10	0	0	0	0	0	0	65	0	0	3	0	0	6	1	1			
12	ALUNO 5	36	1	10	145	57	0-6	1	2	3	4	5	6-45	7	8	9	10-19	11	12	13	14	15			
13							6	0	0	0	0	0	39	0	0	0	9	0	0	0	0	1			
14	ALUNO 6	83	0	14	79	15	0	1-4	2	3	4	5	6	7	8	9-19	10-13	11	12	13	14	15			
15							3									10	3								
16	ALUNO 7	48	6	7	139	83	0-23	1-27	2-10	3-13	4	5	6	7	8-37	9	10	11-19	12	13	14	15			
17							23	26	8	10					29			8							
18	ALUNO 8	82	11	15	145	67	0	1	2-10	3-5	4	5-15	6	7	8-46	9-14	10-19	11	12	13	14	15			
19																									
20	ALUNO 9	88	17	25	292	94	0-27	1-13	2	3	4-66	5	6	7-10	8	9-27	10-19	11-66	12	13	14	15			
21							27	12			62			3		18	9	55							
22	ALUNO 10	141	8	6	123	56	0-3	1-13	2-41	3-13	4	5	6	7	8	9	10	11	12	13	14	15			
23							3	12	39	10															
24	ALUNO 11	49	8	15	111	33	0-10	1	2	3-10	4	5	6	7	8	9	10-19	11-16	12	13	14	15			
25							10			7							9	5							
26	ALUNO 12	57	0	5	122	67	0-6	1	2	3	4	5	6-10	7-61	8	9	10-15	11	12	13-22	14	15			
27							6						4	56			5			9					
28	ALUNO 13	87	3	19	110	45	0-66	1	2-38	3-17	4-27	5-27	6-59	7	8	9	10-66	11	12	13	14	15			
29							66		36	14	23	22	53				56								
30	ALUNO 14	63	25	33	168	51	0-4	1-4	2	3-13	4	5-11	6	7	8	9	10-13	11	12	13	14	15			
31							4	3		10			6				3			8					
32	ALUNO 15	66	0	6	66	42	0-10	1-66	2	3-10	4-17	5	6-45	7	8	9	10-13	11	12	13	14	15			
33							10	65		7	13		39			3	3								
34	ALUNO 16	41	10	5	92	25		1-10	2-10	3	4-6	5-19	6-16	7-72	8	9	10-19	11-24	12	13	14	15			
35								9	8		2	14	10	65			9	13							
36	ALUNO 17	53	4	6	117	57	0-27	1-4	2	3-11	4-69	5	6	7-72	8	9-17	10-13	11-23	12	13	14	15			
37							27	3		8	65			65		8	3	12							
38	ALUNO 18	32	10	12	50	32	0-61	1	2-19	3	4	5-12	6	7	8	9	10-13	11	12	13-19	14	15			
39							61		17			7					3			6					
40	ALUNO 19	40	8	5	108	56	0-13	1	2	3	4	5	6-34	7	8	9	10	11	12-34	13	14	15			
41							13						28	d					22						
42	ALUNO 20	45	3	16	183	75	0-13	1	2-23	3	4	5	6-45	7	8	9	10	11	12	13	14	15			
43							13		21				29												
44	ALUNO 21	71	0	6	111	31	0-13	1	2	3	4	5	6	7	8	9	10-24	11	12	13	14	15			
45							13	d					d	d			14								
46	ALUNO 22	33	0	3	92	67	0-13	1-4	2	3	4-6	5	6-25	7	8	9	10-13	11	12	13	14	15			
47							13	3					19				3								
48	ALUNO 23	26	0	7	107	59	0-7	1	2	3	4	5	6-60	7	8	9	10-13	11	12	13	14	15			
49							7						54				3								
50	ALUNO 24	31	2	7	66	53	0	1	2	3	4	5	6	7-72	8	9	10-13	11	12	13	14	15			
51	ALUNO 25	26	0	4	155	56	0-3	1	2	3	4	5	6-46	7-66	8	9	10-13	11	12	13	14	15			
52							3						40	50			3								
53	ALUNO 26	61	11	5	140	60	0-66	d	1-34	2	3	4-678	5-10	6	7	8	9	10-19	11	12	13	14			
54							66	33				63	5	d	d			9		d					
55	ALUNO 27	128	10	32	248	136	0-32	1	2	3-10	4-32	5	6	7	8	9	10-13	11	12	13	14	15			
56							32			7	28						3								
57	ALUNO 28	82	3	47	158	112	0-13	1	2	3-10	4	5	6	7-10	8	9	10-19	11	12	13	14	15			
58							13			7						9									
59	ALUNO 29	40	14	8	126	41	0-10	1-10	2	3-22	4	5-66	6	7	8	9	10	11	12-20	13	14	15			
60							10	9			19		65						8						
61	ALUNO 30	31	7	8	140	65	0-10	1-25	2	3	4-6	5	6-46	7	8	9	10-14	11	12-20	13-22	14	15			
62							10	24			2		40				4		8	9					
63																									

Figura 15: Dados coletados dos TAs de forma manual

A Figura 15 mostra os dados coletados da turma de 2015/1. Nesta a identificação dos nomes dos alunos foi preservada, os itens coletados foram a quantidade de acesso dos recursos/atividades de cada aluno e também, a dispersão dos valores das arestas de avanço, bem como das arestas de retorno.

A dispersão foi uma opção encontrada para mapear as arestas de origem e destino das interações dos alunos com os recursos/atividades. Para calcular a dispersão primeiramente foi necessário mapear a origem da aresta, que busca sua trajetória de destino a outro vértice, sendo, que para essa representação foram utilizados os identificadores únicos de cada recurso/atividade, armazenados no formato numérico na base de dados do Moodle.

Esse mapeamento pode ser observado na coluna dispersão de valores das arestas de avanço na Figura 15, identificado pela cor preta com destaque em negrito. Após esquematizar a trajetória das arestas de cada aluno foi necessário realizar o cálculo da dispersão das arestas, que consiste em extrair o identificador numérico correspondente ao recurso/atividade de destino menos o de origem (destino - origem).

Este cálculo pode ser visualizado na Figura 15 acima pela cor em vermelho na coluna dispersão de valores das arestas de avanço, o mesmo procedimento foi realizado com as arestas de retorno. Por se tratar de dados extensos, na Figura 16 somente aparece parte dos dados.

Finalizada a coleta dos dados foi realizado uma análise da composição geral das informações por aluno, realizando uma comparação entre eles. Nessa análise, observamos todos os dados dos alunos, principalmente os discentes enfatizados pela professora especialista.

Esses alunos estão destacados em amarelo na Figura 15, percebeu-se que a quantidade de acessos deles é aproximada ou igual dependendo da atividade ou recurso acessado e em relação a dispersão das arestas de avanço quanto de retorno, em certos momentos da disciplina foram iguais ou similares. Ao analisar toda a conjuntura das informações concluiu-se que esses dados iniciais eram bastantes promissores.

No entanto, também foram constatadas algumas limitações na análise realizada, uma delas em relação ao cálculo das dispersões das arestas, que por se tratar de uma análise manual não foi possível detectar todos os valores das dispersões por causa da quantidade excessiva de arestas de avanço e retorno, em que, as suas trajetórias se cruzavam e muitas das vezes, convergiam sendo impossível definir a identificação das arestas.

Outra limitação percebida foi que as TAs dos alunos não eram do mesmo tamanho, ou seja, havia alunos com acesso em todos os vértices e arestas e outros que acessavam somente alguns e outros quase nenhum, criando assim uma disparidade entre os dados, para solucionar essa limitação, foram mapeados todos os 72 recursos/atividades da disciplina anotando a interação, origem e destino, de cada aresta conforme pode ser visualizado também no Figura 15.

Para obter todos os dados do cálculo da dispersão das arestas, resolveu-se coletar os dados de forma automática, por meio da replicação do ambiente Moodle de produção para o ambiente de teste usado pelas turmas investigadas, permitindo realizar uma implementação no próprio LPGraph para coletar a identificação dos alunos e todos os seus dados, permitindo obter suas dispersões das arestas de avanço e retorno.

Com isso, foi possível realizar uma análise de todos os dados e, percebeu-se que esses dados poderiam ser de fato promissores para a finalidade desta proposta. Afinal, durante a análise percebeu-se que havia diversas TAs com trechos parecidos ou similares.

Diante disso, foi realizada mais uma reunião com a professora especialista no mês de agosto de 2016, para apresentar os resultados encontrados com a análise completa, após a verificação dos dados, concluímos que seria necessário verificar como seriam agrupados os alunos com esses dados encontrados e então realizar uma nova análise.

Para realizar os agrupamentos dos alunos, foi utilizada a ferramenta de desenvolvimento Eclipse com a integração da biblioteca do WEKA, com uma cópia da base de dados real do Moodle.

Para os grupos serem gerados no Eclipse, necessitava-se de um algoritmo de clusterização e o algoritmo escolhido foi o *K-Means*, por ser bastante utilizado na literatura na temática de recomendação, ou seja, em 58,33% do total das pesquisas realizadas conforme levantamento da RSL e também por ser elencado nos trabalhos relacionados a esta dissertação.

Outro ponto a ser solucionado é que o *K-Means*, no WEKA, necessitava de um arquivo no formato *arff* para gerar os agrupamentos. Com isso, implementou-se no ambiente de teste do Moodle, mais especificamente no *plugin* LPGraph, uma funcionalidade que gerava os dados de acordo com o formato *arff*. Os dados que foram encontrados na análise inicial foram transformados em atributos, acrescentados o id do aluno, nome do aluno. Após esses dados serem gerados no Moodle, eles foram copiados para um arquivo texto (.txt), renomeado para a extensão *.arff*. Assim, podendo ser utilizado no ambiente Eclipse com WEKA.

Outro requisito a ser verificado era qual métrica de similaridade iria ser usada em conjunto com o *K-Means* para gerar os grupos a serem analisados. A métrica escolhida inicialmente foi a da distância Euclidiana, por ser uma das métricas utilizadas pelos trabalhos relacionados. Com isso foi possível visualizar a primeira versão dos grupos executados no Eclipse conforme pode ser visualizado na Figura 16.

```

8
9
10 public class WekaKMEANSISEUCL {
11
12     public static void main(String[] args) throws Exception{
13         // TODO Auto-generated method stub
14         //importa a base de dados ARFF utilizando classes da Weka
15         FileReader reader= new FileReader ("C:/ECLIPSE-WEKA/turma2015-2/TA2015-2-T.arff");
16         Instances arquivoArff=new Instances(reader);
17
18         //realizando o agrupamento
19         SimpleKMeans agrupamento= new SimpleKMeans();
20         agrupamento.setPreserveInstancesOrder(true);
21         agrupamento.setNumClusters(4);
22         agrupamento.setDisplayStdDevs(true);
23         agrupamento.buildClusterer(arquivoArff);
24
25         int [] assignments=agrupamento.getAssignments();
26         int i=0;
27         for (int clusterNum : assignments){
28             System.out.printf("Instancia: %d para Cluster: %d\n", i, clusterNum);
29             i++;
30         }
31         //Mostrando as estatísticas do agrupamento
32         Instances clusterCenter= agrupamento.getClusterCentroids();
33         Instances clusterStdDev=agrupamento.getClusterStandardDevs();
34         int [] clustersSizes=agrupamento.getClusterSizes();
35         for (int cluster=0;cluster<clusterCenter.numInstances();cluster++){
36             System.out.println("Cluster #"+(cluster)+": " +clustersSizes[cluster]+" alunos");
37             System.out.println("Centroide: "+clusterCenter.instance(cluster));
38             System.out.println("STDEV: "+clusterStdDev.instance(cluster));
39             System.out.println();
40         }
41     }

```

```

<terminated> WekaKMEANSISEUCL [Java Application] C:\Program Files\Java\jre1.8.0_144\bin\javaw.exe (23 de set d
Instancia: 28 para Cluster: 0
Instancia: 29 para Cluster: 0
Instancia: 30 para Cluster: 1
Instancia: 31 para Cluster: 3
Instancia: 32 para Cluster: 1
Instancia: 33 para Cluster: 2
Instancia: 34 para Cluster: 2
Instancia: 35 para Cluster: 2
Instancia: 36 para Cluster: 3
Instancia: 37 para Cluster: 2
Instancia: 38 para Cluster: 0
Instancia: 39 para Cluster: 3
Instancia: 40 para Cluster: 2
Instancia: 41 para Cluster: 2
Instancia: 42 para Cluster: 1
Instancia: 43 para Cluster: 3
cluster #0: 4 alunos
Centroide: 81.14.75,1.25,0.75,0.75,3.25,3.25,5.25,4,0.75,1,11.25,0.5,0,11.75,7,6.75,3.
STDEV: 13.341664,10.626225,0.957427,0.957427,0.5,1.5,3.304038,7.274384,2.828427,0.957

cluster #1: 12 alunos
Centroide: 85.916667,5.25,1.25,0.583333,0.833333,1.083333,0.25,2.833333,2.083333,0.083
STDEV: 18.451205,2.767506,1.484771,0.792961,1.114641,1.621354,0.452267,2.167249,1.928

cluster #2: 17 alunos
Centroide: 81.176471,10.647059,2.294118,0.823529,0.802353,1.294118,1.352941,2.588235,1
STDEV: 13.597956,7.532772,2.312053,1.131111,1.111438,2.143732,1.966633,1.660528,1.536

cluster #3: 11 alunos
Centroide: 105.181818,1.363636,0.545455,0.363636,0.272727,0.181818,0.363636,0.545455,0
STDEV: 43.397738,1.858641,0.8202,0.6742,0.467099,0.48452,0.80904,1.035725,0.301511,0

```

Figura 16: Agrupamento de Alunos usando Eclipse e WEKA

A Figura 16 mostra parte da implementação usada para gerar os clusters e o resultado da execução do código. Com o resultado obtido foi possível visualizar os grupos formados pelos atributos escolhidos inicialmente.

Tabela 1: Resultados dos atributos *K-Means* com Distância Euclidiana

Distância Euclidiana			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
Alunos 29	Alunos 30		Alunos 25, Alunos 26
4 Alunos	12 Alunos	17 Alunos	12 Alunos

Com a análise preliminar dos resultados da Tabela 1 foi possível concluir que os alunos 25 e 26 ficaram no mesmo cluster no caso o 3 e os alunos 29 no cluster 0 e 30 no cluster 1 ficando em clusters diferentes esses alunos correspondem aos alunos especificados pela professora especialista.

Diante da análise do resultado preliminar foi realizada uma nova reunião com a professora especialista, no mês de outubro de 2016 para mostrar os dados obtidos. Durante a

reunião foram discutidos os seguintes pontos: como seria a formação com outras métricas de similaridade e quais métricas utilizar? Será que o resultado chegaria a agrupar os discentes de determinada disciplina no AVA no mesmo cluster? Visto que, com os atributos iniciais houve um resultado promissor.

Diante desse cenário, novos testes seriam necessários e, com isso, outras métricas para fazer comparações. Verificou-se que em outras abordagens sobre agrupamento de alunos encontrados na RSL, usou-se a distância de *Manhattan* e Cosseno com outros algoritmos de clusterização.

Partindo desse fato, as duas distâncias foram implementadas com o algoritmo *K-Means* utilizando os mesmos parâmetros da distância Euclidiana. Na métrica de *Manhattan* é possível perceber que os alunos enfatizados pela professora ficaram nos mesmos clusters em comparação com o resultado da distância Euclidiana, tendo o diferencial na distribuição de número de alunos por cluster correspondendo a cluster 0: 4 alunos, cluster 1: 11 alunos cluster 2: 17 alunos e cluster 3: 11 alunos. Já na Figura 20, que corresponde ao uso da métrica do Cosseno, os alunos também estão nos mesmos clusters, sendo que a divisão de número de clusters também difere das outras duas métricas, cluster 0: 7 alunos, cluster 1: 15 alunos cluster 2: 12 alunos e cluster 3: 10 alunos.

Perante os resultados apresentados foi possível perceber que era preciso refinar os atributos novamente, para tentar agrupar os alunos usados como parâmetro no mesmo cluster. Assim, começou uma nova investigação sobre quais atributos acrescentar ou retirar para serem testados novamente. Para essa nova fase, inseriu-se no arquivo *.arff* já existente, os atributos: quantidade de arestas de avanço, retorno e padrão, quantidade total de vértices (soma de todos os vértices), dispersão das arestas padrão, dispersão geral (cálculo da dispersão considerando todas as arestas) e novamente foi realizado o teste. Os resultados dos agrupamentos com a distância Euclidiana, *Manhattan* e Cosseno encontram-se no (Apêndice C).

Percebe-se pelos resultados apresentados na Tabela 2 (Apêndice C), que os atributos ainda não satisfazem o parâmetro determinado pela professora especialista. Mas um fato que chama a atenção é que os alunos 25 e 26 estão no mesmo cluster faltando somente os alunos 29 e 30. A partir desse fato, reuniu-se com a professora especialista no final do mesmo mês (agosto de 2016), para mostrar os resultados encontrados até o momento. Durante a reunião, mostrou-se os atributos usados até o momento e foi sugerido acrescentar um atributo de média de quantidade de vértices acessados por aluno, afinal ao depender da quantidade de módulos da disciplina e de recursos os quantitativos de atributos do arquivo *arff* por vértice

ficariam gigantescos. Isto é necessário, porque a média é um valor que alivia um pouco o peso do fator quantidade de vértices.

A partir da inclusão do atributo média de quantidade de vértices foi possível ter outras medidas, como as de dispersões estatísticas: exemplo a variância e o desvio padrão. Isso foi necessário, por que percebeu-se que a média pode não ser suficiente para avaliar o conjunto dos dados das TAs.

Então, os atributos variância e desvio padrão foram incluídos no arquivo existente. Estes foram inseridos um de cada vez no arquivo e seus resultados analisados. Após isso, foi realizada uma nova reunião para apresentação dos resultados. A professora especialista considerou a ideia relevante. Então, novamente foram retornadas as análises dos atributos utilizando os dados da variância e do desvio padrão, juntamente com a execução de novos atributos.

O primeiro atributo a ser incluído no teste foi o valor da variância da aresta de avanço representada no arquivo como *@attribute avanço_variancia numeric* e o resultado obtido se encontra na Tabela 3 (Apêndice C).

Após a inclusão do atributo de variância das arestas de avanço, inseriu-se mais um atributo o *@attribute avanço_desvioPadrao numeric* que corresponde ao valor do desvio padrão das arestas de avanço, os resultados com as métricas podem ser visualizados na Tabela 4 (Apêndice C).

Então faltou verificar os resultados somente com o valor do desvio padrão das arestas de retorno. Sendo assim, novamente o teste foi feito com o valor do desvio padrão sem o atributo variância das arestas de avanço. Como pode ser visto na Tabela 5 (Apêndice C).

Outro teste realizado foi da exclusão dos atributos: *@attribute dispersao_geral numeric* e o *@attribute avanço_desvioPadrao numeric*. Assim foram realizados os testes novamente e o resultado pode ser encontrado na Tabela 6 (Apêndice C).

Outro teste realizado foi a inclusão do atributo *@attribute total_vertice_value numeric* (este corresponde a soma da interação com todos os vértices da disciplina) abaixo a Figura 17 apresenta a quantidade atual de atributos utilizados no arquivo *.arff* para execução.

```

@relation 'TrilhasDeAprendizagem-
weka.filters.unsupervised.attribute.Remove-R2-86-
weka.filters.unsupervised.attribute.Remove-R82-83-
weka.filters.unsupervised.attribute.Remove-R2-74'

@attribute aluno_id numeric
@attribute total_vertice_value numeric
@attribute avanco numeric
@attribute padrao numeric
@attribute retorno numeric
@attribute dispersao_avanco numeric
@attribute dispersao_padrao numeric
@attribute dispersao_retorno numeric
@attribute avanco_variancia numeric

@data
86,179,60.344828,6.896552,32.758621,15.342857,1,-23.578947,307
.425306
72,214,46.91358,8.641975,44.444444,18.026316,1,-17.194444,284.
815097
80,191,43.617021,5.319149,51.06383,19.073171,1,-17.645833,402.
458061
90,125,44.262295,14.754098,40.983607,19.888889,1,-22.48,454.39
5062
79,288,47.761194,7.462687,44.776119,20.375,1,-17.866667,419.89
0625
94,338,42.657343,13.286713,44.055944,14.983607,1,-15.539683,26
7.557108
68,215,52.040816,7.142857,40.816327,19.117647,1,-24.65,442.143
022
125,264,43.307087,6.286213,50.393701,20.809091,1,-17.4375,408

```

Figura 17: Arquivo arff refinado.

Como pode ser visualizado na Figura 17, houve uma redução considerável da quantidade de atributos, pois antes cada vértice contido na disciplina era convertido em um atributo do arquivo, com essa redução, diminuiu significativamente o número de atributos a serem processados pelo algoritmo. A Tabela 7 (Apêndice C) mostra os resultados obtidos com esses atributos apresentados na Figura anterior.

Com o resultado obtido com a inserção do atributo total de vértices, percebeu-se que estava havendo um distanciamento do agrupamento dos alunos 25, 26, 29 e 30. Já que na análise manual, estes tinham trilhas semelhantes e hábitos de estudos semelhantes, estas informações foram confirmadas pela professora. A partir disso foi realizado uma análise com os atributos já testados para verificar novos atributos que pudessem ser utilizados, para tentar agrupar todos os alunos em um só cluster ou dividir os alunos em dois clusters.

Assim, após análise dos atributos já testados, verificou-se a possibilidade de inserir como atributos as variâncias das arestas de retorno e padrão, já que havia no arquivo a variância da aresta de avanço e substituir o atributo total de vértices pelo atributo média de acesso dos vértices (valor da soma de todos os acessos aos vértices dividido pela quantidade total de vértices) e acrescentar novamente o atributo dispersão geral. Então o arquivo ficou com os atributos apresentados na Figura 18.

```

@relation 'TrilhasDeAprendizagem-
weka.filters.unsupervised.attribute.Remove-R2,11,13,15,17'

@attribute aluno_id numeric
@attribute media_acesso_vertice numeric
@attribute avanco numeric
@attribute padrao numeric
@attribute retorno numeric
@attribute dispersao_avanco numeric
@attribute dispersao_padrao numeric
@attribute dispersao_retorno numeric
@attribute dispersao_geral numeric
@attribute avanco_variancia numeric
@attribute padrao_variancia numeric
@attribute retorno_variancia numeric

@data
8818,3.325581,54.736842,8.421053,36.842105,10.711538,1,-14.742
857,0.515789,216.820636,?,400.076735
8815,2.616279,55.555556,13.333333,31.111111,10.88,1,-16.714286
,0.977778,138.5056,?,467.204082
9452,2.290698,52.631579,12.280702,35.087719,11.9,1,-16.25,0.68
4211,202.09,?,371.4875
8712,2.790698,49,7,44,13.285714,1,-14.454545,0.22,213.795918,?,
427.747934
8719,2.534884,51.086957,10.869565,38.043478,10.510638,1,-14.08
5714,0.119565,120.249887,?,262.706939
8727,3.790698,45.238095,12.698413,42.063492,15.438596,1,-15.37
7358,0.642857,428.070791,?,404.121752
9562,0.709302,48.648649,8.108108,43.243243,15.611111,1,-13.937
5.1.648649,224.78321,2.150.683564

```

Figura 18: Arquivo arff com média de acesso dos vértices.

Com isso, novamente foi testado o arquivo apresentado na Figura 18, com o Eclipse integrado ao WEKA e o resultado é mostrado na Tabela 8 (Apêndice C). Como mostrado na Tabela 8, os grupos dos alunos enfatizados pela professora especialista estão no mesmo cluster na distância Euclidiana e de *Manhattan*, mostrando a possibilidade de similaridade nas suas TAs.

O que difere entre a distância Euclidiana e de *Manhattan* é que os clusters apesar de ter a mesma divisão de número de alunos, os membros de cada cluster são diferentes em cada resultado das métricas, por exemplo, no resultado com a distância Euclidiana o aluno 09 está no cluster 0, mas no resultado da distância de *Manhattan* ele está no cluster 2.

Apesar desse teste ter agrupado os alunos especificados pela professora, foram realizados mais testes para averiguar as outras possibilidades para serem mostradas na reunião. O próximo teste que foi realizado, adicionou os atributos do desvio padrão das arestas de avanço, retorno e padrão ao arquivo existente e o resultado pode ser visualizado na Tabela 9 (Apêndice C).

Diante dos resultados apresentados na Tabela 9 podemos perceber novamente que o quantitativo dos alunos nos clusters entre as distâncias de *Manhattan* e Euclidiana são iguais, mas o que difere entre estes é a divisão dos membros de cada cluster em cada resultado. Existem membros que são alocados em clusters diferentes em cada resultado e isso ocorreu em todos os resultados coletados até aqui.

Os nove resultados foram impressos para serem analisados pela professora especialista. Então, no dia 6 de outubro de 2016, foi realizada a reunião, quando foram mostrados todos os resultados.

Durante reunião foi verificado que o resultado obtido da Tabela 8, a princípio, foi satisfatório, pois agrupava os alunos indicados, sendo que necessitava verificar a distribuição dos membros nos clusters para averiguar se os membros restantes poderiam ser agrupados realmente.

Partindo dessa perspectiva, a professora especialista analisou e concordou com o agrupamento formado pelo *K-Means* com a distância Euclidiana. E fez a sugestão de pegar novamente esses atributos e executar na turma fechada de 2016/1 que acabara de finalizar, para verificar se os atributos encontrados iriam obter o mesmo resultado, considerando a particularidade de cada turma.

Então, toda a base de dados foi modificada no Moodle Teste para a turma de 2016/1 da disciplina de Matemática Discreta e foram realizados os ajustes na implementação do LPGraph para obter o arquivo no formato .arff da respectiva turma, para ser executado no Eclipse. A Figura 19 abaixo apresenta o arquivo contendo os atributos usados no Eclipse.

```
@attribute aluno_id numeric
@attribute media_acesso_vertice numeric
@attribute avanco numeric
@attribute padrao numeric
@attribute retorno numeric
@attribute dispersao_avanco numeric
@attribute dispersao_padrao numeric
@attribute dispersao_retorno numeric
@attribute dispersao_geral numeric
@attribute avanco_variancia numeric
@attribute padrao_variancia numeric
@attribute retorno_variancia numeric
@data
8818,3.325581,54.736842,8.421053,36.842105,10.711538,1,-14.742
857,0.515789,216.820636,?,400.076735
8815,2.616279,55.555556,13.333333,31.111111,10.88,1,-16.714286
,0.977778,138.5056,?,467.204082
9452,2.290698,52.631579,12.280702,35.087719,11.9,1,-16.25,0.68
4211,202.09,?,371.4875
8712,2.790698,49,7,44,13.285714,1,-14.454545,0.22,213.795918,?
,427.747934
8719,2.534884,51.086957,10.869565,38.043478,10.510638,1,-14.08
5714,0.119565,120.249887,?,262.706939
8727,3.790698,45.238095,12.698413,42.063492,15.438596,1,-15.37
7358,0.642857,428.070791,?,404.121752
9562,0.709302,48.648649,8.108108,43.243243,15.611111,1,-13.937
5,1.648649,224.79321,?,150.683594
8722,0.744186,36.842105,21.052632,42.105263,18.714286,1,-16.06
25,0.342105,450.632653,?,330.433594
9761,0.244186,66.666667,8.333333,25,14.125,1,-16.333333,5.4166
67,114.856275,0.100,0.000000
```

Figura 19: Arquivo arff da turma de Matemática Discreta 2016/1.

A professora especialista escolheu os alunos 34, 35, 42 e 43 da turma de 2016/1, como parâmetro para possíveis formações de grupos de acordo com os seus perfis de aprendizagem. A partir disso, o arquivo acima foi inserido no código e em seguida foi executado no Eclipse e obteve-se os resultados apresentados na Tabela 10 (Apêndice C). Analisando os dados, os alunos especificados estão no mesmo *cluster*, mas separados em duplas, em todos os três resultados das métricas. Diante disso, foi feito novamente uma

solicitação de reunião com a professora especialista, para verificar quais dos resultados poderiam ser utilizados por ela, caso a turma estivesse em andamento.

No dia 26 de outubro, nos reunimos novamente para analisar os resultados, desta vez com a turma de 2016/1. A professora especialista analisou todos os 3 resultados e concluiu que, o mais adequado seria o resultado gerado pela distância *Manhattan*, pois foi realizado uma análise em relação a uma atividade de grupo já realizada por ela e por ter conhecimento sobre a turma. Um dado interessante percebido foi que seis alunos reprovados no semestre de 2015/1, estavam novamente matriculados na turma de 2016/1. Como eles já haviam feito essa disciplina, era possível que formassem grupos durante a disciplina. Então, durante a realização de uma atividade colaborativa, a ferramenta agrupou esses mesmos alunos.

Diante dos fatos, os atributos utilizados nas turmas fechadas de 2015/1 e 2016/1 da disciplina de Matemática Discreta foram promissores. O próximo passo foi implementar o *plugin* para ser integrado com o Moodle, para testar os atributos em uma turma em andamento. A descrição da realização deste estudo de caso está no capítulo 6. A seguir, a descrição do módulo de formação de grupos.

4.2.3.4 Formação dos Grupos

Como componente central da estrutura do *framework* o *plugin* M-CLUSTER é o responsável por gerar os grupos, por meio da alocação dos alunos nos clusters. O M-CLUSTER faz a classificação dos atributos extraídos das TAs descritos no tópico Trilhas de Aprendizagem, a partir do algoritmo *K-Means* integrado com uma métrica de similaridade. O *K-Means* é integrado individualmente com uma métrica de similaridade que pode ser a distância Euclidiana, *Manhattan* e Cosseno. Cada execução dos dados com uma métrica resulta nos agrupamentos que são exibidos ao professor. O M-CLUSTER apresenta ao professor os três resultados obtidos com cada métrica. Mais detalhes sobre o M-CLUSTER são apresentados no capítulo 5.

4.2.3.5 A interface Docente/Tutor

Através desta interface baseada na web, o instrutor pode selecionar os parâmetros que desejar para a formação dos grupos, como seleção de grupos, tipo de exibição, seleção do número de clusters, seleção dos dados, download de arquivo em formato pdf e representação

visual. Para cada parâmetro, o professor pode selecionar a opção desejada, por exemplo, os grupos podem ser visualizados de forma padrão (visão descritiva) ou bolhas (visualização, por meio do gráfico de bolhas) e assim por diante, mais detalhes são apresentados, na seção 5.2 desta dissertação. Após a definição de todos os parâmetros e execução por parte da ferramenta destes é definida para a interface docente/tutor a saída dos grupos. Na próxima seção é apresentada arquitetura geral deste trabalho.

4.3 ARQUITETURA DE TRABALHO

Como a proposta desta pesquisa é auxiliar o professor durante o processo de ensino-aprendizagem, com a recomendação de grupos de aprendizes na realização das atividades colaborativas, é necessário buscar informações sobre os alunos dentro do AVA. Essas informações são obrigatórias para caracterizar os alunos por meio das suas trilhas de aprendizagem. Assim, decidimos pelo uso do Moodle, pois além de ser um dos AVAs mais utilizados em diversas instituições de ensino (SANTOS; BOTELHO; BICHET, 2014), é utilizado também na Universidade Federal do Amazonas (UFAM), e, portanto, foi o ambiente adequado para implementação da pesquisa. A seguir, a Figura 20 apresenta a arquitetura utilizada nesta pesquisa.

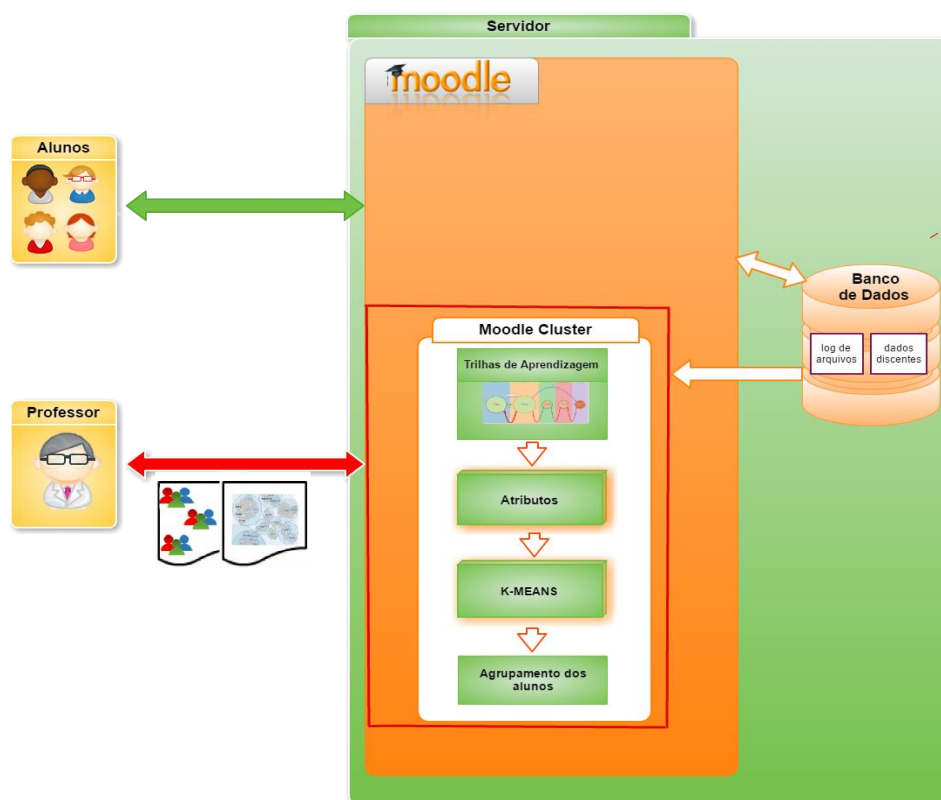


Figura 20: Arquitetura do Trabalho.

Resumidamente, apresentamos a arquitetura do trabalho:

- Coleta as informações das interações dos usuários com o AVA, por meio dos *logs* de acesso dos usuários na plataforma Moodle. Essa coleta de dados, permitiu caracterizar os aprendizes. A partir dessa caracterização foi composto o cenário das trilhas de aprendizagem. As trilhas são representadas por estruturas de grafo, onde as arestas correspondem à navegação do usuário entre os recursos e atividades e os vértices representam os próprios recursos/atividades.

- A partir dos grafos das TAs, são extraídos os seguintes dados: id do aluno, média de acesso por vértice, quantidade, variância e dispersão de arestas de avanço, retorno e padrão, para compor os atributos que foram necessários para serem processados pelo algoritmo *K-Means*.

- O *K-Means* classifica os dados dos aprendizes utilizando as distâncias Euclidiana, Cosseno e *Manhattan*, sendo que são obtidos três resultados de agrupamentos levando-se em consideração a métrica usada.

- Após o processamento são gerados os agrupamentos dos aprendizes podendo-se obter duas representações, uma descritiva e uma visual por meio do gráfico de bolhas, esses resultados são exibidos ao professor.

Ao final, criou-se duas representações uma descritiva e uma visual, para que o professor possa verificar as sugestões de recomendações dos grupos para realizar as atividades colaborativas de acordo com as suas trilhas de aprendizagem na disciplina. Os detalhes sobre o uso do banco de dados do Moodle, o desenvolvimento do *plugin* e as representações descritivas e visual dos agrupamentos dos alunos estão descritos no Capítulo 5, sobre o desenvolvimento da ferramenta M-CLUSTER.

4.4 FERRAMENTAS TECNOLÓGICAS

Para o desenvolvimento do *framework* e da ferramenta deste trabalho, foram necessárias, as seguintes ferramentas:

- **Apache:** é um servidor web criado pela Apache Software Foundation e teve como objetivo executar os scripts PHP e incorporar o Moodle;
- **Eclipse IDE:** Ambiente para desenvolvimento de software, porém suporta várias linguagens de programação. No caso, foi utilizada a linguagem JAVA, pois a

biblioteca do WEKA está nesta linguagem;

- **HTML 5:** uma linguagem de marcação de hipertexto utilizada para publicar conteúdo na Web;
- **Moodle:** um software para apoio de aprendizagem em cursos *online* (SANTOS; BOTELHO; BICHET, 2014). A versão do Moodle usada neste trabalho foi a 3.0.
- **PHP:** uma linguagem de programação para Web;
- **MySQL:** um sistema de gerenciamento de banco de dados (SGBD), que utiliza a linguagem *Structured Query Language* (SQL) como padrão.
- **Data-Driven Documents (D3):** biblioteca JavaScript para a criação de representação visual a partir de dados texto.
- **WEKA:** um conjunto de ferramentas e algoritmos de aprendizagem de máquina e mineração de dados.

5 O DESENVOLVIMENTO DA FERRAMENTA M-CLUSTER

O Moodle é um Ambiente Virtual de Aprendizagem que foi desenvolvido na Linguagem PHP, que suporta vários tipos de bases de dados e é uma aplicação baseada na Web, logo a ferramenta que realiza recomendações para formação de grupos em AVAs denominada de *Moodle Cluster* (M-CLUSTER), é capaz de se incorporar ao PHP e rodar num servidor Apache, e no caso desta pesquisa, a ferramenta foi integrada ao Moodle no formato de *plugin*.

Por esse motivo, parte do código-fonte segue as especificações de desenvolvimento do Moodle, principalmente no que diz respeito ao formato de apresentação da ferramenta.

Embora a ferramenta pudesse ser desenvolvida separadamente do Moodle, com interface própria e apenas acessando o banco de dados do AVA diretamente, a escolha pelo formato de *plugin* se justifica pelo fato deste trabalho unir aos demais do grupo de pesquisa que também utilizam o *plugin* como base de desenvolvimento de suas ferramentas e pela facilidade de uso proporcionada ao usuário que já está familiarizado com o Moodle.

5.1 O BLOCO M-CLUSTER

O Moodle é um sistema modular e isso permite a criação de novos *plugins* que podem ser instalados conforme a necessidade. Na página de tipos de *plugins* (MOODLE, 2017b) do site oficial do Moodle, existe uma lista com 46 tipos de *plugins* diferentes que podem ser criados, sendo um deles, o bloco.

Blocos são pequenas áreas retangulares que podem ser adicionadas na lateral esquerda ou direita de qualquer página do Moodle (MOODLE, 2017a). Alguns exemplos de blocos são: calendário, usuários online, eventos recentes, administração, navegação, entre outros.

O *plugin* M-CLUSTER, constitui-se de um bloco que ao ser clicado abre a ferramenta de recomendação de grupos. Na Figura 21, é apresentado o bloco do M-CLUSTER, a partir dele o professor/tutor tem acesso à interface principal do *plugin* e desta, pode realizar as configurações que julgar necessárias para obter as sugestões dos agrupamentos dos discentes.

O *plugin* só pode ser visualizado, a princípio, pelo professor/tutor, pois ele que vai definir se usará as sugestões geradas pela ferramenta ou não. E caso escolha a sugestão da ferramenta, ele poderá disponibilizar o resultado aos alunos.



Figura 21: Bloco M-CLUSTER

5.2 LÓGICA DE CAMADAS DO M-CLUSTER

A lógica da ferramenta foi dividida em quatro partes: a identificação das TAs, lista de atributos da TA, agrupamento dos alunos e representação descritiva e visual.

A identificação das TAs, foi realizada a partir dos dados do ambiente virtual, onde foram coletadas todas as informações e transformadas no formato de grafo conforme descrito na seção 2.7.1 desta dissertação.

A partir disso, obteve-se uma lista de atributos relevantes para cada aluno estes apresentados na seção 4.2.3, obtidos por meio das TAs. Com os atributos definidos, esses foram usados e classificados com algoritmo *K-Means*, resultando nas sugestões de agrupamentos;

Para apresentar as sugestões de agrupamento ao docente, este pode escolher uma forma de representação dos dados descritiva ou visual. Um resumo da distribuição em camadas pode ser observado na Figura 22.



Figura 22: Camadas do M-CLUSTER

O *plugin* M-CLUSTER constitui um bloco que, ao ser clicado, abre a ferramenta de cluster de aprendizagem. No canto inferior esquerdo da Figura 23, é apresentado o bloco de acesso ao M-CLUSTER e ao centro está a tela do M-CLUSTER que está dividida em: seleção

de grupos, tipo de exibição, seleção do número de clusters, seleção dos dados, download de arquivo em formato pdf e representação visual/descritiva.

O professor/tutor escolhe os parâmetros necessários para que o algoritmo possa recomendar os grupos, tais parâmetros são descritos a seguir.

Em seleção de grupos, deve-se selecionar um grupo (turma), o Moodle permite ao professor dividir os alunos de uma disciplina em turmas, selecionar um grupo irá retornar os alunos do grupo dentro do item Seleção de Dados.

Caso não haja nenhum grupo, todos os alunos da disciplina são listados. Abaixo, encontra-se o Tipo de Exibição, onde é escolhida a forma de exibição dos clusters, que pode ser apenas descritivo, ou em forma de gráfico de bolhas. Este último está exemplificado no item Representação Visual na Figura 23.

Em Seleção de Número de Clusters, o professor/tutor seleciona a quantidade de clusters que deseja como resultado do *plugin*. De posse dessa informação, o algoritmo realiza o agrupamento e gera a quantidade de agrupamentos escolhida.

Em Seleção de Dados, o professor/tutor escolhe quais alunos participarão da recomendação de grupos. Para isso, ele faz a seleção dos dados no campo usuário e, em seguida, clica no botão adicionar, ao lado direito do campo.

Com isso, os alunos escolhidos compõem o campo Alunos para Recomendação. Somente os alunos deste campo, serão analisados e classificados na ferramenta.

Caso seja necessário remover um aluno deste campo, existe o botão remover, que retira o item que foi selecionado. Isso é útil, por exemplo, para remover alunos evadidos das atividades.

Após as escolhas realizadas pelo professor/tutor dentro da ferramenta, o mesmo deve clicar no botão Sugerir Grupos na parte de Seleção de Dados.

O processo de análise e classificação do algoritmo *K-Means* juntamente com as métricas de similaridade é realizado, e, em seguida é apresentado na tela a representação dos grupos de acordo opção escolhida, visual ou descritiva.

Caso o professor/tutor queira realizar o download em formato pdf da representação escolhida, o mesmo deve clicar no botão “Download pdf”, para baixar o conteúdo contendo a descrição de todos os grupos recomendados pela ferramenta, assim, o arquivo pdf pode ser postado juntamente com a tarefa a ser realizada.

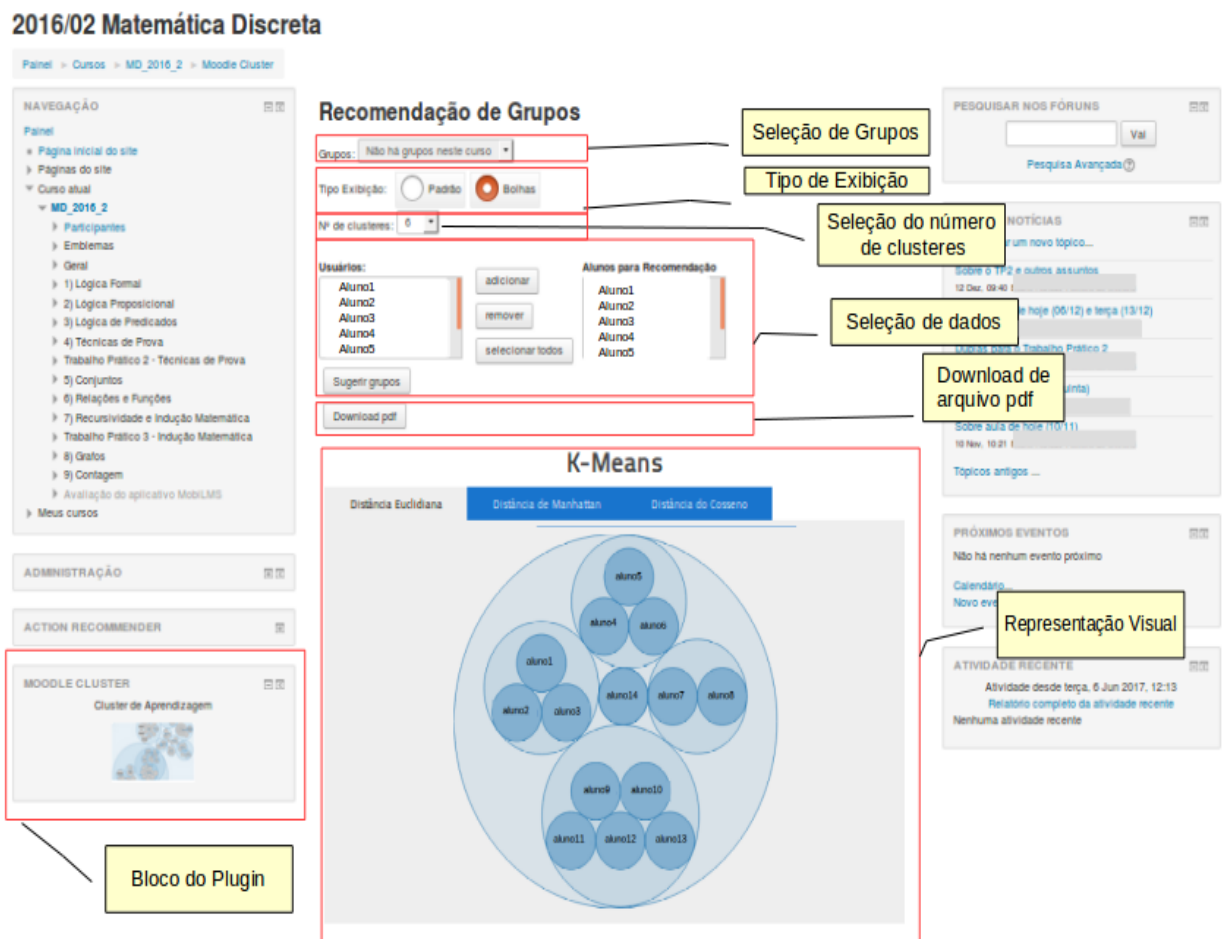


Figura 23: Parte da tela do M-CLUSTER.

5.3 IMPLEMENTAÇÃO DO NÚCLEO DO M-CLUSTER

Com as camadas definidas da ferramenta M-CLUSTER, o passo seguinte consistiu em selecionar a linguagem de programação e as ferramentas usadas em cada etapa.

A linguagem de programação escolhida para manipulação da base de dados e implementação do M-CLUSTER foi a PHP, por ser a linguagem utilizada na implementação do Moodle, e também para desenvolvimento de *plugins* para Moodle.

Utilizando funções próprias do Moodle, é possível ter acesso ao banco de dados e assim coletar os registros necessários para a identificação e extração das TAs. Os dados são coletados da tabela de logs, usuário e das tabelas dos recursos/atividades. A modelagem das TAs segue o proposto por Ramos *et al.* (2016a).

Na sequência, as TAs são processadas para a aquisição dos atributos, mencionados na arquitetura, que compõe a lista a ser avaliada pelo algoritmo de agrupamento.

Cada aluno, portanto, é representado agora como uma lista de atributos, que em PHP é armazenada no formato de vetor contendo: identificação (id) do aluno, nome, média de acesso por vértice, percentual de arestas de avanço, percentual de arestas padrão, percentual de arestas de retorno, dispersão das arestas de avanço, dispersão das arestas padrão, dispersão das arestas de retorno, dispersão geral das arestas (total), variância da dispersão das arestas de avanço, variância da dispersão das arestas padrão e variância da dispersão das arestas de retorno.

Para a realização do agrupamento foi utilizado o algoritmo *K-Means*. Após receber a lista de atributos de todos os alunos, para cada execução foram utilizadas as métricas de distância Euclidiana, *Manhattan* e Cosseno.

Vale ressaltar que os atributos de identificação e nome do aluno são ignorados pelo *K-Means*. Assim, tem-se três possibilidades de agrupamentos que são utilizados para a exibição dos resultados no Moodle. A seguir são descritas as representações descritiva e visual do M-CLUSTER.

5.4 IMPLEMENTAÇÃO DA REPRESENTAÇÃO DESCRITIVA

Ao final da execução da fase de descoberta dos grupos, tem-se três resultados de agrupamentos distintos. Cada resultado é uma lista (vetor) de grupos, onde cada grupo armazena a lista de atributos de cada aluno daquele grupo.

Na Figura 24, pode-se visualizar um exemplo da representação descritiva. Nesta representação, os grupos recebem uma numeração sequencial, iniciando em zero, em seguida, é extraído da lista de atributos, o nome de cada aluno e apresentado no Moodle.

Isso é feito para todos os grupos sugeridos pelo *K-Means*. Em cada aba, do item Representação visual, ficam os agrupamentos sugeridos de acordo com as métricas utilizadas.



Figura 24: Exemplo de Visualização da Representação Descritiva

5.5 IMPLEMENTAÇÃO DA REPRESENTAÇÃO VISUAL

Para mostrar a representação visual, foi utilizado o D3.js, ou simplesmente, D3 (Bostock, 2015). D3 é uma biblioteca *JavaScript* (Bostock, Ogievetsky e Heer, 2011), que neste trabalho, foi utilizada para gerar o gráfico de bolhas, mas que pode ser utilizada para diversas formas de representação visual.

O D3 é capaz de manipular dados no formato JSON, assim, para a representação dos grupos como um gráfico de bolhas. Primeiramente, converteu-se a lista de vetores, a mesma usada na representação descritiva, em um objeto no formato JSON. No objeto JSON, foi armazenado um objeto para cada grupo, contendo somente o nome do grupo, nome dos alunos do grupo e um valor constante que representa o diâmetro das bolhas que exibem o nome dos alunos.

Dessa forma, usando o objeto JSON e o algoritmo de renderização escrito usando a biblioteca D3, criou-se o gráfico de bolhas. A Figura 25 mostra essa representação. Para cada

grupo foi criada uma bolha de fundo mais claro, e dentro de cada bolha, foram criadas outras bolhas menores com os nomes dos alunos, com fundo mais escuro.

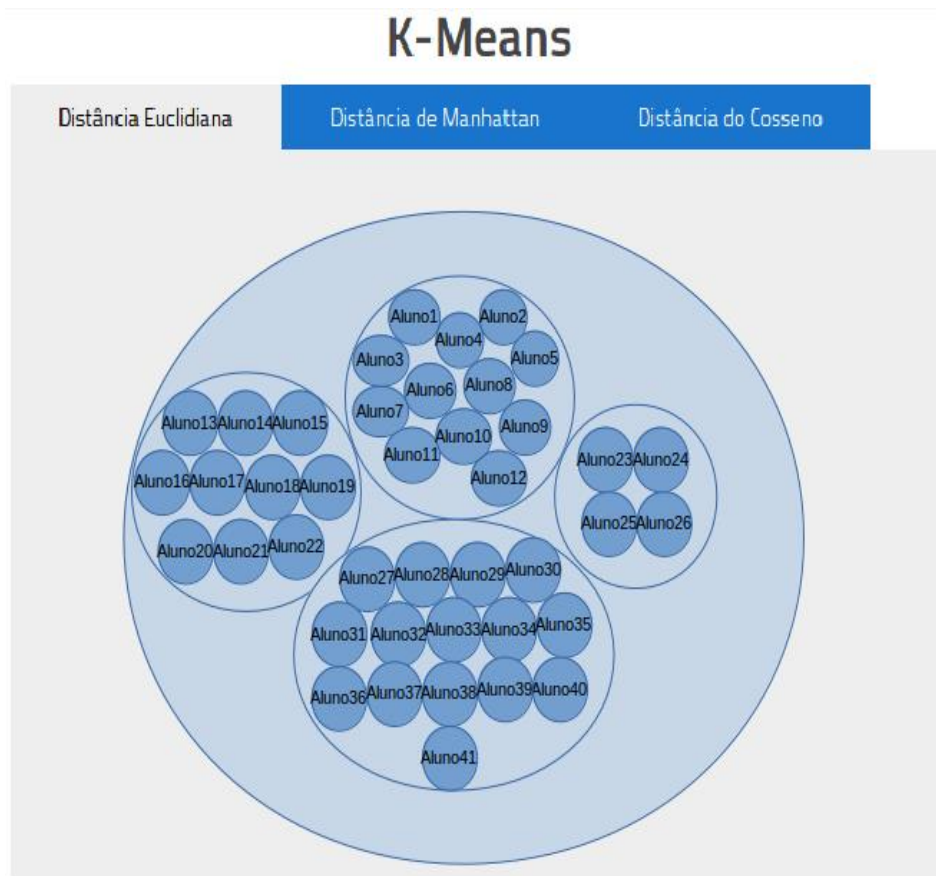


Figura 25: Exemplo de Visualização do Gráfico de Bolhas

6 VALIDAÇÃO INICIAL E RESULTADOS DA PROPOSTA

Neste capítulo, são descritos os testes realizados para validar os resultados obtidos com o M-CLUSTER.

Para validar a ferramenta, foi realizado um estudo de caso em uma turma real em andamento, da disciplina de Matemática Discreta, do ano de 2016/2, do curso de Ciência da Computação da UFAM, na modalidade presencial e que usava o AVA como suporte às atividades pedagógicas.

Participaram do estudo de caso 41 alunos matriculados, um docente e um tutor, onde todos os alunos assinaram o termo de consentimento livre e esclarecido (Apêndice B).

No estudo de caso, o M-CLUSTER utilizou os atributos relatados na seção 4.2.3 e os classificou com o *K-Means*, obtendo-se três resultados de agrupamentos homogêneos, um para cada métrica.

O docente analisou e escolheu, dentre as sugestões geradas pela ferramenta, qual opção seria disponibilizada aos alunos. Durante a realização do estudo de caso, o aluno 41 desistiu da disciplina, portanto, a distribuição dos alunos nos grupos foram feitas com 40 alunos.

Para validar os grupos sugeridos, foram realizadas duas atividades durante o semestre, com o mesmo grau de dificuldade, para que os seus resultados fossem analisados.

Na primeira atividade, o docente solicitou que os alunos realizassem uma atividade em dupla, sendo que estes, puderam escolher seus pares e na segunda, o professor utilizou a ferramenta para obter sugestões de clusters, em que os alunos pudessem formar as duplas.

Para gerar os agrupamentos foi necessário escolher a quantidade de clusters e o professor definiu que a ferramenta sugerisse os grupos de alunos com quatro clusters.

O tipo de exibição escolhido pelo professor durante a realização do estudo de caso foi a do gráfico de bolhas. A seguir, as Figuras 26, 27 e 28 apresentam as visualizações do gráfico de bolhas disponibilizados ao professor durante a realização do estudo de caso.

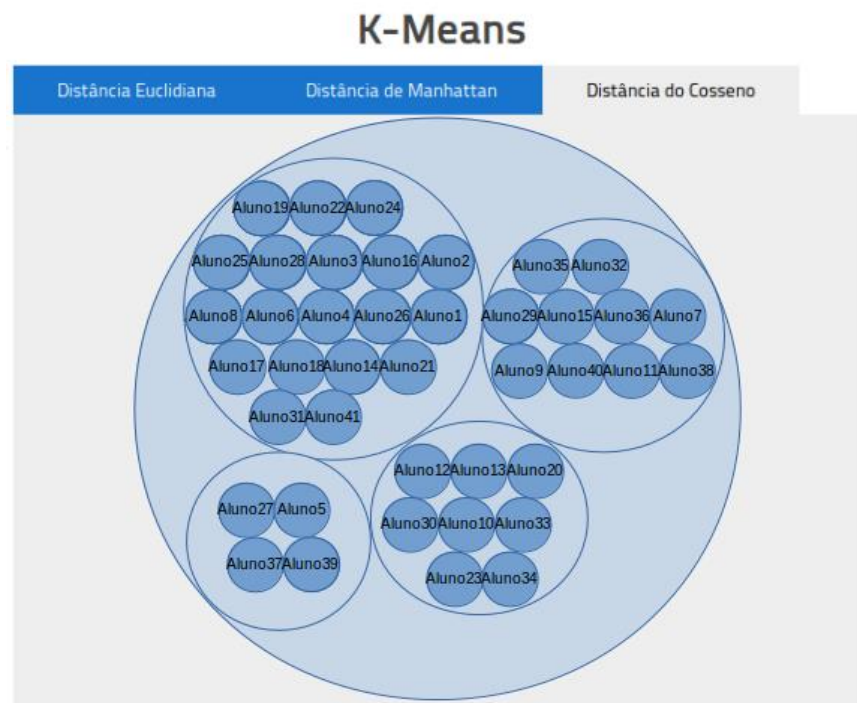


Figura 26: Resultado da Distância do Cosseno

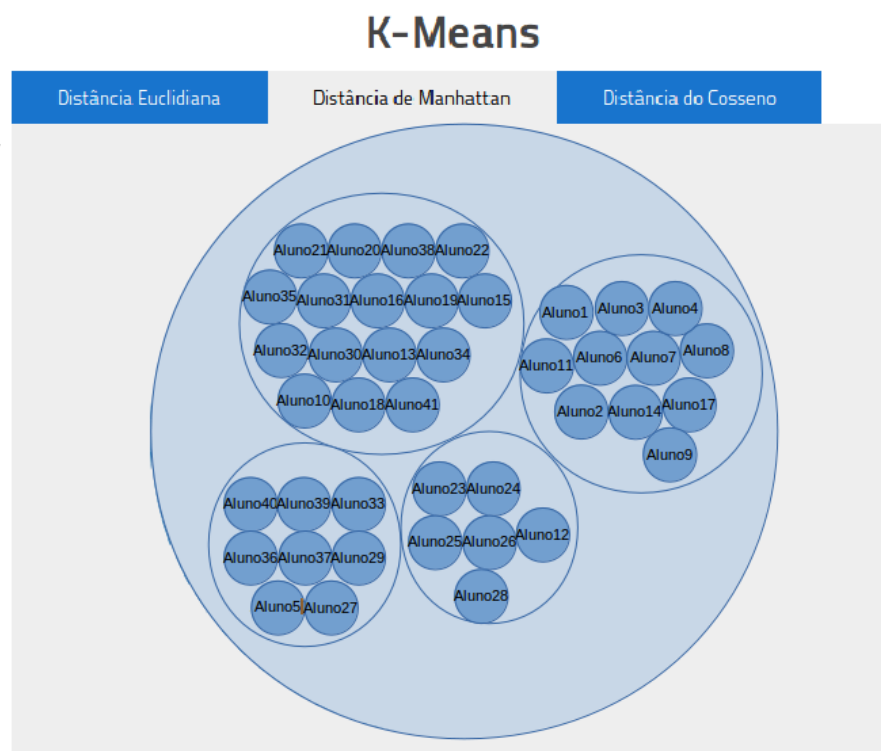


Figura 27: Resultado da Distância de *Manhattan*

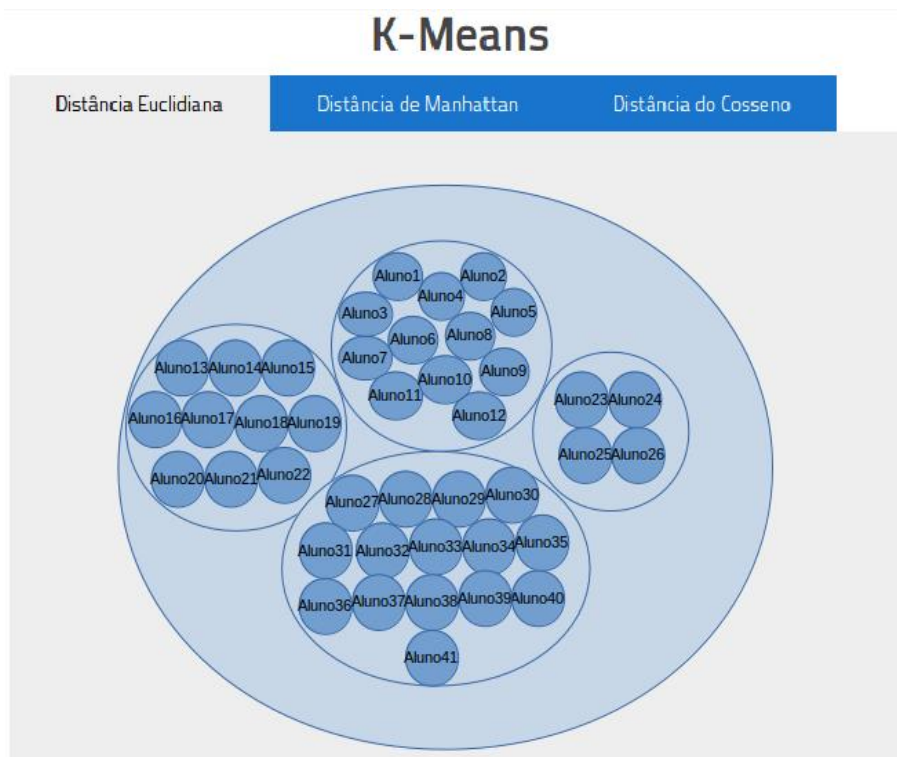


Figura 28: Resultados da Distância Euclidiana

A partir dos parâmetros inseridos pelo docente na ferramenta, a sugestão mais adequada, conforme a visão do docente, foi do algoritmo *K-Means* com a distância Euclidiana, este resultado é apresentado na Figura 28. Após a escolha da sugestão, o professor disponibilizou a mesma aos alunos no ColabWeb.

Ao disponibilizar os clusters aos alunos, os mesmos tiveram a orientação que só poderiam formar as duplas com os componentes que estivessem no mesmo cluster. Assim as duplas foram formadas e postadas no ColabWeb pelos alunos para o professor.

Os resultados mostraram que, de acordo com a professora, a sugestão gerada pela ferramenta obteve resultados satisfatórios, onde 75% dos alunos aproximaram-se ou superaram as suas notas individuais (essas adquiridas ao formar a dupla para executar as atividades), em relação às notas alcançadas na primeira atividade conforme pode ser visualizado na Figura 29.

Na Figura 29, pode-se visualizar as notas individuais dos alunos de Matemática Discreta. Os alunos desistentes foram excluídos do gráfico.

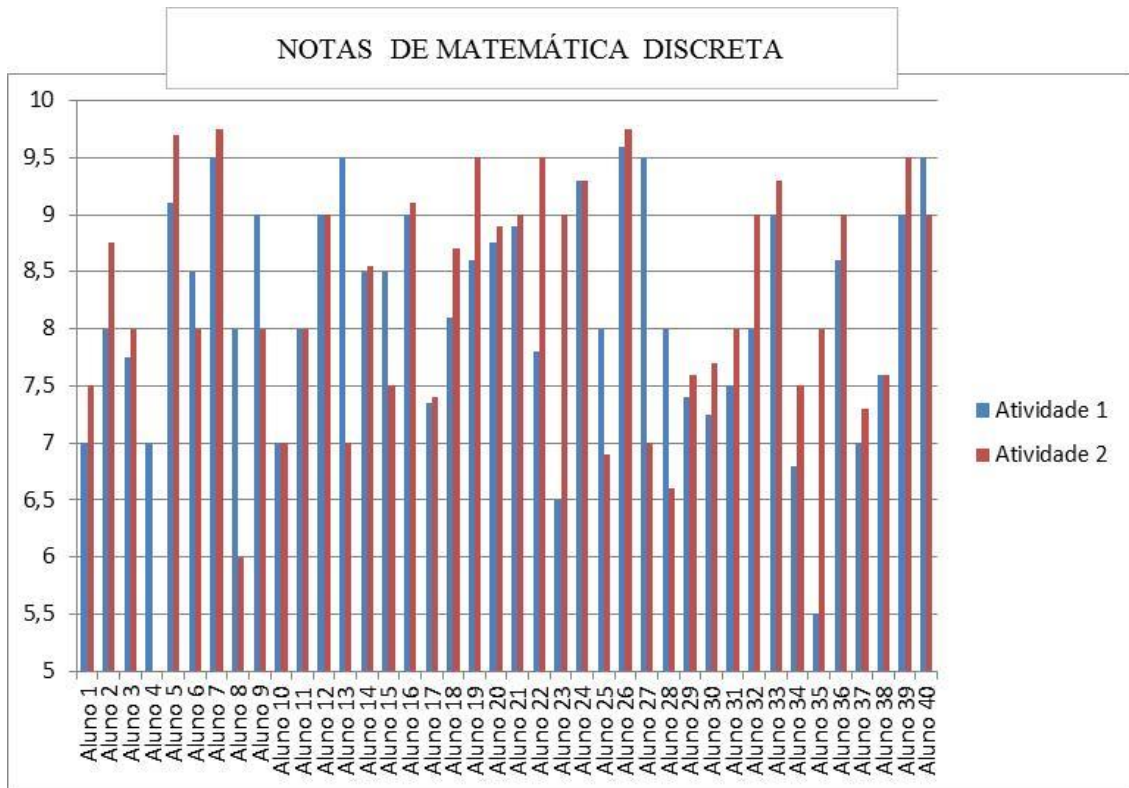


Figura 29: Notas dos alunos de Matemática Discreta.

Foi constatado também, que do total de grupos formados, 30% destes formaram duplas idênticas a da primeira formação (os mesmos obtiveram bons resultados nas duas atividades). Esses resultados ainda não são conclusivos, mas já são um ponto de partida para análises futuras mais detalhadas.

7 CONSIDERAÇÕES FINAIS

Esta pesquisa desenvolveu um *framework* conceitual de recomendação para formação de grupos visando promover a aprendizagem colaborativa no AVA, utilizando a caracterização dos aprendizes.

Essa caracterização foi baseada em dados das trilhas de aprendizagem, estas em formato de grafo, para auxiliar o professor no processo de ensino-aprendizagem. Para se chegar no desenvolvimento do *framework*, foi realizada uma Revisão Sistemática da Literatura sobre recomendação de grupos para atividades colaborativas nos ambientes virtuais.

Essa revisão serviu para identificar trabalhos sobre agrupamentos de alunos em ambientes virtuais. Como resultado, obtivemos um levantamento do estado da arte sobre agrupamentos de alunos em AVAs existentes na literatura, visando identificar as seguintes características: quais técnicas de mineração de dados foram utilizadas, quais informações foram usadas para compor o perfil dos aprendizes, qual abordagem foi usada para adquirir as informações sobre as preferências dos discentes e quais os tipos de AVAs utilizados.

De posse das informações obtidas da RSL, foi possível estruturar um *framework* conceitual, que possibilitasse a formação de grupos de aprendizes baseada nas suas trilhas de aprendizagem, onde foram extraídos os atributos mais relevantes de acordo com os professores especialistas para realizar o agrupamento dos aprendizes.

Para validar o *framework*, criou-se o M-CLUSTER no formato de *plugin*. A criação da ferramenta trouxe uma formação de grupos, onde se observou que os grupos formados a priori com os atributos utilizados, nas turmas analisadas, são do tipo homogêneo, possibilitando extrair das trilhas padrões de comportamento que possam auxiliar o docente no ensino a distância.

Tendo em vista a complexidade do problema abordado por esta dissertação, pode-se afirmar que o *framework* proposto, foi adequado ao processo de recomendação de grupos – a partir da possibilidade de constatar com a ferramenta – por utilizar uma sugestão de recomendação que aprimorou os resultados (notas) dos discentes nas atividades colaborativas realizadas durante o estudo de caso.

A arquitetura geral do *framework* foi proposta pensando em possíveis inclusões de mais algoritmos de clusterização e métricas de similaridades para possibilitar mais tipos de agrupamentos, como: homogêneos, heterogêneos e híbridos. Assim, o docente poderá escolher qual tipo de agrupamento deseja utilizar para determinada atividade colaborativa.

Os atributos utilizados para gerar os agrupamentos dos alunos, passaram por diversos refinamentos, até chegar em um conjunto de atributos promissores para recomendação. Para assim, serem testados em uma turma em andamento.

A validação do framework passou por duas etapas: a primeira foi uma análise de duas turmas reais fechadas e a outra por um estudo de caso exploratório com uma turma em andamento.

A primeira etapa consistiu em realizar uma análise das TAs de duas turmas fechadas da disciplina de Matemática Discreta de 2015/1 e 2016/1. Essa análise possibilitou detectar e encontrar os atributos que foram utilizados para sugerir os grupos de alunos nas turmas fechadas validados pela professora, para que estes fossem usados no *plugin* M-CLUSTER para serem aplicados na turma em andamento.

Já o estudo de caso exploratório foi realizado em uma turma em andamento que esteve disponível no ColabWeb por um semestre. Assim, a professora teve a oportunidade de obter as sugestões dos agrupamentos e utilizá-los na atividade que pretendia, assim como de realizar grupos de alunos sem necessitar ter um contato presencial com os mesmos. Nesse sentido, a utilização da abordagem se torna promissora para realizar as sugestões de agrupamentos de alunos em AVAs.

Após o uso da ferramenta para agrupar os alunos, percebeu-se que houve alunos da turma em andamento que melhoraram suas notas nas atividades colaborativas. Durante o processo, os alunos tiveram a oportunidade de escolher com quem queriam realizar a atividade, pois os clusters sugeridos pela ferramenta permitiam que os alunos escolhessem seus pares dentro do mesmo cluster.

Como limitações desta proposta, ocorre que em alguns casos, os resultados dos clusters geram um grupo com um único membro, impossibilitando-o de se agrupar com os outros alunos. Outro fato é que não há o controle sobre os números de componentes que ficarão em cada grupo, de forma que, pode ocorrer grupos com menos componentes do que o determinado pelo professor. Por último, os agrupamentos gerados pela ferramenta, por enquanto, são somente do tipo homogêneo. A seção a seguir detalhará as contribuições proporcionadas pela abordagem defendida.

7.1 CONTRIBUIÇÕES

O agrupamento dos alunos de acordo com as suas trilhas de aprendizagem baseadas no próprio Moodle, a partir desse ponto, as TAs deram subsídios para extrair os atributos

mais relevantes para serem usados no algoritmo de clusterização, permitindo gerar as sugestões dos agrupamentos.

A identificação de atributos baseados nas TAs em formato de grafo, até então não foram observados na literatura sobre recomendação de grupos de alunos em AVAs. A terceira contribuição dessa pesquisa é a criação de um *framework* conceitual para detecção automática de grupos de alunos que explora os atributos identificados, servindo de base para pesquisas futuras que queiram aprimorar a investigação sobre recomendação de grupos em AVAs.

Como quarta contribuição deste trabalho, apresenta-se a implementação da métrica de similaridade da distância do cosseno para ser usada no algoritmo de *Clustering K-Means* dentro do WEKA e no M-CLUSTER. Essa implementação não existia nos parâmetros de desenvolvimento do WEKA, permitindo enriquecer ainda mais essa pesquisa e essa implementação no *plugin*.

Finalmente, o desenvolvimento da ferramenta M-CLUSTER no formato de *plugin* para o *Moodle* contribui como um exemplo prático da aplicação, podendo o *plugin* ser integrado em qualquer AVA.

O M-CLUSTER não beneficia apenas a EaD, mas qualquer modalidade de ensino que utiliza o Moodle ou outro AVA como ambiente de ensino, aprendizagem, comunicação e interação entre professor, aluno e conteúdo

Com a validação do *framework* realizada via M-CLUSTER, foi possível perceber que o docente pode ter uma alternativa automatizada para formar grupos de alunos de acordo com as suas interações dentro do ambiente. O professor pode escolher usar ou não a ferramenta para atividades que requerem a formação de grupos.

A iniciativa dessa pesquisa visou a melhoria do processo de ensino-aprendizagem dentro de qualquer modalidade de curso que use um AVA para recomendar grupos de alunos para atividades colaborativas. Esse trabalho pode ser aprimorado e estendido, conforme apresentado na seção a seguir.

7.2 TRABALHO FUTUROS

A partir do que foi observado durante a pesquisa, pretende-se para trabalhos futuros: analisar a ferramenta em uma quantidade maior de turmas, com perfis diferentes, como em cursos totalmente a distância e semipresenciais.

Além disso, pretende-se desenvolver novas funcionalidades, como por exemplo, criar a opção de formar tanto grupos homogêneos como heterogêneos, dando ao docente a possibilidade de agrupar indivíduos que tenham conhecimentos complementares.

Outra possibilidade de trabalho futuro é o de integrar a ferramenta com as técnicas de aprendizagem colaborativa, pois possibilitará formar os grupos de acordo com a técnica selecionada pelo professor. Finalizando, pretende-se associar de forma automática o resultado do grupo gerado pela ferramenta com a atividade postada no módulo do curso, de forma que seja possível a disponibilização automática dos grupos como anexo na descrição da atividade.

Todos esses trabalhos futuros podem ser realizados para potencializar a ferramenta. Assim, a pesquisa apresentada pode se tornar uma nova fonte para formação de grupos para atividades colaborativas em ambientes virtuais de aprendizagem.

REFERÊNCIAS

- ABED. Censo EAD.BR: **Relatório Analítico da Aprendizagem a Distância no Brasil 2015**. Curitiba: InterSaberes, 2016. Disponível em: <http://abed.org.br/arquivos/Censo_EAD_2015_POR.pdf> Acesso: 01 Jan.2016.
- ABNAR, S.; OROOJI, F.; TAGHIYAREH, F. **An evolutionary algorithm for forming mixed groups of learners in web based collaborative learning environments**. In: Proceedings of 2012 IEEE International Conference on Technology Enhanced Education, ICTEE 2012, 2012.
- ALMEIDA, C. Z.; VIEIRA M. B.; LUCIANO, A. N. **Ambiente Virtual de Aprendizagem: uma proposta para autonomia e cooperação na disciplina de informática**. In: XII Simpósio Brasileiro de Informática na Educação, SBIE, 2001. p. 431–439.
- ALTOÉ, A.; FUGIMOTO, S. M. A. **Computador na educação e os desafios educacionais**. In: IX Congresso Nacional de Educação, EDUCERE, 2009. p. 164–175.
- AMARAL, F. **Aprenda Mineração de Dados: Teoria e Prática**. Rio de Janeiro: Alta Books, 2016.
- ANTUNES, O. C. **Um mecanismo para apoiar a análise da interação e do desempenho de alunos em Ambientes Virtuais de Aprendizagem**. 2016. 152p. Dissertação (Mestrado em Informática), Universidade Federal do Amazonas, Manaus, 2016.
- ANTUNES, O.; LUCENA, K. K. T.; OLIVEIRA, E. H. T DE.; GUSMÃO, T. H. Y.; NOZAWA, E. H.; BRODBECK, W. **LMS Monitor: An application for Academic and performance monitoring Of Students in LMS**. Nuevas Ideas em Informática Educativa-TISE, 2016. p.45-52.
- ÁVILA, G. E. Geometria e Fundamentos. Revista do Professor de Matemática, 45, 2001.
- BALABANOVIC, M.; SHOHAM, Y. **Fab: Content-Based Collaborative Recommendation. Communication of the ACM**. New York. v. 40, n. 3, 1997. p. 66-72.
- BARKLEY, E. F.; MAJOR, C. H.; CROSS, K. P. **Collaborative Learning Techniques**. 2ed. Estados Unidos da América: Jossey- Bass, 2014. 417p.
- BASILI, V. R.; CALDIERA, G.; ROMBACH, H. D. **The Experience Factory**. Encyclopedia of Software Engineering. New York: John Wiley & Sons. 1994.
- BEAN, J.C.(1996) **Engaging ideas: The professor's guide to integrating writing, critical thinking, and active learning in the classroom**. San Francisco: Jossey-Bass.
- BORATTO, L.; CARTA, S. **State-of-the-art in group recommendation and new approaches for automatic identification of groups**. Studies in Computational Intelligence, Springer Verlag. v. 324, 2010. p. 1–20.
- BOSTOCK, M. Home · mbostock/d3 Wiki · GitHub. **GitHub**, 2015. Disponível em: <<https://github.com/mbostock/d3/wiki/Gallery>>. Acesso em: 17 dez. 2015.

CARVALHO, L. A. M. C.; MACEDO, H. T. **Introdução aos sistemas de recomendação para grupos**. Revista de Informática Teórica e Aplicada. v. 21, 2014. p. 78–109.

CAZELLA, S. C.; NUNES, M.; REATEGUI, E. **A Ciência da Opinião: Estado da arte em Sistemas de Recomendação**. In: XXX Congresso da SBC Jornada de Atualização de Informática - JAI, 2010, p. 161–216.

CAZELLA, S. C.; REATEGUI, E. B.; MACHADO, M. **Recomendação de Objetos de Aprendizagem Empregando Filtragem Colaborativa e Competências**. In: Anais do Simpósio Brasileiro de Informática na Educação, v. 1, 2009.

CITADIN, J. R.; KEMCZINSKI, A.; MATOS, A. V. **Formação de Grupos para Aprendizagem Colaborativa: Um mapeamento sistemático da literatura**. In: XIX Conferência Internacional sobre Informática na Educação, v. 10, 2014. p. 46-54.

DARADOUMIS, T.; GUITERT, M.; GIMENEZ, F.; MARQUES, J. M.; LORET, T. **Supporting the composition of effective virtual groups for collaborative learning**. In: International Conference on Computers in Education, v. 1, n. 1, 2002. p. 332–336.

DINIZ, C. A.; LOUZADA, N. F. **Data Mining: uma introdução**. São Paulo: ABE, 2000.

FERRO, M. R. C. **Um Modelo de Sistema de Recomendação de Materiais Didáticos para Ambientes Virtuais de Aprendizagem**. In: Anais do Simpósio Brasileiro de Informática na Educação, v. 1, n. 1, 2011. p. 810–819.

GABARDO, P.; QUEVEDO S.; ULBRICHT V. R. **Estudo Comparativo das plataformas de ensino-aprendizagem**. In: XVI Congresso Internacional de Educação a Distância, 2010.

GAERTNER, C. A. **Desenvolvimento de Framework e Processo para Justificação Econômico-Financeira de Tecnologias Avançadas de Manufatura (AMT)**. Dissertação de Mestrado. Curitiba: PUC Paraná, 2005.

GOLDBARG, M.; GOLDBARG E. **Grafos: conceitos, algoritmos e aplicações**. Rio de Janeiro: Elsevier, 2012.

GOLDSCHMIT, R.; BEZERRA, E.; PASSOS, E. **Data Mining: conceitos, técnicas, orientações e aplicações**. 2ed. Rio de Janeiro: Elsevier, 2015.

HERLOCKER, J. L. **Understanding and Improving Automated Collaborative Filtering Systems**. 2000. Tese de Doutorado (Doutorado em Ciência da Computação), University of Minnesota, Minnesota, 2000.

INEP. **Censo Escolar da Educação Básica 2013**. Brasília, DF: Ministério da Educação/Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, 2013. Disponível em:
<http://download.inep.gov.br/download/superior/censo/2013/resumo_tecnico_censo_educacao_superior_2013.pdf> Acesso em: 05 Mar. 2016.

JAGADISH, D. **Grouping in collaborative e-learning environment based on interaction among students**. In: International Conference on Recent Trends in Information Technology, IEEE, 2014.

JAIN, K. A. **Data Clustering 50 Years Beyond K-Means**. 2008. Disponível em: < <http://www.cs.utah.edu/~piyush/teaching/kmeans50yrs.pdf> >. Acessado em: 02 Mar. 2017.

KITCHENHAM, B. **Procedures for Performing Systematic Reviews**. Department of Computer Science Keele University. Keele, Reino Unido. 2004.

KITCHENHAM, B.; CHARTES, S. **Guidelines for performing systematic literature reviews in software engineering**. EBSE Technical Report EBSE-2007-01, Software Engineering Group Department of Computer Science Keele University, 2007.

LATCHI, R. L.; ROCHA, H. V. **Aspectos básicos de clustering: Conceitos e Técnicas**. Relatório Técnico. Campinas, 2005. 26p.

LAVRADOR, I. M. B. F. **Análise de Redes Sociais no Estudo de Interações em Fóruns de discussão: o caso de uma disciplina de mestrado em regime de E-Learning**. 2015. 67p. Dissertação (Mestrado em Educação), Universidade de Lisboa, Lisboa, 2015.

LIMA, D.; GADELHA, B.; NETTO, J.F.; BREMGARTNER, V. **MoodleGroups: Um Aplicativo para identificar as Relações Sociais entre Alunos em Ambientes Virtuais de Aprendizagem**. In: XI Simpósio Brasileiro de Sistemas Colaborativos (SBSC 2014), 2014. p. 6-9.

MARINHO, E. C.; SILVA, M. F.; SCHMITZ, E.; ALENCAR, A. J. **A Brazilian initiative towards Digital Literacy: assessing Motivational Factors of e-Proinfo using TAM**. In: XXV Simpósio Brasileiro de Informática na Educação, SBIE, 2014. p. 1–10.

MARTINS, H. N. J.; COSTA, E. C.; OLIVEIRA, T. T. M.; SILVA, A. P.; BITTENCOURT, I. I. **Sistema de Recomendação Híbrido para Bibliotecas Digitais que Suportam o Protocolo OAI-PMH**. In: XXII Simpósio Brasileiro de Informática na Educação, SBIE, 2011. p. 140–149.

MEDEIROS, I. R. G. **Estudo sobre Sistemas de Recomendação Colaborativos**. 2013. 45p. Monografia apresentada como requisito parcial para Conclusão de Curso de Bacharel em Ciência da Computação, Universidade Federal de Pernambuco, Recife, 2013.

MENEZES, C. S.; VALE, K. S. B.; CRUZ, P. O. **AICAPA - Um Ambiente Inteligente e Colaborativo para Apoio à Produção Acadêmica**. Revista Novas Tecnologias na Educação. v.11, 2013. p.1-10.

MESSA, V. C. **Utilização de Ambientes Virtuais De Aprendizagem – AVAs: a Busca Por Uma Aprendizagem Significativa**. Revista Brasileira de Aprendizagem Aberta e a Distância, v. 9, 2010. p. 1–49.

MOODLE. Blocks. **Blocks - MoodleDocs**, 2017a. Disponível em: <<https://docs.moodle.org/30/en/Blocks>>. Acesso em: 01 set. 2017.

MOODLE. Plugin types. **Plugin types - MoodleDocs**, 2017b. Disponível em: <https://docs.moodle.org/dev/Plugin_types>. Acesso em: 01 set. 2017.

MONTAZER, G. A.; REZAEI, M. S. **A new approach in E-Learners grouping using hybrid clustering method**. In: International Conference on Education and e-Learning Innovations, v. 77, n. 380, 2013. p. 841–847.

MORAN, M. J. **Ensino e aprendizagem inovadores com o apoio de tecnologias in: Novas Tecnologias e mediação pedagógica**. Orgs. José Manuel Moran, Maria T. Masetto, Maria Aparecida Behrens – 21ª edição. Campinas, SP: Papirus, 2013. (Coleção Papirus da Educação).

NASCIMENTO, M. G. F.; BRANDÃO, L. O. **Novo núcleo para o Moodle: promovendo um modelo hierárquico e compartilhável para disciplinas virtuais e materiais digitais**. In: Anais do Simpósio Brasileiro de Informática na Educação, 2010.

NETO, M. A. M. **Estratégia Híbrida para recomendação personalizada utilizando o Guia de Programação Eletrônico**. 2011. 106p. Dissertação (Mestrado Acadêmico em Ciência da Computação), Universidade Estadual do Ceará, Fortaleza, 2011.

NETTO, P.O.B. **Grafo: teorias, modelos, algoritmos**. São Paulo: Blucher, 2011.

NUNES, F. B.; PRASS F. S. **Desenvolvimento de um Ambiente Virtual de Aprendizagem**. In: Anais do 23º Simpósio Brasileiro de Informática na Educação, 2012.

OTSUKA J. L.; ROCHA H. V. **Avaliação formativa em ambientes de EaD**. In: Anais do XIII Simpósio Brasileiro de Informática na Educação, 2002. p. 146–156.

OUNNAS, A.; DAVIS, H. C.; MILLARD, D. E. **Towards semantic group formation**. In: Proceedings of The 7th IEEE International Conference on Advanced Learning Technologies, ICAIT 2007, 2007. p. 825–827.

OUNNAS, A.; DAVIS, H. C.; MILLARD, D. E. **A framework for Semantic Group Formation in Education**. Educational Technology & Society, 12 (4), 2009. p. 43-55.

PRENSKY, M. (2001). **Digital Natives, Digital Immigrants**. Disponível em: <<http://bit.ly./mprenskyndi>>. Acesso em: 03 Jan. 2015.

RAMOS, D. B; RAMOS, I. M. M; NASCIMENTO, P. B; OLIVEIRA, E. H. T. **Ferramenta Baseada em Grafo para Identificação e Visualização de Trilhas de Aprendizagem**. In: XXI Congresso Internacional de Informática Educativa, 2016a. p.237-243.

RAMOS, D. B. **Uma ferramenta baseada em grafo para identificação e visualização de trilhas de aprendizagem**. 2015. 93p. Dissertação (Mestrado em Informática), Instituto de Computação, Universidade Federal do Amazonas, Manaus, 2016.

RAMOS, I.; RAMOS, D.; NASCIMENTO, P.; GADELHA, B.; E OLIVEIRA, E. **Sistema de Combinação Social em um Ambiente Virtual de Aprendizagem**. In: XXI Congresso Internacional de Informática na Educativa , 26, 2016b. p.310-319.

RAMOS, D. B.; OLIVEIRA, E. H. T.; RAMOS, I. M. M.; OLIVEIRA, K. M. T. **Trilhas de Aprendizagem em Ambientes Virtuais de Ensino- aprendizagem: Uma Revisão Sistemática da Literatura**. In: Simpósio Brasileiro de Informática de Educação, 2015. p. 338–347.

RAMOS, I. M. M.; OLIVEIRA, E. H. T.; RAMOS, D. B. **Recomendação de grupos para atividades colaborativas utilizando a caracterização dos aprendizes baseada em trilhas de aprendizagem**. In: Simpósio Brasileiro de Sistemas Colaborativos - (SBSC 2015), 2015. p. 213–219.

RAMOS, I. M. M.; OLIVEIRA, E. H. T. **Relatório Técnico da Revisão Sistemática de Literatura sobre Recomendação de Grupos para atividades colaborativas utilizando a caracterização dos aprendizes baseada em trilhas de aprendizagem**. Universidade Federal do Amazonas. Manaus. 2015. Disponível em: <http://mobmoodle.icomp.ufam.edu.br/projeto/?page_id=10>. Acesso em: 01 dez. 2015.

SALAZAR, A. E. C. **Mineração de estruturais musicais e composição automática utilizando redes complexas**. 2015. 171p. Tese (Doutorado em Ciência de Computação e Matemática Computacional), Universidade de São Paulo, São Carlos, 2015.

SANTOS, L.; CASTRO, J. DE; CASTRO, T. **Alteração no Modelo de Grupos do Moodle para Apoiar a Colaboração**. In: Anais do XVIII Simpósio Brasileiro de Informática na Educação, 2007. p. 123–134.

SANTOS, R. A. P.; BOTELHO, S. S. C.; BICHET, M. A. A. **Ambientes Físico-Virtuais de Aprendizagem**. In: Congresso Brasileiro de Informática na Educação, 2014. p. 70–79.

SCHAFER, J. B.; KONSTAN, J.; RIEDL, J. **E-Commerce Recommendation Applications**. In: Data Mining and Knowledge Discovery, Higham, MA, USA, Vol 5, Kluwer Academic Publishers, 2001.

SHEHABUDDEEN.; PROBERT D.; PHAAL R.; PLATTS K. **Representing and approaching complex management issues: part 1-role and definition** (Working Paper), Institute for Manufacturing, University of Cambridge, UK, 2000.

SILVA, T.A.E. **Avaliação da Usabilidade de Interfaces Web segundo Critérios Ergonômicos de Bastien e Scapin: Pesquisa com os Ambientes Virtuais de Educação à distância Aulanet, E-Proinfo e TelEduc**. 2007. 124p. Dissertação (Mestrado), Universidade Federal do Rio Grande do Norte, Natal, 2007.

SOMMERVILLE, I. **Engenharia de Software**. 9. ed. São Paulo: Pearson Prentice Hall, 2011.

TAFNER, E. P.; TOMELIN, J. F; MULLER, R. B. **Trilhas de aprendizagem: uma nova concepção nos ambientes virtuais de aprendizagem (AVA)**. In: 18º Congresso Internacional de Educação a Distância, 2012.

TEIXEIRA, C. H. C. **Similaridade de Grafos via Hashing**. 2011. 82p. Dissertação (Mestrado), Universidade Federal de Minas Gerais, Belo Horizonte, 2011.

TESTA, G. G. **Uma Abordagem Híbrida para Recoemndação de Parceiros em Ambientes Virtuais Colaborativos de Composição Musical**. 2013. 77p. Dissertação (Mestrado em Ciência da Computação), Universidade Federal do Rio Grande do Sul, Porto Alegre, 2013.

TRAVASSOS, G. H.; DOS SANTOS, P. S. M.; MIAN, P. G.; NETO, A. C. D.; BIOLCHINI, J. C. A. **An Environment to Support Large Scale Experimentation in Software Engineering**. In: Proceedings of 13th IEEE International Conference on the Engineering of Complex Computer Systems (ICECCS), p. 193-202, Belfast, United Kingdom, 2008.

VARGAS, R. R. **Uma Nova forma de Calcular os Centros dos Clusters em Algoritmos de Agrupamento tipo Fuzzy -C-Means**. 2012. 97p. Tese (Doutorado), Universidade Federal do Rio Grande do Norte, Natal, 2012.

YATHONGCHAI, C.; ANGSKUN, T.; YATHONGCHAI, W.; ANGSKUN, J. **Leaner Classification Based on Learning Behavior and Performance**. In: IEEE Conference on Open Systems (ICOS), 2013. p. 66–70.

APÊNDICE A – PUBLICAÇÕES PRODUZIDAS NO MESTRADO

As publicações apresentadas abaixo foram obtidas durante a realização do Mestrado, confirmando a viabilidade da pesquisa desenvolvida neste trabalho.

RAMOS, I. M. M.; OLIVEIRA, E. H. T.; RAMOS, D. B. **Recomendação de grupos para atividades colaborativas utilizando a caracterização dos aprendizes baseada em trilhas de aprendizagem.** In: Simpósio Brasileiro de Sistemas Colaborativos - (SBSC 2015), 2015. p. 213–219.

RAMOS, D. B.; OLIVEIRA, E.H.T.; RAMOS, I.M.M.; OLIVEIRA, K. M.T. **Trilhas de Aprendizagem em Ambientes Virtuais de Ensino- aprendizagem: Uma Revisão Sistemática da Literatura.** In: Simpósio Brasileiro de Informática de Educação, 2015. p. 338–347.

RAMOS, I. M. M.; RAMOS, D. B.; NASCIMENTO, P. B.; GADELHA, B. F.; E OLIVEIRA, E.H.T. **Formação de Grupos em Ambientes Virtuais de Aprendizagem: Uma Revisão Sistemática da Literatura.** In: XXVII Simpósio Brasileiro de Informática na Educação , 26, 2016. p.310-319.

RAMOS, I. M. M.; RAMOS, D. B.; NASCIMENTO, P. B.; GADELHA, B. F.; E OLIVEIRA, E. H. T. **Sistema de Combinação Social em um Ambiente Virtual de Aprendizagem.** In: XXI Congresso Internacional de Informática na Educativa , 26,2016. p.310-319.

RAMOS, I. M. M.; RAMOS, D. B.; NASCIMENTO, P. B.; AMARAL, G.S. ; GADELHA, B. F.; E OLIVEIRA, E. H. T. **M-Cluster: Uma ferramenta de Recomendação para Formação de Grupos em Ambientes Virtuais de Aprendizagem.** In: Simpósio Brasileiro de Informática de Educação, 2017. p. 1657–1666.

APÊNDICE B – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

PODER EXECUTIVO
MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO AMAZONAS
INSTITUTO DE COMPUTAÇÃO



TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Você está sendo convidado(a) a participar da Pesquisa sobre Recomendação de Grupos para Atividades Colaborativas Utilizando a Caracterização dos Aprendizizes Baseada nas Trilhas de Aprendizagem, sob a responsabilidade da pesquisadora Ilmara Monteverde Martins Ramos. Essa pesquisa pretende fornecer dados para uma ferramenta que irá sugerir ao professor grupos de alunos para realizar atividades colaborativas no Moodle.

Sua participação é voluntária e se dará por meio da utilização do ambiente virtual Moodle, no IComp ela foi customizada com o nome ColabWeb, seja pelo computador ou dispositivos móveis. A sua interação com o Moodle é registrada pelo próprio sistema e esses dados serão utilizados pela ferramenta desenvolvida pelo pesquisador David Ramos e os dados coletados por essa ferramenta servirão de suporte para o desenvolvimento dessa pesquisa. Esses dados serão utilizados pela ferramenta desenvolvida por esta pesquisadora. A ferramenta sugere ao professor grupos de alunos baseados nas trilhas de aprendizagem para realização de atividades colaborativas. Se você aceitar participar, estará contribuindo para a melhoria de aprendizado em ambientes virtuais, particularmente o Moodle, a partir do uso de uma ferramenta que permite ao professor ter sugestões de agrupamentos por meio do conhecimento do comportamento de seus alunos e da turma no Ambiente.

Se depois de consentir em sua participação você desistir de continuar participando, tem o direito e a liberdade de retirar seu consentimento em qualquer fase da pesquisa, seja antes, durante ou depois da coleta dos dados, independente do motivo e sem prejuízo a sua pessoa. Você não terá qualquer despesa e também não receberá nenhuma remuneração. Em nenhum momento você será identificado. Os resultados da pesquisa serão publicados e ainda assim a sua identidade será preservada.

Para qualquer outra informação adicional ou esclarecimentos, você poderá entrar em contato com a pesquisadora no endereço ilmaramonteverde@gmail.com.

Consentimento Pós-Informação

Eu, _____, fui informado sobre o que a pesquisadora quer fazer e porque precisa da minha colaboração, e entendi a explicação. Por isso, eu concordo em participar do projeto, sabendo que não vou ganhar nada e que posso sair quando quiser. Este documento é emitido em duas vias que serão ambas assinadas por mim e pelo pesquisador, ficando uma via com cada um.

Manaus, de Janeiro de 2017.

Assinatura do participante

Assinatura do Pesquisador Responsável

APÊNDICE C – RESULTADOS DAS ANÁLISES DOS ATRIBUTOS

Tabela 2: Resultados dos atributos versão um com todas as distâncias turma 2015/1.

Distância Euclidiana			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
	Alunos 30	Alunos 29	Alunos 25, Alunos 26
Quantitativo de Alunos por Cluster			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
3 Alunos	15 Alunos	15 Alunos	11 Alunos
Distância Manhattan			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
	Alunos 30	Alunos 29	Alunos 25, Alunos 26
Quantitativo de Alunos por Cluster			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
3 Alunos	15 Alunos	15 Alunos	11 Alunos
Distância Cosseno			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
Alunos 29	Alunos 30		Alunos 25, Alunos 26
Quantitativo de Alunos por Cluster			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
8 Alunos	12 Alunos	14 Alunos	10 Alunos

Tabela 3: Resultados dos atributos versão dois.

Distância Euclidiana			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
Alunos 29	Alunos 30		Alunos 25, Alunos 26
Quantitativo de Alunos por Cluster			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
4 Alunos	10 Alunos	19 Alunos	11 Alunos
Distância Manhattan			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
Alunos 29	Alunos 30		Alunos 25, Alunos 26
Quantitativo de Alunos por Cluster			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
4 Alunos	10 Alunos	19 Alunos	11 Alunos
Distância Cosseno			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
Alunos 29	Alunos 30		Alunos 25, Alunos 26
Quantitativo de Alunos por Cluster			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
7 Alunos	15 Alunos	12 Alunos	10 Alunos

Tabela 4: Resultados dos atributos versão três

Distância Euclidiana			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
Alunos 29	Alunos 30		Alunos 25, Alunos 26
Quantitativo de Alunos por Cluster			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
4 Alunos	12 Alunos	17 Alunos	11 Alunos
Distância Manhattan			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
Alunos 29	Alunos 30		Alunos 25, Alunos 26
Quantitativo de Alunos por Cluster			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
4 Alunos	12 Alunos	17 Alunos	11 Alunos
Distância Cosseno			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
Alunos 29	Alunos 30		Alunos 25, Alunos 26
Quantitativo de Alunos por Cluster			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
7 Alunos	15 Alunos	12 Alunos	10 Alunos

Tabela 5: Resultados dos atributos versão quatro.

Distância Euclidiana			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
	Alunos 30	Alunos 29	Alunos 25, Alunos 26
Quantitativo de Alunos por Cluster			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
3 Alunos	15 Alunos	15 Alunos	11 Alunos
Distância Manhattan			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
	Alunos 30	Alunos 29	Alunos 25, Alunos 26
Quantitativo de Alunos por Cluster			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
3 Alunos	15 Alunos	15 Alunos	11 Alunos
Distância Cosseno			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
Alunos 29	Alunos 30		Alunos 25, Alunos 26
Quantitativo de Alunos por Cluster			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
8 Alunos	12 Alunos	14 Alunos	10 Alunos

Tabela 6: Resultados dos atributos versão cinco.

Distância Euclidiana			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
Alunos 29	Alunos 30		Alunos 25, Alunos 26
Quantitativo de Alunos por Cluster			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
4 Alunos	10 Alunos	19 Alunos	11 Alunos
Distância Manhattan			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
Alunos 29	Alunos 30		Alunos 25, Alunos 26
Quantitativo de Alunos por Cluster			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
4 Alunos	10 Alunos	19 Alunos	11 Alunos
Distância Cosseno			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
Alunos 29	Alunos 30		Alunos 25, Alunos 26
Quantitativo de Alunos por Cluster			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
7 Alunos	15 Alunos	12 Alunos	10 Alunos

Tabela 7: Resultados dos atributos versão seis.

Distância Euclidiana			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
Alunos 29	Alunos 30	Alunos 26	Alunos 25
Quantitativo de Alunos por Cluster			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
10 Alunos	20 Alunos	11 Alunos	3 Alunos
Distância Manhattan			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
Alunos 29	Alunos 30	Alunos 25	Alunos 26
Quantitativo de Alunos por Cluster			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
10 Alunos	20 Alunos	11 Alunos	3 Alunos
Distância Cosseno			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
Alunos 29	Alunos 30		Alunos 25, Alunos 26
Quantitativo de Alunos por Cluster			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
7 Alunos	14 Alunos	13 Alunos	10 Alunos

Tabela 8: Resultados dos atributos versão sete.

Distância Euclidiana			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
	Alunos 26,29		Alunos 25,26
Quantitativo de Alunos por Cluster			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
2 Alunos	28 Alunos	12 Alunos	2 Alunos
Distância Manhattan			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
	Alunos 26,29	Alunos 25,30	
Quantitativo de Alunos por Cluster			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
2 Alunos	28 Alunos	12 Alunos	2 Alunos
Distância Cosseno			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
Alunos 29	Alunos 30		Alunos 25, Alunos 26
Quantitativo de Alunos por Cluster			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
7 Alunos	14 Alunos	13 Alunos	10 Alunos

Tabela 9: Resultados dos atributos versão oito.

Distância Euclidiana			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
	Alunos 26	Alunos 25,30	Aluno 29
Quantitativo de Alunos por Cluster			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
4 Alunos	10 Alunos	19 Alunos	11 Alunos
Distância Manhattan			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
	Alunos 26	Alunos 25,30	Aluno 29
Quantitativo de Alunos por Cluster			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
4 Alunos	10 Alunos	19 Alunos	11 Alunos
Distância Cosseno			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
Alunos 29	Alunos 30	Aluno 25, Aluno 26	
Quantitativo de Alunos por Cluster			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
7 Alunos	15 Alunos	12 Alunos	10 Alunos

Tabela 10: Resultado dos agrupamentos da turma 2016/1.

Distância Euclidiana			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
	Alunos 42,43		Alunos 34,35
Quantitativo de Alunos por Cluster			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
9 Alunos	13 Alunos	10 Alunos	21 Alunos
Distância Manhattan			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
		Alunos 34,35	Alunos 42,43
Quantitativo de Alunos por Cluster			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
9 Alunos	13 Alunos	10 Alunos	21 Alunos
Distância Cosseno			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
Alunos 42,43			Alunos 34,35
Quantitativo de Alunos por Cluster			
Cluster 0	Cluster 1	Cluster 2	Cluster 3
16 Alunos	1 Aluno	11 Alunos	25 Alunos

APÊNDICE D – ARTIGO

Este apêndice apresentará um artigo autoral que mais foi citado no texto e que serviu de base para a concepção das trilhas de aprendizagem.

Trilhas de Aprendizagem em Ambientes Virtuais de Ensino-aprendizagem: Uma Revisão Sistemática da Literatura

David Brito Ramos^{1,3}, Elaine Harada Teixeira de Oliveira¹, Ilmara M. M. Ramos^{1,3}, Kelson M. T. Oliveira²

¹Instituto de Computação – Universidade Federal do Amazonas (UFAM) – Manaus – AM – Brasil

²Instituto de Ciências Exatas – Universidade Federal do Amazonas (UFAM) – Manaus – AM – Brasil

³Instituto Federal de Educação, Ciência e Tecnologia do Amazonas (IFAM) – Campus Parintins – Parintins – AM – Brasil

davidness@gmail.com, elaine@icomp.ufam.edu.br, ilmaramonteverde@gmail.com, kelsonmota@ufam.edu.br

Abstract. *The research presented in this paper is the result of a Systematic Literature Review that aimed to identify how learning paths are represented within the e-learning systems. It was also verified if these representations include not only data structure and modeling of e-learning systems, but also the visualization of learning paths by teachers and students.*

Resumo. *A pesquisa apresentada neste trabalho é resultado de uma Revisão Sistemática da Literatura que buscou identificar como as trilhas de aprendizagem são representadas nos sistemas e-learning. Foi verificado se essas representações contemplam apenas a estrutura de dados e modelagem dos sistemas de e-learning, ou se também há visualização das trilhas de aprendizagem pelos professores e alunos.*

1 INTRODUÇÃO

O uso da Internet como facilitador da comunicação tem se mostrado útil em diversas áreas de conhecimento. Com o melhoramento das conexões e a ampliação do acesso à Internet, muitas pesquisas têm voltado seus esforços para o melhoramento dos ambientes virtuais de aprendizagem. Para que estas mudanças de melhoria aconteçam, está claro que é necessário conhecer o aluno. Nesse sentido muitas técnicas estão sendo aplicadas, entretanto para a maioria delas é necessária a solicitação explícita de dados do aluno, tendo como fonte direta de informações unicamente o próprio aluno, como nos trabalhos de Ferreira, Barbosa e Gluz (2013), de Müller e Silveira (2013) e Gomes e Pimentel (2013). Assim os dados coletados podem não corresponder à situação real do aluno. Para evitar esse problema, existem

trabalhos que buscam monitorar, de forma transparente, as atividades dos alunos no intuito de encontrar padrões de comportamento que possam elucidar possíveis problemas ou melhoras no aprendizado. Uma das formas de monitoramento do aluno pode ser realizada através da observação das ações que ele executa dentro do sistema, sendo que estas formam caminhos que são conhecidos como trilhas de aprendizagem. Nesse sentido, esta Revisão Sistemática busca identificar como essas trilhas são representadas, uma vez que elas descrevem de forma natural o comportamento do aluno.

A seguir, a seção 2 define o contexto da pesquisa. A seção 3 aborda o método. A seção 4 apresenta o processo de condução da revisão. Na seção 5 são realizadas a análise e discussão dos dados. Na seção 6 são apresentadas as limitações da revisão e por fim são feitas as considerações finais.

2 CONTEXTO DA PESQUISA

2.1 AMBIENTES VIRTUAIS DE APRENDIZAGEM

Um Ambiente Virtual de Aprendizagem ou Ensino-Aprendizagem (AVA), também conhecido como LMS (*Learning Management System*) ou VLE (*Virtual Learning Environment*), de acordo com Ninoriya, Chawan e Meshram (2011), é uma ferramenta que permite o gerenciamento de cursos online, registro de alunos e a avaliação de seus resultados de aprendizagem. Segundo Silva (2011), um AVA busca permitir a construção do conhecimento por meio da interatividade e interação entre pessoas ou grupos e não apenas disponibilizar conteúdo.

2.2 TRILHAS DE APRENDIZAGEM

De acordo com Tafner, Tomelin e Müller (2012), as trilhas de aprendizagem são caminhos virtuais para o desenvolvimento intelectual que promovem e desenvolvem competências.

No contexto de desenvolvimento de ferramentas educacionais, nos trabalhos da literatura relacionados aos AVAs, as trilhas de aprendizagens, são consideradas sob dois pontos de vista: a trilha é a sequência de conteúdo e atividades definida pelo professor, ao planejar sua disciplina e, posteriormente, disponibilizá-la no ambiente virtual; ou a trilha é aquela percorrida pelo aluno, durante a sua interação com recursos disponibilizados no ambiente virtual. Trilhas de aprendizagem também são referenciadas como: caminhos de aprendizagem, percurso de aprendizagem e em língua inglesa: *learning path*, *learning route*, *learning itineraries*.

3 METODOLOGIA

Normalmente, as pesquisas iniciam com uma revisão da literatura, uma vez que o estudo de trabalhos relacionados ajuda no processo de criação de um novo conhecimento. O resultado de uma revisão abrangente é, indubitavelmente, fator determinante no desenvolvimento da pesquisa. Para Kitchenham e Charters (2007) a pesquisa deve ser minuciosa e para ter valor científico, sendo esta a principal razão para a realização de Revisões Sistemáticas da Literatura (RSL).

Trabalhos recentes dentro do escopo de ensino-aprendizagem têm sido publicados como RSL. Pontes *et al* (2014), apresenta uma RSL sobre filtragens de recomendação de Objetos de Aprendizagens publicados no SBIE. O trabalho de Zem-Lopes *et al* (2013) busca identificar como tecnologias da Web Semântica são utilizadas em ambientes educacionais e quais as principais técnicas empregadas. O trabalho de Fassbinder, Delamaro e Barbosa (2014), apresenta uma RSL sobre *Massive Open Online Courses* (MOOCs) que busca definir, categorizar e caracterizar o ambiente dos MOOCs.

Este trabalho está baseado em um estudo aprofundado sobre trilhas de aprendizagem, realizado por meio de uma RSL, seguindo as orientações de Kitchenham (2004) (o resultado completo está registrado em Ramos e Oliveira (2015)). A revisão realizada pode ser classificada como uma revisão *quasi*-sistemática, uma vez que não houve comparações entre publicações, pois buscou-se realizar uma pesquisa de cunho exploratório. Durante o processo da RSL utilizou-se uma ferramenta especialmente para este fim, chamada de *StArt*, desenvolvida na Universidade de Federal de São Carlos, pelo Laboratório de Pesquisa em Engenharia de Software (LaPES) e está disponível no endereço lapes.dc.ufscar.br. A versão 2.0 foi utilizada para auxiliar a pesquisa.

3.1 OBJETIVO E QUESTÃO DE PESQUISA

O objetivo da RSL foi identificar como as trilhas de aprendizagem são representadas dentro dos sistemas *e-learning* e se existe uma representação visual dessas trilhas para os professores e alunos. A descrição do objetivo conforme o paradigma *Goal-Question-Metric* de Basili, Caldiera e Rombach (1984), pode ser observada abaixo:

Analisar publicações científicas através de um estudo baseado em revisão sistemática, **com o propósito de** identificar como são representadas as trilhas de aprendizagem em Ambientes Virtuais de Aprendizagem **com relação à** representação dessas trilhas para os sistemas, professores e alunos **do ponto de vista dos** pesquisadores **no contexto** acadêmico ou industrial.

A RSL buscou responder à questão principal: *como as trilhas de aprendizagem são representadas dentro dos sistemas de e-learning?*

3.2 MÉTODO DE BUSCA DAS PUBLICAÇÕES

Utilizou-se as máquinas de buscas específicas, das bibliotecas selecionadas, sendo consultados os conjuntos de palavras-chave pré-estabelecidos na pesquisa.

Apenas fontes digitais foram consideradas para consulta, sendo estas: *IEEE Computer Science Digital Library* <<http://ieeexplore.ieee.org>> e *Scopus* <<http://www.scopus.com>>. A *Scopus* destaca-se por armazenar publicações de diversas fontes como, por exemplo: *Springer, ACM, ScienceDirect/Elsevier, British Computer Society*.

3.3 EXPRESSÃO DE BUSCA

A expressão de busca foi definida após a execução de testes, auxiliados com base em artigos de controle, composta por palavras-chave que representam a população (P) e a intervenção (I) (maiores detalhes estão descritos em Ramos e Oliveira (2015)). A composição básica da expressão de busca possui a seguinte estrutura: (P) AND (I). Para a busca na Scopus, as aspas ("") foram substituídas por chaves ({}).

A população é descrita por um conjunto de palavras relacionadas a *Learning Management Systems* e *e-Learning*, enquanto a intervenção descreve palavras relacionadas a *Learning Path*. A Tabela 1, a seguir, apresenta as palavras-chave da expressão de busca.

Tabela 1 - Termos de busca

População
"learning management system", "learning management systems", "lms", "virtual learning environment", "virtual learning environments", "vle", "course management system", "course management systems", "intelligent tutoring system", "intelligent tutoring systems", "coaching assisted system", "coaching assisted systems", "moodle", "distance course", "distance education", "distance learning", "distance training course", "online learning", "distance courses", "distance training courses", "e-learning framework", "e-learning system", "e-learning systems", "e-learning module", "e-learning modules", "e-learning recommender system", "e-learning recommender systems", "e-learning program", "e-learning programs", "e-learning adaptive system", "e-learning adaptive systems", "e-learning path", "e-learning paths", "e-learning setting", "e-learning settings", "distant learning system" e "distant learning systems"
Intervenção
"learning path", "learning itinerary", "learning route", "learning workflow", "individual path", "learning paths", "learning itineraries", "learning routes", "learning workflows" e "individual paths"

3.4 CRITÉRIOS DE INCLUSÃO E EXCLUSÃO

Os critérios de inclusão (CI) e de exclusão (CE), listados na Tabela 11, foram elaborados a partir dos artigos de controle.

Tabela 11 - Critérios de Inclusão e Exclusão

Critérios de Inclusão

(CI1) Artigos que apresentem representações de modelos, visuais ou não, sobre trilhas de aprendizagem.

(CI2) Artigos que apresentem propostas de adaptação de trilhas de aprendizagem em ambientes virtuais.

CrITÉRIOS de ExclusÃO

(CE1) Não serão selecionadas publicações em que as palavras-chave da busca não apareçam no título, resumo e/ou palavras-chave.

(CE2) Não serão selecionadas publicações em que as siglas equivalentes não signifiquem: LMS - "Learning Management System, VLE – "Virtual Learning Environment".

(CE3) Serão excluídas publicações que não tem disponibilidade de conteúdo para leitura (por exemplo, onde os trabalhos são pagos ou não disponibilizados pelas máquinas de buscas).

(CE4) Artigos que apenas apresentam relatos de experiência de algum modelo, técnica ou ferramenta.

(CE5) Artigos que apenas citam sistemas de e-learning (LMS, VLE, etc.), mas não o abordam como foco principal.

(CE6) Não serão selecionadas publicações que não relacionam Trilhas de Aprendizagem e Ambientes Virtuais de Aprendizagem (LMS, CMS, ITS, LCMS).

(CE7) Não serão selecionadas publicações que apresentam ambientes virtuais de aprendizagem, mas não descrevam trilhas de aprendizagem.

(CE8) Não serão selecionadas publicações que descrevam trilhas de aprendizagem, mas não relacionados a ambientes virtuais de aprendizagem.

3.5 EXTRAÇÃO DE DADOS

Foram extraídos dados básicos da publicação como título, autores, local de publicação, ano de publicação, palavras-chave e também dados relativos às características que auxiliaram na resposta da questão de pesquisa e também na caracterização das publicações. São eles: técnicas utilizadas no trabalho, ambiente virtual utilizado, tipo AVA, público-alvo no trabalho, abrangência do trabalho, tipo da pesquisa, momento de criação das trilhas de aprendizagem, como as trilhas são representadas, se a ferramenta da publicação apresentada recomenda trilhas de aprendizagem e quais os nós das trilhas.

4 CONDUÇÃO DE REVISÃO

Após o processo de planejamento da revisão, realizou-se a fase de execução. É importante informar que a execução da expressão de busca ocorreu em junho de 2014. Primeiramente, os estudos preliminares foram identificados através da execução da expressão de busca. Ao todo foram retornados 279 artigos, sendo 201 artigos da base da Scopus e 78 da base da IEEE, os 60 trabalhos duplicados foram identificados posteriormente via software e revisão manual, portanto foram analisadas, no primeiro filtro, 219 publicações.

No primeiro filtro, os trabalhos foram analisados de acordo com o título, resumo e palavras-chave. Publicações cujo conteúdo para leitura não pode ser obtido diretamente da fonte de pesquisa (Scopus e IEEE), foram pesquisadas via máquina de busca (*google.com*)

e/ou quando possível, os autores eram contatados via e-mail ou pelo portal *researchgate.net*. Esgotando-se as possibilidades, os trabalhos foram classificados conforme critérios estabelecidos na fase de planejamento (Figura 30).

Na fase do Segundo Filtro, somente os 127 trabalhos aceitos na fase anterior foram lidos na íntegra. Também foram adotados os mesmos critérios de inclusão e exclusão do primeiro filtro. Portanto, ao todo 76 trabalhos foram aceitos nesta última fase e formaram a base para a extração de dados (Figura 31).

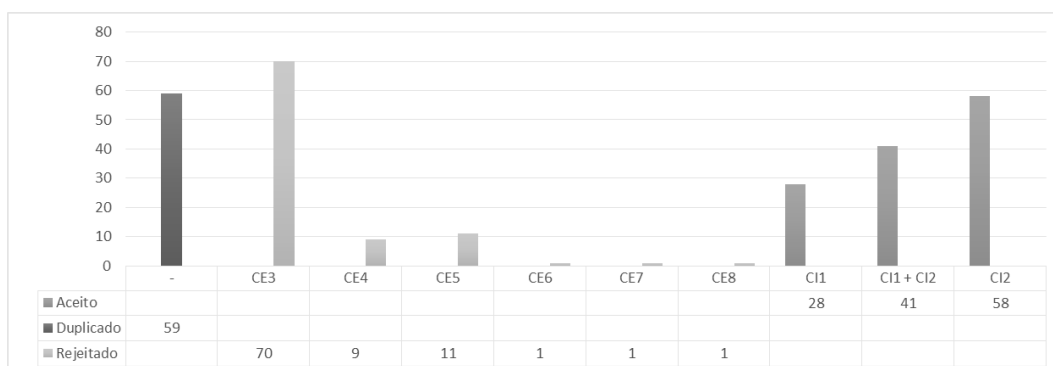


Figura 30 - Resultado do Primeiro Filtro por Critério

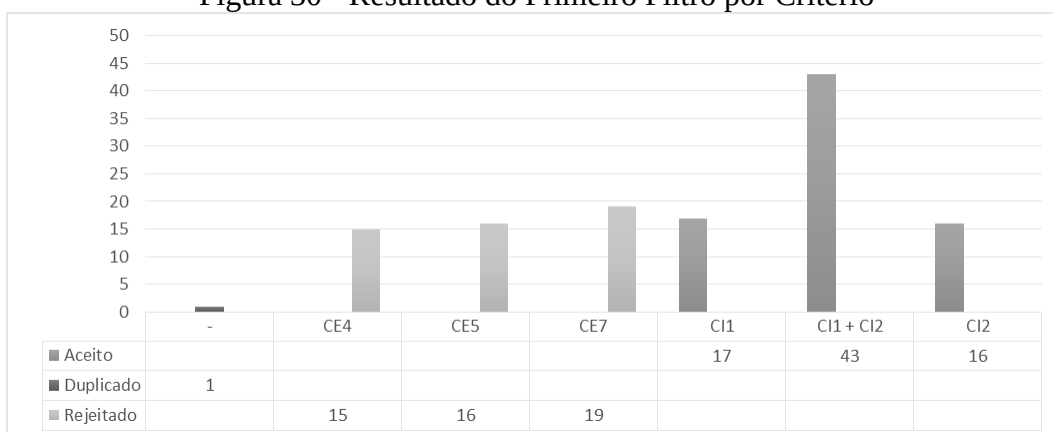


Figura 31 – Resultado do Segundo Filtro por Critério

5 ANÁLISE E DISCUSSÃO DOS DADOS

5.1 RESPOSTA À QUESTÃO PRINCIPAL

A partir dos dados das publicações aceitas foi possível estabelecer como as trilhas de aprendizagem são representadas nos ambientes virtuais, quais nós as compõem, em que momentos são geradas e se há recomendação de trilhas de aprendizagem.

Os métodos de representação de trilhas de aprendizagem mais utilizados são Grafos (39,47%) e Sequenciamento (30,26%). Trabalhos que utilizam grafos podem explorar uma flexibilização em tempo real da aprendizagem. O aprendiz pode, ao longo de todo o processo de estudo e em cada parte dele, mudar suas características e isso pode ser avaliado

constantemente em seu perfil. Já o sequenciamento implica que, em um determinado intervalo do curso, o aprendiz deverá percorrer uma sequência de atividades que não poderá ser mudada, ou seja, o aluno é avaliado periodicamente, mas somente entre as sequências, de forma que as características como conhecimento adquirido, estilo de aprendizagem, perfil cognitivo e outros dados sobre o aprendiz sejam atualizados e utilizados como parâmetros para a geração da sequência de atividades seguintes. A Tabela 12 apresenta todas as representações de trilhas de aprendizagem encontradas na revisão.

Tabela 12 - Representação das trilhas de aprendizagem

Como as trilhas são representadas	Total	Como as trilhas são representadas	Total
Grafo	30	Cadeia de Cromossomos	2
Sequenciamento	23	Árvore	1
Rede de Petri	4	Vetor	1
Ontologia	4	Mapa 3d_SLN	1
Fluxo de trabalho (workflow)	3	Espaço de Estados de Aprendizagem	1
Gráfico	3	Plano - Case-Based planning (CBP)	1
Relacionamento Conceito-Efeito	2		

Os termos mais citados com relação aos nós das trilhas de aprendizagem são: Objetos de Aprendizagem (OA), Conceitos, Módulos e Atividades. Observa-se que esses nós mais citados possuem uma visão macro do processo de aprendizagem, assim não representam especificamente um recurso, mais sim um agregado de recursos estáticos e/ou dinâmicos representados de forma única. Uma atividade envolve, por exemplo, o contato com os conhecimentos teóricos e a aplicação em um ou mais exercícios. A Tabela 13 apresenta todos os termos citados nas publicações analisadas, com destaque para dois trabalhos que citam a intervenção do professor/tutor como um nó da trilha de aprendizagem. Caso a publicação tenha afirmado trabalhar com OA e descreve alguns destes, por exemplo, como fóruns, vídeos e tarefas então os nós das trilhas de aprendizagem são registrados na extração de dados como dos tipos OA, fóruns, vídeos e tarefas e não apenas como OA. A mesma ideia vale para Conceitos, Módulos e Atividades. Esta forma de registro foi necessária por causa dos diferentes níveis detalhes das publicações que não podem ser desconsiderados.

Quando o aprendiz percorre um caminho dentro do ambiente de aprendizagem, ele executa uma série de ações, como assistir a um vídeo, acessar uma página externa, ler um conteúdo, realizar um teste, postar em um fórum, entre outras que lhe são permitidas. O conjunto de ações possíveis pode ser especificado antes mesmo do aprendiz interagir com o AVA, neste, cria-se um caminho padrão e uma série de caminhos alternativos, dessa forma é possível prever a próxima ação que será disponibilizada para o aprendiz. Para sistemas de *e-learning* com estas características, denominou-se o momento da geração das trilhas de

aprendizagem como pré-estabelecido. Existem 36 publicações, conforme Tabela 14, com esta característica. As demais publicações foram classificadas como de tempo real, isto significa que, apesar do ambiente virtual possuir em sua base de dados os recursos e atividades que serão disponibilizados aos aprendizes, não é determinado, previamente, qual atividade ou recurso será fornecido após a conclusão da ação anterior.

Tabela 13 - Nós das trilhas de aprendizagem

Tipos de nó	Total	Tipos de nó	Total	Tipos de nó	Total
Objetos de Aprendizagem	31	Material Web/Conteúdo Externo	4	Laboratório virtual	2
Conceitos	22	Página Web	4	Seminário	1
Atividades	16	Comentário das questões	3	Falar em publico	1
Módulos	16	Fórum	2	Curso alto nível	1
Questionários	12	Chat	2	Animações 3D	1
Tarefas	11	Intervenção Professor/Tutor	2	Biblioteca virtual	1
Objetos Multimídia	9	Notas Suplementares	2	Workshop virtual	1
Assistir um vídeo	4	Livros Recomendados	2	Colaboração virtual	1
		Resumo	2		

Tabela 14 - Geração das Trilhas de Aprendizagem

Momento de criação das trilhas de aprendizagem	Total
Tempo real	40
Pré-estabelecido	36

5.2 PUBLICAÇÕES

Além do levantamento sobre as questões diretamente relacionadas às trilhas de aprendizagem, também foi possível identificar informações referentes às publicações. A Figura 32, apresenta o número de publicações por ano, considerando apenas os artigos que foram aceitos dentro dos critérios pré-estabelecidos, o que revela um aumento do interesse dos pesquisadores nos últimos dez anos e, conseqüentemente, a crescente relevância do tema.



Figura 32- Número de publicações por ano

Outro destaque é que as publicações estão distribuídas por várias fontes cujos temas não são exatamente sobre Educação, como, por exemplo, Engenharia Elétrica ou Inteligência

Artificial. A Tabela 15, apresenta uma visão das principais fontes por contagem de publicações considerando apenas os artigos aceitos.

Tabela 15 - Principais locais de publicação

Local de Publicação	Total
Advanced Learning Technologies (ICALT), IEEE International Conference on	9
Computers and Education	5
Expert Systems with Applications	4
Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)	4
Education, IEEE Transactions on	2
Electronics, Circuits and Systems (ICECS), IEEE International Conference on	2
Multimedia Computing and Systems (ICMCS), International Conference on	2
Neural Networks (IJCNN), IEEE International Joint Conference on. (IEEE World Congress on Computational Intelligence).	2

Também verificou-se a abordagem de pesquisa das publicações aceitas para a revisão. Conforme é apresentado na Tabela 16, trinta e seis publicações foram classificadas como pesquisa de validação, ou seja, as técnicas não foram aplicadas na prática, os experimentos desses trabalhos são realizados em laboratório. Vinte e um são pesquisas de avaliação, nesse caso o trabalho possui uma ferramenta que foi aplicada na prática onde as consequências foram avaliadas. Por fim, dezenove publicações são propostas de solução onde foram defendidas a aplicabilidade e possíveis benefícios, mas o trabalho não possui uma ferramenta pronta ou ela está em fase inicial de desenvolvimento e consequente não possui resultados de experimentos.

Tabela 16 - Abordagem da Pesquisa

Abordagem da Pesquisa	Total
Validação	36
Avaliação	21
Proposta de solução	19

5.3 CONCLUSÕES DA ANÁLISE

Grafo é o método mais utilizado para representar trilhas de aprendizagem (39,47%). Os termos mais citados para os nós de uma TA são: Objetos de Aprendizagem, Conceitos, Módulos e Atividades. Em geral, as TAs são geradas em tempo real. O Moodle é citado por 13% dos trabalhos como AVA utilizado. O grande foco de pesquisa nos trabalhos é a adaptação do ambiente virtual de aprendizagem à individualidade do aluno. Verificou-se também que 55,26% dos trabalhos analisados foram publicados somente nos últimos cinco anos, sendo que o trabalho mais antigo foi publicado em 1994. Dentre todos os artigos aceitos para estudo, somente quatro apresentaram as trilhas de forma visual ao usuário (aprendiz, docente), conforme Tabela 17.

Tabela 17 - Publicações que exibem as trilhas visualmente aos aprendizes e docentes

Artigos que apresentam as trilhas visualmente aos usuários	Ano
How to See Training Paths in Learning Management Systems?	2010
eGraph tool: graphing the learning process in LMSs	2014
A Business Process Management Based Virtual Learning Environment : Customised Learning Paths	2011
Learning maps usage of cartographic metaphors for orientation in e-learning courses	2010

6 LIMITAÇÕES DA REVISÃO

A RSL é realizada por meio de protocolo robusto e objetivo que rege todo o processo de revisão, entretanto isso não garante a eliminação por completo da subjetividade inerente aos pesquisadores. Concomitantemente, a exclusão de algum trabalho relevante, eliminado por não corresponder a alguns dos critérios de seleção, é sempre uma possibilidade premente. Nesse sentido, é possível que a execução do primeiro filtro possa ter excluído alguns trabalhos relevantes. Isso porque as publicações que não continham os termos de busca nas palavras-chave, títulos ou resumos são descartadas nessa fase e nenhuma leitura completa é realizada. Outra limitação decorre de decisões subjetivas ocorridas durante o processo de leitura e compreensão dos trabalhos, pois alguns estudos não apresentaram uma descrição clara, dificultando a aplicação direta dos critérios de inclusão e exclusão.

7 CONSIDERAÇÕES FINAIS

Os estudos mostraram uma grande heterogeneidade para a representação das trilhas de aprendizagem, com destaque para o nível de detalhes de trabalhos que citavam especificamente páginas Web, fórum e chats, por exemplo, ou genericamente, como conceitos/módulos/curso e outros trabalhos que descreviam vários níveis de representação. A maioria dos trabalhos trata as trilhas de aprendizagem como resultado de sua criação e não do processo natural do aluno de percorrer o conteúdo do *e-learning*, mas também há trabalhos que consideram a ótica do aluno. Esta RSL também revela que o grande foco de pesquisa nos trabalhos estudados é a adaptação do ambiente virtual de aprendizagem à individualidade do aluno, seja ela caracterizada por seu perfil cognitivo, estilo de aprendizagem, habilidade ou nível de conhecimento. Há também um crescente aumento de pesquisas que utilizam técnicas de Inteligência Artificial, tais como: agentes, otimização de colônia de formiga, algoritmos genéticos, redes neurais. Muitos sistemas propostos são anunciados como sistema de recomendação, porém os conteúdos selecionados são impostos ao aluno.

De posse das informações obtidas, está sendo desenvolvido um modelo de identificação e representação de trilhas de aprendizagem que apresenta uma visualização de ações realizadas no ambiente virtual Moodle.

8 AGRADECIMENTOS

Os autores agradecem o apoio financeiro da Fundação de Amparo à Pesquisa do Estado do Amazonas (FAPEAM) através do Edital N° 016/2013, Projeto PROTI-Pesquisa sob o processo N° 062.00597/2014 e bolsa de pesquisa. Além disso, parte dos resultados apresentados neste trabalho foi obtida através de atividades de P&D do "Programa de Qualificação em Grande Escala em Tecnologias Móveis - PROMOBILE", projeto patrocinado pela Samsung Eletrônica da Amazônia Ltda. nos termos da lei federal brasileira n° 8.248/91.

9 REFERÊNCIAS

- Basili, V. R., Caldiera, G. e Rombach, H. D. (1984). The Experience Factory. Em J. Marciniak (Ed.), In: Encyclopedia of Software Engineering. New York: John Wiley & Sons.
- Fassbinder, A., Delamaro, M. E. e Barbosa, E. F. (2014). Construção e Uso de MOOCs: Uma Revisão Sistemática Aracele. In: XXV Simpósio Brasileiro de Informática na Educação (SBIE 2014), 25, (pp. 332-341). Dourados.
- Ferreira, L. G., Barbosa, J. L. e Gluz, J. C. (2013). Um Modelo de Recomendação Ubíqua de Conteúdo para Grupos de Aprendizes. In: XXIV Simpósio Brasileiro de Informática na Educação (SBIE 2013), 24. Campinas.
- Gomes, E. H. e Pimentel, E. P. (2013). Personalização do E-Learning Baseado no Nível de Aquisição de Conhecimentos do Aprendiz. In: XXIV Simpósio Brasileiro de Informática na Educação (SBIE 2013), 24. Campinas.
- Kitchenham, B. (2004). Procedures for Performing Systematic Reviews. Technical Report TR/SE 0401, Departament of Computer Science Keele University, Keele, Reino Unido.
- Kitchenham, B. e Charters, S. (2007). Guidelines for performing Systematic Literature Reviews in Software Engineering. Technical Report EBSE-2007-01, School of Computer Science and Mathematics, Keele, Reino Unido.
- Müller, L. e Silveira, M. S. (2013). Podes me ajudar? Apoiando a formação de pares em sistemas de ajuda em pares através de técnicas de recomendação. In: XXIV Simpósio Brasileiro de Informática na Educação (SBIE 2013), 24. Campinas.
- Ninoriya, S., Chawan, P. e Meshram, B. (2011). CMS, LMS and LCMS For eLearning. In: International Journal of Computer Science Issues (IJCSI), 8, pp. 644-647.
- Pontes, W. L., França, R. M., Costa, A. P. e Behar, P. (2014). Filtragens de Recomendação de Objetos de Aprendizagem: uma revisão sistemática do CBIE. In: XXV Simpósio Brasileiro de Informática na Educação (SBIE 2014), 25, (pp. 549-558). Dourados.
- Ramos, D. B. e Oliveira, E. H. T. (2015). Relatório Técnico da Revisão Sistemática de Literatura sobre Trilhas de Aprendizagem em Ambientes Virtuais de Aprendizagem. Universidade Federal do Amazonas, Manaus. Fonte: http://mobmoodle.icomp.ufam.edu.br/projeto/?page_id=10
- Silva, R. S. (2011). Moodle para autores e tutores. 2 ed. São Paulo: Novatec.

- Tafner, E. P., Tomelin, J. F. e Müller, R. B. (2012). Trilhas de aprendizagem: uma nova concepção nos ambientes virtuais de aprendizagem – AVA. In: Congresso Internacional de Educação a Distância, 18. São Luís.
- Zem-Lopes, A. M., Pedro, L. Z., Andrade, F. R., Borges, S. S. e Isotani, S. (2013). Tecnologias da Web Semântica em Ambientes Educacionais: uma Revisão Sistemática. In: XXIV Simpósio Brasileiro de Informática na Educação (SBIE 2013), 24, (pp. 977-981). Campinas.