



Criação e análise de perfis de usuários em plataformas educacionais

Trabalho de Conclusão de Curso
Engenharia da Computação

Aluno: Mariana Gomes da Motta Macedo
Orientador: Prof. Dr. Carmelo José Albanez Bastos Filho



Universidade de Pernambuco
Escola Politécnica de Pernambuco
Graduação em Engenharia de Computação

Mariana Gomes da Motta Macedo

**CRIAÇÃO E ANÁLISE DE PERFIS DE USUÁRIOS
EM PLATAFORMAS EDUCACIONAIS**

Monografia apresentada como requisito parcial para obtenção do diploma de Bacharel em Engenharia de Computação pela Escola Politécnica de Pernambuco – Universidade de Pernambuco.

Recife, junho de 2016

MONOGRAFIA DE FINAL DE CURSO

Avaliação Final (para o presidente da banca)*

No dia 21 de 6 de 2016, às 13:00 horas, reuniu-se para deliberar a defesa da monografia de conclusão de curso do discente MARIANA GOMES DA MOTTA MACEDO, orientado pelo professor Carmelo José Albanez Bastos Filho, sob título Criação e análise de perfis de usuários em plataformas educacionais, a banca composta pelos professores:

Alexandre Magno Andrade Maciel

Carmelo José Albanez Bastos Filho

Após a apresentação da monografia e discussão entre os membros da Banca, a mesma foi considerada:

Aprovada Aprovada com Restrições* Reprovada

e foi-lhe atribuída nota: 9,5 (NOVE E MEIO)

*(Obrigatório o preenchimento do campo abaixo com comentários para o autor)

O discente terá 8 dias para entrega da versão final da monografia a contar da data deste documento.

ALEXANDRE MAGNO ANDRADE MACIEL

CARMELO JOSÉ ALBANEZ BASTOS FILHO

O conhecimento é uma ferramenta, e como todas as ferramentas,
o seu impacto está nas mãos de quem o usa. - **Dan Brown**

Agradecimentos

Agradeço pelo conhecimento adquirido em todo o processo acadêmico que originou este trabalho, a Deus e todos os santos que me abençoaram e proveram todas as oportunidades.

Agradeço pela inspiração e dedicação transmitida pelos professores Carmelo Filho, Fernando Buarque, Alexandre Maciel, Bruno Fernandes e Sérgio Oliveira.

Agradeço pelo imensurável amor e apoio dos meus pais, Emilio Macedo e Yelva Sá, das minha avós, Eunice Gomes e Maria Angela Macedo, das minhas irmãs, Luciana Macedo e Ana Paula Macedo, do meu namorado, Leonardo Carréra e dos meus sogros, Carlos Leal e Márcia Carréra.

Agradeço pelos conselhos e experiências transmitidas dos meus amigos Diego Pinheiro, Paulo Roger, Arthur Padilha, David Iseppon, Paula Giestosa, Vanessa Codeceira, Isabela Albuquerque, José Lucas Silva, Felipe Farias e Matheus Dornelas.

Agradeço a todas as pessoas que indiretamente e diretamente auxiliam na minha construção como profissional.

Agradeço especialmente ao Prof. Carmelo José Albanez Bastos Filho pelo excepcional apoio realizado como orientador e a Diego Marconi Pinheiro Ferreira Silva pela paciência de me inspirar a navegar em um mar desconhecido de conhecimento.

Resumo

O intenso uso diário de plataformas educacionais acarreta em alto volume de dados. As necessidades reais dos usuários deixam de ser melhor aperfeiçoadas pois muitas vezes não são percebidas. Por isso, este trabalho utilizou uma base de dados educacional para identificar e caracterizar os perfis dos estudantes. O agrupamento de dados foi realizado pelos algoritmos *K-means* e *C-means*. Além disso, o problema de escolha de número de perfis para esses algoritmos foi resolvido pelo uso da Métrica de Davies-Bouldin e da Estatística do Gap. O uso do algoritmo *K-means* não foi eficiente para o conjunto de dados. No entanto, o algoritmo *C-means* alcançou resultados satisfatórios que foram comprovados pelo Coeficiente de Correlação de Spearman.

Palavras-Chave: Processo de descoberta de conhecimento, Mineração de dados, Agrupamento de dados, *cluster*, dissimilaridades, Métrica de Davies-Bouldin, Estatística do Gap, Coeficiente de Correlação de Spearman.

Abstract

The intense daily use of educational platforms results in high volumes of data. Amassed in this gigantic pile of data, the users' needs remain unnoticed and thus are not enhanced. This study characterized students' profiles based on an educational database. In order to achieve these goals, K-means and C-means algorithms were applied. The number of profiles was chosen according to Davies-Bouldin Metric and Gap Statistic. After experiments, K-means revealed to be inefficient for the database, on the other hand, C-means reached a satisfactory result, sustained by Spearman's rank correlation coefficient.

Keywords: Knowledge Discovery in Databases, Data Mining, Clustering, cluster, dissimilarities, Davies-Bouldin Metric, Gap Statistic, Spearman's rank correlation coefficient.

Sumário

Sumário	i
Lista de Figuras	ii
Lista de Tabelas	iii
1 Introdução	1
1.1 Motivação e caracterização do problema	1
1.2 Objetivo	2
1.2.1 Objetivo específico	3
1.3 Estrutura da monografia	3
2 Referencial Teórico	4
2.1 Processo de Descoberta do Conhecimento	4
2.1.1 Seleção de dados	4
2.1.2 Pré-Processamento de dados	4
2.1.3 Transformação de dados	5
2.1.4 Mineração de dados	5
2.1.5 Interpretação de dados	6
2.2 Agrupamento de dados	6
2.2.1 Técnicas	6
2.2.2 Tipos	7
2.2.3 Medidas de similaridades e dissimilaridades	9
2.2.4 <i>K-means</i>	11
2.2.5 <i>Fuzzy C-means</i>	12

3	Proposta para agrupamento de usuários	15
3.1	Seleção, Pré-processamento e Transformação	16
3.2	Mineração de dados	17
3.2.1	<i>K-means</i>	17
3.2.2	<i>C-means</i>	18
3.3	Interpretação	19
3.3.1	Métricas	19
4	Proposta de perfis de usuários	23
4.1	Análise de convergência entre os algoritmos <i>K-means</i> e <i>C-means</i>	24
4.1.1	Cenário 1: Análise de convergência do grupo com 947 usuários e 17 disciplinas	25
4.1.2	Cenário 2: Análise de convergência do grupo com 947 usuários e 11 disciplinas	32
4.2	Análise da diferença entre as convergências com 11 e 17 disci- plinas utilizando <i>C-means</i> para 6791 usuários	38
4.3	Análise para escolha do número de grupos	42
4.3.1	Cenário 1: Análise para escolha do número de grupos do Subconjunto G947-11 e Subconjunto G947-22	43
4.3.2	Cenário 2: Análise para escolha do número de grupos do Subconjunto G3241-11 e Subconjunto G3241-22	45
4.4	Análise da proposta dos perfis de usuários	47
4.4.1	Cenário 1: Análise da auto-correlação entre as efici- ências das 11 disciplinas dos Subconjunto G947-11 e Subconjunto G947-22	48
4.4.2	Cenário 2: Análise da auto-correlação entre as efici- ências das 11 disciplinas dos Subconjunto G3241-11 e Subconjunto G3241-22	50
5	Considerações Finais	52
	Referências bibliográficas	55
	Apêndice	60

Lista de Figuras

2.1	Dendograma	8
4.1	Gráfico do <i>K-means</i> e <i>C-means</i> utilizando Subconjunto G947-17 com 2 grupos	27
4.2	Gráfico do <i>K-means</i> e <i>C-means</i> utilizando Subconjunto G947-17 com 3 grupos	28
4.3	Gráfico do <i>K-means</i> e <i>C-means</i> utilizando Subconjunto G947-17 com 4 grupos	28
4.4	Gráfico do <i>K-means</i> e <i>C-means</i> utilizando Subconjunto G947-17 com 5 grupos	29
4.5	Gráfico do <i>K-means</i> e <i>C-means</i> utilizando Subconjunto G947-17 com 6 grupos	29
4.6	Gráfico do <i>K-means</i> e <i>C-means</i> utilizando Subconjunto G947-17 com 7 grupos	30
4.7	Gráfico do <i>K-means</i> e <i>C-means</i> utilizando Subconjunto G947-17 com 8 grupos	30
4.8	Gráfico do <i>K-means</i> e <i>C-means</i> utilizando Subconjunto G947-17 com 9 grupos	31
4.9	Gráfico do afastamento entre as dissimilaridades internas e externas no <i>K-means</i>	31
4.10	Gráfico do afastamento entre as dissimilaridades internas e externas no <i>C-means</i>	32
4.11	Gráfico do <i>K-means</i> e <i>C-means</i> utilizando Subconjunto G947-11 com 2 grupos	33

4.12	Gráfico do <i>K-means</i> e <i>C-means</i> utilizando Subconjunto G947-11 com 3 grupos	33
4.13	Gráfico do <i>K-means</i> e <i>C-means</i> utilizando Subconjunto G947-11 com 4 grupos	34
4.14	Gráfico do <i>K-means</i> e <i>C-means</i> utilizando Subconjunto G947-11 com 5 grupos	34
4.15	Gráfico do <i>K-means</i> e <i>C-means</i> utilizando Subconjunto G947-11 com 6 grupos	35
4.16	Gráfico do <i>K-means</i> e <i>C-means</i> utilizando Subconjunto G947-11 com 7 grupos	35
4.17	Gráfico do <i>K-means</i> e <i>C-means</i> utilizando Subconjunto G947-11 com 8 grupos	36
4.18	Gráfico do <i>K-means</i> e <i>C-means</i> utilizando Subconjunto G947-11 com 9 grupos	36
4.19	Gráfico do afastamento entre as dissimilaridades internas e externas no <i>K-means</i>	37
4.20	Gráfico do afastamento entre as dissimilaridades internas e externas no <i>C-means</i>	37
4.21	Gráfico do <i>C-means</i> de 30 simulações e 100 iterações utilizando Subconjunto G6791-17 com 2 e 3 grupos	39
4.22	Gráfico do <i>C-means</i> de 30 simulações e 100 iterações utilizando Subconjunto G6791-17 com 4 e 5 grupos	39
4.23	Gráfico do <i>C-means</i> de 30 simulações e 100 iterações utilizando Subconjunto G6791-17 com 6 e 7 grupos	40
4.24	Gráfico do <i>C-means</i> de 30 simulações e 100 iterações utilizando Subconjunto G6791-17 com 8 e 9 grupos	40
4.25	Gráfico do <i>C-means</i> de 30 simulações e 100 iterações utilizando Subconjunto G6791-11 com 2 e 3 grupos	41
4.26	Gráfico do <i>C-means</i> de 30 simulações e 100 iterações utilizando Subconjunto G6791-11 com 4 e 5 grupos	41
4.27	Gráfico do <i>C-means</i> de 30 simulações e 100 iterações utilizando Subconjunto G6791-11 com 6 e 7 grupos	42

4.28	Gráfico do <i>C-means</i> de 30 simulações e 100 iterações utilizando Subconjunto G6791-11 com 8 e 9 grupos	42
4.29	Gráfico das dissimilaridades externas do <i>C-means</i> utilizando 30 simulações, 2000 iterações e grupos de 2 a 5 utilizando Subconjunto G947-11 e Subconjunto G947-22	44
4.30	Gráfico da métrica de Davies-Bouldin do <i>C-means</i> utilizando 30 simulações, 2000 iterações e grupos de 2 a 5 utilizando Subconjunto G947-11 e Subconjunto G947-22	44
4.31	Gráfico da Estatística do Gap do <i>C-means</i> utilizando 30 simulações, 2000 iterações e grupos de 2 a 5 utilizando Subconjunto G947-11 e Subconjunto G947-22	45
4.32	Gráfico das dissimilaridades externas do <i>C-means</i> utilizando 30 simulações, 2000 iterações e grupos de 2 a 5 utilizando Subconjunto G3241-11 e Subconjunto G3241-22	46
4.33	Gráfico da métrica de Davies-Bouldin do <i>C-means</i> utilizando 30 simulações, 2000 iterações e grupos de 2 a 5 utilizando Subconjunto G3241-11 e Subconjunto G3241-22	46
4.34	Gráfico da Estatística do Gap do <i>C-means</i> utilizando 30 simulações, 2000 iterações e grupos de 2 a 5 utilizando Subconjunto G3241-11 e Subconjunto G3241-22	47
5.1	Autocorrelação do perfil com notas mais elevadas encontrados pelo grupo de 3241 usuários com 11 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Começo do Bloco	61
5.2	Autocorrelação do perfil com notas mais elevadas encontrados pelo grupo de 3241 usuários com 11 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Continuação do Bloco	62
5.3	Autocorrelação do perfil com notas mais baixas encontrados pelo grupo de 3241 usuários com 11 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Começo do Bloco	63

5.4	Autocorrelação do perfil com notas mais baixas encontrados pelo grupo de 3241 usuários com 11 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Continuação do Bloco	64
5.5	Autocorrelação do perfil com notas mais elevadas encontrados pelo grupo de 3241 usuários com 22 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Começo do Bloco	65
5.6	Autocorrelação do perfil com notas mais elevadas encontrados pelo grupo de 3241 usuários com 22 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Continuação do Bloco	66
5.7	Autocorrelação do perfil com notas mais baixas encontrados pelo grupo de 3241 usuários com 22 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Começo do Bloco	67
5.8	Autocorrelação do perfil com notas mais baixas encontrados pelo grupo de 3241 usuários com 22 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Continuação do Bloco	68
5.9	Autocorrelação do perfil com notas mais elevadas encontrados pelo grupo de 947 usuários com 11 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Começo do Bloco	69
5.10	Autocorrelação do perfil com notas mais elevadas encontrados pelo grupo de 947 usuários com 11 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Continuação do Bloco	70
5.11	Autocorrelação do perfil com notas mais baixas encontrados pelo grupo de 947 usuários com 11 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Começo do Bloco	71

5.12	Autocorrelação do perfil com notas mais baixas encontrados pelo grupo de 947 usuários com 11 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Continuação do Bloco	72
5.13	Autocorrelação do perfil com notas mais baixas encontrados pelo grupo de 947 usuários com 22 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Começo do Bloco	73
5.14	Autocorrelação do perfil com notas mais baixas encontrados pelo grupo de 947 usuários com 22 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Continuação do Bloco	74
5.15	Autocorrelação do perfil com notas mais elevadas encontrados pelo grupo de 947 usuários com 22 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Começo do Bloco	75
5.16	Autocorrelação do perfil com notas mais elevadas encontrados pelo grupo de 947 usuários com 22 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Continuação do Bloco	76
5.17	Correlação entre os dois perfis encontrados pelo grupo de 3241 usuários com 11 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Começo do Bloco	77
5.18	Correlação entre os dois perfis encontrados pelo grupo de 3241 usuários com 11 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Continuação do Bloco	78
5.19	Correlação entre os dois perfis encontrados pelo grupo de 3241 usuários com 22 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Começo do Bloco	79
5.20	Correlação entre os dois perfis encontrados pelo grupo de 3241 usuários com 22 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Continuação do Bloco	80

5.21	Correlação entre os dois perfis encontrados pelo grupo de 947 usuários com 11 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Começo do Bloco	81
5.22	Correlação entre os dois perfis encontrados pelo grupo de 947 usuários com 11 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Continuação do Bloco	82
5.23	Correlação entre os dois perfis encontrados pelo grupo de 947 usuários com 22 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Começo do Bloco	83
5.24	Correlação entre os dois perfis encontrados pelo grupo de 947 usuários com 22 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Continuação do Bloco	84

Lista de Tabelas

3.1	Símbolos utilizados pelas equações do Capítulo 3.	16
4.1	Subconjuntos utilizados para os experimentos no Capítulo 4. .	24
4.2	Tabela demonstrativa das quantidades de correlações encontradas em intervalos de valores no Perfil 0 e 1 Subconjunto G947-11 e Subconjunto G947-22	49
4.3	Correlações mais altas entre Disciplinas encontrados pelo Perfil 1 (Notas mais elevadas) pelo Subconjunto G947-11 e Subconjunto G947-22	49
4.4	Correlações mais baixas entre Disciplinas encontrados pelo Perfil 1 (Notas mais elevadas) pelo Subconjunto G947-11 e Subconjunto G947-22	50
4.5	Tabela demonstrativa das quantidades de correlações encontradas em intervalos de valores no Perfil 0 e 1 pelo Subconjunto G3241-11 e Subconjunto G3241-22	51
4.6	Correlações mais altas entre Disciplinas encontrados pelo Perfil 1 (Notas mais elevadas) pelo Subconjunto G3241-11 e Subconjunto G3241-22	51

Capítulo 1

Introdução

1.1 Motivação e caracterização do problema

O primeiro registro brasileiro de ensino à distância foi publicado pelo Jornal do Brasil em 1904, oferecendo curso de datilografia por correspondência. Nas décadas de 1970 e 1980, iniciativas privadas e sem participações governamentais ofereceram cursos a distância via satélite. No entanto, somente em 1996 foi criada a Secretaria de Educação à Distância (SEED) pelo Ministério da Educação, colaborando para oficializar a educação à distância. Com esta regulamentação, a partir de 2000 houve o surgimento de novas iniciativas de ensino à distância. Posteriormente, foram regulamentados os cursos superiores de graduação independente das modalidades presenciais ou à distância e suas diretrizes e bases [1][2].

Com o desenvolvimento da computação e da Internet, surgiu o conceito de educação *online*, vindo do inglês, *E-learning*, ou seja, utilização de suporte tecnológico como objeto de aprendizagem que permite ao aluno visualizar e participar mais do conteúdo. Incorporando a ideia de educação à distância e *E-learning*, os Sistemas de Gestão da Aprendizagem (LMS, do inglês *Learning Management Systems*), são utilizados como ambiente de compartilhamento e comunicação de conteúdos. Os Ambientes Virtuais de Aprendizagem (AVA) aproximam as relações entre professores e alunos via Internet nos LMS [3]. Outro tipo de sistema são os Ambientes Pessoais de Aprendizagem (APA),

do inglês *Personal Learning Environment* (PLE), cujo objetivo é oferecer conteúdo personalizado para seus usuários controlarem seu aprendizado [4] [5]. Todos esses sistemas diariamente geram dados sobre mudanças na plataforma, usuários e conteúdos originando um banco de dados complexo. Diante de tantos dados, as necessidades reais e hipotéticas passam despercebidas. No entanto, as plataformas educacionais objetivam satisfazer as necessidades de estudo do usuário. Por isso, é importante identificar as características da plataforma que melhor auxiliam no bom desempenho dos estudantes. Além disso, a personalização do conteúdo de acordo com o nível do estudante pode motiva-los.

Portanto, o projeto se propõe a utilizar o Processo de Descoberta do Conhecimento (KDD, do inglês *Knowledge-Discovery in Databases*), para encontrar conhecimento em bancos de dados [6]. O KDD seleciona os dados que são de interesse de análise, pré-processa e os transforma para evitar ruídos e ambiguidades, posterior Mineração de dados, (DM, do inglês *Data Mining*), para que sejam processados e, por último, interpreta e avalia resultados para conclusão do conhecimento.

Na Mineração de dados é possível escolher diversos métodos para cada problema [7] [8] [9]. Assim, foi escolhido Agrupamento de dados (do inglês *Clustering*), tendo por objetivo de rotular os usuários de acordo com suas características, usando o mesmo rótulo para indivíduos com características semelhantes, caracterizando um grupo (*cluster*) [10] [11]. Esta divisão objetiva identificar as dificuldades em comum de alunos que possuem desempenho acadêmico insuficiente. Com este diagnóstico sobre os estudantes, torna-se possível modificar o perfil utilizado pelo usuário no sistema para que seu desempenho melhore ao longo do processo de formação.

1.2 Objetivo

Este trabalho tem como principal objetivo identificar e implementar as técnicas de Agrupamento de dados, K-means e C-means, escolhidas para Mineração de dados educacionais, bem como comparar o desempenho destas técnicas aplicadas ao problema de definição de perfil dos usuários. Além

disso, este trabalho visa relatar as decisões tomadas no KDD.

1.2.1 Objetivo específico

Para atingir os objetivos principais, foram alcançadas os seguintes objetivos específicos:

- Pesquisar, implementar, aplicar e analisar métricas para determinação da escolha do número de grupos que melhor separa os perfis de usuários;
- Pesquisar, aplicar e analisar métricas de correlação entre as disciplinas dos usuários da plataforma educacional para identificar as relações encontradas nos perfis de usuários;
- Identificar as características em comum dos estudantes da plataforma educacional para cada agrupamento de usuários.

1.3 Estrutura da monografia

A monografia esta dividida em 5 capítulos. No Capítulo 2, a fundamentação teórica aborda todo o processo utilizado para descoberta de conhecimento, suas etapas, seus algoritmos e métodos e as métricas utilizadas. No Capítulo 3, as especificidades do trabalho no KDD são evidenciadas. Em seguida, no Capítulo 4 são apresentados os resultados e análises estatísticas resultantes dos experimentos. Concluindo no Capítulo 5, as considerações finais serão expostas.

Capítulo 2

Referencial Teórico

2.1 Processo de Descoberta do Conhecimento

O Processo de Descoberta do Conhecimento investiga relevância e magnitude em volumes de dados que crescem rapidamente com o uso diário de sistemas tecnológicos [12] [13]. O KDD contempla como os dados são armazenados, acessados, processados, extraídos, modelados e analisados. Com isso, este processo se divide em 5 etapas que serão explicadas nas Subseções 2.1.1, 2.1.2, 2.1.3, 2.1.4 e 2.1.5 a seguir.

2.1.1 Seleção de dados

Na primeira etapa, Seleção, os dados devem ser escolhidos dentre todo o conjunto utilizado pelo sistema. Esta escolha é realizada com a ajuda de especialistas nas bases de dados. Este subconjunto escolhido deve conter as principais características dos sistemas que serão avaliadas para encontrar soluções aos problemas pré-definidos.

2.1.2 Pré-Processamento de dados

A etapa de Pré-processamento procura identificar e analisar variáveis que possuem valores discrepantes, ruídos e informações ausentes ou cronológicas. A importância dessa etapa está na remoção de possíveis ambiguidades

nas etapas posteriores. Portanto, valores que não caracterizam a realidade do sistema são removidos para que não prejudique o evidenciamento de características relevantes.

2.1.3 Transformação de dados

Cada problema e sistema possui suas especificidades. As peculiaridades precisam ser representadas da forma mais adequada. A fase de transformação aglutina dados via médias, medianas, normalizações ou funções, e, reduz variáveis desnecessárias. Nesta etapa portanto são substituídas as representações dos dados para que tenham a coerência necessária ao problema.

2.1.4 Mineração de dados

A mineração de dados (DM, *Data Mining*) é a etapa responsável por processar os dados para alcançar os objetivos pré-definidos, ou seja, estabelecer os métodos para busca de padrões [14] [15] . Os modelos podem ser preditivo ou descrito, ou seja, respectivamente podem objetivar inferência ou caracterização. As tarefas mais populares na literatura que podem ser utilizadas são :

- i Classificação: mapeamento de dados para categorias pré-definidas modelando uma função para todo o conjunto e novos dados;
- ii Regressão: previsão de variáveis dependentes com base em independentes;
- iii Agrupamento: distinguir classes ou grupos que caracterizem subconjuntos;
- iv Sumarização: encontrar uma descrição compacta para conjuntos de dados;
- v Modelagem dependente: identificar relações entre variáveis;
- vi Detecção de mudanças e desvios: constatar modificações significantes.

2.1.5 Interpretação de dados

Após os dados serem processados pela fase de mineração, é necessário compreender e explorar os resultados encontrados. A interpretação dos mesmos esclarece a necessidade de reiteração ou não das fases anteriores. Portanto, nesta etapa são elaborados arquivos para documentar todo o processo.

2.2 Agrupamento de dados

O Agrupamento de dados (*Clustering*) é determinado pela relação de similaridade e/ou dissimilaridade. Dados com padrões semelhantes tendem a pertencer ao mesmo grupo. No entanto, dados distintos, com dissimilaridade mais alta, devem estar classificados diferentemente. Essa técnica de *Data Mining* é um processo de aprendizado não supervisionado. Isso é verdade pela ausência de classes que redirecionem os resultados, ou seja, não há resposta pré-definida. Portanto, o agrupamento de dados encontra o número ótimo de grupos e as características de cada um. Um grupo pode ser denominado como *cluster*, expressão em inglês, porém muito utilizada na literatura brasileira.

2.2.1 Técnicas

O Agrupamento de dados possui uma série de técnicas. Cada técnica detém um conjunto de aspectos que podem ser encontrados na literatura:

- i. Aglomerativo ou divisionista: A aglomeração é evidenciada pela junção de grupos semelhantes a partir de *clusters* menores. Em contra partida, o algoritmo divisionista possui grandes grupos que serão divididos enquanto houver dissimilaridade entre os mesmos.
- ii. Monotético ou politético: Monoteticamente, o algoritmo processa um elemento da estrutura de dados por vez. Politeticamente, são utilizadas múltiplas características por iteração.

- iii. Clássico ou *fuzzy* : Um dado pode pertencer a um só grupo ou relacionar-se a vários grupos por graus de pertinência, respectivamente, clássico ou *fuzzy*. Além disso, é possível combinar os dois aspectos. Primeiramente, utiliza-se o relacionamento com vários grupos ao mesmo tempo. Ao término da técnica pode-se utilizar o aspecto clássico para determinar um grupo de maior similaridade para cada dado.
- iv. Determinístico ou estocástico: O método determinístico utiliza técnicas tradicionais como médias e medianas para escolha e mudança de variáveis. Estocástico, uso de variáveis aleatórias, atua na inicialização ou pesquisa de espaço da técnica. Esses aspectos são relevantes na minimização da taxa de erro, pois influenciam nos valores encontrados em curto prazo.
- v. Incremental ou Não-Incremental: O reuso de um algoritmo pode ser realizado n vezes. O não reuso, ou seja, o uso de apenas uma iteração do algoritmo determina o aspecto não incremental.

2.2.2 Tipos

De acordo com a literatura científica, os tipos de algoritmos de *Clustering* são:

- i. Algoritmo de agrupamento hierárquico (*Hierarchical Clustering Algorithms*): produz uma representação em dendograma, exposto na Figura 2.1. Nesta imagem é possível visualizar níveis. Cada nível possui sua similaridade. A altura da mesma significa que quanto mais alto menor a similaridade. Portanto, deve-se definir o quão similar é necessário para a resolução dos objetivos. Os exemplos deste tipo de algoritmo são *Maximum or complete-linkage clustering* [16] e *Minimum or single-linkage clustering* [17].

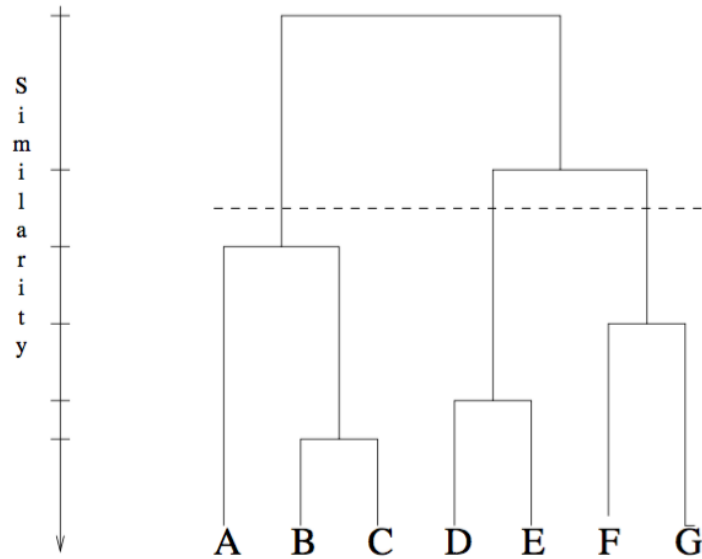


Figura 2.1: Dendrograma

Fonte: [[14]]

- ii. Algoritmo baseado em particionamento (*Partitional Algorithms*): a divisão de grupos é estipulada pela otimização de uma função definida por características locais ou globais. Esse tipo de algoritmo é reiterado por n vezes para a escolha da melhor configuração. Pela ausência do uso de estruturas como o dendrograma, é considerado uma ótima escolha para problemas com número de dados elevado. *K-means* e *Squared Error Clustering* são exemplos deste tipo de algoritmo [18].
- iii. Algoritmo de resolução mista e modo de busca (*Mixture-Resolving and Mode-Seeking Algorithms*): construção de distribuições a partir dos padrões encontrados. O objetivo é a identificação parâmetros e seus valores. Muitos trabalhos utilizam a distribuição Gaussiana. O algoritmo mais utilizado na literatura deste tipo é *Expectation Maximization* [19].
- iv. Agrupamento dos vizinhos mais próximos (*Nearest Neighbor Clustering*): junção de padrões dos vizinhos mais próximos semelhantes. Neste tipo de algoritmo os grupos crescem ao redor de seus vizinhos [20].

- v. Agrupamento utilizando lógica difusa (*Fuzzy Clustering*): cada padrão pertencerá a todos os grupos por graus de pertinências. Portanto, cada dado dispõe uma porcentagem de similaridade para cada grupo. Um exemplo de algoritmo é o *C-means* [21].
- vi. Representação de grupos (*Representation of Clusters*): uso de estruturas para análise da composição de todo conjunto. O uso de centroides para identificar grupos é a mais utilizada [22].
- vii. Redes Neurais Artificiais (*Artificial Neural Networks*): uso de interligações dos neurônios artificiais para aprendizado com sucessão de padrões. A susceptibilidade aos erros é dada pela escolha de inicialização dos parâmetros necessários à rede. Uma das redes neurais artificiais citadas na literatura é Mapas Auto-Organizáveis (*Self Organizing Maps*) [23].
- viii. Algoritmos evolucionários (*Evolutionary Algorithms*): uso de operadores evolucionários com uma população de soluções para obtenção do ótimo global. Os operadores mais utilizados são o de seleção, recombinação e mutação. Enquanto não satisfizer um critério de parada, haverá eleição, fusão e modificação dos melhores resultados [24].
- ix. Algoritmos de busca (*Search-Based Algorithms*): combinação de ótimas referências locais encontradas, soluciona globalmente o problema [25].

2.2.3 Medidas de similaridades e dissimilaridades

O estudo do agrupamento de dados utiliza a similaridade entre os dados para agrupá-los e a dissimilaridade entre grupos para manter distante os dados que possuem distinções mais altas. Por isso, é importante compreender que é possível analisar somente a similaridade ou apenas a dissimilaridade entre dados e grupos. O uso da ideia de similar ou dissimilar são opcionais quanto a nomenclatura se for utilizada com o mesmo objetivo, ou seja, objetivando a minimização da dissimilaridade interna e a maximização da dissimilaridade externa ou a maximização da similaridade interna e a minimização da similaridade externa. Portanto, a determinação de similaridade

ou dissimilaridade é fundamental para o agrupamento de dados [26]. Essa determinação é representada pela distância entre dados e deve cumprir as seguintes propriedades:

- i. A distância entre dois dados deve ser finita e não negativa;
- ii. Se a distância entre dois dados é igual a zero, então os dados são iguais;
- iii. Se a distância de um dado x a um dado y for igual a distância contrária, então os dados são simétricos;
- iv. Se a distância entre dois dados for menor ou igual ao somatório das distâncias desses dois a um dado entre os mesmos, então há uma desigualdade triangular;
- v. O cálculo da similaridade não pode ter interações externas;
- vii. O resultado das distâncias precisam ter significância independentemente do número de dimensões.

Obedecendo essas as propriedades, algumas métricas podem ser utilizadas para calcular a dissimilaridade utilizando x_i e x_j como dois pontos quaisquer. A métrica Euclidiana é calculada pela Equação (2.1). Sua utilização é adequada para dados que possuem escalas próximas. O uso de variáveis com escalas muito diferentes ocasiona em uma influência maior de valores mais altos. Além disso, a distância Euclidiana é um caso especial em que o p é igual a dois para a distância de Minkowski na Equação (2.2). O aumento dos valores de p aumentam o intervalo da distância de Minkowski.

$$d(x_i, x_j) = \sqrt{\sum_{p=1}^{dim} (x_{i,p} - x_{j,p})^2}, \quad (2.1)$$

$$d(x_i, x_j) = \sqrt[p]{\sum_{p=1}^{dim} (x_{i,p} - x_{j,p})^p}. \quad (2.2)$$

A distância Manhattan calcula a diferença absoluta entre dois pontos como apresentado na Equação (2.3) sendo mais utilizada quando é necessário escalas diferentes.

$$d(x_i, x_j) = \sum_{p=1}^{dim} |x_{i,p} - x_{j,p}| \quad (2.3)$$

2.2.4 *K-means*

O algoritmo *K-means* permanece popular pela sua simplicidade, eficiência computacional, baixo consumo de memória e bom desempenho em volume elevado de dados [27] [28]. As versões que impulsionaram o uso dessa ferramenta foram desenvolvidas por Forgy [29] e Lloyd [30], MacQueen [31] e Hartigan e Wong [32].

Versão clássica do algoritmo *K-means*

Os resultados encontrados pelo *K-means* são otimizados pela minimização da Equação (2.4) em que k é o número de grupos, n_k é o número de padrões no *cluster* k , c_k é o centro do *cluster* k , i é a posição do padrão no vetor de padrões, x_i é o padrão na posição i e K é o número máximo de padrões.

$$\epsilon = \sum_{k=1}^K \sum_{i=1}^{n_k} |x_i^k - c_k|^2. \quad (2.4)$$

Todo o processo do *K-means* pode ser resumido pelo Algoritmo 1. A entrada para o algoritmo são o vetor de dados com suas dimensões e os números de grupos para particionar, e sua saída é um vetor de dados agrupados. O primeiro passo do algoritmo é a inicialização dos centroides randomicamente. Os passos 2, 3 e 4 são iterados até que o passo 4 satisfaça o critério de parada. A segunda etapa calcula para cada padrão a distância entre o mesmo e os centros. O terceiro passo atualiza cada centroide pela Equação 2.5.

$$c_k = \frac{1}{n_k} \sum_{i=1}^{n_k} x_i^k \quad (2.5)$$

Algoritmo 1: Pseudo-Algoritmo Clássico K-means

- 1: **Entrada:** Vetor de i dados com j dimensões
 - 2: e número máximo de agrupamentos;
 - 3: **Saída:** Vetor com os dados agrupados
 - 4: **Passo 1:** Inicialização randômica dos centros c_k dos c grupos;
 - 5: **Passo 2:** Atribuição de grupo com menor distância para cada x_i ;
 - 6: **Passo 3:** Cálculo de novos centros na Equação (2.5);
 - 7: **Passo 4:** Se critério de parada não estiver satisfeito **então**
 - 8: volte para o Passo 2.
-

2.2.5 *Fuzzy C-means*

O algoritmo *Fuzzy C-means* (FCM) foi escolhido para este trabalho pelas características semelhantes ao *K-means*, porém com a vantagem da inspiração em Lógica Nebulosa (*Fuzzy Logic*) [33]. O *K-means* não estabelece bem os agrupamentos para dados que estão em fronteiras de grupos. Por isso, o *FCM* será utilizado principalmente por agrupar dados pela pertinência os grupos, sendo assim mais de acordo com a realidade.

O *Fuzzy C-means* foi idealizado por James Bezdek [34] e aprimorado juntamente com Robert Ehrlich e William Full [35]. Este algoritmo é utilizado em diferentes problemas como processamento de imagem [36], baseado em estratégias diferentes como a Entropia [37] e com aprimoramentos diversos como inicializações especiais [38]. Portanto, cada problema terá suas especificidades adequadas.

Versão Clássica do Algoritmo C-means

A cada iteração do algoritmo clássico de *Fuzzy C-means*, a Equação (2.6) será minimizada onde N é o número de dados máximo, K é o número máximo de grupos, μ_{ij} é a pertinência do exemplo i com grupo j , x_i é o exemplo na posição i no vetor do conjunto todo de dados (X) e c_j é o centro do grupo j .

$$J_m = \sum_{i=1}^N \sum_{j=1}^K \mu_{ij}^m \|x_i - c_j\|^2. \quad (2.6)$$

A soma das pertinências de um exemplo com n grupos deve ser igual a 1. Além disso, o valor individual de cada pertinência em um grupo de um exemplo deve ser no intervalo de 0 a 1.

O *C-means* inicializa randomicamente a matriz de pertinência μ . Enquanto o critério de parada não for satisfeito, o algoritmo continuará recalculando os centroides e a matriz de pertinência. O pseudo-algoritmo clássico do *C-means* está exposto no Algoritmo 2. Esse algoritmo possui como entrada um vetor com os dados em suas dimensões e outro vetor com os números de grupos para partição dos dados. A saída do *C-means* é o melhor agrupamento identificado pela diferença entre as matrizes de pertinências e os dados agrupados. Os centroides dos grupos são determinados pelo somatório da multiplicação das inclusões dos exemplos divididos pelo somatório de suas inclusões como apresentado na Equação (2.7). A atualização das pertinências devem conter apenas números menores que um. Por isso, na Equação (2.8) dividi-se 1 pelo resultado do somatório das dissimilaridades entre grupos elevados a uma razão que representará o tamanho da mudança do algoritmo. O coeficiente de *Fuzziness* (m) é responsável pela velocidade e o quão detalhista será o algoritmo. Quanto menor o valor de m , mais lento será o algoritmo porém mais detalhista. O valor de m deve ser escolhido com cuidado, pois um valor alto não é capaz de encontrar mínimos globais e sim locais.

$$c_j = \frac{\sum_{i=1}^N \mu_{ij}^m \cdot x_i}{\sum_{i=1}^N \mu_{ij}^m}, \quad (2.7)$$

$$\mu_{ij}^m = \frac{1}{\sum_{k=1}^C \frac{\|x_i - c_j\|^{\frac{2}{m-1}}}{\|x_i - c_k\|^{\frac{2}{m-1}}}}. \quad (2.8)$$

Algoritmo 2: Pseudo-Algoritmo Clássico C-means

- 1: **Entrada:** Vetor de i dados com j dimensões
 - 2: e vetor dos números de agrupamentos;
 - 3: **Saída:** Vetor com os dados agrupados e o número de grupos escolhido
 - 4: **Passo 1:** Inicialização randômica da matriz de pertinência de c grupos;
 - 5: **Passo 2:** Cálculo de novos centros na Equação (2.7);
 - 6: **Passo 3:** Cálculo das pertinências na Equação (2.8);
 - 7: **Passo 4:** **Se** critério de parada não estiver satisfeito **então**
 - 8: volte para o Passo 2
-

Capítulo 3

Proposta para agrupamento de usuários

Em uma plataforma educacional existem vários tipos e características de usuários. A hipótese é de que o uso diário na mesma acarrete em uma divisão de perfis desconhecida *a priori*. Este trabalho iniciou o estudo sobre usuários, tipos e características da plataforma educacional cedida em confidencialidade. Essa plataforma complexa possui várias funcionalidades que objetivam ensinar aos alunos conteúdos específicos de exames e provas. Para este trabalho foi escolhido analisar somente a funcionalidade de resolução das questões para que fosse mapeado o comportamento dos alunos para uma provável eficiência nos exames. Assim, uma vez agrupados os usuários é possível analisar quais funcionalidades os agrupamentos utilizam e como. Porém, para esse trabalho, as análises das outras funcionalidades não podem ser publicadas. Por isso, todo processo de análise será voltado para resolução de questões e posteriormente são analisados as correlações entre disciplinas e perfis de usuários encontrados.

Neste capítulo é explicado todo o processo realizado para a descoberta de perfis de usuários utilizando os algoritmos *K-means* e *C-means* na etapa de Mineração de Dados no KDD. Além disso, é explicado quais métricas são utilizadas para escolha da quantidade de perfis de usuários. Para todas as seções, os símbolos em comum utilizados, seus significados e intervalos são

descritos na tabela 3.1.

Tabela 3.1: Símbolos utilizados pelas equações do Capítulo 3.

Símbolo	Significado	Intervalo/Tamanho
i	identificador da posição do usuário	$[0, \infty]$
r	quantidade de grupos	$[1, \infty]$
R	número máximo de grupos	$[1, \infty]$
N_r	número total de usuários no grupo r	$[0, \infty]$
N_d	número total de distâncias	$[0, \infty]$
J	número de características (dimensões)	$[0, \infty]$
X	vetor de todos os usuários	$[0, \infty]$
Z	vetor dos usuários classificados	$[0, \infty]$
x_{i_j}	usuário na posição i da dimensão j	$[0, 1]$
c_{r_j}	centro do grupo r da dimensão j	$[0, 1]$

3.1 Seleção, Pré-processamento e Transformação

Os dados de todos os usuários foram selecionados com as respectivas quantidades de questões feitas e as respondidas corretamente de todas as disciplinas. O número de usuários chega a mais de 300 mil e a plataforma educacional possui 17 disciplinas. Selecionado esses números, cada usuário teve sua eficiência estabelecida pela divisão das questões corretas pelo número total das respondidas, onde esse valor já está normalizado, pois possui intervalo entre 0 e 1.

Devido à divergência detectada em uma das iterações do KDD, foi necessário remover os usuários que não obtinham uma frequência de interação na plataforma. Os alunos que não tinham respondido pelo menos a média de questões requeridas em exames relacionados ao conteúdo de todas as disciplinas foram descartados.

3.2 Mineração de dados

Como o objetivo é encontrar grupos de usuários com semelhanças, então a tarefa adequada é Agrupamento de dados e os algoritmos escolhidos foram *K-means* e *C-means* pelas vantagens de serem propícios para volume de dados elevado, de simples utilização e populares na área educacional [27] [28]. Para escolher o melhor algoritmo entre *K-means* e *C-means*, foram realizados testes de convergência para encontrar qual algoritmo agrupava dados mais similares entre si e mais dissimilares entre grupos diferentes. Na Seção 3.2.1 e 3.2.2 será explicado como foi utilizado respectivamente *K-means* e *C-means*.

3.2.1 *K-means*

O *K-means* utilizado pelo estudo foi implementado na linguagem *Python*. O algoritmo implementado utiliza o paralelismo para economia de tempo. No intervalo de 2 a 9 grupos, os elementos são reclassificados para os centros mais próximos. As distâncias trabalhadas foram as euclidianas pela ausência de escalas diferentes. Todos os dados estão normalizados entre 0 e 1.

Em relação ao algoritmo, enquanto o centro atual não for idêntico ao anterior ou não iterar um número máximo, os centros serão recalculados assim como as distâncias e classes. A classificação de um elemento é feita para o centro mais próximo. O novo centro é computado pela Equação (3.1) onde R é o valor máximo de grupos, N_r é o número total de elementos no grupo r , i é a posição do elemento no vetor de elementos X do grupo r com d características. O pseudo-algoritmo do processo realizado pelo *K-means* é mostrado no Algoritmo 3.

A inicialização randômica foi escolhida pela simplicidade e bom desempenho descritos por Peña, Lozano e Larrañaga [39]. Apesar do uso da técnica de Kaufman (KA) [40], obter resultados superiores aos randômicos sua complexidade é superior tornando não interessante ao problema.

$$\text{novoCentro} = \sum_{r=2}^R \frac{1}{N_r} \sum_{i \in X_r} x_{i_d}. \quad (3.1)$$

Algoritmo 3: Pseudo-Algoritmo K-means

- 1: **Entrada:** $\mathbf{X} = \{x_i\}_{i=1}^N$, \mathbf{k} grupos;
 - 2: **Saída:** $\mathbf{Z} = \{z_k\}_{k=1}^K$;
 - 3: Em paralelo, execute de 2 a \mathbf{k} grupos;
 - 4: Inicialização randômica dos centros de \mathbf{k} grupos;
 - 5: *Enquanto* novoCentro \neq centroAntigo ou numeroMaximoIteracoes $<$ numeroIteracoes :
 - 6: centroAntigo = novoCentro;
 - 7: distancias = calcularDistancias(novoCentro);
 - 8: classes = classificar(distancias);
 - 9: novoCentro = calcularCentro(classes).
-

3.2.2 C-means

O *C-means* também foi paralelizado pelos mesmos motivos que o *K-means*. No Algoritmo 4 está descrito o processo de funcionamento do pseudo-código. Para cada número de grupo, enquanto a quantidade máxima de iterações não for alcançado será recalculado centro e matriz de pertinência dos elementos. Antes de entrar no *loop*, os centroides são inicializados randomicamente e a partir dos mesmos será calculado as pertinências correspondentes. O cálculo do novo centro está na Equação (2.7) e o da matriz de pertinências está na Equação (2.8) demonstrados no Capítulo 2. Não foi utilizado o critério de parada das diferenças das matrizes de pertinências pois o objetivo é analisar a melhor configuração para todos os números de grupos.

Algoritmo 4: Pseudo-Algoritmo C-means

- 1: **Entrada:** $\mathbf{X} = \{x_i\}_{i=1}^N$, \mathbf{c} grupos;
 - 2: **Saída:** $\mathbf{Z} = \{z_c\}_{c=1}^K$;
 - 3: Inicialização randômica dos centros de \mathbf{c} grupos;
 - 4: Atualização da matriz de pertinência;
 - 5: Em paralelo, execute de 2 a \mathbf{c} grupos:
 - 6: *Enquanto* numeroMaximoIteracoes $<$ numeroIteracoes:
 - 7: novoCentro = calcularCentro(matrizPertinencia);
 - 8: matrizPertinencia = atualizarMatriz(novoCentro);
-

3.3 Interpretação

Após os agrupamentos dos dados realizados pela fase da Mineração de dados, é necessário analisar os resultados adquiridos e identificar qual a quantidade adequada de perfis de usuários para a base utilizada. Na Seção 3.3.1 são explicadas as métricas utilizadas.

3.3.1 Métricas

Como a tarefa de agrupamento tem como principal objetivo a dissimilaridade baixa interna e alta externa, é analisado primeiramente se isso está sendo cumprido pelo algoritmo. Além disso, a medida de separação criada por Davies e Bouldin será utilizada por indicarem grupos esféricos como resposta [41]. Por fim, a estatística de Gap é aplicada por demonstrar eficiência na escolha de *cluster* em base de dados diversas encontradas na literatura.

Dissimilaridades euclidianas internas e externas

A dissimilaridade euclidiana interna é calculada como mostra a Equação (3.2). Essa equação resulta no somatório de todas as distâncias euclidianas entre dados e centroides ($d_{ii'}$) divididos pelo número total de dados no grupo r . Internamente, o agrupamento deve conter elementos próximos ao valor do centroide. Por isso, o cálculo soma todas as distâncias entre elementos e o centro do grupo a que pertence procurando o resultado que minimize essa equação.

$$d_{in} = \frac{1}{N_r} \sum_{i,i' \in C_r}^{N_r} d_{ii'} \quad (3.2)$$

Apresentadas na Equação (3.3), as distâncias externas calculadas entre os centros $dc_{ii'}$ são divididas pelo número de total das distâncias encontradas N_d . Diferentemente da dissimilaridade interna, as dissimilaridades entre grupos deve ser a maior possível.

$$d_{out} = \frac{1}{N_d} \sum_{i,i' \in C_r}^{N_r} dc_{ii'} \quad (3.3)$$

Davies-Bouldin

A medida de separação proposta por Davies-Bouldin (db) engloba os dois cálculos interno e externo das dissimilaridades em busca de encontrar em apenas uma função a melhor opção de escolha para o número de *clusters* [41] [42]. A métrica cumpre as seguintes regras:

- i. A similaridade deve ser não negativa;
- ii. A propriedade de simetria deve ser verdadeira;
- iii. O aumento das distâncias entre *clusters* sem mudança na dispersão ocasiona na diminuição de similaridade entre os grupos;
- iv. O aumento da dispersão dos *clusters* sem mudança das distâncias entre eles ocasiona no aumento de similaridade entre os grupos;

Para a métrica de Davies-Bouldin, a dissimilaridade entre grupos é calculada pela distância entre dois centroides demonstrada na Equação 3.4. A dissimilaridade em um grupo, na Equação 3.5, é escrita pela divisão do somatório de todas as distâncias euclidianas entre cada ponto (x_i) e seu centroide (c_r) com o número de quantidade de dados no grupo r . O expressão matemática $\|\cdot\|$ representa a norma euclidiana dos dois elementos em suas dimensões.

$$d_{ce}(l, r) = \|c_r - c_l\|. \quad (3.4)$$

$$S_c(r) = \frac{\sum_i \|x_i - c_r\|}{N_r}, \quad (3.5)$$

A relação entre as dissimilaridades interna e externa é vista pela Equação 3.6 onde a maximização da Equação 3.7 é escolhida. Logo, a melhor combinação será onde a relação entre dois grupos r e l distintos obtiver o valor mais baixo para a pior forma de agrupamento (valores mais altos de dissimilaridades).

$$db = \frac{1}{R} \sum_{r=1, i=1}^R \max(db(l, r)), \quad (3.6)$$

$$db(l, r) = \frac{S_c(r) + S_c(l)}{d_{ce}(l, r)}. \quad (3.7)$$

Estatística do *Gap*

A estatística do *Gap* é apontada como uma métrica eficiente para encontrar o número adequado de grupos em *Clustering*. Essa medida foi comparada com Hartigan [43], KL [44], CH [45] e Silhouette [40]. Na maioria dos testes realizados por Tibshirani, Walther e Hastie, *Gap* teve resultados melhores [46].

A maior deficiência do *Gap* está em agrupamentos com *overlap* de probabilidade p , ou seja, independente da possibilidade que grupos tenham de se sobrepor, há equívocos no acerto de escolha de agrupamentos. No entanto, apesar de aprimoramentos realizados a técnica, no artigo "A comparison of *Gap* statistic definitions with and without logarithm function"[47] é afirmado a ausência de melhora para este caso ruim.

Como a base de dados possui uma formação bem densa e a probabilidade de *overlap* seria alta, então foi escolhido utilizar a Estatística do *Gap* apresentada em [46]. Portanto, apesar das versões de estatística do *Gap* não acertar em todas as simulações para grupos sobrepostos a escolha de quantidade de *clusters*, é apontada ainda como a melhor métrica pela literatura.

Na Equação (3.8) calcula-se o somatório da distância euclidiana ($d_{ii'}$) entre dois pontos posicionados em i e i' encontrados no vetor de todos os pontos classificados no grupo r . Calculada a distância euclidiana de todos os pontos classificados em cada grupo, a Equação (3.9) soma todas essas distâncias divididas cada uma pelo dobro do número de quantidade de distâncias de cada agrupamento ($2n_r$).

Então, a medida calculada na Equação (3.10) compara o valor esperado por uma amostra aleatória (E_n^*) utilizando os limites encontrados pela base de dados e o valor retirado do conjunto em estudo. O logaritmo das amostras é calculado para que os valores encontrados diminuam, e assim, torne-se mais próximas suas diferenças. Por isso, o uso de logaritmo é opcional.

Portanto, o melhor valor de agrupamento escolhido pelo *Gap* é um valor

que se distancia entre valores esperados e calculados. A escolha do melhor valor é afirmada pela expressão matemática $Gap(k) \geq Gap(k + 1) - s_{k+1}$ onde o símbolo s_{k+1} é o desvio padrão encontrado pelo grupo $k + 1$.

$$D_r = \sum_{i,i' \in C_r} d_{ii'}, \quad (3.8)$$

$$W_k = \sum_{r=1}^R \frac{1}{2n_r} D_r, \quad (3.9)$$

$$Gap_n(k) = E_n^*(\log(W_k)) - \log(W_k). \quad (3.10)$$

Capítulo 4

Proposta de perfis de usuários

Os subconjuntos retirados da base de dados que foram utilizados nos experimentos deste capítulo estão demonstrados na Tabela 4.1. A funcionalidade escolhida foi apenas a de resolução das questões pois o objetivo requisitado pela plataforma era identificar possíveis eficiências em exames correlacionados pelas disciplinas publicadas pela interação do usuário na plataforma. A escolha dos subconjuntos foi determinada pelos números mínimos de questões respondidas pelos estudantes da plataforma educacional. Uma questão é determinada por uma pergunta ou afirmação que pode ser respondida por múltiplas escolhas sendo verdadeiro ou falso ou 2 a 5 itens como opções.

O menor grupo foi formado pelos usuários que responderam a menor quantidade total de questões publicadas de uma disciplina pela plataforma, 16. O maior subconjunto escolhido foi formado pelos alunos que responderam a metade do número estabelecido pelo primeiro subconjunto. Portanto, os subconjuntos possuem usuários que responderam pelo menos todas ou metade das questões de cada disciplina.

Além disso, cada subconjunto foi experimentado com 11, 17 e 22 dimensões. O primeiro número escolhido foi 17 dimensões pois é a quantidade total de disciplinas na plataforma educacional. O segundo número foi escolhido por uma das iterações no KDD. Com a experimentação das 17 disciplinas, pode-se perceber a dificuldade de convergência. Por isso, foi separado um

subconjunto somente com as disciplinas que mais são requeridas em exames relacionados ao conteúdo.

A escolha das 22 dimensões foram adicionadas pelo fato das 11 disciplinas não possuírem um peso adequado para o número de questões feitas e sim acertadas, ou seja, com 11 dimensões estão representadas a proporção de acerto nas disciplinas. Portanto, as 22 dimensões equivalem a um vetor com as eficiências das 11 disciplinas concatenadas com o número de questões normalizados para cada disciplina. Assim, a hipótese dessa adição é que os alunos podem ter aprendido com os erros ao responderem mais questões, por isso merecem uma ponderação quanto a isso.

Tabela 4.1: Subconjuntos utilizados para os experimentos no Capítulo 4.

Subconjunto	Significado
G947-11	Subconjunto de 947 usuários e 11 dimensões
G947-17	Subconjunto de 947 usuários e 17 dimensões
G947-22	Subconjunto de 947 usuários e 22 dimensões
G3241-11	Subconjunto de 3241 usuários e 11 dimensões
G3241-22	Subconjunto de 3241 usuários e 22 dimensões
G6791-11	Subconjunto de 6791 usuários e 11 dimensões
G6791-17	Subconjunto de 6791 usuários e 17 dimensões

4.1 Análise de convergência entre os algoritmos *K-means* e *C-means*

Nesta seção é evidenciado o algoritmo que possui melhor desempenho para o problema proposto. A análise dessa seção é realizada abordando os aspectos seguintes:

- i. A dissimilaridade dentro dos grupos devem ser valores mais baixos, ou seja, os dados de um grupo devem ter a maior semelhança possível;
- ii. A dissimilaridade entre grupos devem ser valores mais altos, ou seja, os dados de um grupo devem ter a menor semelhança possível;

- iii. O módulo da diferença das distâncias internas e externas deve ser a maior possível, ou seja, significa melhor agrupamento;
- iv. Os desvios-padrões devem possuir os menores valores possíveis pois quanto menor o desvio padrão, menor a probabilidade de incerteza da resposta.

Portanto, para esta seção são analisados dois cenários. O **Cenário 1** analisa como os algoritmos *K-means* e *C-means* está convergindo, comparando as dissimilaridades encontradas dentro e entre grupos. Para o **Cenário 1**, utiliza o **Subconjunto G947-17** observando então como são agrupados estudantes com bastante interação na plataforma para todas as disciplinas disponíveis. O **Cenário 2** compara os dois algoritmos porém utilizando o **Subconjunto G947-11** que possui melhor convergência que o **Subconjunto G947-17** pois o **Subconjunto G947-11** contem apenas as disciplinas de mais impacto nos exames correlacionados.

4.1.1 Cenário 1: Análise de convergência do grupo com 947 usuários e 17 disciplinas

Nas figuras 4.1 a 4.8, os agrupamentos foram realizados com *K-means* e *C-means* por 30 simulações e 100 iterações de 2 a 9 grupos respectivamente. Nesse cenário são explicadas as figuras relacionadas a cada algoritmo e suas comparações. Além disso, os experimentos foram realizados com **Subconjunto G947-17**. Essa amostra foi selecionada por ser um grupo pequeno de apenas 947 alunos com acessos a plataforma mais constantes, sendo assim usuários com desempenhos não tão divergentes entre si. Para esse cenário, serão expostos os resultados utilizando todas as 17 disciplinas encontradas na plataforma para que o entendimento geral pelos alunos esteja sendo aplicado.

Na Figura 4.1, *K-means* é observado a convergência do *K-means* que inicializa randomicamente seus centros e em menos de 20 iterações encontra um espaço próximo do valor onde agruparia com menor dissimilaridade interna e maior dissimilaridade entre grupos. Observa-se que há poucas iterações com

desvios padrões, ou seja, o agrupamento está encontrando valores semelhantes com 30 simulações. A dissimilaridade dentro dos grupos possui valores acima e próximos de 0.4, ao contrário, a dissimilaridade entre grupos possui valores abaixo e próximos de 0.4. Além disso, pode-se perceber que os valores internos estão diminuindo e os externos aumentando, provando, então, que o algoritmo está melhorando ambos valores. Na figura 4.2. *K-means* observa-se a presença de mais valores de desvios padrões em mais iterações. Então, nessas figuras o *K-means* demonstrou mais incerteza com o aumento de grupos. Continuando a aumentar o número de grupos para o algoritmo *K-means*, nas Figuras 4.3. *K-means* a 4.8. *K-means*, os desvios padrões aparecem com maior frequência e com valores maiores porém possuem dissimilaridades similares independente do número de agrupamentos.

Observando a convergência do algoritmo *C-means* na Figura 4.1. *C-means* os valores das dissimilaridades internas se aproximam de 0.8 e as externas se aproximam de 0.5, e, com poucas iterações é encontrado valores nessas aproximações. O aumento de número de grupos nesta técnica nas Figuras 4.1. *C-means* e 4.8. *C-means* acarreta na frequente presença de desvios padrões nas dissimilaridades externas, significando que a incerteza para encontrar os centroides está sendo dificultada pela diferenciação entre *clusters* e não a similaridade dos dados. Além disso, para a técnica *C-means* suas imagens demonstram uma melhor convergência com 2 e 4 grupos. Apesar da presença de desvios padrões pela Figura 4.3. *C-means* com 4 grupos, o aumento de iterações diminui a incerteza da escolha de centroides significando então que o aumento de iterações pode diminuir ainda mais a incerteza.

Portanto, neste conjunto de imagens é possível constatar que os valores encontrados da dissimilaridade entre grupos são maiores utilizando o algoritmo *C-means*. No entanto, as dissimilaridades internas dos grupos são menores no algoritmo *K-means*. Por isso, para compreender então qual a técnica melhor adapta os valores requeridos das dissimilaridades é a distância calculada entre as dissimilaridades dentro e entre grupos nas imagens deste conjunto.

O afastamento entre os valores de dissimilaridades corresponde a um agrupamento em que os dados estão muito juntos entre si e com grupos

distantes ou dados muito separados entre si mas com centroides próximos. Então, se a dissimilaridade entre grupos é bem mais alta que a distância dentro dos grupos significa que os dados que se encontram em *clusters* diferentes são menos similares que os dados que estão dentro do grupo. Nas Figuras 4.9 e 4.10 observa-se o afastamento de ambos os algoritmos entre os valores das dissimilaridades encontrados para todos os números de grupos utilizados. A Figura 4.9 mostra que o intervalo entre o afastamento dos dois valores tem intervalo entre 0 e 0.12. No entanto, a Figura 4.10 tem intervalo mais alto entre 0 e 0.40. Portanto, a diferença entre as dissimilaridades externas e internas deve ser a maior possível sendo verificada pelo algoritmo *C-means*.

Figura 4.1: Gráfico do *K-means* e *C-means* utilizando **Subconjunto G947-17** com 2 grupos

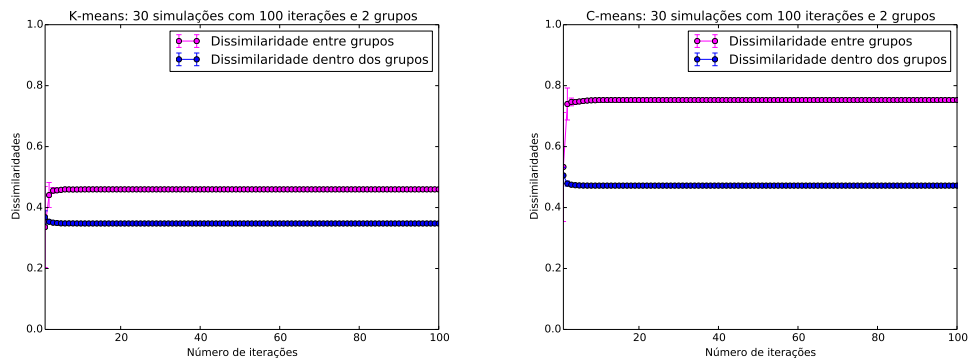


Figura 4.2: Gráfico do *K-means* e *C-means* utilizando Subconjunto G947-17 com 3 grupos

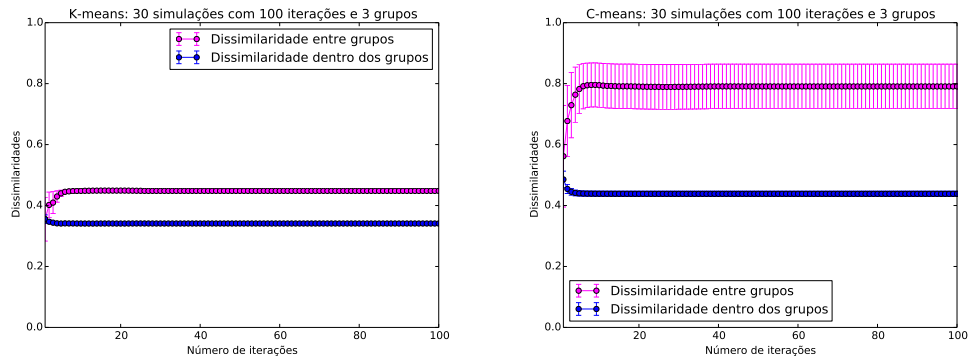


Figura 4.3: Gráfico do *K-means* e *C-means* utilizando Subconjunto G947-17 com 4 grupos

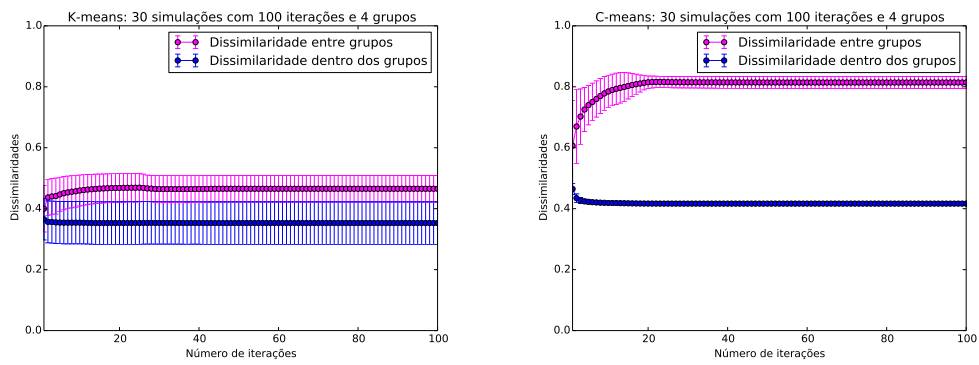


Figura 4.4: Gráfico do *K-means* e *C-means* utilizando Subconjunto G947-17 com 5 grupos

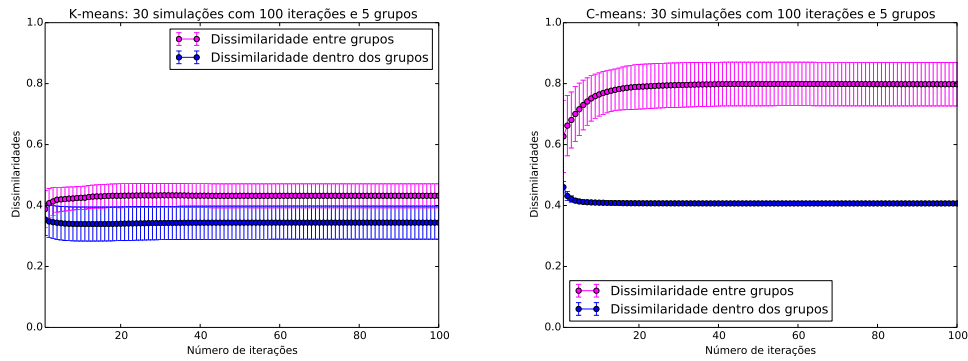


Figura 4.5: Gráfico do *K-means* e *C-means* utilizando Subconjunto G947-17 com 6 grupos

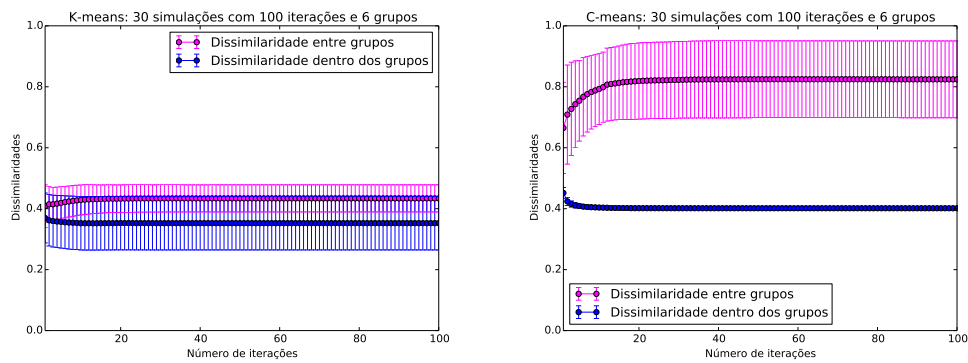


Figura 4.6: Gráfico do *K-means* e *C-means* utilizando Subconjunto G947-17 com 7 grupos

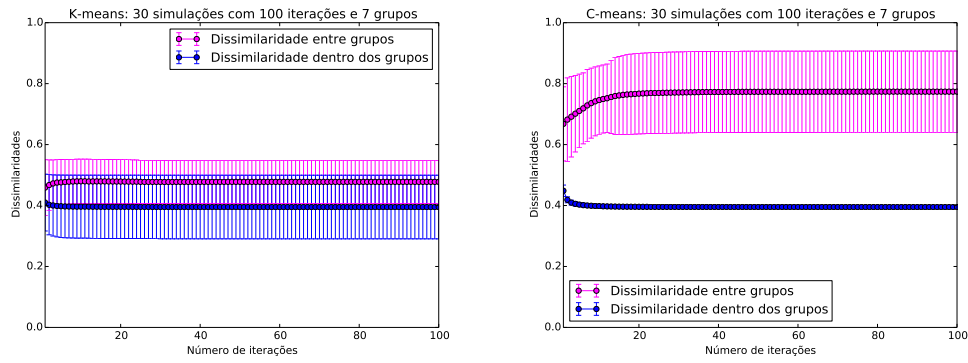


Figura 4.7: Gráfico do *K-means* e *C-means* utilizando Subconjunto G947-17 com 8 grupos

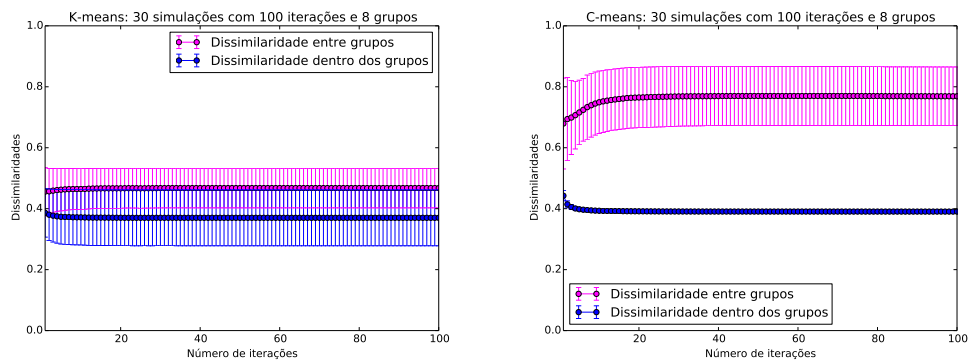


Figura 4.8: Gráfico do *K-means* e *C-means* utilizando Subconjunto G947-17 com 9 grupos

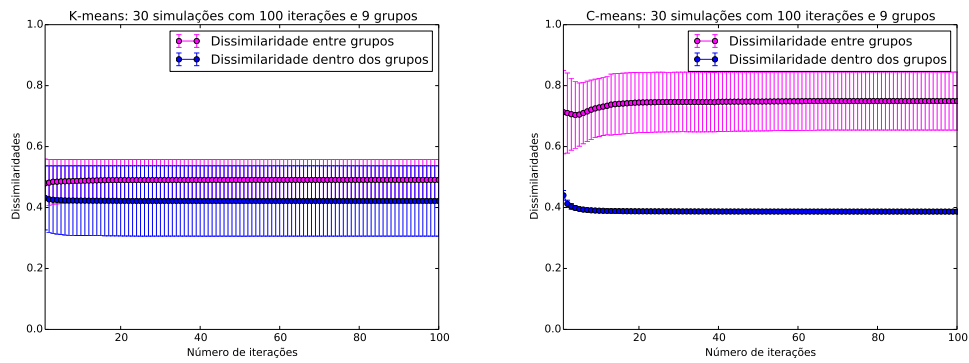


Figura 4.9: Gráfico do afastamento entre as dissimilaridades internas e externas no *K-means*

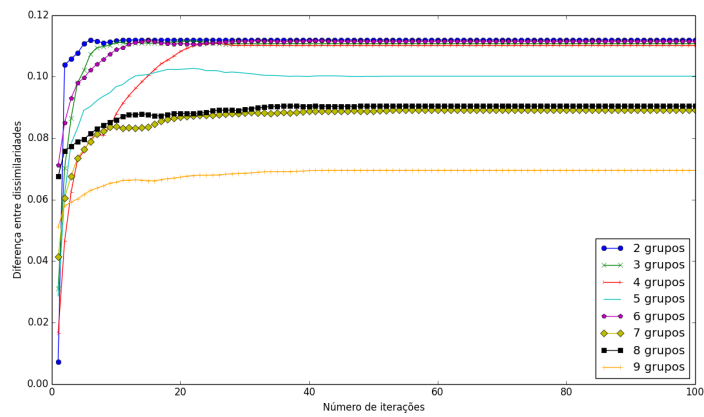
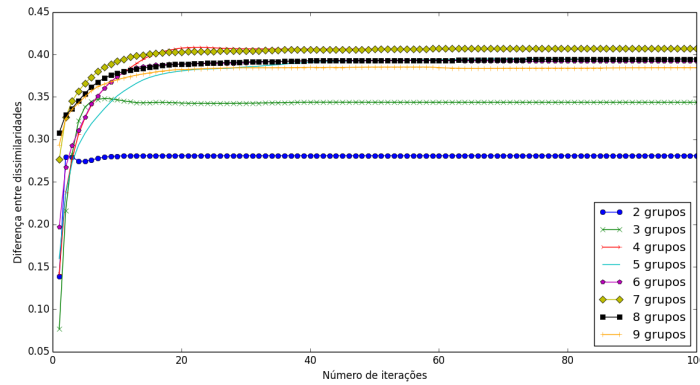


Figura 4.10: Gráfico do afastamento entre as dissimilaridades internas e externas no *C-means*



4.1.2 Cenário 2: Análise de convergência do grupo com 947 usuários e 11 disciplinas

Para este experimento será utilizado **Subconjunto G947-11** pois foi evidenciado a presença frequente de desvios padrões na Subseção 4.1.1. Portanto, esta nova amostra será formada somente pelas disciplinas que possuem mais peso em exames relacionados aos conteúdos oferecidos pela plataforma. Com essa escolha, a amostra da base de dados dispõe menos divergências entre si, ou seja, desempenhos mais semelhantes.

As Figuras 4.11 a 4.18 mostram os valores das dissimilaridades retiradas dos algoritmos *K-means* e *C-means* com 100 iterações e 30 simulações utilizando o intervalo de 2 a 9 grupos respectivamente. As conclusões retiradas pela Subseção 4.1.1 também sustentam essas argumentações. Os resultados retirados pelo *K-means* demonstram pequeno distanciamento entre as dissimilaridades e alta frequência de desvios padrões. Nas Figuras 4.19 e 4.20, os valores de afastamento entre as dissimilaridades utilizando o algoritmo *K-means* possui valores menores que utilizando o algoritmo *C-means*. Por isso, o algoritmo *C-means* nas duas amostras, demonstrado pelos gráficos, é um agrupamento mais similar internamente e mais dissimilar entre grupos.

Figura 4.11: Gráfico do *K-means* e *C-means* utilizando Subconjunto G947-11 com 2 grupos

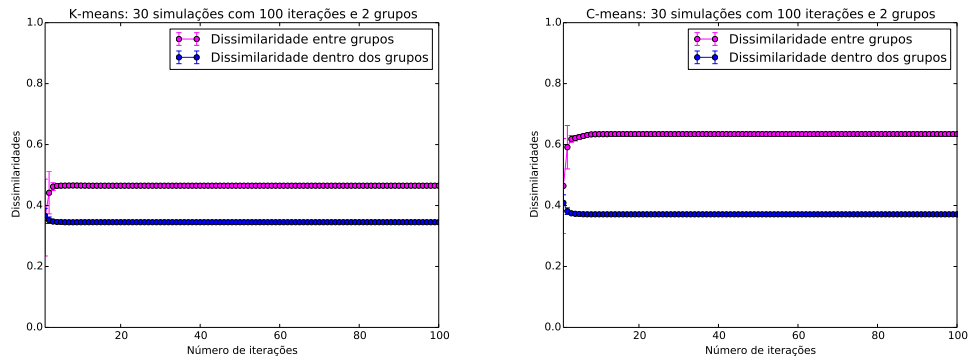


Figura 4.12: Gráfico do *K-means* e *C-means* utilizando Subconjunto G947-11 com 3 grupos

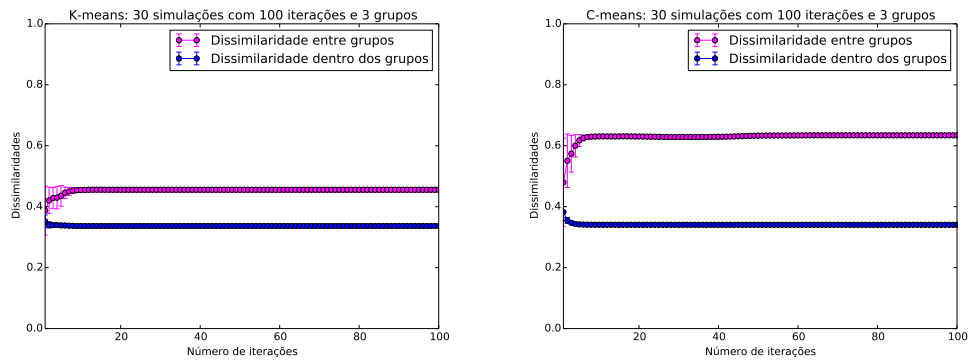


Figura 4.13: Gráfico do *K-means* e *C-means* utilizando Subconjunto G947-11 com 4 grupos

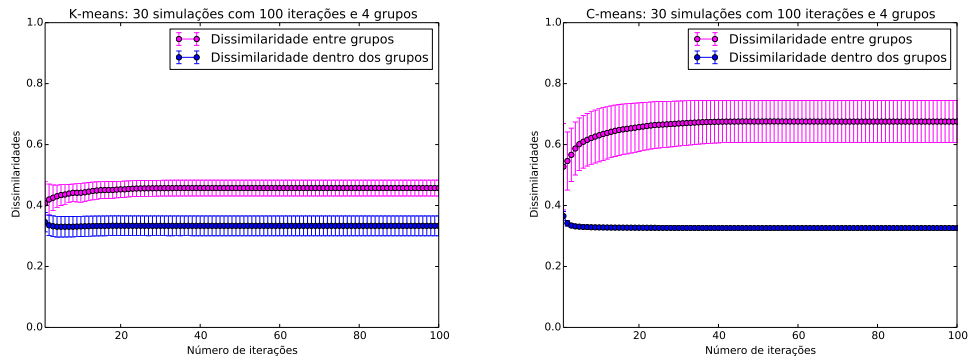


Figura 4.14: Gráfico do *K-means* e *C-means* utilizando Subconjunto G947-11 com 5 grupos

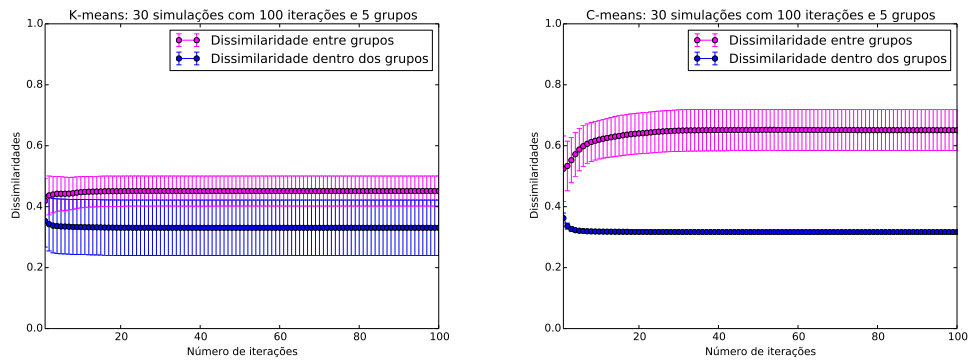


Figura 4.15: Gráfico do *K-means* e *C-means* utilizando Subconjunto G947-11 com 6 grupos

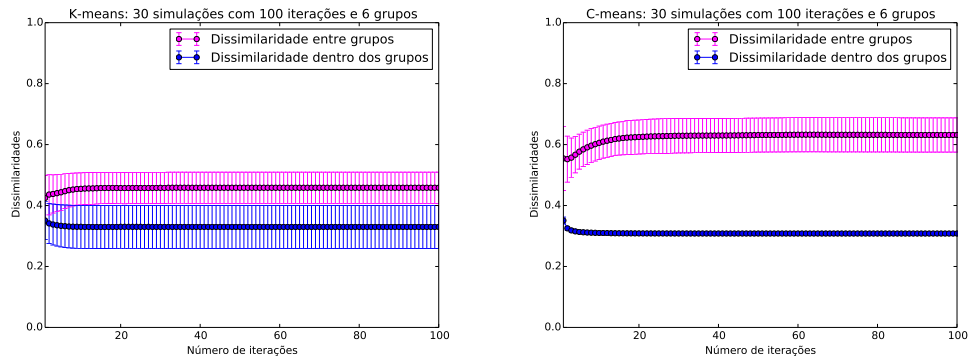


Figura 4.16: Gráfico do *K-means* e *C-means* utilizando Subconjunto G947-11 com 7 grupos

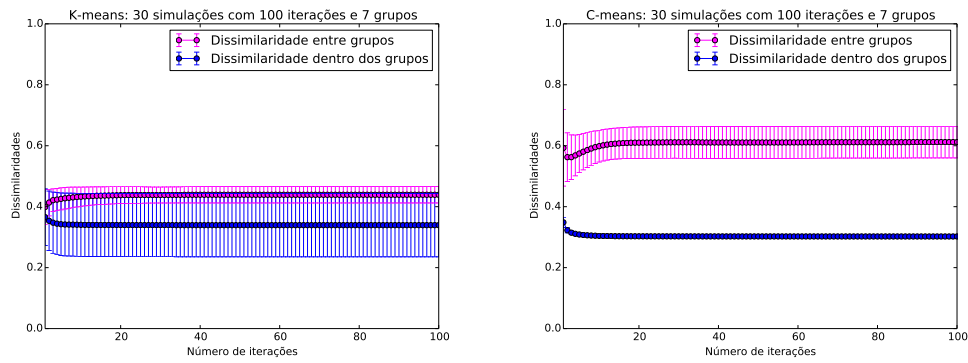


Figura 4.17: Gráfico do *K-means* e *C-means* utilizando Subconjunto G947-11 com 8 grupos

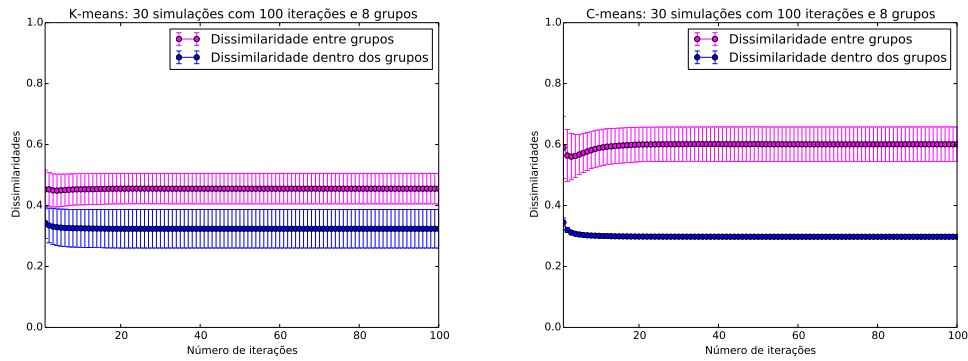


Figura 4.18: Gráfico do *K-means* e *C-means* utilizando Subconjunto G947-11 com 9 grupos

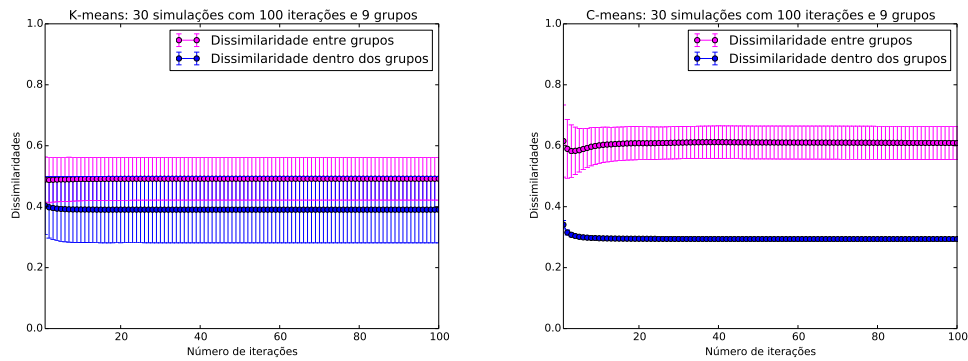


Figura 4.19: Gráfico do afastamento entre as dissimilaridades internas e externas no K -means

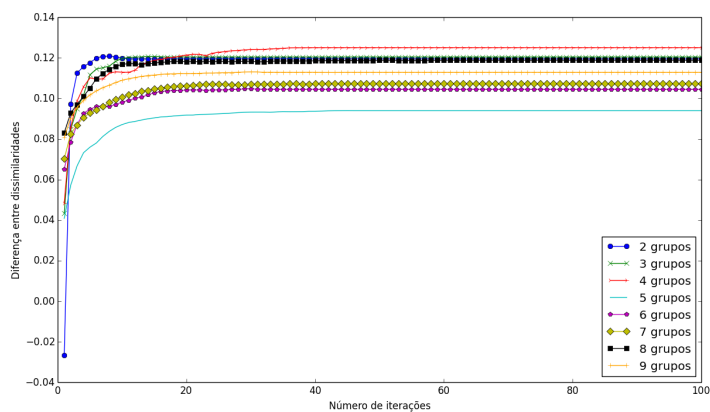
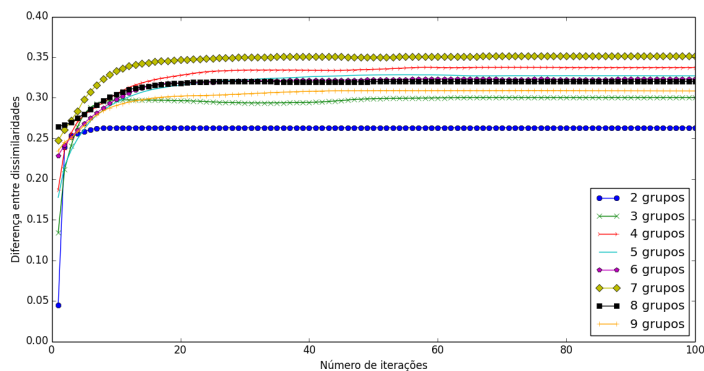


Figura 4.20: Gráfico do afastamento entre as dissimilaridades internas e externas no C -means



Então, os aspectos abordados nas subseções 4.1.1 e 4.1.2 demonstram que o algoritmo C -means forma *clusters* com menos incerteza e maior dissimilaridade entre grupos em relação ao algoritmo K -means. Além disso, é possível perceber uma tendência a formações mais similares e com menos incertezas utilizando os números 2, 3 e 4. Esses números são reafirmados pelo fato de que a partir de 5 grupos, o agrupamento de nenhum elemento

pelo algoritmo *C-means* é passível de ocorrer.

4.2 Análise da diferença entre as convergências com 11 e 17 disciplinas utilizando *C-means* para 6791 usuários

Ao concluir que o algoritmo *C-means* resultava em grupos com centroides mais distantes, os experimentos não serão realizados com *K-means* nas próximas seções. Na Seção 4.1 utilizou-se um grupo menor com mais similaridade e constatou-se que a amostra com 11 disciplinas obteve menor frequência e valor de incerteza comparado à amostra com 17 disciplinas. Essa seção considera o aumento do número de estudantes para verificar se a amostra com 11 disciplinas permanece tendo menos incerteza que a de 17 disciplinas. O **Subconjunto G6791-11** e **Subconjunto G6791-17** utilizado é formado pelos usuários que possuem um desempenho na plataforma maior que a média das questões realizadas em exames relacionados ao conteúdo. A média foi escolhida pois a nota para obter aprovação nesses exames é metade da nota máxima.

As Figuras 4.21 a 4.24 estão representando a amostra com 17 disciplinas. Nessas imagens é possível identificar a convergência do algoritmo com desvios padrões mais frequentes no fim das iterações quando o número de grupos for maior ou igual a 6. Nas Figuras 4.21 e 4.22, os desvios padrões com o aumento de iterações desaparecem. Apesar da demora para encontrar os centroides mais propícios para agrupamento, o algoritmo consegue convergir tardiamente sem desvios padrões. Portanto, essa amostra demonstra que a quantidade para melhor agrupamento é menor que 6.

As Figuras 4.25 a 4.28 representam a amostra com 11 disciplinas. Ao contrário da amostra utilizada com 17 disciplinas, o **Subconjunto G6791-11** não apresenta valores altos de desvios padrões demonstrando uma incerteza menor. Além disso, na análise de convergência de **Subconjunto G6791-11** não é possível descartar nenhum número de grupos.

Sendo assim, com relação a incerteza para encontrar os centroides que

melhor configurem grupos dissimilares entre si, porém com elementos internos similares, a configuração utilizada pela amostra com 11 disciplinas obteve resultados mais interessantes. Por isso, nas próximas seções foi utilizada a configuração com 11 disciplinas.

Figura 4.21: Gráfico do *C-means* de 30 simulações e 100 iterações utilizando Subconjunto G6791-17 com 2 e 3 grupos

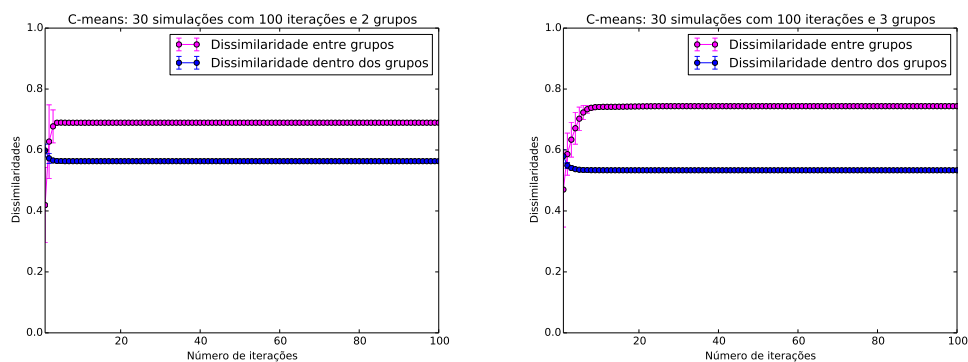


Figura 4.22: Gráfico do *C-means* de 30 simulações e 100 iterações utilizando Subconjunto G6791-17 com 4 e 5 grupos

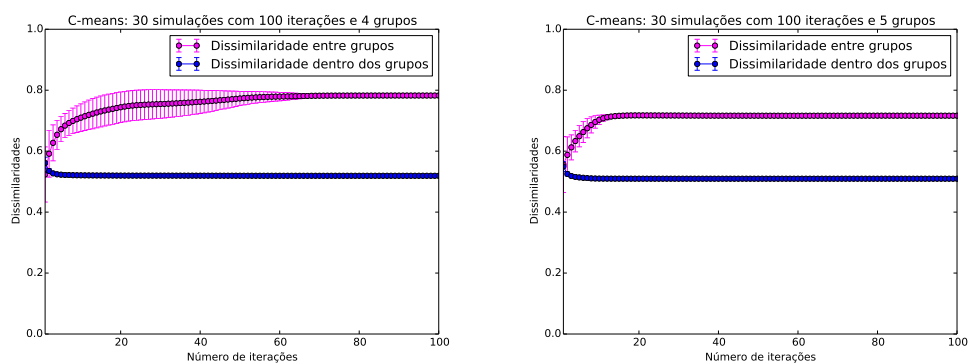


Figura 4.23: Gráfico do *C-means* de 30 simulações e 100 iterações utilizando Subconjunto G6791-17 com 6 e 7 grupos

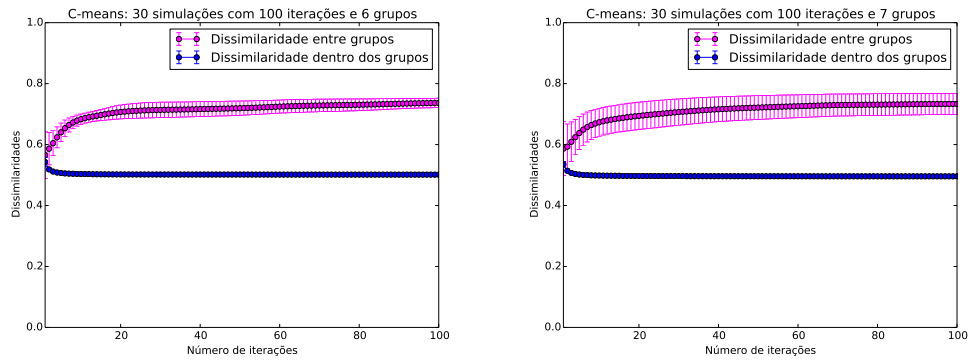


Figura 4.24: Gráfico do *C-means* de 30 simulações e 100 iterações utilizando Subconjunto G6791-17 com 8 e 9 grupos

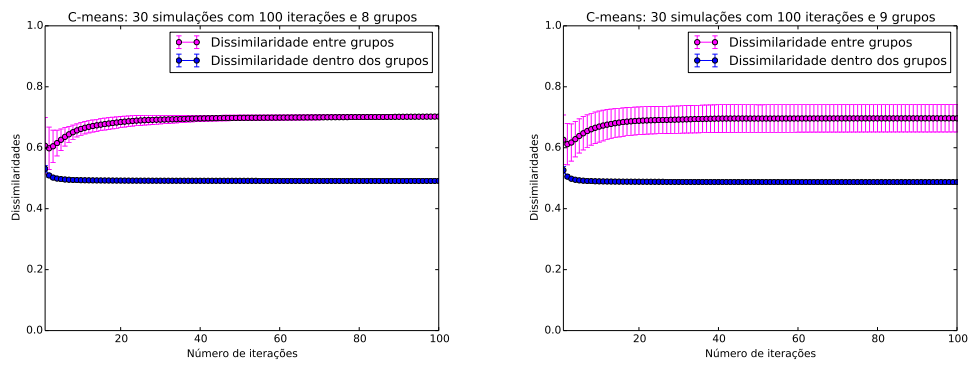


Figura 4.25: Gráfico do *C-means* de 30 simulações e 100 iterações utilizando Subconjunto G6791-11 com 2 e 3 grupos

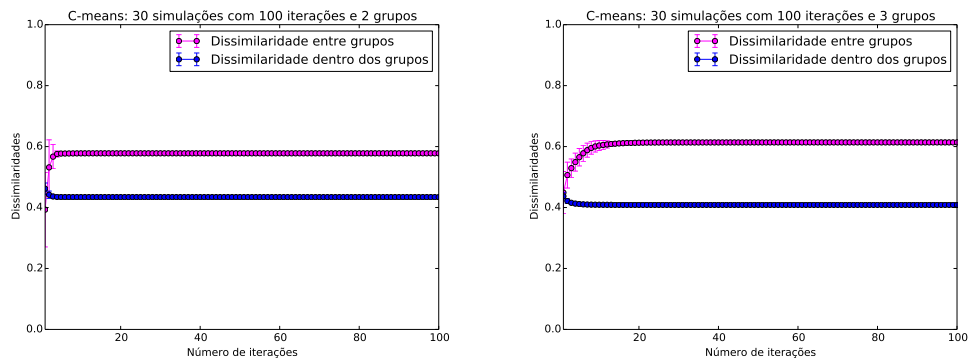


Figura 4.26: Gráfico do *C-means* de 30 simulações e 100 iterações utilizando Subconjunto G6791-11 com 4 e 5 grupos

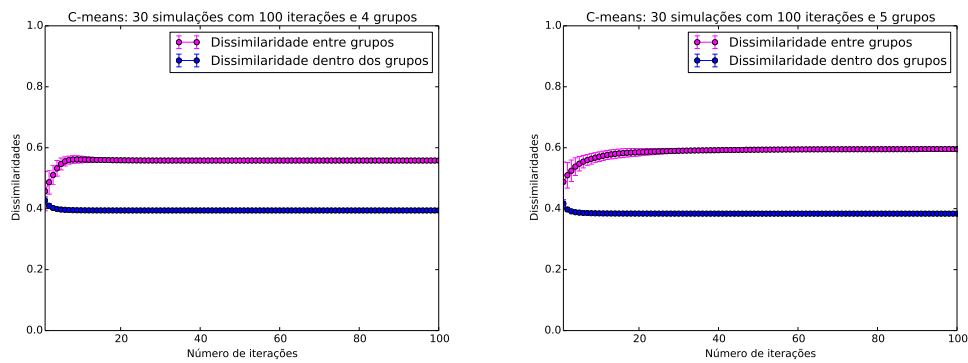


Figura 4.27: Gráfico do *C-means* de 30 simulações e 100 iterações utilizando Subconjunto G6791-11 com 6 e 7 grupos

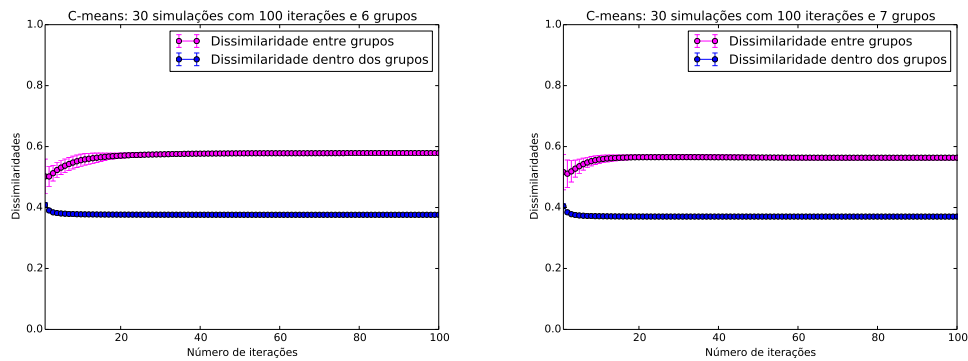
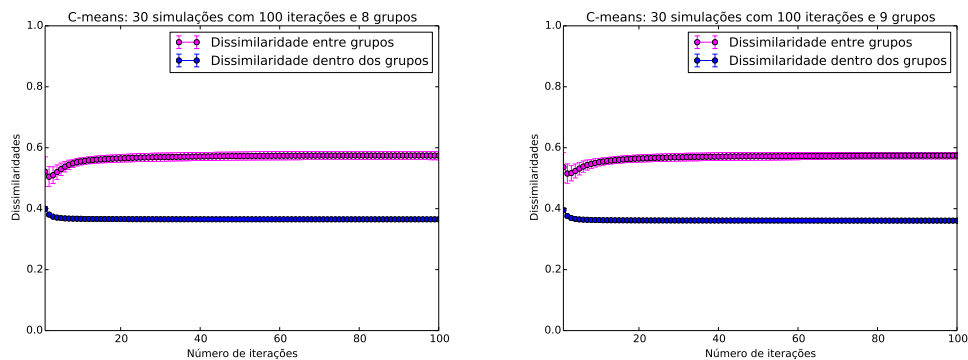


Figura 4.28: Gráfico do *C-means* de 30 simulações e 100 iterações utilizando Subconjunto G6791-11 com 8 e 9 grupos



4.3 Análise para escolha do número de grupos

A escolha do número de grupos foi realizada pelas dissimilaridades externas, a métrica de Davies-Bouldin e estatística do *Gap*. Como as dissimilaridades internas são bastante semelhantes em valor independente do experimento, esta métrica não foi utilizada para esta análise. Para essa se-

ção, dois cenários da base de dados foram utilizados. O **Cenário 1** utiliza os agrupamentos com o **Subconjunto G947-11** e **Subconjunto G947-22** e o **Cenário 2** utiliza os agrupamentos com o **Subconjunto G3241-11** e **Subconjunto G3241-22**. O primeiro e segundo subconjunto representam, respectivamente, 947 e 3241 usuários utilizando 11 e 22 dimensões pelos motivos justificados pela introdução desse capítulo. Assim, o agrupamento dos estudantes foi realizado para aqueles que obtiveram pelo menos um desempenho mediano na plataforma.

4.3.1 Cenário 1: Análise para escolha do número de grupos do Subconjunto G947-11 e Subconjunto G947-22

Os resultados do experimento **Subconjunto G947-11** e **Subconjunto G947-22** estão demonstrados pelas Figuras 4.31, 4.30 e 4.29. As dissimilaridades entre os grupos devem ser os valores mais altos possíveis, ou seja, os grupos devem possuir centroides mais distantes. As Figuras 4.29.a e 4.29.b mostram que o número de grupos igual a 3 possui um valor maior considerando o desvio padrão. A métrica de Davies-Bouldin nas Figuras 4.30.a e 4.30.b é minimizada pelo agrupamento em 2 para ambos experimentos. A estatística do *Gap* para esses subconjuntos é maximizada em 2. Portanto, para esses dois experimentos o agrupamento com melhor formação seria dividir em 2 *clusters* pelas dissimilaridades.

Figura 4.29: Gráfico das dissimilaridades externas do *C-means* utilizando 30 simulações, 2000 iterações e grupos de 2 a 5 utilizando **Subconjunto G947-11** e **Subconjunto G947-22**

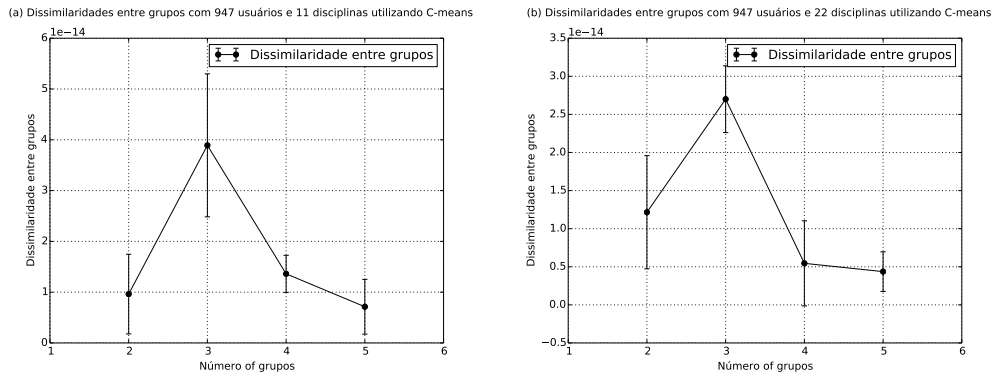


Figura 4.30: Gráfico da métrica de Davies-Bouldin do *C-means* utilizando 30 simulações, 2000 iterações e grupos de 2 a 5 utilizando **Subconjunto G947-11** e **Subconjunto G947-22**

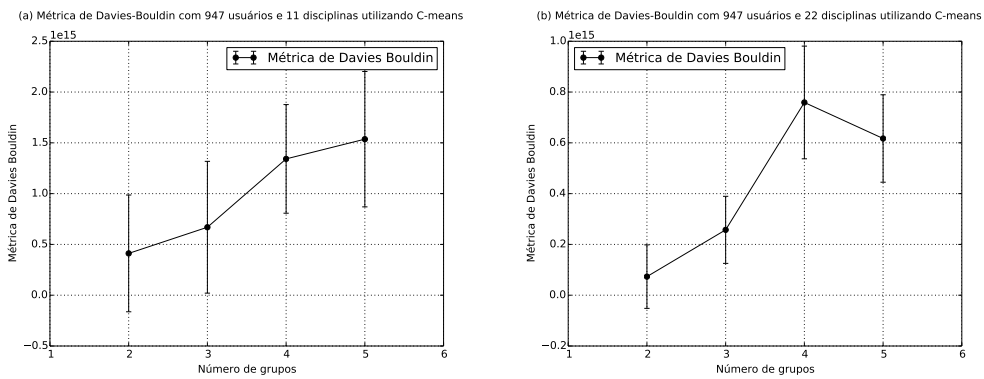
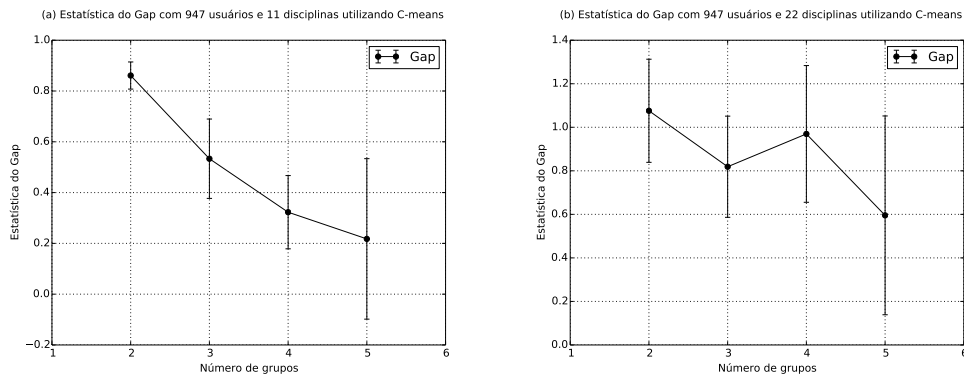


Figura 4.31: Gráfico da Estatística do Gap do *C-means* utilizando 30 simulações, 2000 iterações e grupos de 2 a 5 utilizando **Subconjunto G947-11** e **Subconjunto G947-22**



4.3.2 Cenário 2: Análise para escolha do número de grupos do Subconjunto G3241-11 e Subconjunto G3241-22

Os experimentos **Subconjunto G3241-11** e **Subconjunto G3241-22** estão demonstrados pelas Figuras 4.34, 4.33 e 4.32. As Figuras 4.32.a e 4.32.b evidenciam que o número de grupos igual a 5 possui mais distanciamento entre os centroides. A métrica de Davies-Bouldin é minimizada considerando o desvio padrão pelo agrupamento em 2 para ambos experimentos. A estatística do *Gap* para esses 2 experimentos maximiza em 2. Portanto, para esses subconjuntos o agrupamento mais adequado seria em 2 *clusters*.

Figura 4.32: Gráfico das dissimilaridades externas do *C-means* utilizando 30 simulações, 2000 iterações e grupos de 2 a 5 utilizando **Subconjunto G3241-11** e **Subconjunto G3241-22**

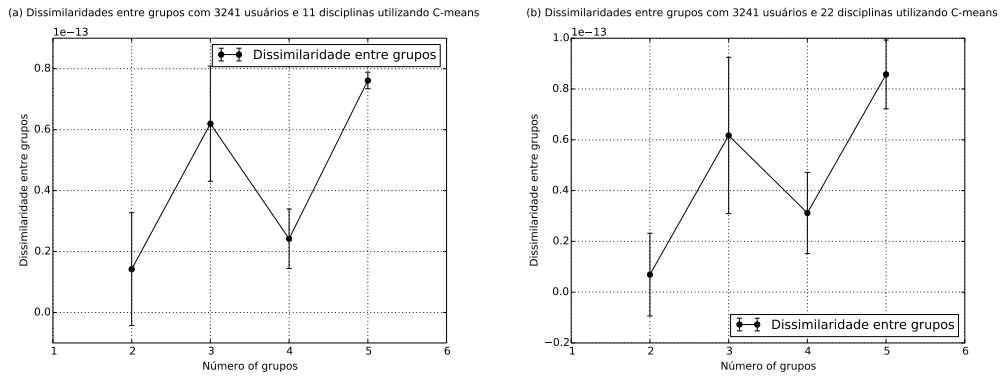


Figura 4.33: Gráfico da métrica de Davies-Bouldin do *C-means* utilizando 30 simulações, 2000 iterações e grupos de 2 a 5 utilizando **Subconjunto G3241-11** e **Subconjunto G3241-22**

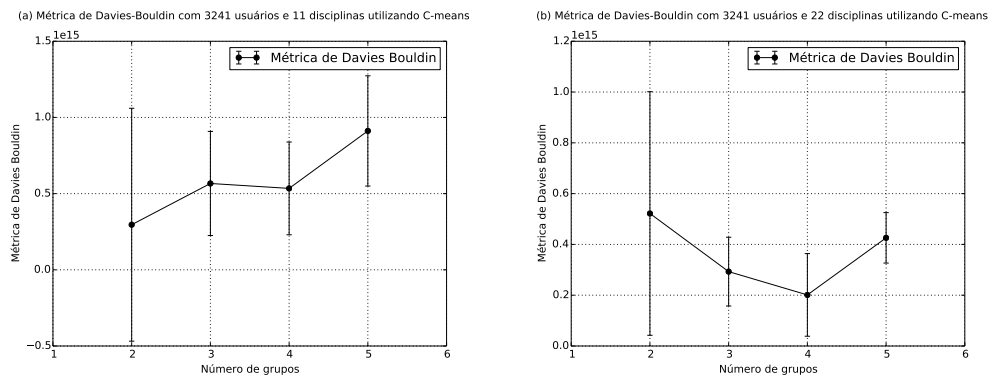
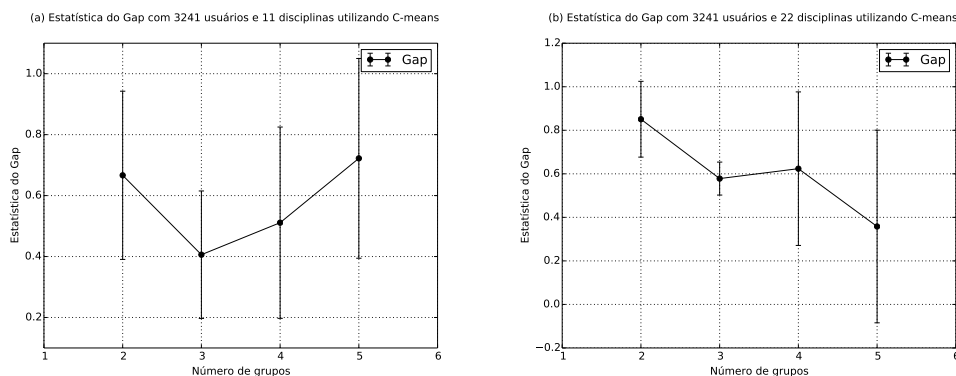


Figura 4.34: Gráfico da Estatística do Gap do *C-means* utilizando 30 simulações, 2000 iterações e grupos de 2 a 5 utilizando **Subconjunto G3241-11** e **Subconjunto G3241-22**



4.4 Análise da proposta dos perfis de usuários

As métricas utilizadas pelo trabalho, Davies-Bouldin e *Gap*, revelaram que o agrupamento seria 2 *clusters* para que as dissimilaridades internas e entre centroides sejam as adequadas. Portanto, para esta seção foram analisados dois cenários. O **Cenário 1** utiliza os agrupamentos com o **Subconjunto G947-11** e **Subconjunto G947-22** e o **Cenário 2** utiliza os agrupamentos com o **Subconjunto G3241-11** e **Subconjunto G3241-22**.

Para identificar as relações entre disciplinas em um grupo foram realizados testes do coeficiente de correlação de Spearman (Spearman's rank correlation coefficient). Essa correlação não requer que a relação entre variáveis seja linear [48].

No Apêndice, as figuras 5.1 a 5.16 mostram as correlações de Spearman encontradas em todos os subconjuntos. A análise de todas as imagens originou as Tabelas 4.2 e 4.5. Nestas tabelas foram computadas as quantidades encontradas com os intervalos especificados. Além disso, todas as autocorrelações entre a própria disciplina foram descartadas, ou seja, não foram contados por exemplo a relação da Disciplina 1 com ela mesma.

No Apêndice, as figuras 5.17 a 5.24 mostram valores muito próximos de zero. Esse resultado revela que a correlação cruzada entre os perfis são baixíssimas significando que os perfis encontrados não se assemelham, ou seja, o algoritmo utilizado *C-means* agrupou bem os usuários.

4.4.1 Cenário 1: Análise da auto-correlação entre as eficiências das 11 disciplinas dos Subconjunto G947-11 e Subconjunto G947-22

A Tabela 4.2 mostra que o Perfil 1 possui mais correlações altas e o Perfil 0 possui mais correlações baixas. Além disso, o Perfil 1 possui notas mais elevadas. Portanto, o Perfil 1 representa os alunos mais estudiosos que conseqüentemente conseguem correlacionar melhor as disciplinas, ou seja, a eficiência em determinadas disciplinas acarreta na eficiência em outras, mostrando então a possível relação entre disciplinas.

Na Tabela 4.3, é possível perceber que as disciplinas mais correlacionadas para um bom desempenho estão determinadas para o primeiro cenário. Portanto, a junção das disciplinas encontradas na mesma linha significa que o desempenho de alunos em uma disciplina acarreta em um bom ou mau desempenho na seguinte disciplina.

Em contrapartida, na Tabela 4.4 as correlações das disciplinas demonstram a falta de ligação entre elas. Sendo assim, as disciplinas relacionadas por linha não possuem consequência direta no desempenho entre si.

Para o **Cenário 1**, os usuários possuem características bem marcantes que poderão ser analisados mais profundamente. Sabendo que estes usuários estão apresentando uma semelhança alta internamente e que o perfil dos melhores estudantes na plataforma educacional consegue relacionar melhor as disciplinas, é possível mapear o comportamento geral destes estudantes em todas as funcionalidades para que a plataforma encontre estratégias relevantes para auxiliar os alunos com notas mais baixas.

Tabela 4.2: Tabela demonstrativa das quantidades de correlações encontradas em intervalos de valores no Perfil 0 e 1 **Subconjunto G947-11** e **Subconjunto G947-22**.

Experimento	Intervalo de Correlação	Perfil 0	Perfil 1
G947-11	[0.00, 0.25]	63	10
G947-11	[0.25, 0.40]	71	119
G947-11	[0.40, 1.00]	2	7
G947-22	[0.00, 0.25]	65	13
G947-22	[0.25, 0.40]	69	117
G947-22	[0.40, 1.00]	2	6

Tabela 4.3: Correlações mais altas entre Disciplinas encontrados pelo Perfil 1 (Notas mais elevadas) pelo **Subconjunto G947-11** e **Subconjunto G947-22**.

Disciplina	Disciplina
1	5
1	14
3	7
5	10
5	16
10	11
12	13

Tabela 4.4: Correlações mais baixas entre Disciplinas encontrados pelo Perfil 1 (Notas mais elevadas) pelo **Subconjunto G947-11** e **Subconjunto G947-22**.

Disciplina	Disciplina
1	17
2	10
2	11
4	16
4	17
5	17
6	16
6	17
7	16
10	17
11	17
12	17
13	17

4.4.2 Cenário 2: Análise da auto-correlação entre as eficiências das 11 disciplinas dos Subconjunto G3241-11 e Subconjunto G3241-22

Para o **Cenário 2**, as altas auto-correlações entre disciplinas são menos frequentes do que no **Cenário 1**, ou seja, quando aumenta-se o valor de usuários a dissimilaridade aumenta e não é possível identificar as relações fortes que os usuários mais estudiosos e de melhor desempenho possuem.

Na Tabela 4.5, é possível identificar alta quantidade de relações fracas entre disciplinas. Além disso, é possível perceber que o grupo de Perfil 1 que possui notas mais elevadas possui apenas 2 altas correlações descritas na Tabela 4.6.

Portanto, o **Cenário 2** não apresentou resultados relevantes para que a plataforma educacional analise seus usuários mais profundamente. Assim sendo, a análise dos estudantes deverá ser realizada pelos resultados do **Cenário 1**.

Tabela 4.5: Tabela demonstrativa das quantidades de correlações encontradas em intervalos de valores no Perfil 0 e 1 pelo **Subconjunto G3241-11** e **Subconjunto G3241-22**.

Experimento	Intervalo de Correlação	Perfil 0	Perfil 1
G3241-11	[0.00, 0.25]	97	49
G3241-11	[0.25, 0.40]	39	87
G3241-11	[0.40, 1.00]	0	2
G3241-22	[0.00, 0.25]	93	70
G3241-22	[0.25, 0.40]	39	64
G3241-22	[0.40, 1.00]	0	2

Tabela 4.6: Correlações mais altas entre Disciplinas encontrados pelo Perfil 1 (Notas mais elevadas) pelo **Subconjunto G3241-11** e **Subconjunto G3241-22**.

Disciplina	Disciplina
1	5
12	13

Em conclusão, para a base de dados utilizada neste trabalho foi verificada pelos experimentos que o algoritmo *C-means* obteve 2 perfis de usuários bem diferentes. Para o Perfil 1, de desempenho melhor na plataforma, correlações entre disciplinas foram fortemente ou fracamente comprovadas. No entanto, para o Perfil 0, de desempenho razoável, a falta de correlação evidenciou a necessidade de estudo na plataforma.

Capítulo 5

Considerações Finais

As plataformas educacionais, preocupadas em ensinar de diversas formas aos alunos os conteúdos propostos, investigam em meio a volumes grandes de dados fatos que auxiliem na melhora do sistema. Para este trabalho, foi desenvolvido a análise em uma base de dados educacionais com 300 mil estudantes e 17 disciplinas.

Objetivando entender as principais características dos usuários, o agrupamento de dados evidenciou dois perfis bem diferentes entre si de acordo com os testes realizados com o Coeficiente de Correlação de Spearman. Portanto, o algoritmo escolhido *C-means* conseguiu separar bem os usuários e as métricas utilizadas Davies-Bouldin e *Gap Statistic* foram importantes na escolha do número de perfis.

No Processo de Descoberta do Conhecimento diversas decisões tiveram que ser realizadas. Em uma das iterações do KDD, os alunos que não tinham uma interação de pelo menos 5 questões na plataforma educacional foram retirados devido a dificuldade de convergência nos agrupamentos. Em seguida, foi identificado que a configuração com todas as disciplinas disponíveis na plataforma também dificultava a convergência para os mesmos centroides no final. Com isso, o uso de 17 disciplinas, número máximo, acarretava em uma frequência constante de desvios padrões. Ao retirar as disciplinas que possuíam entre si comportamentos fracos semelhantes dos usuários, tanto em quantidade de questões feitas quanto de eficiência, foi possível obter conver-

gências com menos incerteza.

Para analisar os agrupamentos do usuários, foi escolhido duas amostras. A primeira amostra deveria ter respostas com menos incerteza e usuários que interagiam mais na plataforma. Por isso, foram escolhidos os alunos que responderam a quantidade total de questões da disciplina com menor quantidade significando então que pelo menos uma disciplina o aluno teria feito todas as questões. Sendo assim, esse grupo traria grupos com resultados da eficiência de alunos na plataforma para prováveis exames relacionados aos conteúdos, ou seja, quais grupos de alunos precisava estudar mais e como o grupo que estava tirando maiores notas com uma boa interação estaria estudando na plataforma.

A segunda amostra selecionou os alunos que responderam metade das questões da primeira amostra. Essa metade significa que os estudantes teriam que responder pelo menos metade das questões de menor quantidade na plataforma da disciplina. Conseqüentemente, essa amostra teria estudantes que tinham respondidos um valor mediano para baixo de questões de cada disciplina. Portanto, a segunda amostra poderia agrupar grupos que interagiram com a plataforma com mais dissimilaridade que a primeira amostra porém com interação mediana ou bem próximo na plataforma educacional, agrupando grupos mais eficientes distante de grupos com melhor desempenho.

Para as amostras escolhidas, foi possível identificar um grupo com mais similaridade na primeira amostra onde demonstrava um desempenho elevado e interação elevado. Neste grupo, as alta correlações entre disciplinas analisadas mostraram as principais matérias relacionadas que os estudantes possuíam facilidade e dificuldade. Além disso, ao analisar mais os alunos, é possível comparar suas interações em outras funcionalidades para identificar quais funcionalidades é mais utilizada pelos mesmos e como estavam utilizando a plataforma educacional para estudo.

A utilização de apenas uma funcionalidade foi realizada pelo propósito de encontrar similaridades e correlações entre estudantes através de uma possível eficiência em exames correlacionados nos conteúdos disponíveis pela plataforma. O objetivo então geral do trabalho de encontrar características

em comum de agrupamentos com usuários bem semelhantes entre si e grupos bem diferentes pela eficiência e interação na plataforma foi alcançado.

Como trabalhos futuros, poderia ser investigado os usuários do perfil classificado como menos eficiente. Essa investigação poderia trazer ao sistema novas ideias de correlações e ineficiências. Além disso, seria interessante avaliar a eficiência dos alunos utilizando a *Bayesian Knowledge Tracing* e Teoria de Resposta ao Item onde respectivamente estudam a forma dinâmica que o conhecimento é adquirido pelo estudante a cada interação a funcionalidades e a cada período de um conjunto de interações no sistema.

Referências Bibliográficas

- [1] Gomes JH. Educação a distância e participação política sustentável: O contributo da universidade aberta de Portugal no contexto internacional. Dissertação de Mestrado em Cidadania Ambiental e Participação – Universidade Aberta; 2014.
- [2] Alves L. Educação a distância: conceitos e história no Brasil e no mundo [Periódico]. Associação Brasileira de Educação a Distância - ABED; 2011.
- [3] Holliman R, E S. *Investigating Cooperation and Collaboration in Near Synchronous Computer Mediated Conferences*. Computers and Education. 2006;46:322–335.
- [4] Alves GMT, John E, Costa F. Proposta de avaliação para a aprendizagem invisível com o uso das *Personal Learning Environment (PLEs)*, (i), 273–282; 2014.
- [5] Dabbagh N, Kitsantas APLE. *Social media and self-regulated learning: A natural formula for connecting formal and informal learning, The Internet and Higher Education*. 2012 Jan;15.
- [6] Fayyad UM, Piatetsky-Shapiro G, Smyth P. *The KDD Process for Extracting Useful Knowledge from Volumes of Data*. Communications of ACM. 1996;39(11):27–34.
- [7] Dias MM, Filho LAS, Lino ADP, Favero EL, Ramos EMLS. Aplicação de técnicas de mineração de dados no processo de aprendizagem na

- Educação a Distância. Anais do Simpósio Brasileiro de Informática na Educação; 2008.
- [8] Wu X, Kumar V, Quinlan JS, Ghosh J, Yang Q, Motoda H, et al. Top 10 algorithms in data mining. *Journal of Knowledge and Information Systems*. 2008;14(1).
- [9] Merceron A, Yacef K. Educational Data Mining: a Case Study. *Artificial Intelligence in Education*; 2005.
- [10] Huei-Tse H. A case study of online instructional collaborative discussion activities for problem-solving using situated scenarios: an examination of content and behavior cluster analysis. *Computers and Education*. 2011;56(3).
- [11] Agrupamento DEADE, Minera NA, Educacionais DED. Use of Clustering Algorithms in the Educational Data Mining, 1–10; 2006.
- [12] Fayyad U, Piatetsky-shapiro G, Smyth P. From Data Mining to Knowledge Discovery in. 1996;17(3):37–54.
- [13] Romero C, Ventura S. Educational Data Mining: A Review of the State of the Art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*. 2010;40(6):601–618.
- [14] Jain AK, Murty MN, Flynn PJ. Data Clustering : A Review. *ACM Computing Surveys*. 2000 September;31(3).
- [15] Pernambuco UFD. Conceitos , técnicas , ferramentas e aplicações de Mineração de Dados para gerar conhecimento a partir de bases de dados. 2006;.
- [16] Ghaemmaghami H, Dean D, Vogt R, Sridharan S. Speaker attribution of multiple telephone conversations using a complete-linkage clustering approach. In: 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2012. p. 4185–4188.

- [17] Zhang J, Lu Y, Wang L, Zhang H, Li R, Wang L, et al. Single Linkage Clustering Fit for Establishing Theoretical MAPK/Erk Signaling Pathway in Human Soft Tissue Sarcoma Samples. In: 2009 2nd International Conference on Biomedical Engineering and Informatics; 2009. p. 1–4.
- [18] Celebi ME, Kingravi HA, Vela PA. Expert Systems with Applications A comparative study of efficient initialization methods for the k-means clustering algorithm. *Expert Systems With Applications*. 2013;40(1):200–210.
- [19] Bogarín A, Romero C, Cerezo R, Sánchez-Santillán M. Clustering for improving educational process mining. *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge - LAK '14*. 2014;p. 11–15.
- [20] Bubeck S, von Luxburg U. Nearest Neighbor Clustering: A Baseline Method for Consistent Clustering with Arbitrary Objective Functions. *Journal of Machine Learning Research*. 2009;10:657–698.
- [21] Devi R, Kannan SR. Fuzzy Clustering Systems in Analyzing High Dimensional Database. 2015;p. 15–18.
- [22] Xu F, He J, Zhang S. Assessment on uncertainty of multi-scale representation of point cluster. In: 2010 18th International Conference on Geoinformatics; 2010. p. 1–6.
- [23] Saadatdoost R, Jafarkarimi H. Application of self organizing map for knowledge discovery based in higher education data. 2011 International Conference on Research and Innovation in Information Systems. 2011;p. 1–6. Available from: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6125693>.
- [24] Hruschka ER, Campello RJGB, Freitas AA, de Carvalho ACPLF. A Survey of Evolutionary Algorithms for Clustering. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*. 2009 March;39(2):133–155.

- [25] Jianping S, Ya Z. Personalized Search Based on Hybrid Clustering Algorithms. In: Education Technology and Computer Science, 2009. ETCS '09. First International Workshop on. vol. 2; 2009. p. 21–25.
- [26] Witten IH, Frank E. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann; 2005.
- [27] Lan X, Qian L, Zheng Y. Density K-means A New Algorithm for Centers Initialization for K-means;.
- [28] Peña-Ayala A. Educational data mining: A survey and a data mining-based analysis of recent works. Expert Systems with Applications. 2014;41(4 PART 1):1432–1462.
- [29] Forgy E. Cluster analysis of multivariate data: efficiency vs. interpretability of classification, Biometrics. 1965;21(768.).
- [30] Lloyd SP. Least squares quantization in PCM, IEEE Transactions on Information Theory. 28, 129–136; 1982.
- [31] MacQueen J. Some methods for classification and analysis of multivariate observations, Proceedings of the Fifth Berkeley Symposium On Mathematical Statistics and Probabilities. 1; 1967.
- [32] Hartigan JA, Wong MA. Algorithm AS 136 A K- Means Clustering Algorithm, Applied Statistics, 28(1). 100-108; 1979.
- [33] Ghosh S, Dubey SK. Comparative Analysis of K-Means and Fuzzy C-Means Algorithms. 2013;4(4):35–39.
- [34] Bezdek JC. “Pattern Recognition with Fuzzy Objective Function Algorithms”. New York: Plenum Press; 1981.
- [35] Full W, Ehrlich R, Bezdek J. *FCM : The Fuzzy C-means clustering algorithm*. 1984;10(2):191–203.
- [36] Yong Y, Chongxun Z, Pan L. “A Novel Fuzzy C-Means Clustering Algorithm for Image Thresholding”, Measurement Science Review. 2004;4(1).

- [37] Chattopadhyay S, Pratihar D, Sarkar S. A Comparative Study of Fuzzy C-Means Algorithm and Entropy-Based Fuzzy Clustering Algorithms. *Computing and Informatics*. 2012;30:701–720.
- [38] Stetco A, Zeng Xj, Keane J. Expert Systems with Applications Fuzzy C-means ++ : Fuzzy C-means with effective seeding initialization. *Expert Systems With Applications*. 2015;42(21):7541–7548.
- [39] Peña JM, Lozano JA, Larrañaga P. *An empirical comparison of four initialization methods for the K -Means algorithm*. 1999;20.
- [40] Kaufman L, Rousseeuw PJ. *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley, Canada; 1990.
- [41] Vesanto J, Alhoniemi E, Member S. Clustering of the Self-Organizing Map. 2000;11(3):586–600.
- [42] Davies DL, Bouldin DW. “A cluster separation measure. ” *IEEE Trans Patt Anal*. 1979 Apr;PAMI-1:224–227.
- [43] Hartigan J. *Clustering algorithms*. John Wiley and Sons, Inc; 1975.
- [44] Krzanowski W, Lai Y. A criterion for determining the number of groups in a data set using sum-of-squares clustering. *Biometrics* 44(1), 23–34; 1988.
- [45] Calinski T, Harabasz J. A dendrite method for cluster analysis. *Communications in Statistics-Theory and Methods* 3(1), 1–27; 1974.
- [46] Tibshirani R, Walther G, Hastie T. Estimating the number of clusters in a data set via the gap statistic. *J R Statist Soc B*. 2001;63(2).
- [47] Mohajer M, Englmeier Kh, Schmid VJ. A comparison of Gap statistic definitions with and without logarithm function. *arXiv preprint arXiv:11034767*. 2011;p. 1–11.
- [48] Teixeira MF. *Agrupamento e Visualização de Dados: Estudo e Implementações para a Ferramenta YADMT*. 2013;.

Apêndice

Figura 5.1: Autocorrelação do perfil com notas mais elevadas encontrados pelo grupo de 3241 usuários com 11 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Começo do Bloco

	D1	D2	D3	D4	D5	D6	D7	D8	D9
D1	1.0	0.429782421059	0.424385844888	0.376289900127	0.614754827292	0.360697357739	0.364909780203	0.402916278168	0.375310497265
D2	0.429782421059	1.0	0.458270552711	0.302122460928	0.346760344229	0.315857754086	0.450333144122	0.414783386142	0.405369648412
D3	0.424385844888	0.458270552711	1.0	0.456525730078	0.439154708999	0.389223401324	0.533231661139	0.423545899255	0.404093338298
D4	0.376289900127	0.302122460928	0.456525730078	1.0	0.412709995871	0.317348902619	0.425973683844	0.351168499236	0.340529716371
D5	0.614754827292	0.346760344229	0.439154708999	0.412709995871	1.0	0.308799463323	0.341490635104	0.425844565082	0.359815668633
D6	0.360697357739	0.315857754086	0.389223401324	0.317348902619	0.308799463323	1.0	0.352429612458	0.338673443016	0.315378980958
D7	0.364909780203	0.450333144122	0.533231661139	0.425973683844	0.341490635104	0.352429612458	1.0	0.405435893244	0.410123991428
D8	0.402916278168	0.414783386142	0.423545899255	0.351168499236	0.425844565082	0.338673443016	0.405435893244	1.0	0.473527945504
D9	0.375310497265	0.405369648412	0.404093338298	0.340529716371	0.359815668633	0.315378980958	0.410123991428	0.473527945504	1.0
D10	0.361465313235	0.241573552168	0.318853713168	0.358050520064	0.50774912623	0.276255602736	0.32765471572	0.293899429058	0.280641062804
D11	0.355832766196	0.247524877331	0.279804375655	0.458931023731	0.425613120536	0.266241432972	0.35319078927	0.322073735295	0.323651262119
D12	0.455722254992	0.29145319105	0.40733464493	0.338982275581	0.490156463757	0.275730337466	0.37750186126	0.320900209162	0.32197835023
D13	0.444025878604	0.299037154457	0.33712329276	0.455579037947	0.437860389269	0.28467895107	0.36439770437	0.328841992577	0.315833253249
D14	0.544237545296	0.434687069221	0.407797154696	0.420680329786	0.482883689242	0.377762054344	0.430102293813	0.408017001679	0.402504889502
D15	0.312953361609	0.309237112011	0.363371651144	0.271136898097	0.282289873505	0.340290240151	0.360096833568	0.368652674414	0.415728800143
D16	0.427486435319	0.262299605632	0.375190656641	0.23438071647	0.537860896487	0.264307643008	0.256824234654	0.377829331161	0.31815663965
D17	0.21473913953	0.357093128647	0.290442490602	0.200671708089	0.199319389975	0.188837860583	0.329851538542	0.359102633445	0.296222671727

Figura 5.2: Autocorrelação do perfil com notas mais elevadas encontrados pelo grupo de 3241 usuários com 11 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Continuação do Bloco

	D10	D11	D12	D13	D14	D15	D16	D17
D1	0.361465313235	0.355832766196	0.455722254992	0.444025878604	0.544237545296	0.312953361609	0.427486435319	0.21473913953
D2	0.241573552168	0.247524877331	0.29145319105	0.299037154457	0.434687069221	0.309237112011	0.262299605632	0.357093128647
D3	0.318853713168	0.279804375655	0.40733464493	0.33712329276	0.407797154696	0.363371651144	0.375190656641	0.290442490602
D4	0.358050520064	0.458931023731	0.338982275581	0.455579037947	0.420680329786	0.271136898097	0.23438071647	0.200671708089
D5	0.50774912623	0.425613120536	0.490156463757	0.437860389269	0.482883689242	0.282289873505	0.537860896487	0.199319389975
D6	0.276255602736	0.266241432972	0.275730337466	0.28467895107	0.377762054344	0.340290240151	0.264307643008	0.188837860583
D7	0.32765471572	0.35319078927	0.37750186126	0.36439770437	0.430102293813	0.360096833568	0.256824234654	0.329851538542
D8	0.293899429058	0.322073735295	0.320900209162	0.328841992577	0.408017001679	0.368652674414	0.377829331161	0.359102633445
D9	0.280641062804	0.323651262119	0.32197835023	0.315833253249	0.402504889502	0.415728800143	0.31815663965	0.296222671727
D10	1.0	0.584824822843	0.479036273259	0.409525499611	0.361392063365	0.351433766339	0.368620781827	0.0887065129758
D11	0.584824822843	1.0	0.392865528237	0.432813342435	0.35120164454	0.302240340449	0.284567655942	0.101299757142
D12	0.479036273259	0.392865528237	1.0	0.660493830085	0.435943485927	0.295697797046	0.428958153859	0.111651128322
D13	0.409525499611	0.432813342435	0.660493830085	1.0	0.445347465836	0.297539904135	0.320809842237	0.11811128298
D14	0.361392063365	0.35120164454	0.435943485927	0.445347465836	1.0	0.274550068858	0.369752999242	0.285621119112
D15	0.351433766339	0.302240340449	0.295697797046	0.297539904135	0.274550068858	1.0	0.279129015282	0.216188956599
D16	0.368620781827	0.284567655942	0.428958153859	0.320809842237	0.369752999242	0.279129015282	1.0	0.252905388941
D17	0.0887065129758	0.101299757142	0.111651128322	0.11811128298	0.285621119112	0.216188956599	0.252905388941	1.0

Figura 5.3: Autocorrelação do perfil com notas mais baixas encontrados pelo grupo de 3241 usuários com 11 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Começo do Bloco

	D1	D2	D3	D4	D5	D6	D7	D8	D9
D1	1.0	0.261546135002	0.365464420312	0.348939653996	0.531175012602	0.327941809956	0.228601950248	0.345877474103	0.17819925841
D2	0.261546135002	1.0	0.31220968578	0.213722350831	0.205720195496	0.333079666543	0.266954338706	0.328094960854	0.199820919025
D3	0.365464420312	0.31220968578	1.0	0.425539899836	0.290470254102	0.382836327921	0.333931160627	0.325346544275	0.236182458131
D4	0.348939653996	0.213722350831	0.425539899836	1.0	0.269176803892	0.334662264769	0.286860685961	0.281160083078	0.191298981937
D5	0.531175012602	0.205720195496	0.290470254102	0.269176803892	1.0	0.229319556728	0.170131712857	0.303355468328	0.17644155866
D6	0.327941809956	0.333079666543	0.382836327921	0.334662264769	0.229319556728	1.0	0.345693880057	0.363448039326	0.262420193258
D7	0.228601950248	0.266954338706	0.333931160627	0.286860685961	0.170131712857	0.345693880057	1.0	0.249275603745	0.268653805471
D8	0.345877474103	0.328094960854	0.325346544275	0.281160083078	0.303355468328	0.363448039326	0.249275603745	1.0	0.189312974711
D9	0.17819925841	0.199820919025	0.236182458131	0.191298981937	0.17644155866	0.262420193258	0.268653805471	0.189312974711	1.0
D10	0.255605154739	0.184812742568	0.280924648383	0.296044771612	0.29168659062	0.180614288501	0.151254948279	0.163134953278	0.21966594602
D11	0.275994248535	0.238271935503	0.29692692129	0.356708856505	0.277427925607	0.226657417828	0.221882882668	0.234230633664	0.117630645977
D12	0.24946238256	0.227576528758	0.262525946069	0.233057545774	0.402274784976	0.235914692285	0.232282142186	0.167530094824	0.18115983577
D13	0.28449186269	0.141637633283	0.237721837006	0.333448521237	0.362072046254	0.24695890056	0.237541325771	0.180589788991	0.132058244651
D14	0.368225699307	0.26919114913	0.300158878751	0.335856495268	0.37857160659	0.312333813707	0.306000862181	0.300377890773	0.268446245227
D15	0.234605484541	0.265033250271	0.285099550723	0.25926529771	0.226567034588	0.221844493336	0.250961419476	0.286750661566	0.215427731556
D16	0.424075380512	0.189811506057	0.332272431404	0.307067003605	0.493987293907	0.292328594843	0.18484602071	0.318333923854	0.241182031275
D17	0.131971306944	0.243344911345	0.166604224553	0.176038336491	0.0911070669015	0.178167550743	0.137612994832	0.233748758452	0.125249097828

Figura 5.4: Autocorrelação do perfil com notas mais baixas encontrados pelo grupo de 3241 usuários com 11 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Continuação do Bloco

	D10	D11	D12	D13	D14	D15	D16	D17
D1	0.255605154739	0.275994248535	0.24946238256	0.28449186269	0.368225699307	0.234605484541	0.424075380512	0.131971306944
D2	0.184812742568	0.238271935503	0.227576528758	0.141637633283	0.26919114913	0.265033250271	0.189811506057	0.243344911345
D3	0.280924648383	0.29692692129	0.262525946069	0.237721837006	0.300158878751	0.285099550723	0.332272431404	0.166604224553
D4	0.296044771612	0.356708856505	0.233057545774	0.333448521237	0.335856495268	0.25926529771	0.307067003605	0.176038336491
D5	0.29168659062	0.277427925607	0.402274784976	0.362072046254	0.37857160659	0.226567034588	0.493987293907	0.0911070669015
D6	0.180614288501	0.226657417828	0.235914692285	0.24695890056	0.312333813707	0.221844493336	0.292328594843	0.178167550743
D7	0.151254948279	0.221882882668	0.232282142186	0.237541325771	0.306000862181	0.250961419476	0.18484602071	0.137612994832
D8	0.163134953278	0.234230633664	0.167530094824	0.180589788991	0.300377890773	0.286750661566	0.318333923854	0.233748758452
D9	0.21966594602	0.117630645977	0.18115983577	0.132058244651	0.268446245227	0.215427731556	0.241182031275	0.125249097828
D10	1.0	0.404474555248	0.40506919971	0.309357162401	0.276697498557	0.312714540621	0.365154079393	0.0703419162248
D11	0.404474555248	1.0	0.208597677383	0.314109111137	0.24593043942	0.233802235846	0.28486572135	0.034160214279
D12	0.40506919971	0.208597677383	1.0	0.516551319306	0.312260376637	0.258925230315	0.413271747477	-0.00382812224873
D13	0.309357162401	0.314109111137	0.516551319306	1.0	0.341444156943	0.208640263764	0.381007255189	0.00382950023095
D14	0.276697498557	0.24593043942	0.312260376637	0.341444156943	1.0	0.213667552626	0.327499809787	0.130344700027
D15	0.312714540621	0.233802235846	0.258925230315	0.208640263764	0.213667552626	1.0	0.260402142235	0.162749384154
D16	0.365154079393	0.28486572135	0.413271747477	0.381007255189	0.327499809787	0.260402142235	1.0	0.180254814525
D17	0.0703419162248	0.034160214279	-0.00382812224873	0.00382950023095	0.130344700027	0.162749384154	0.180254814525	1.0

Figura 5.5: Autocorrelação do perfil com notas mais elevadas encontrados pelo grupo de 3241 usuários com 22 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Começo do Bloco

	D1	D2	D3	D4	D5	D6	D7	D8	D9
D1	1.0	0.427278959503	0.430328639383	0.375047429347	0.614093934362	0.351186890256	0.362884847427	0.399528350275	0.374685595095
D2	0.427278959503	1.0	0.473697857898	0.294097619303	0.34202153753	0.3172030181	0.460809333727	0.424463382199	0.416171189976
D3	0.430328639383	0.473697857898	1.0	0.46293700156	0.428062857768	0.38553090719	0.534277884957	0.421314327633	0.423614455559
D4	0.375047429347	0.294097619303	0.46293700156	1.0	0.405046468913	0.310963272862	0.423288101272	0.344633816709	0.340472373355
D5	0.614093934362	0.34202153753	0.428062857768	0.405046468913	1.0	0.284428499212	0.323339059788	0.41061351893	0.346330787226
D6	0.351186890256	0.3172030181	0.38553090719	0.310963272862	0.284428499212	1.0	0.343270365055	0.321003699954	0.318237862856
D7	0.362884847427	0.460809333727	0.534277884957	0.423288101272	0.323339059788	0.343270365055	1.0	0.395220202969	0.420399442441
D8	0.399528350275	0.424463382199	0.421314327633	0.344633816709	0.41061351893	0.321003699954	0.395220202969	1.0	0.476783122523
D9	0.374685595095	0.416171189976	0.423614455559	0.340472373355	0.346330787226	0.318237862856	0.420399442441	0.476783122523	1.0
D10	0.356695521219	0.250813491371	0.317327020398	0.360904211201	0.49030773148	0.267120166285	0.319161143019	0.278489568838	0.283319928719
D11	0.354962721253	0.241472224669	0.277029110211	0.455935569984	0.415530708368	0.257250213732	0.343480624162	0.312907208119	0.320721909095
D12	0.45775029611	0.306354984505	0.408537069394	0.340032187651	0.476898480561	0.267873158826	0.376288147329	0.313688546463	0.324161621491
D13	0.432676506924	0.29556834428	0.336287265951	0.454096843467	0.421178082937	0.270225961014	0.357934842424	0.305854129951	0.305822641368
D14	0.538036971541	0.420874940085	0.393324808359	0.413050781455	0.473775944877	0.367754767818	0.412428132415	0.396753039489	0.391231517282
D15	0.304485210686	0.326920746519	0.370604524194	0.273778240144	0.278297172308	0.350392102473	0.373348184121	0.365760227589	0.436671332451
D16	0.416863459334	0.261878819996	0.369640982671	0.224606994925	0.527660304077	0.249837821259	0.2456192258	0.366961888753	0.31346638419
D17	0.233160827936	0.385963316693	0.32451486696	0.221110708164	0.212784120609	0.206877046514	0.354627022097	0.377201089091	0.334733956782

Figura 5.6: Autocorrelação do perfil com notas mais elevadas encontrados pelo grupo de 3241 usuários com 22 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Continuação do Bloco

	D10	D11	D12	D13	D14	D15	D16	D17
D1	0.356695521219	0.354962721253	0.45775029611	0.432676506924	0.538036971541	0.304485210686	0.416863459334	0.233160827936
D2	0.250813491371	0.241472224669	0.306354984505	0.29556834428	0.420874940085	0.326920746519	0.261878819996	0.385963316693
D3	0.317327020398	0.277029110211	0.408537069394	0.336287265951	0.393324808359	0.370604524194	0.369640982671	0.32451486696
D4	0.360904211201	0.455935569984	0.340032187651	0.454096843467	0.413050781455	0.273778240144	0.224606994925	0.221110708164
D5	0.49030773148	0.415530708368	0.476898480561	0.421178082937	0.473775944877	0.278297172308	0.527660304077	0.212784120609
D6	0.267120166285	0.257250213732	0.267873158826	0.270225961014	0.367754767818	0.350392102473	0.249837821259	0.206877046514
D7	0.319161143019	0.343480624162	0.376288147329	0.357934842424	0.412428132415	0.373348184121	0.2456192258	0.354627022097
D8	0.278489568838	0.312907208119	0.313688546463	0.305854129951	0.396753039489	0.365760227589	0.366961888753	0.377201089091
D9	0.283319928719	0.320721909095	0.324161621491	0.305822641368	0.391231517282	0.436671332451	0.31346638419	0.334733956782
D10	1.0	0.583997676206	0.471758657937	0.402291191405	0.349602179022	0.368621641822	0.354311626575	0.110396543601
D11	0.583997676206	1.0	0.387462452853	0.429952631	0.345378192803	0.303691319278	0.275602963337	0.123038253967
D12	0.471758657937	0.387462452853	1.0	0.658097068902	0.425944174303	0.313410990688	0.42349843592	0.138954459143
D13	0.402291191405	0.429952631	0.658097068902	1.0	0.43408703615	0.300219901704	0.302189947161	0.129446394002
D14	0.349602179022	0.345378192803	0.425944174303	0.43408703615	1.0	0.264663406704	0.354856466147	0.289041417662
D15	0.368621641822	0.303691319278	0.313410990688	0.300219901704	0.264663406704	1.0	0.270782138827	0.226611609353
D16	0.354311626575	0.275602963337	0.42349843592	0.302189947161	0.354856466147	0.270782138827	1.0	0.263436020722
D17	0.110396543601	0.123038253967	0.138954459143	0.129446394002	0.289041417662	0.226611609353	0.263436020722	1.0

Figura 5.7: Autocorrelação do perfil com notas mais baixas encontrados pelo grupo de 3241 usuários com 22 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Começo do Bloco

	D1	D2	D3	D4	D5	D6	D7	D8	D9
D1	1.0	0.271602648408	0.364650830335	0.339622009242	0.522467904627	0.327458403559	0.221220643017	0.334666769726	0.196450368076
D2	0.271602648408	1.0	0.318865423074	0.227644009459	0.207182653446	0.334093946431	0.265855092082	0.322928338855	0.208101230349
D3	0.364650830335	0.318865423074	1.0	0.425801689156	0.280442907262	0.378400225607	0.338690286688	0.319349186481	0.236545100575
D4	0.339622009242	0.227644009459	0.425801689156	1.0	0.252861372561	0.327829966152	0.28903387419	0.269062480732	0.205221632932
D5	0.522467904627	0.207182653446	0.280442907262	0.252861372561	1.0	0.224227048247	0.156055312864	0.293304593052	0.189951814505
D6	0.327458403559	0.334093946431	0.378400225607	0.327829966152	0.224227048247	1.0	0.34667855693	0.364151198282	0.263967120733
D7	0.221220643017	0.265855092082	0.338690286688	0.28903387419	0.156055312864	0.34667855693	1.0	0.240400968129	0.268864610633
D8	0.334666769726	0.322928338855	0.319349186481	0.269062480732	0.293304593052	0.364151198282	0.240400968129	1.0	0.194962537888
D9	0.196450368076	0.208101230349	0.236545100575	0.205221632932	0.189951814505	0.263967120733	0.268864610633	0.194962537888	1.0
D10	0.247747548255	0.179029029392	0.272102083022	0.28102356102	0.282599200321	0.170109984004	0.147546919376	0.157026870084	0.221662901455
D11	0.268063098632	0.248354174179	0.293496155623	0.357732147372	0.265242602415	0.220263716228	0.223246900422	0.223037368452	0.129802940969
D12	0.244541912846	0.216166682588	0.260968596547	0.228938304826	0.398568406832	0.233920659457	0.228489042891	0.160549232115	0.175735956781
D13	0.273983961676	0.137256741245	0.22048814822	0.316684632137	0.348970643171	0.238055630618	0.217910685506	0.175372667296	0.139236746486
D14	0.360153852517	0.277571104139	0.300923138449	0.326456494277	0.357579814008	0.302179911252	0.294861383701	0.283793441217	0.274788704719
D15	0.249880640123	0.266479446011	0.286053818595	0.264995743462	0.230359242227	0.225100244956	0.241022708541	0.292395748516	0.218692755166
D16	0.4162569742	0.186262019702	0.322818329972	0.292791427275	0.486576850124	0.287924083955	0.170528871487	0.310460174669	0.245534291532
D17	0.156155484553	0.25557976433	0.178511997728	0.193618404789	0.105726774194	0.194789060397	0.14722723535	0.240377442693	0.144036282618

Figura 5.8: Autocorrelação do perfil com notas mais baixas encontrados pelo grupo de 3241 usuários com 22 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Continuação do Bloco

	D10	D11	D12	D13	D14	D15	D16	D17
D1	0.247747548255	0.268063098632	0.244541912846	0.273983961676	0.360153852517	0.249880640123	0.4162569742	0.156155484553
D2	0.179029029392	0.248354174179	0.216166682588	0.137256741245	0.277571104139	0.266479446011	0.186262019702	0.25557976433
D3	0.272102083022	0.293496155623	0.260968596547	0.22048814822	0.300923138449	0.286053818595	0.322818329972	0.178511997728
D4	0.28102356102	0.357732147372	0.228938304826	0.316684632137	0.326456494277	0.264995743462	0.292791427275	0.193618404789
D5	0.282599200321	0.265242602415	0.398568406832	0.348970643171	0.357579814008	0.230359242227	0.486576850124	0.105726774194
D6	0.170109984004	0.220263716228	0.233920659457	0.238055630618	0.302179911252	0.225100244956	0.287924083955	0.194789060397
D7	0.147546919376	0.223246900422	0.228489042891	0.217910685506	0.294861383701	0.241022708541	0.170528871487	0.14722723535
D8	0.157026870084	0.223037368452	0.160549232115	0.175372667296	0.283793441217	0.292395748516	0.310460174669	0.240377442693
D9	0.221662901455	0.129802940969	0.175735956781	0.139236746486	0.274788704719	0.218692755166	0.245534291532	0.144036282618
D10	1.0	0.391534869829	0.411427176169	0.293201539181	0.258072807631	0.30575980803	0.363481383083	0.085096163814
D11	0.391534869829	1.0	0.205216111571	0.298610221472	0.231255793312	0.238117417328	0.271367192086	0.0505599859677
D12	0.411427176169	0.205216111571	1.0	0.505631679557	0.299628926572	0.245767782515	0.409177904186	0.00122058196374
D13	0.293201539181	0.298610221472	0.505631679557	1.0	0.326877960148	0.206302274235	0.373172270535	0.0179806886473
D14	0.258072807631	0.231255793312	0.299628926572	0.326877960148	1.0	0.221795270265	0.312917549768	0.149270549799
D15	0.30575980803	0.238117417328	0.245767782515	0.206302274235	0.221795270265	1.0	0.264189422197	0.178268852926
D16	0.363481383083	0.271367192086	0.409177904186	0.373172270535	0.312917549768	0.264189422197	1.0	0.191037255382
D17	0.085096163814	0.0505599859677	0.00122058196374	0.0179806886473	0.149270549799	0.178268852926	0.191037255382	1.0

Figura 5.9: Autocorrelação do perfil com notas mais elevadas encontrados pelo grupo de 947 usuários com 11 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Começo do Bloco

	D1	D2	D3	D4	D5	D6	D7	D8	D9
D1	1.0	0.254231653017	0.372442097037	0.342936685078	0.51308895611	0.223923530885	0.31333551289	0.312152618851	0.259939568053
D2	0.254231653017	1.0	0.289905139191	0.187762900256	0.189107950179	0.191259653847	0.275864122705	0.267391990381	0.252290408148
D3	0.372442097037	0.289905139191	1.0	0.384535505922	0.369013054436	0.253489078153	0.425526352319	0.305449266682	0.29209921797
D4	0.342936685078	0.187762900256	0.384535505922	1.0	0.372084313844	0.238703519209	0.32880636792	0.21299088702	0.231645851758
D5	0.51308895611	0.189107950179	0.369013054436	0.372084313844	1.0	0.210610211448	0.29913111563	0.302476893973	0.269424007176
D6	0.223923530885	0.191259653847	0.253489078153	0.238703519209	0.210610211448	1.0	0.223293846401	0.190177511026	0.162997578686
D7	0.31333551289	0.275864122705	0.425526352319	0.32880636792	0.29913111563	0.223293846401	1.0	0.259882268191	0.311220961821
D8	0.312152618851	0.267391990381	0.305449266682	0.21299088702	0.302476893973	0.190177511026	0.259882268191	1.0	0.286825366648
D9	0.259939568053	0.252290408148	0.29209921797	0.231645851758	0.269424007176	0.162997578686	0.311220961821	0.286825366648	1.0
D10	0.338047344281	0.163899539561	0.333999531026	0.333757712944	0.459952286311	0.159000434756	0.254045704837	0.201660498588	0.210514277621
D11	0.261424306743	0.153010417756	0.255064052824	0.379463002715	0.321799386295	0.158306799724	0.233350228421	0.183758306283	0.147918808795
D12	0.351810902334	0.164064260126	0.346550513924	0.268644260856	0.446265150534	0.162238334449	0.310790515627	0.239012587719	0.262060302667
D13	0.322975405335	0.210640343889	0.324628195347	0.371476726833	0.394393194846	0.198583662268	0.326249562733	0.220454612221	0.216714692764
D14	0.445091346471	0.2625053956	0.364221827121	0.358286255417	0.413319964004	0.229335828429	0.386509852486	0.283579250038	0.281942720764
D15	0.241020264113	0.231380081778	0.291247959064	0.230881050091	0.258191219667	0.182641551108	0.26756247255	0.265196173428	0.272945206814
D16	0.374581518184	0.189030045644	0.355218290113	0.241869558005	0.470264828642	0.206842561598	0.25432223209	0.307487313474	0.283464870574
D17	0.271341853881	0.276142685752	0.251977906736	0.192557517451	0.239143674265	0.152630999508	0.301214239337	0.282423285011	0.225719136367

Figura 5.10: Autocorrelação do perfil com notas mais elevadas encontrados pelo grupo de 947 usuários com 11 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Continuação do Bloco

	D10	D11	D12	D13	D14	D15	D16	D17
D1	0.338047344281	0.261424306743	0.351810902334	0.322975405335	0.445091346471	0.241020264113	0.374581518184	0.271341853881
D2	0.163899539561	0.153010417756	0.164064260126	0.210640343889	0.2625053956	0.231380081778	0.189030045644	0.276142685752
D3	0.333999531026	0.255064052824	0.346550513924	0.324628195347	0.364221827121	0.291247959064	0.355218290113	0.251977906736
D4	0.333757712944	0.379463002715	0.268644260856	0.371476726833	0.358286255417	0.230881050091	0.241869558005	0.192557517451
D5	0.459952286311	0.321799386295	0.446265150534	0.394393194846	0.413319964004	0.258191219667	0.470264828642	0.239143674265
D6	0.159000434756	0.158306799724	0.162238334449	0.198583662268	0.229335828429	0.182641551108	0.206842561598	0.152630999508
D7	0.254045704837	0.233350228421	0.310790515627	0.326249562733	0.386509852486	0.26756247255	0.25432223209	0.301214239337
D8	0.201660498588	0.183758306283	0.239012587719	0.220454612221	0.283579250038	0.265196173428	0.307487313474	0.282423285011
D9	0.210514277621	0.147918808795	0.262060302667	0.216714692764	0.281942720764	0.272945206814	0.283464870574	0.225719136367
D10	1.0	0.445437498529	0.383746404799	0.353246703299	0.303684321904	0.280326070631	0.329150787192	0.152955638352
D11	0.445437498529	1.0	0.260166397095	0.320966836369	0.28751745965	0.22546696646	0.204208290471	0.0946212004549
D12	0.383746404799	0.260166397095	1.0	0.576187489381	0.361914389868	0.23496202886	0.431499023469	0.133078259246
D13	0.353246703299	0.320966836369	0.576187489381	1.0	0.375383741812	0.268295004728	0.35196501557	0.112378521504
D14	0.303684321904	0.28751745965	0.361914389868	0.375383741812	1.0	0.26037091445	0.324249038119	0.248421034101
D15	0.280326070631	0.22546696646	0.23496202886	0.268295004728	0.26037091445	1.0	0.273507198587	0.192261498372
D16	0.329150787192	0.204208290471	0.431499023469	0.35196501557	0.324249038119	0.273507198587	1.0	0.239189788834
D17	0.152955638352	0.0946212004549	0.133078259246	0.112378521504	0.248421034101	0.192261498372	0.239189788834	1.0

Figura 5.11: Autocorrelação do perfil com notas mais baixas encontrados pelo grupo de 947 usuários com 11 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Começo do Bloco

	D1	D2	D3	D4	D5	D6	D7	D8	D9
D1	1.0	0.190849202153	0.319353695486	0.291951870908	0.453073628104	0.267173730716	0.189045225799	0.248192638565	0.162032958134
D2	0.190849202153	1.0	0.169590973436	0.126606491215	0.119927005712	0.203174139818	0.109709388091	0.198296515033	0.117788094718
D3	0.319353695486	0.169590973436	1.0	0.324079568066	0.273665335475	0.239774346053	0.250308865428	0.226591594968	0.143000082509
D4	0.291951870908	0.126606491215	0.324079568066	1.0	0.244833028348	0.211659797812	0.209362620496	0.187917959597	0.0452298365875
D5	0.453073628104	0.119927005712	0.273665335475	0.244833028348	1.0	0.152680236787	0.161230646539	0.260767898488	0.139812600178
D6	0.267173730716	0.203174139818	0.239774346053	0.211659797812	0.152680236787	1.0	0.174902627864	0.236001931232	0.130308134302
D7	0.189045225799	0.109709388091	0.250308865428	0.209362620496	0.161230646539	0.174902627864	1.0	0.189418084351	0.133804169222
D8	0.248192638565	0.198296515033	0.226591594968	0.187917959597	0.260767898488	0.236001931232	0.189418084351	1.0	0.147798986011
D9	0.162032958134	0.117788094718	0.143000082509	0.0452298365875	0.139812600178	0.130308134302	0.133804169222	0.147798986011	1.0
D10	0.24625914491	0.073166183056	0.219501443301	0.225181987652	0.312274729097	0.111508097493	0.0934254068355	0.132053905588	0.108876037203
D11	0.217662687296	0.103862901189	0.253463355547	0.290618118825	0.24562220293	0.110744974265	0.151123426428	0.128312350019	0.0523001191049
D12	0.277207444278	0.129234633964	0.274440518526	0.184021690719	0.352459115646	0.185813022968	0.200749913164	0.140147383746	0.134609608376
D13	0.312961116431	0.112023811726	0.227179286455	0.290332849197	0.349085553654	0.160005912859	0.202700127299	0.140247767817	0.0943594740028
D14	0.384555993109	0.18324266312	0.268083185778	0.249457719598	0.336794224398	0.187008751996	0.203813479705	0.248877426788	0.153126854509
D15	0.224131925445	0.167354762754	0.26074438449	0.231752383805	0.214047037484	0.21676464389	0.215818848951	0.213845244531	0.129967711646
D16	0.409877015845	0.117053820327	0.31792565273	0.269089238787	0.487100053758	0.24467407332	0.165618249487	0.264696962729	0.16611328487
D17	0.191174132695	0.177345753094	0.216491673111	0.169213313352	0.141985830881	0.211453296513	0.08435744373	0.174345479341	0.0999868799923

Figura 5.12: Autocorrelação do perfil com notas mais baixas encontrados pelo grupo de 947 usuários com 11 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Continuação do Bloco

	D10	D11	D12	D13	D14	D15	D16	D17
D1	0.24625914491	0.217662687296	0.277207444278	0.312961116431	0.384555993109	0.224131925445	0.409877015845	0.191174132695
D2	0.073166183056	0.103862901189	0.129234633964	0.112023811726	0.18324266312	0.167354762754	0.117053820327	0.177345753094
D3	0.219501443301	0.253463355547	0.274440518526	0.227179286455	0.268083185778	0.26074438449	0.31792565273	0.216491673111
D4	0.225181987652	0.290618118825	0.184021690719	0.290332849197	0.249457719598	0.231752383805	0.269089238787	0.169213313352
D5	0.312274729097	0.24562220293	0.352459115646	0.349085553654	0.336794224398	0.214047037484	0.487100053758	0.141985830881
D6	0.111508097493	0.110744974265	0.185813022968	0.160005912859	0.187008751996	0.21676464389	0.24467407332	0.211453296513
D7	0.0934254068355	0.151123426428	0.200749913164	0.202700127299	0.203813479705	0.215818848951	0.165618249487	0.08435744373
D8	0.132053905588	0.128312350019	0.140147383746	0.140247767817	0.248877426788	0.213845244531	0.264696962729	0.174345479341
D9	0.108876037203	0.0523001191049	0.134609608376	0.0943594740028	0.153126854509	0.129967711646	0.16611328487	0.0999868799923
D10	1.0	0.317583778709	0.3423095143	0.287872821419	0.236496046356	0.200543419394	0.352489677231	0.112517051316
D11	0.317583778709	1.0	0.198598062724	0.253698357495	0.207499186244	0.185775730592	0.257346909907	0.0464272782517
D12	0.3423095143	0.198598062724	1.0	0.498707126221	0.298496620015	0.21124880522	0.41086113203	0.1015996481
D13	0.287872821419	0.253698357495	0.498707126221	1.0	0.304928610001	0.211586486942	0.395310777032	0.0757359536941
D14	0.236496046356	0.207499186244	0.298496620015	0.304928610001	1.0	0.1982136437	0.310556897642	0.156362095706
D15	0.200543419394	0.185775730592	0.21124880522	0.211586486942	0.1982136437	1.0	0.272006670459	0.118863951071
D16	0.352489677231	0.257346909907	0.41086113203	0.395310777032	0.310556897642	0.272006670459	1.0	0.188472085204
D17	0.112517051316	0.0464272782517	0.1015996481	0.0757359536941	0.156362095706	0.118863951071	0.188472085204	1.0

Figura 5.13: Autocorrelação do perfil com notas mais baixas encontrados pelo grupo de 947 usuários com 22 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Começo do Bloco

	D1	D2	D3	D4	D5	D6	D7	D8	D9
D1	1.0	0.191409198137	0.344378927071	0.32567597953	0.474780583699	0.331022547473	0.202026536266	0.262196415282	0.152066095506
D2	0.191409198137	1.0	0.174612855099	0.124742664894	0.119647113187	0.234355745956	0.100992667439	0.194808058414	0.115086392407
D3	0.344378927071	0.174612855099	1.0	0.335804277285	0.290869571597	0.29050271336	0.255286823941	0.232798063967	0.136529934094
D4	0.32567597953	0.124742664894	0.335804277285	1.0	0.286233290255	0.278820587667	0.218963119652	0.200346859798	0.0457350824668
D5	0.474780583699	0.119647113187	0.290869571597	0.286233290255	1.0	0.228916264267	0.16636871142	0.277811580585	0.138363857869
D6	0.331022547473	0.234355745956	0.29050271336	0.278820587667	0.228916264267	1.0	0.213818415627	0.277808055595	0.151979944648
D7	0.202026536266	0.100992667439	0.255286823941	0.218963119652	0.16636871142	0.213818415627	1.0	0.17680484341	0.114314417766
D8	0.262196415282	0.194808058414	0.232798063967	0.200346859798	0.277811580585	0.277808055595	0.17680484341	1.0	0.132350404926
D9	0.152066095506	0.115086392407	0.136529934094	0.0457350824668	0.138363857869	0.151979944648	0.114314417766	0.132350404926	1.0
D10	0.282916704216	0.0710234619609	0.234537622334	0.252429752988	0.334567784988	0.167684961044	0.102554055049	0.147837331468	0.0963482029447
D11	0.267286463363	0.113998304462	0.267339138748	0.315872501223	0.284842276388	0.169425509172	0.16888504381	0.155479116875	0.0651100420133
D12	0.249992140327	0.114000467741	0.247852300907	0.170018690742	0.323364386286	0.199743363399	0.16010209276	0.120702927695	0.103487428603
D13	0.293452017361	0.0952163674408	0.215376211414	0.292086260904	0.331735016676	0.19191326524	0.188016338581	0.117848720926	0.0579611667066
D14	0.330837127568	0.135021907054	0.230282182192	0.236112052162	0.292376560171	0.194684213945	0.150715108416	0.195438640457	0.0920301351528
D15	0.211837931777	0.161340152215	0.250245375332	0.23090434238	0.198238385074	0.21654087992	0.190293430222	0.200702201906	0.089691226077
D16	0.40235163806	0.0932898294407	0.302386882287	0.267612362492	0.478445142523	0.276267060122	0.150206869712	0.256149045132	0.125336475468
D17	0.152911945769	0.13308069356	0.170059764933	0.137738602715	0.108033550792	0.195942518705	0.0518023953548	0.130318935568	0.0398284390759

Figura 5.14: Autocorrelação do perfil com notas mais baixas encontrados pelo grupo de 947 usuários com 22 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Continuação do Bloco

	D10	D11	D12	D13	D14	D15	D16	D17
D1	0.282916704216	0.267286463363	0.249992140327	0.293452017361	0.330837127568	0.211837931777	0.40235163806	0.152911945769
D2	0.0710234619609	0.113998304462	0.114000467741	0.0952163674408	0.135021907054	0.161340152215	0.0932898294407	0.13308069356
D3	0.234537622334	0.267339138748	0.247852300907	0.215376211414	0.230282182192	0.250245375332	0.302386882287	0.170059764933
D4	0.252429752988	0.315872501223	0.170018690742	0.292086260904	0.236112052162	0.23090434238	0.267612362492	0.137738602715
D5	0.334567784988	0.284842276388	0.323364386286	0.331735016676	0.292376560171	0.198238385074	0.478445142523	0.108033550792
D6	0.167684961044	0.169425509172	0.199743363399	0.19191326524	0.194684213945	0.21654087992	0.276267060122	0.195942518705
D7	0.102554055049	0.16888504381	0.16010209276	0.188016338581	0.150715108416	0.190293430222	0.150206869712	0.0518023953548
D8	0.147837331468	0.155479116875	0.120702927695	0.117848720926	0.195438640457	0.200702201906	0.256149045132	0.130318935568
D9	0.0963482029447	0.0651100420133	0.103487428603	0.0579611667066	0.0920301351528	0.089691226077	0.125336475468	0.0398284390759
D10	1.0	0.351846565918	0.320306026858	0.267386375404	0.194946777126	0.188434963922	0.338518006962	0.0753547807752
D11	0.351846565918	1.0	0.183647304316	0.243830080804	0.194203977697	0.183390811235	0.264827135085	0.0336002953562
D12	0.320306026858	0.183647304316	1.0	0.474291180825	0.240538743331	0.19207420831	0.377953328689	0.0429793808553
D13	0.267386375404	0.243830080804	0.474291180825	1.0	0.264198453846	0.194502212192	0.366914449532	0.0279335840293
D14	0.194946777126	0.194203977697	0.240538743331	0.264198453846	1.0	0.153624046708	0.250496284963	0.0827503967298
D15	0.188434963922	0.183390811235	0.19207420831	0.194502212192	0.153624046708	1.0	0.250263625321	0.0882252345473
D16	0.338518006962	0.264827135085	0.377953328689	0.366914449532	0.250496284963	0.250263625321	1.0	0.152090725865
D17	0.0753547807752	0.0336002953562	0.0429793808553	0.0279335840293	0.0827503967298	0.0882252345473	0.152090725865	1.0

Figura 5.15: Autocorrelação do perfil com notas mais elevadas encontrados pelo grupo de 947 usuários com 22 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Começo do Bloco

	D1	D2	D3	D4	D5	D6	D7	D8	D9
D1	1.0	0.269987203893	0.391021533081	0.370571736162	0.533869035974	0.291560272856	0.322339505854	0.335905127014	0.26178224919
D2	0.269987203893	1.0	0.292232706523	0.216080914557	0.204844273517	0.22047427979	0.274160120322	0.269231680191	0.225897799314
D3	0.391021533081	0.292232706523	1.0	0.406465459362	0.387046819581	0.321118429335	0.415091316028	0.304909273452	0.26974372476
D4	0.370571736162	0.216080914557	0.406465459362	1.0	0.376505735786	0.30210297011	0.327850512759	0.240214048181	0.234709560642
D5	0.533869035974	0.204844273517	0.387046819581	0.376505735786	1.0	0.261557241981	0.301335827148	0.305601271829	0.25724422263
D6	0.291560272856	0.22047427979	0.321118429335	0.30210297011	0.261557241981	1.0	0.270466486724	0.244055042109	0.210322558174
D7	0.322339505854	0.274160120322	0.415091316028	0.327850512759	0.301335827148	0.270466486724	1.0	0.266202751539	0.297286459669
D8	0.335905127014	0.269231680191	0.304909273452	0.240214048181	0.305601271829	0.244055042109	0.266202751539	1.0	0.282330540982
D9	0.26178224919	0.225897799314	0.26974372476	0.234709560642	0.25724422263	0.210322558174	0.297286459669	0.282330540982	1.0
D10	0.352922708058	0.172716170532	0.346192623995	0.349434281506	0.464040641821	0.227072436246	0.251954085015	0.205929750595	0.214544858219
D11	0.302280668074	0.179556056557	0.309289574027	0.417463638698	0.355556017803	0.241058751321	0.248859096931	0.21576670272	0.156660734851
D12	0.335709800112	0.143226546904	0.325710345662	0.25953256792	0.440962051412	0.181494314823	0.297325153377	0.204295727267	0.222883635156
D13	0.314987203841	0.192169743935	0.300731747205	0.357848800856	0.384133743081	0.213778625579	0.304244265927	0.202540661413	0.187612214881
D14	0.4247898093	0.238372601956	0.309114569049	0.309911029284	0.393925360555	0.233137107218	0.338143703666	0.249662254731	0.222661767248
D15	0.24088264417	0.206287437221	0.272237595786	0.218338633012	0.245079747128	0.216531641992	0.25824918117	0.251395437311	0.264321023465
D16	0.372369607796	0.184813418512	0.342149021246	0.232334082192	0.467880969134	0.224300756428	0.226420691741	0.287621620338	0.263551463265
D17	0.22685458788	0.259513811226	0.195872595144	0.147522893763	0.178313449076	0.147889104051	0.244471712303	0.235117406342	0.177690593846

Figura 5.16: Autocorrelação do perfil com notas mais elevadas encontrados pelo grupo de 947 usuários com 22 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Continuação do Bloco

	D10	D11	D12	D13	D14	D15	D16	D17
D1	0.352922708058	0.302280668074	0.335709800112	0.314987203841	0.4247898093	0.24088264417	0.372369607796	0.22685458788
D2	0.172716170532	0.179556056557	0.143226546904	0.192169743935	0.238372601956	0.206287437221	0.184813418512	0.259513811226
D3	0.346192623995	0.309289574027	0.325710345662	0.300731747205	0.309114569049	0.272237595786	0.342149021246	0.195872595144
D4	0.349434281506	0.417463638698	0.25953256792	0.357848800856	0.309911029284	0.218338633012	0.232334082192	0.147522893763
D5	0.464040641821	0.355556017803	0.440962051412	0.384133743081	0.393925360555	0.245079747128	0.467880969134	0.178313449076
D6	0.227072436246	0.241058751321	0.181494314823	0.213778625579	0.233137107218	0.216531641992	0.224300756428	0.147889104051
D7	0.251954085015	0.248859096931	0.297325153377	0.304244265927	0.338143703666	0.25824918117	0.226420691741	0.244471712303
D8	0.205929750595	0.21576670272	0.204295727267	0.202540661413	0.249662254731	0.251395437311	0.287621620338	0.235117406342
D9	0.214544858219	0.156660734851	0.222883635156	0.187612214881	0.222661767248	0.264321023465	0.263551463265	0.177690593846
D10	1.0	0.461015009817	0.368888154807	0.344912134757	0.269217655629	0.264732840569	0.327895683348	0.102613842073
D11	0.461015009817	1.0	0.267335569189	0.327261508519	0.260211737579	0.231762050295	0.214617608572	0.055493575458
D12	0.368888154807	0.267335569189	1.0	0.558100394569	0.322543612197	0.197536497391	0.410308634816	0.0743562044102
D13	0.344912134757	0.327261508519	0.558100394569	1.0	0.318223314873	0.235088850118	0.328406831305	0.0453632791074
D14	0.269217655629	0.260211737579	0.322543612197	0.318223314873	1.0	0.210157297986	0.279848625235	0.178372821537
D15	0.264732840569	0.231762050295	0.197536497391	0.235088850118	0.210157297986	1.0	0.243856481629	0.130164254531
D16	0.327895683348	0.214617608572	0.410308634816	0.328406831305	0.279848625235	0.243856481629	1.0	0.171823102761
D17	0.102613842073	0.055493575458	0.0743562044102	0.0453632791074	0.178372821537	0.130164254531	0.171823102761	1.0

Figura 5.17: Correlação entre os dois perfis encontrados pelo grupo de 3241 usuários com 11 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Começo do Bloco

	D1	D2	D3	D4	D5	D6	D7	D8	D9
D1	-0.0147046614788	-0.0316155154122	0.00977990135927	-0.0141868636533	-0.068283169689	0.0773426403733	0.0129889685385	-0.075325634621	-0.00445801369351
D2	0.00990159441488	-0.0459705600006	-0.00992168782842	0.0335567760598	-0.0660387091634	0.0105077193552	0.109321386161	-0.0123571727698	-0.0524405183435
D3	-0.00458320553669	0.0795162669415	0.0103011460859	0.00989434200526	-0.0797870245978	0.117667872045	0.0759318487425	-0.0311047906025	-0.0283065799759
D4	-0.00756707724757	-0.0143210853127	0.0306696237689	0.0280264314893	-0.0876312661065	0.090027157096	0.0472985295421	0.0230856933582	0.0539500945615
D5	0.022502923669	0.0485083728215	0.0370346411643	-0.0445994879022	-0.0151507395743	0.0190816498622	-0.068523595758	-0.0175629575244	0.0242393511471
D6	0.0692185513095	0.00890293658312	0.0387153274129	0.0942536453855	0.0231538829182	0.062043582507	0.082997128275	0.0655887136794	0.0643410149989
D7	0.0069250607595	-0.015407318868	-0.0260321921825	0.0143629826232	-0.022325995483	0.0570050463582	0.040359403286	0.0653487019564	0.00285679845852
D8	-0.0183619736545	-0.0896901729745	-0.0382051549969	0.0187153223535	-0.00439181342599	0.0124930651733	-0.119333048672	-0.0675023376548	-0.0863671773046
D9	0.0051450095909	0.0600020320598	0.042620558774	0.00912631208783	-0.0640302811777	0.037476883665	-0.00604453726571	-0.0112800566482	-0.0122799287129
D10	-0.0143323690875	-0.0297949593291	-0.0663419759865	-0.0893376272717	-0.0617477686011	0.0616092645802	-0.0319117971915	-0.0730395352044	0.0171109652283
D11	0.0368913510666	-0.099472473064	-0.0560307352638	-0.00737988058479	-0.0128170115227	0.0537441406547	0.0133170292689	-0.0434472546482	-0.0429268135689
D12	-0.0359607675413	0.0729764403772	-0.0166815186214	-0.085916352056	-0.0439534163259	0.0551612952661	-0.0126052216796	-0.0219815815375	0.0171514146388
D13	0.0674984014705	-0.0105772781616	0.014106313275	0.0527787210482	0.0146778509449	0.0703757772124	-0.0519867046222	0.0316648328519	0.0606494494819
D14	0.0574099982458	-0.0257215261532	0.0117097219728	0.0946934049663	-0.000242816487503	0.106746528372	0.0866931535143	-0.0101994812122	0.00392778502496
D15	0.0441660494585	0.00961850444	-0.120100479471	-0.0173856346437	0.0390289296529	0.0428151198225	0.0178327130088	-0.0446439956526	-0.0401281454498
D16	0.0543483153082	-0.0121838092551	-0.0327667770758	0.0159081729908	-0.00708275233067	0.0177588178131	0.0286923518191	0.0253632326611	0.00971049975057
D17	0.0087788726554	-0.0336539048403	-0.0575940628825	-0.0359110118171	0.000835833108144	-0.0024159179435	0.00088052960072	0.0374617031728	-0.0470712222459

Figura 5.18: Correlação entre os dois perfis encontrados pelo grupo de 3241 usuários com 11 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Continuação do Bloco

	D10	D11	D12	D13	D14	D15	D16	D17
D1	-0.0654570549547	0.00833919329132	-0.05196257016	0.0735164787393	-0.0665182075316	0.061232746761	-0.0196831455232	0.0250220483391
D2	-0.00540747865453	0.000482734601638	-0.0296027080541	-0.00307209049738	-0.0242725844841	0.0481312091396	-0.101718335607	0.0305607723057
D3	-0.0201021153888	0.00272109698198	0.0206072812651	-0.0325642705404	-0.0146929419573	0.196974809613	-0.0318580014597	-0.000918730246563
D4	-0.0350731117973	0.0241860351741	-0.0327080460566	0.0514739261552	-0.0266617069226	0.107605749855	0.0249291406271	-0.0409944572471
D5	-0.0414298447215	-0.0172434992611	-0.0132738445007	0.0515784183653	-0.0770344731637	0.0654454919203	0.0152054233813	0.0110420162273
D6	-0.0279415422175	-0.0109205384805	0.0124575110375	0.0028765660361	0.0146827868168	0.0748392854007	0.00241668283021	0.0985030167757
D7	-0.0514512428098	0.0365397140681	-0.0258495156108	-0.0276502560719	-0.0263638180945	0.0735276908638	0.0139966677398	0.007423510346
D8	0.021532303098	-0.0197937981342	-0.00274593833521	-0.034764570439	-0.0747299971204	0.0268076264563	-0.000599059722564	0.0185922398183
D9	-0.00283323065493	0.0118557598258	0.01582019188	-0.0410624920071	-0.0053540122577	0.0642319443329	-0.040395774343	0.043973146779
D10	0.0086798345088	0.0055688885526	-0.0320416901671	0.046860492511	-0.0850633600969	0.00328537731775	-0.0505426366909	0.00468308372102
D11	0.0341497374053	0.0382016314676	0.009372735411	0.0743208570123	-0.0492021617891	0.0255303743814	0.034491957682	0.0909734138021
D12	-0.0551840980146	-0.00310057802039	0.0259673762083	0.0400647347992	-0.0803387649696	0.0115539420231	0.0370456743629	0.00911455176075
D13	-0.00156540952167	0.0381627280349	-0.0265299948092	0.0458178556901	-0.113598994981	0.0843790502892	0.0841641113319	0.0998885216256
D14	-0.0354719669549	0.0958164639981	0.00544467910727	0.049986973041	0.0440010489906	0.0783818510877	-0.0335133188655	-0.00754608184207
D15	0.0195440645839	0.0566821659164	0.0124580543261	0.0526656668164	-0.0369052960961	0.10980019476	-0.0185631150777	0.00150549859997
D16	-0.00368356938527	-0.0054272589435	0.0106005467068	0.0495216611724	-0.00453927222049	0.0575296165684	0.045859726372	0.0107686181411
D17	0.0373343788659	-0.000550637308656	0.00368226250289	0.0032820996203	-0.00312691893836	0.0897378611115	0.0200717365861	0.0251683069046

Figura 5.19: Correlação entre os dois perfis encontrados pelo grupo de 3241 usuários com 22 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Começo do Bloco

	D1	D2	D3	D4	D5	D6	D7	D8	D9
D1	0.0821490617037	0.00336914679367	0.0171447201839	0.0546219275343	0.0706994231682	-0.017347655387	0.0159548570375	0.0315991459745	0.063464054759
D2	-0.0183384078738	-0.0173218615584	-0.0567847034153	0.00407482250628	-0.0303874834971	0.0694035656899	-0.0217308396396	-0.00250472608688	0.0157863688119
D3	-0.0663296273111	-0.0558865091684	-0.0849618755776	-0.014283923592	0.0219025453847	-0.00416583999818	-0.0478730999401	-0.065109253021	0.00733004315486
D4	-0.00241788779496	0.00973111458393	0.0544776267455	0.00291894526643	-0.00987919398363	-0.0146646294864	0.00799656368659	0.00557716536151	0.0622606807385
D5	0.0329706362098	-0.000346736617945	-0.0164553485508	-0.0269674257012	0.040437890702	-0.0575036499776	0.035256097303	-0.00786049165819	-0.023468589218
D6	-0.0195559424921	-0.0770028925346	-0.128667986437	-0.070784791161	-0.0356757406827	-0.0348351274283	-0.118470037646	-0.0750588616676	-0.0620900095938
D7	0.0896785194323	-0.0255819616005	-0.065500021996	0.0402990597785	0.0686264694398	0.0403377972372	-0.00234304779517	0.00512656598055	-0.0222438229163
D8	0.0827636574324	-0.0315200841812	-0.0601569410331	-0.0381882629751	0.0664045404039	-0.0596615607491	-0.0296478231689	-0.0178002835563	0.00318425815979
D9	0.0331647725034	-0.0504923431953	-0.0813106241768	0.0244652348687	0.0501156656866	0.039519435331	0.0406949636922	0.0358708567015	0.0641824827594
D10	0.0265515465446	0.0244527678277	-0.0200906205429	0.00196883382925	0.00161572986267	-0.0694939368355	-0.0735768537998	0.0319275733033	0.0367679566771
D11	0.0431905023802	-0.0141351091027	0.061622924496	0.0790016682187	0.0771385389065	-0.0313449153782	0.024822536616	0.0437547397577	0.0647786575973
D12	0.00259402739947	-0.0626356384306	-0.0104291459765	-0.0223667171453	0.0446314503822	-0.0682379587383	-0.0798096974411	0.0464681684373	-0.0251219152098
D13	-0.0193941966812	-0.0762875091955	-0.107028576621	-0.0493610124142	-0.0303813738783	-0.0555981086228	-0.0419437671775	0.0210230183821	-0.0783323991926
D14	0.0137491508828	0.0322820956939	0.00166964802078	0.033250678875	0.0173420143456	-0.0170217182227	-0.039768652817	0.0589790360405	0.0103979932092
D15	0.028393542045	-0.00619738645868	-0.0226834456732	-0.0211378267084	-0.0120764306825	-0.020957679506	-0.0931588395854	-0.0415506253444	0.0796609949613
D16	0.00917683589805	0.0216574931654	-0.0125754809879	-0.0273705857522	0.0214644807432	-0.0263215946659	-0.0373224866979	-0.0127212068283	-0.0389891009035
D17	-0.0201131385181	-0.0534594580575	0.0358888933101	-0.0392795093402	-0.0425321595182	-0.0207028273266	-7.67845636446e-05	-0.11870165678	0.0750996650239

Figura 5.20: Correlação entre os dois perfis encontrados pelo grupo de 3241 usuários com 22 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Continuação do Bloco

	D10	D11	D12	D13	D14	D15	D16	D17
D1	-0.0249863435385	0.0421559942726	0.0137711560536	0.0501144729216	0.0511582303191	0.0441989710712	0.0709100112735	-0.00440050314082
D2	-0.0053036197199	-0.00327397303434	0.00331634912202	-0.0037054588117	-0.0339730084776	0.0362272592864	0.00375793027164	0.0386184122539
D3	-0.0382847775421	-0.0355594198698	-0.0503682714131	-0.00106344475002	0.0244543055591	-0.10278939882	0.0182441096282	0.0131031990881
D4	-0.0384201465404	-0.069059609323	-0.00501258995402	0.0117548368893	0.0164335997898	0.0648949240282	0.0152954334425	0.071445575556
D5	-0.0156172801658	0.0100207465134	-0.0215257598624	-0.0131109469827	0.00458586386723	0.011807031798	0.0671909041177	-0.0105913265739
D6	-0.0752058611873	-0.109362784998	-0.0114216492805	0.0188449412477	-0.108267233999	-0.0310216589046	-0.0605786491911	-0.0802840963358
D7	0.0502050237081	0.0203027968982	0.0220651484368	0.0400916713237	0.0222425853719	0.0481554249506	0.0141414657859	0.0313699798562
D8	-0.0100101917778	-0.0173071039949	0.0127173449031	0.0135296775983	-0.00779375946618	-0.0353009224093	0.0241753118785	-0.0133751767647
D9	0.0501764713379	0.0948492118109	0.0126987576558	0.0495895916819	0.0129298266679	-0.0723955113897	-0.0109407821519	-0.00143706535753
D10	0.0080852778117	0.011858845898	-0.0704118453384	-0.0460220456404	-0.0535423257834	0.0176198796549	0.0421011918395	0.0391979383568
D11	0.0306700819073	0.0245544888818	0.00415746531755	0.0653638311779	0.0785961496784	0.0661765825789	0.0848780400187	0.0480330146304
D12	0.0460397591642	0.0823157569074	-0.00311648796656	-0.0628301520979	-0.039608054412	0.0202521735113	0.0635597457098	-0.0254976834639
D13	-0.0975146443068	0.0462935348754	-0.0143014964635	0.00199978361473	-0.0642563100981	0.00969253901286	0.0241367615518	-0.0161369903339
D14	-0.033478179419	0.02795022264	-0.024318231068	-0.0354550117865	0.0642275377908	0.0371419442148	0.084800623983	0.0502698912151
D15	0.0404419273049	-0.0376559425697	-0.0143724491158	-0.018981923944	-0.0171915343846	0.0193863176062	0.0310112100759	0.0388722471935
D16	0.00202943282289	0.0444386005417	-0.0352335505439	0.0113251883188	-0.0383084547033	0.0233441967473	0.0631974285594	-0.0177292128018
D17	-0.032786945317	-0.0112852992731	-0.0598665243254	-0.0788224992545	-0.0248725586229	0.0200379526627	0.0145843329481	0.0522881126393

Figura 5.21: Correlação entre os dois perfis encontrados pelo grupo de 947 usuários com 11 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Começo do Bloco

	D1	D2	D3	D4	D5	D6	D7	D8	D9
D1	0.0265235584158	0.0302035777486	0.0439850822408	0.0724419241019	0.0154690192112	0.0249118329985	0.052505477247	0.0387038951395	0.0113275667893
D2	-0.0215321466527	-0.0527442826172	-0.0280654045284	-0.0457175900752	0.0134482568324	-0.128306969594	0.0158579639524	-0.0679353858238	-0.0674054837188
D3	-0.0522232992176	-0.0262157586188	0.00147021384904	0.0199415311424	0.00482469839073	-0.0400123908708	0.005311770183	-0.0134301837274	-0.0894524940839
D4	0.0249536254295	0.0422813438178	-0.0219104562157	0.0194925648014	0.0554348996486	0.0257455467546	0.0266992784146	0.0268463389086	-0.045900575944
D5	-0.0206407047175	-0.0767831737193	0.0117717339451	0.0374457397882	-0.029116640093	0.0109850212232	0.0410475816947	-0.00388301271703	0.0328074401464
D6	-0.0341829233757	-0.0172050288262	-0.0305628848566	-0.0387569965007	-0.00306439825787	-0.0464608070771	-0.00883523947859	0.0513260005317	-0.0454901990375
D7	-0.0355077981902	0.0215083610937	0.0348432176615	-0.0247769698181	0.00916839654142	-0.00987106029096	0.0272595651866	-0.0192459254748	-0.038279232664
D8	-0.0778098271846	0.0159356841659	-0.013462858939	-0.0563744979531	0.0192992841288	-0.0209040509869	0.00539013798585	0.0173553808281	-0.0766366208208
D9	-0.0162674277931	0.0457292797838	0.0253754071114	-0.0336168563814	-0.00200820306984	-0.0636368652679	-0.0055825544243	0.0469008541288	-0.0440666486756
D10	0.0453927739541	-0.0565708435317	-0.0409457746732	-0.00142304941612	0.0437502830104	-0.0399901800444	-0.00514456418546	0.0283221771484	0.0512618959791
D11	-4.64071534893e-06	-0.0147883589585	-0.0513803996967	-0.038037762977	0.0515204608222	-0.102835977574	0.0130367902249	-0.0104401571762	0.00558223842648
D12	-0.032653635788	0.023793530092	0.0350858162992	-0.00585637028513	0.065908220516	-0.0120753041921	0.0851351789774	0.0174631154547	0.0354009592179
D13	-0.0577486803869	-0.00999675860056	0.0190772647671	-0.0238470969501	0.0320985121885	0.00324748823467	-0.000710550945634	0.0256163051587	0.0643017261022
D14	-0.0168915398487	0.0567481543189	0.0257059139424	0.0279998708107	-0.0435746633825	-0.0544408064157	0.0685645476828	-0.023129889092	0.0064467359727
D15	0.00441303829662	0.024118697745	-0.0285459076295	-0.0303214285486	0.0749498800442	-0.0667151675119	0.00745839138643	0.0330931507195	-0.0200549736281
D16	-0.0653928651364	-0.0486717757117	-0.0300757451941	0.00581120799874	-0.0458506231166	-0.0337844217596	0.0347887221587	-0.0473114086592	-0.0115991043604
D17	0.0180458324508	0.0217905192169	0.0296264673093	0.029676700849	0.0152091869805	-0.05322791855	-0.0201109414708	-0.0101100124945	-0.0324247765416

Figura 5.22: Correlação entre os dois perfis encontrados pelo grupo de 947 usuários com 11 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Continuação do Bloco

	D10	D11	D12	D13	D14	D15	D16	D17
D1	-0.00165381682418	0.0929244701297	0.0160863081505	0.0321720837062	-0.0206431808452	0.0925011073808	0.0563308872669	0.0567927819358
D2	0.0444029236048	0.00752798383849	0.0290030956405	-0.0647322195577	-0.0411951170392	0.0496793430787	0.00424207689658	-0.0437242423589
D3	0.0140947853397	0.0332491069943	-0.042723681201	-0.0647105291281	-0.0313459527592	-0.0145051534077	-0.028600974802	-0.0378979072263
D4	0.0435807624991	0.112536622044	-0.0686637110349	-0.0127097573571	-0.0241701130305	0.0739854205271	0.0107883864115	0.0291104030672
D5	-0.0589630696591	-0.0311122073786	-0.0725126804399	-0.0785698270967	-0.00133717062459	0.0417399947344	-0.0692347208237	0.0468514920887
D6	-0.0462426748735	0.0514584556349	-0.0598091542472	-0.00931993907565	0.0111060890109	0.101777452448	-0.0494419362166	-0.0815941859852
D7	0.035980005697	0.0708791526366	-0.0233712806013	-0.00931234096355	0.00657533730313	0.0151987657572	0.0500321966551	-0.00918665670525
D8	0.0966615059037	0.025032603766	0.0371725010987	-0.0681078459053	-0.0419851089119	0.0894055105685	-0.011655609136	0.0144612607679
D9	0.0657920075331	-0.0235706322355	-0.0451104018136	-0.0683963742494	-0.034145217116	0.0283028699691	-0.00899795622241	0.100869097155
D10	-0.00806063377122	0.0263226592397	0.000198424649847	-0.0715614476838	-0.0403562841641	0.0300532237899	-0.0195835129638	0.0396508344372
D11	-0.0467855865659	0.0425600643064	-0.0666780825064	-0.0342521041494	-0.0354101004451	-0.0061310119521	-0.0453889707085	0.0496804956752
D12	0.029749247762	0.0334468713804	0.0095830948724	-0.0338453733982	-0.00258838031087	0.0895657023163	0.0153035445347	-0.0359839989083
D13	0.0165786371497	0.032027959788	-0.0778066569169	-0.100437835724	-0.0791654982901	0.0452327298305	-0.0140673994473	0.0534200583161
D14	0.00224159458282	-0.00862407481768	-0.04594229382	-0.0348796618042	-0.0525710575844	0.0493094735926	0.00524456929991	-0.0552946051973
D15	-0.00997977943458	-0.053923131868	0.0300804271993	-0.0550901677524	0.0160696521299	0.0625442044386	-0.0081300069861	-0.000907930867021
D16	-0.0237569461937	-0.0174678552225	-0.0269228605745	-0.0104149078488	-0.0618106812282	0.0146133356944	-0.00664345586831	0.0206477503472
D17	0.135698884556	0.0195674594759	0.0106339726817	0.0178393469276	-0.035245468317	-0.00575767913878	0.0254081707529	-0.0704519878702

Figura 5.23: Correlação entre os dois perfis encontrados pelo grupo de 947 usuários com 22 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Começo do Bloco

	D1	D2	D3	D4	D5	D6	D7	D8	D9
D1	0.00284919636139	-0.0509041268873	-0.00219360109376	-0.0023839040911	0.0248986647637	-0.0830397480619	-0.106821432374	-0.00145349668684	-0.0682749004955
D2	0.0506979692343	-0.00240196844913	0.0388010267029	0.064588414874	0.0856679786341	-0.0376810069577	-0.0547579983626	0.0118186115825	0.0160240002893
D3	0.00165016416565	-0.0232777977897	0.0186223623796	0.0306133168183	0.0171000548067	-0.0483174322989	-0.0842750012845	-0.00625215580753	0.0243789648179
D4	0.0216704467557	0.0124034658185	0.0520916729141	-0.00492943472411	-0.00105817521453	-0.100395363791	-0.109300242083	0.0380720712942	-0.0574833405981
D5	0.004626813185	-0.0848391129711	-0.041750310683	-0.0238353276851	-0.0157997682575	-0.16261453145	-0.121952957376	-0.00635678346364	-0.058942249602
D6	-0.0284009907362	-0.0276840357366	0.0258538157107	0.0140940574571	-0.0425897167123	-0.0332848066221	0.0326201063266	0.0264303692978	-0.0277905046643
D7	0.0046113403638	-0.0390649592033	0.0795365608439	-0.00251649062158	0.0398200681197	-0.108009309993	-0.0884698769049	0.0412722480299	-0.00272698834814
D8	0.062389137275	-0.0755884596032	0.00944431889357	-0.0015329966239	0.0477791319321	-0.0687287147396	-0.0131825853012	0.0356648739418	-0.0615105020633
D9	0.0640067150556	-0.0664532805647	-0.0309112787156	0.0100215830763	-0.0176904416084	-0.0319921329011	-0.0293567580889	-0.0286155272381	0.0240117127531
D10	-0.0294873393456	-0.0635454066623	-0.0128565934056	-0.0438466399103	0.000486941935737	-0.0868685174838	-0.0523296667156	0.000894290562856	-0.0453498173729
D11	-0.0263846773474	0.0294523716002	0.00177468191409	-0.0411368759467	-0.0176008744377	-0.0651051443574	-0.10161580216	0.0615565854358	-0.0534912752084
D12	0.0231702416266	0.0418063603167	0.0297968039219	0.0121212635425	-0.0411250392936	-0.0725937654825	-0.0785498216165	-0.0218958988539	0.0454749211635
D13	-0.00483842955653	-0.0509842045372	0.0186760531164	-0.012748608435	-0.00619352841512	-0.138801957356	-0.090026644043	-0.0232687828601	-0.0531559978301
D14	0.0504494961284	-0.0142915061402	0.0497234251504	-0.0358156208226	0.0419989258954	-0.0837922310901	-0.100882623235	0.0103817879885	-0.0444110724627
D15	0.0235400769809	-0.0325490866102	0.0201057460959	-0.016352530337	-0.0100069367366	-0.0217547645325	0.0715048435349	0.0321439179628	0.0416815313278
D16	0.0364793143266	-0.0042230738711	-0.054293695762	-0.0813909790568	0.0115160579995	-0.0964301984184	-0.0564289592695	-0.119926606291	-0.0447742643
D17	0.0368533976591	-0.00580077076872	0.0291279462755	0.0487521085432	0.0933976187747	-0.00848338775622	-0.0137277441285	-0.0127915024853	0.0196345183054

Figura 5.24: Correlação entre os dois perfis encontrados pelo grupo de 947 usuários com 22 disciplinas utilizando Coeficiente de correlação de postos de Spearman - Continuação do Bloco

	D10	D11	D12	D13	D14	D15	D16	D17
D1	0.0242702481604	-0.0259283714506	-0.0287522439366	-0.0200210819292	0.0436225621148	0.0127732952752	-0.0656843118561	-0.120652870398
D2	-0.0193243484179	0.0330774133808	0.00462474781295	0.0517861311277	0.0678965864984	0.0161152386703	-0.0501649932979	-0.0586530710405
D3	0.00290253582819	0.0499301129651	-0.0420242418292	-0.101557547646	-0.00382612266347	0.0317042890169	-0.0780781297396	-0.0520584064976
D4	-0.0113442760384	-0.00581630138491	-0.045735105118	-0.0584175644954	0.0599811896086	-0.0247693422329	-0.0073874417806	-0.11494055621
D5	0.0159221712119	-0.0305840089811	-0.0315167676161	-0.039908269698	0.00265085736728	-0.0582560201335	-0.151910850567	-0.0935730189411
D6	-0.0500777576064	-0.0285023340123	0.00293512668413	0.0210446510688	0.0269878035086	-0.0200616541067	-0.0670783032368	-0.0527441553987
D7	0.082901121106	0.0434405459534	0.0467807454589	0.0272450749839	-0.00707887849667	0.0179405088623	0.0198361962813	-0.0552381552007
D8	-0.0268645422742	-0.0418029502859	0.00384616051727	0.0554567189096	0.00876610153933	-0.0468069832327	0.00867621039931	-0.0449297653274
D9	-0.0408670655492	0.0123517822953	-0.0548980043832	-0.0466918128375	0.0432946971041	-0.0912216013059	-0.0454002536977	-0.100048069456
D10	-0.0943503138161	-0.0303871623434	-0.039261080292	-0.00953609322699	0.0508858169353	-0.0930276048385	-0.0535188711952	-0.0155225091285
D11	-0.0896271135086	-0.0419152012333	-0.00851463092711	-0.00200442116433	0.0236142012683	-0.0568814178001	-0.0540081488631	0.0138070914453
D12	0.00857553702694	-0.0233234656497	0.0280552901606	-0.0673708310701	0.0244955438336	-0.00235012558849	-0.0565212876991	-0.0208081209605
D13	0.028342198161	-0.0482809576411	0.0456273083401	-0.016375328513	-0.00484174503074	-0.038345340475	-0.0714881669946	-0.0861662427711
D14	0.0452474211481	-0.01529796582	0.0434848092018	-0.0330303463274	-0.0119558127964	-0.0335784979851	-0.0659647348449	-0.0896195978144
D15	-0.083034684185	-0.0045757555134	-0.0778027233176	-0.0287713933599	0.0480971993373	-0.0404851697172	-0.00746113580732	-0.0635565559898
D16	-0.0671735318237	-0.0555036797904	0.0113266898746	-0.018250850011	-0.034035926791	-0.056001657385	-0.109420645598	-0.0356265105865
D17	0.0942751358125	0.00377427710871	0.098831721961	0.107475611939	0.0149324124945	0.0690203547909	0.0341372811058	0.0304448944881