

EduVis: Visualização Interactiva de Dados Educaçãoais

Vilma Rodrigues Jordão

Dissertação para a obtenção do Grau de Mestre em

Engenharia Informática e de Computadores

Júri
Presidente
Orientador
Vogal

Setembro de 2014

Agradecimentos

Em primeiro lugar aos meus pais, Odete Jordão e João Jordão, sem eles não seria possível ter chegado até aqui.

Ao Pedro Lopes pelo seu suporte e motivação constantes. Pela sua presença nos momentos difíceis e por acreditar sempre.

Ao professor Daniel Gonçalves que me orientou e guiou, mostrando-me o caminho a seguir.

À Sandra pela disponibilidade, apoio e espírito crítico

Às minhas queridas amigas Rita, Joana e Cecília, por estarem sempre presentes.

Este trabalho foi financiado por fundos nacionais através da FCT - Fundação para a Ciência e a Tecnologia, no projeto Educare - PTDC/EIA-EIA/110058/2009, e pelo financiamento plurianual do INESC-ID - PEst-OE/EEI/LA0021/2013

RESUMO

Uma análise efectiva dos resultados do percurso escolar efectuado pelos alunos no ensino tem uma grande relevância por proporcionar ferramentas úteis à comunidade académica, na identificação de problemas, e situações atípicas. Actualmente através de aplicações de técnicas de *data mining* é possível identificar padrões, pontos fracturantes e fazer previsões do possível desempenho que os estudantes possam vir a ter. No entanto avaliar estes resultados torna-se muito complexo, por consistirem, numa situação real, num conjunto muito amplo de dados, que facilmente atingem grandezas na ordem dos milhares, dificultando ou impossibilitando a sua visão de conjunto. A este cenário, acresce também o facto de ser necessário que muitas vezes o utilizador tenha de ter conhecimento de algoritmos matemáticos e estatísticos complexos, de modo a entender e avaliar os padrões obtidos. Desta forma, e dado o potencial já demonstrado da visualização para representar grandes quantidades de informação, nesta tese é proposto e implementado um mecanismo de visualização multi-matricial que permite observar o trajecto académico de alunos, bem como depreender as várias relações e precedências das disciplinas existentes. A solução apresentada, encontra-se no âmbito do projecto Educare, e utiliza como base padrões extraídos através de *data mining* de dados recolhidos do curso de Engenharia Informática e de Computadores do Instituto Superior Técnico.

Neste contexto existe uma visualização previamente desenvolvida, que o mecanismo agora proposto, pretende complementar com novas funcionalidades, criando uma solução integrada, cujo objectivo final é possibilitar uma resposta eficaz para a visualização de padrões educacionais.

Palavras-chave: Visualização de informação, padrões educacionais, análise de dados, interacção.

ÍNDICE

1	INTRODUÇÃO	1
1.1	Objectivo	2
1.2	Publicações	3
1.3	Estrutura do Documento	3
2	TRABALHO RELACIONADO	5
2.1	Aplicações desenvolvidas para para representação de dados educacionais.	5
2.1.1	Generation of Graphical Representations of Student Tracking Data in Course Management Systems	5
2.1.2	Monitoring an Online Course with the GISMO Tool: A Case Study 8	8
2.1.3	Work In Progress - Curriculum Visualization	10
2.2	Técnicas e métodos de visualização de informação aplicados a dados educa- cionais	12
2.2.1	Visual Analysis of College Student's Scores in English Test	12
2.2.2	Visual Exploration for Time Series Data Using Multivariate Analysis Method	14
2.2.3	Visualizing Student Histories Using Clustering and Composition	15
2.2.4	Visualizing Trends in Student Performance across Computer Science Courses	19
2.2.5	A Multidimensional Data Visualization Method Based On Parallel Coordinates and Enhanced Ring	20
2.2.6	Discussão	22
3	PROJETO EDUCARE	25
3.1	Padrões Educacionais	26
4	SOLUÇÃO PROPOSTA	28
4.1	Protótipos de baixa fidelidade	29
4.2	EduVis	30
4.3	Visualização multi-matricial	32
4.4	Estado Inicial	32
4.5	Interação	33
4.5.1	Mouseover	34
4.5.2	Seleção de Disciplinas	34
4.6	Implementação	36
4.6.1	Processamento de padrões Educacionais	37
4.6.2	Evolução da Visualização	39
4.7	Integração no Educare	41
5	AVALIAÇÃO	45
5.1	Descrição do Protocolo	45
5.2	Tarefas	46
5.3	Descrição dos Resultados	47
5.4	Discussão de Resultados	49
5.4.1	Análise Imediata	49
5.4.2	Exploração da Visualização	50

5.4.3	Discussão dos Objectivos Alcançados	55
6	CONCLUSÕES FINAIS E DISCUSSÃO	57
6.1	Trabalho Futuro	58

LISTA DE FIGURAS

Figura 2.1	Representação gráfica sobre threads e debates	6
Figura 2.2	Cognitive Matrix	7
Figura 2.3	Assiduidade dos alunos	7
Figura 2.4	Gráfico de comportamento	8
Figura 2.5	Acessos dos estudantes ao curso	9
Figura 2.6	Número de acessos globais feitos por um estudante aos recursos do curso	10
Figura 2.7	Detalhes dos acessos feitos por um aluno aos recursos do curso	10
Figura 2.8	Gráfico dos acessos de todos os estudantes a todos os recursos do curso	11
Figura 2.9	Visualização do currículo do curso de ciências da computação	11
Figura 2.10	Dados em coordenadas paralelas	12
Figura 2.11	Classificações dos alunos	13
Figura 2.12	Comparação das médias e Boxplot	13
Figura 2.13	Correlação entre leitura e escrita	13
Figura 2.14	Regra da associação entre leitura e escrita	14
Figura 2.15	Agrupamento de estudantes	15
Figura 2.16	Comparação de vários parâmetros	15
Figura 2.17	Representação bi-dimensional do histórico dos estudantes . .	16
Figura 2.18	Representação bi-dimensional a cores do histórico dos estudantes	17
Figura 2.19	Exemplo do algoritmo <i>Level/ Set</i> quando a média da trajetória é usada como condição inicial.	17
Figura 2.20	color blending e o método weaving	18
Figura 2.21	Método weaving com uma escala de cores Brewer	18
Figura 2.22	Diferenças nas notas entre os dois géneros.	19
Figura 2.23	Representação em grafos do percurso dos estudantes.	19
Figura 2.24	Percurso de um estudante.	20
Figura 2.25	Gráfico em anel normal, usa o mesmo conjunto de dados para os três anéis.	21
Figura 2.26	Enhanced Ring Chart.	21
Figura 2.27	Visualização dos dados sobre as notas dos alunos usando PCER	22
Figura 3.1	Visualização Multi-Matricial	25
Figura 3.2	Exploração das vertentes temporal e dinâmica	26
Figura 4.1	Pbf : grafo de forças	30
Figura 4.2	Pbf : coordenadas paralelas	31
Figura 4.3	Pbf : eixos cartesianos	31
Figura 4.4	Pbf matriz	32
Figura 4.5	<i>Heatmap</i> , estado inicial	33
Figura 4.6	Variação da cor	33
Figura 4.7	Relações da disciplina <i>am1</i>	34
Figura 4.8	Dupla selecção	35
Figura 4.9	Contorno de uma disciplina seleccionada	35
Figura 4.10	Elipses com o nome das disciplinas seleccionadas	35
Figura 4.11	Uma possível restrição dos padrões da disciplina <i>fex</i>	36

Figura 4.12	Restrição dos padrões comuns às cadeiras de <i>fex</i> e <i>fisica1</i> . . .	37
Figura 4.13	Visão de células sobrepostas nos padrões das disciplinas de <i>sdig</i> e <i>aled</i>	37
Figura 4.14	<i>Blend</i> das disciplinas <i>al</i> , <i>am1</i> e <i>fex</i>	37
Figura 4.15	Porcentagem de padrões em cada suporte	38
Figura 4.16	Divisão vertical de células	40
Figura 4.17	Divisão diagonal de células	41
Figura 4.18	EduVis: Versão final, caso demonstrativo da selecção de uma disciplina e a sua repercussão nas duas visualizações	43
Figura 5.1	Distribuição da média dos tempos de resposta	48
Figura 5.2	Distribuição da média dos erros cometidos	48
Figura 5.3	Boxplot com o numero de erros cometidos na tarefa 1	52
Figura 5.4	Boxplot do tempo demorado para realizar a tarefa 1	52
Figura 5.5	Boxplot com o numero de erros cometidos na tarefa 2	53
Figura 5.6	Boxplot do tempo demorado para realizar a tarefa 2	53
Figura 5.7	Boxplot com o numero de erros cometidos na tarefa 3	53
Figura 5.8	Boxplot do tempo demorado para realizar a tarefa 3	53
Figura 5.9	Boxplot com o numero de erros cometidos na tarefa 4	54
Figura 5.10	Boxplot do tempo demorado para realizar a tarefa 4	54
Figura 5.11	Boxplot com o numero de erros cometidos na tarefa 5	54
Figura 5.12	Boxplot do tempo demorado para realizar a tarefa 5	54
Figura 5.13	Boxplot com o numero de erros cometidos na tarefa 6	55
Figura 5.14	Boxplot do tempo demorado para realizar a tarefa 6	55
Figura 5.15	Boxplot com o numero de erros cometidos na tarefa 7	55
Figura 5.16	Boxplot do tempo demorado para realizar a tarefa 7	55

LISTA DE TABELAS

Tabela 1	Sumarização das características dos diferentes trabalhos.	23
Tabela 2	Tempos de resposta da Visualização ao deslocamento do cursor e à selecção de uma célula	39
Tabela 3	Sumarização das características da visualização multi-camada e multi-matricia	42
Tabela 4	Género dos inquiridos	47
Tabela 5	Idade dos inquiridos	47
Tabela 6	Grau de escolaridade	47
Tabela 7	Tempo e número de erros médio	47
Tabela 8	Sumarização dos erros cometidos em cada tarefa	52

LISTA DE SIGLAS E ABREVIATURAS

CMS Course Management Systems

EDM Educational Data Mining

LMS Learning Management System

MOOC Cursos Abertos Massivos Online

INTRODUÇÃO

O número de alunos inscritos no ensino tradicional e online tem vindo a aumentar nas últimas décadas. Relativamente ao ensino universitário a nível global, a taxa de alunos que o frequentam teve um crescimento de 16.2% até 2006, relativamente à década de 70, cujo valor rondava os 8.5% [2]. Quanto ao ensino online, a criação e a popularização de MOOC (Cursos Abertos Massivos Online) em que podem participar alunos de todo o mundo, contribuiu grandemente para aumento elevado do número de estudantes. Com este crescimento popularizou-se o uso de CMS e de LMS, ferramentas desenhadas para prestar assistência às turmas e mover o ensino para a internet [6]. Com a difusão deste tipo de ferramentas, e com o aumento da taxa de escolarização [2], emerge uma grande quantidade de dados em bruto sobre o currículo dos alunos, pelo que esta informação está cada vez mais acessível para ser recolhida e analisada. Fazê-lo é essencial por proporcionar à comunidade académica um meio para analisar e descobrir padrões, tornando exequível o ganho de conhecimento e compreensão sobre o desempenho académico de estudantes.

Para tratar este tipo de informação são utilizadas técnicas de *data mining*, que neste contexto assume a denominação EDM e que actualmente tem vindo a ganhar terreno, por desenvolver técnicas para exploração e análise de grandes quantidades de dados provenientes do contexto educacional. Assim, é possível fazer previsões sobre a performance dos alunos a cadeiras futuras, com base na análise da combinação de resultados a outras disciplinas semelhantes. Proporciona também a descoberta de elementos que se destacam num conjunto de dados [23], permitindo identificar alguma situação problemática que se evidencie, ou pelo contrário, acontecimentos positivos dos quais é possível retirar aprendizagens.

Contudo e apesar da relevância dos padrões obtidos não é possível fazer uma análise directa dos mesmos, por virem sob a forma de extensos padrões textuais, difíceis de ler e interpretar, e como tal, muitas vezes não é possível tomar consciência da sua extensão e de como se propagam, bem como pelo facto de ser necessário em muitos casos, ter conhecimentos de algoritmos matemáticos e estatísticos complexos, para se poder inferir algum tipo de conclusão.

Neste sentido, para complementar a prospecção de dados, é necessário envolver o utilizador no processo da exploração e combinar criatividade, flexibilidade e conhecimento geral [7]. Nesse campo, a visualização de informação tem vindo a ganhar destaque na representação de dados. O seu objectivo é ajudar à compreensão de grandes quantidades de dados, aumentando a capacidade do sistema visual humano de descobrir tendências, padrões e pontos discrepantes [5]. Assim, uma vez que através da visão se obtém mais informação que todos os outros sentidos combinados [21], se houver uma visualização bem desenhada, esta pode substituir mecanismos cognitivos apenas com deduções perceptuais, tornando possível a descoberta de propriedades emergentes que eram então desconhecidas, mostrando padrões e sub-problemas [15], em que olhando apenas para um conjunto de estatísticas, seriam mais difíceis de identificar. Permite também que problemas com os próprios dados se tornem imediatamente aparentes, sendo que erros na informação recolhida se tornam evidentes.

Tendo em conta as necessidades que existem para a representação de grandes quantidades de dados acima descritas, e dadas as características da visualização de informação, decidiu-se usar as vantagens que esta oferece para criar uma metáfora visual que representa efectivamente padrões educacionais, tornando a sua análise e avaliação clara e eficaz, de modo a trazer benefícios para os coordenadores de cursos, professores e alunos.

Desta forma, um professor pode por exemplo construir um programa de estudos mais eficiente com base nas dificuldades detectadas, um coordenador pode fazer alterações à estrutura curricular de um curso para que este seja mais adequado às necessidades dos estudantes e os alunos podem por exemplo, observar os pontos mais complicados do currículo, podendo precaver-se e ter consciência da importância de determinadas disciplinas.

Nesse sentido, no âmbito do projecto Educare(PTDC/EIAEIA/110058/2009), em que um dos principais focos é trabalhar um conjunto de técnicas de descoberta e visualização de informação, foram recolhidos dados sobre os três primeiros anos dos percursos académicos da Licenciatura em Engenharia informática e Computadores pela equipa da Prof^a Dr^a Cláudia Antunes, do Instituto Superior Técnico.

Utilizando o método de EDM, foi possível extrair padrões, que fornecem o número de chumbos e sucessos a determinadas cadeiras e as relações de precedência e simultaneidade entre as mesmas. Um exemplo de precedência seria, os alunos que não conseguiram aprovação à cadeira A num semestre, também não conseguiram aprovação à cadeira B no semestre seguinte. Uma relação de simultaneidade pode ser, 5% dos alunos que foram aprovados na disciplina C num determinado semestre, também foram aprovados na cadeira D no mesmo semestre. Apesar de estas relações estarem identificadas, como foi mencionado anteriormente, analisá-las é um processo moroso e difícil, por os dados estarem representados no formato textual e por terem uma dimensão muito elevada, impossibilitando o utilizador de ter uma visão global dos mesmos. Neste caso, seria quase impossível a demonstração de qualquer tipo de percurso feito pelos estudantes, ou descoberta de problemas existentes.

Assim, nesta dissertação procurámos endereçar os problemas referidos, desenvolvendo uma solução que vem completar o primeiro mecanismo de visualização desenvolvido no projecto Educare, permitindo uma interpretação fácil, correcta e completa dos dados educacionais, de forma a que seja possível mostrar todas as vistas dos dados recolhidos, sem perda de informação.

1.1 OBJECTIVO

Este trabalho tem como objectivo usar a visualização de informação para tornar possível a análise e consequente avaliação dos percursos académicos dos estudantes, a partir dos padrões educacionais obtidos no contexto do projecto Educare.

Desse modo, este trabalho ambiciona facilitar a aquisição e assimilação de conhecimentos, sobre o comportamento dos alunos às diferentes cadeiras. Comportamentos esses que acabam por definir uma trajectória, e por destacar quais as cadeiras com mais e menos dependências, bem como, as que possuem maior ou menor taxa de aprovações. Assim, é possível munir o corpo académico de ferramentas que ajudam à percepção de padrões e problemas que passariam despercebidos, passando a ter a possibilidade de tomar decisões para encontrar as soluções mais adequadas, ajudando os alunos, que assim recebem um maior apoio por parte dos docentes e têm os cursos mais adaptados às suas necessidades.

Para desenvolver um sistema adequado e que explore a informação que é possível extrair dos padrões, na sua totalidade, foi pensado e implementado um modelo, que vem completar a visualização já existente.

O primeiro mecanismo desenvolvido consiste em várias camadas de círculos, e utiliza conectores visuais para demonstrar relações entre cadeiras bem como a quantidade de alunos presente nessas relações. No entanto, não cumpre todos os requisitos, mostrando apenas algumas vistas de toda a informação disponibilizada. Surgiu então a necessidade de criar uma nova visualização.

Este novo modelo é composto por várias matrizes, que evidenciam quais as cadeiras com mais ou menos precedências, e que estão presentes em mais padrões. Assim, ao criar um mecanismo que complementa a visualização já existente, pretende-se tornar possível a interpretação das várias relações e dependências, existentes entre as diferentes cadeiras nos vários semestres. Deve também mostrar de forma objectiva quais as cadeiras envolvidas em mais padrões, ou seja, que possuem muitas relações de precedência ou simultaneidade, e que portanto podem ser consideradas cadeiras base, cuja aprovação é necessária para se ter sucesso a cadeiras futuras. Por fim, pretende-se que seja ainda possível observar a quantidade de alunos envolvidos em cada padrão.

1.2 PUBLICAÇÕES

O trabalho desenvolvido nesta tese deu origem a duas publicações. A primeira é referente a um poster, publicado na conferência internacional, "8th Nordic Conference on Human-Computer Interaction" (**NordiCHI2014**). A segunda publicação é um artigo, submetido na conferência nacional EPCG 2014 - 21 Encontro de Português de Computação Gráfica, e que está a aguardar resposta. As referências para as publicações são as seguintes:

1. Vilma Jordão, Sandra Gama e Daniel Gonçalves, EduVis: Visualização Interativa de Dados Educacionais, EPCG 2014 - 21 Encontro de Português de Computação Gráfica.
2. Vilma Jordão, Sandra Gama e Daniel Gonçalves, EduVis: Visualizing Educational Information, 8th Nordic Conference on Human-Computer Interaction, Helsinki, October 2014

As publicações focam-se em apresentar a solução proposta neste documento para a visualização de padrões educacionais, referindo o resultado dos testes efectuados, que avaliam a usabilidade do EduVis.

1.3 ESTRUTURA DO DOCUMENTO

Este documento encontra-se estruturado da seguinte forma:

No capítulo 2 está descrito o trabalho relacionado, mostrando o que tem vindo a ser feito na área em estudo. São tidas em conta as fontes de recolha e análise de grandes quantidades de dados, provenientes de meios educacionais, e os métodos usados para os visualizar. São analisados trabalhos que usam diferentes abordagens na visualização de informação e é feita uma discussão no final do capítulo, dos resultados observados, de modo a ter uma base para desenvolver os requisitos a implementar. O capítulo 3 explica o que é o projecto Educare, e em que fase se encontra do seu desenvolvimento. É descrita a primeira visualização do projecto, as suas características e é justificada a necessidade de a complementar com outra visualização. É também analisada a estrutura dos dados e a sua proveniência. No capítulo 4 é apresentada a solução desenvolvida para o desafio proposto, apresentando o caminho seguido desde a prototipagem até à visualização final. São descritas as funcionalidades do mecanismo, e as várias fases da sua implementação. O capítulo 5 apresenta os resultados obtidos no contexto do presente estudo. São descritas as métricas adoptadas para medir o desempenho do utilizador, e os resultados dos vários testes efectuados, bem como a

sua análise. Por fim, no capítulo 6 são enunciadas as conclusões que foram retiradas do presente trabalho, bem como as possíveis melhorias a efectuar, para um futuro trabalho de investigação, que poderá advir deste.

TRABALHO RELACIONADO

Com a difusão de cursos online bem como o aumento do número de alunos no ensino tradicional, emerge uma grande quantidade de dados em bruto dos seus currículos. Para os representar são usadas diferentes técnicas e metodologias de visualização de informação, de forma a sustentar a tomada de decisões, por parte do corpo docente, relativamente a problemas que tenham sido detectados. Nesse âmbito foram pesquisados os trabalhos relevantes desenvolvidos na área de visualização de informação de padrões educacionais, relacionados com o percurso e desempenho académico dos alunos. A secção 2.1 descreve que vantagens podem ser obtidas ao representar graficamente dados sobre o desempenho de alunos em ensino tradicional e à distância, através de aplicações desenvolvidas para tal. Na secção 2.2 são descritas técnicas de representação de dados e as vantagens que existem na visualização da informação recolhida.

2.1 APLICAÇÕES DESENVOLVIDAS PARA PARA REPRESENTAÇÃO DE DADOS EDUCACIONAIS.

Existem algumas aplicações desenvolvidas, que surgiram devido à necessidade de representar informação, proveniente de determinadas ferramentas, e que usam técnicas de visualização de forma a tornarem os dados fáceis de ler e interpretar. Nos pontos seguintes estão descrições de algumas dessas aplicações, sobre como funcionam e os benefícios que trazem.

2.1.1 *Generation of Graphical Representations of Student Tracking Data in Course Management Systems*

Os CMS, são uma ferramenta que permite criar salas de aula virtuais onde os alunos e professores podem partilhar informação. Desta forma é possível participar em discussões e gerir turmas à distância. Deste tipo de interacção resulta uma grande quantidade de informação que os CMS ainda não tratam da devida forma e não dão o apoio devido aos professores na hora de analisar o desempenho dos alunos. De forma a conseguir colmatar esta falha e gerir a informação dos estudantes, tal como os trabalhos desenvolvidos, as suas notas e o número de vezes que acederam ao CMS, foi desenvolvida uma ferramenta, CourseVis [1], que utiliza informação multidimensional e gera representações gráficas produzidas com a ferramenta de visualização OpenDX¹ de modo a que os professores consigam compreender melhor o que se passa em cursos de ensino à distância. O CourseVis é usado como uma extensão aos CMS, pois tal como acima mencionado, a informação fornecida pelos mesmos está num formato que é muito difícil de ler e compreender, o que complica muito assim a tarefa de conseguir analisar bem a informação e descobrir padrões e tendências, fazendo com que raramente se utilizem os dados gerados pela ferramenta em questão. Anteriormente já se tinham utilizado técnicas de visualização de informação para representar os dados dos

¹ Open visualization data explorer, 2004. <http://www.openDX.org>.

alunos, mas não permitiam que o utilizador manipulasse os dados e os pudesse explorar, limitando-se apenas a representá-los. A ferramenta CourseVis surgiu então no âmbito de tentar solucionar esse problema. Um exemplo da utilização do CourseVis é por exemplo a representação da quantidade de alunos que participaram na thread de um fórum, em que são utilizadas esferas com tamanhos diferentes para representar o número de alunos. Nesta visualização o professor tem a possibilidade de fazer zoom e pan à imagem de modo a conseguir analisar o envolvimento dos estudantes na thread, Figura 2.1. Na primeira imagem é possível ver os alunos que iniciaram as threads e na segunda imagem, depois de rodar o gráfico, observam-se os debates relacionados.

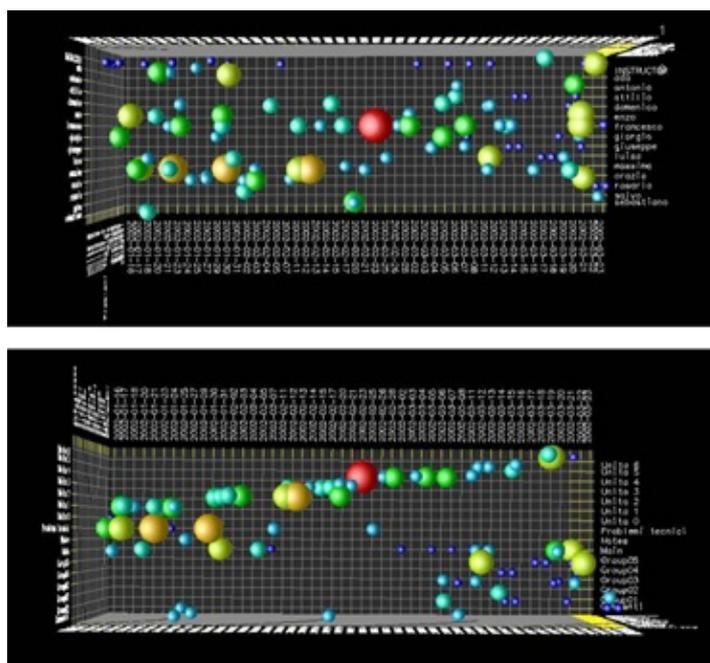


Figura 2.1: Representação gráfica sobre threads e debates

Outro tipo de visualização que o CourseVis proporciona designa-se por Cognitive Matrix, que utiliza os dados do desempenho dos alunos para avaliar o entendimento dos mesmos sobre as disciplinas. No eixo horizontal estão os nomes dos alunos e no eixo vertical estão os conceitos do curso. O resultado do seu desempenho é indicado através de uma escala de cores, Figura 2.2.

É possível monitorizar a assiduidade dos alunos, que é bastante importante para haver uma gestão eficiente do curso. De cada vez que um estudante acede ao CMS, este regista internamente a data e a duração do seu acesso. Visualizar estes dados permite aos professores observarem e identificarem padrões sobre o número de acessos numa data em particular, ou quais as datas em que houve uma maior utilização do CMS por parte dos alunos, Figura 2.3.

Provou-se também, através de pesquisas junto dos professores que é importante saber quando é que os estudantes lêem o material que têm ao seu dispôr, quando fazem avaliações e quando participam em debates. A Figura 2.4 demonstra a utilização do método *Single-axis composition* para representar um grande número de variáveis num espaço métrico a 2D.

De forma a provar a usabilidade e eficiência da ferramenta CourseVis foi efectuado um estudo com professores que utilizaram o CMS e com professores que utilizaram o CMS com a ferramenta integrada, concluindo-se que a representação gráfica permite identificar com

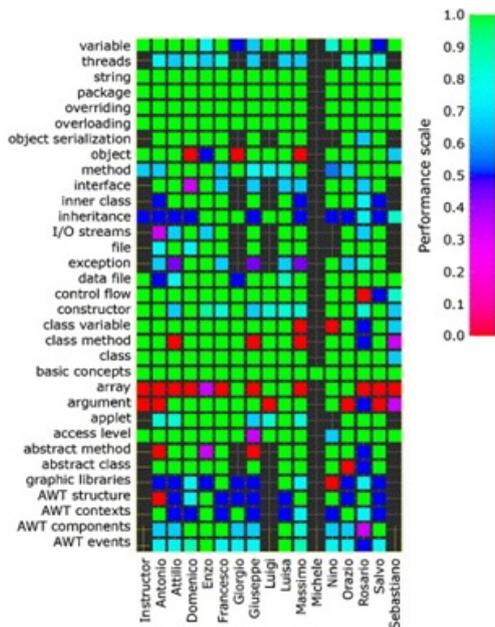


Figura 2.2: Cognitive Matrix

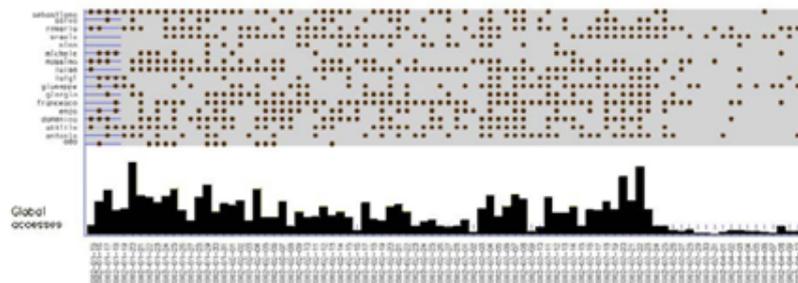


Figura 2.3: Assiduidade dos alunos

mais antecedência e até prevenir, alguns dos problemas causados pelo ensino à distância, pois permite aos docentes obter informação de forma mais rápida e precisa sobre os aspectos comportamentais e cognitivos dos estudantes. Esta ferramenta para visualização de dados representa o objectivo da visualização do desempenho dos estudantes nos diferentes cursos, conseguindo evidenciar e facilitar a análise de problemas, apesar de as técnicas de representação escolhidas nem sempre serem intuitivas e a informação apresentada por vezes ser difícil de ler.

Por exemplo na Figura 2.1 a representação usada tem como desvantagem o facto de haver sobreposição nas esferas, em que esferas maiores escondem as mais pequenas, tornando uma análise à primeira vista mais difícil e menos óbvia. Outro caso é demonstrado na Figura 2.2. Nesta situação teria sido preferível usar variações de brilho a partir de uma cor ao longo dos dois eixos ao invés de seguir uma ordenação de cores, o que complica a leitura do gráfico e o torna mais confuso e menos intuitivo [12]. Por fim na Figura 4 os dados estão representados de uma forma pouco clara e a imagem acaba por apresentar algum ruído fazendo com que o leitor perca mais tempo a conseguir decodificar a informação apresentada [12]. É relevante mencionar que foi desenvolvido um sucessor do CourseVis, GISMO(Graphical In-

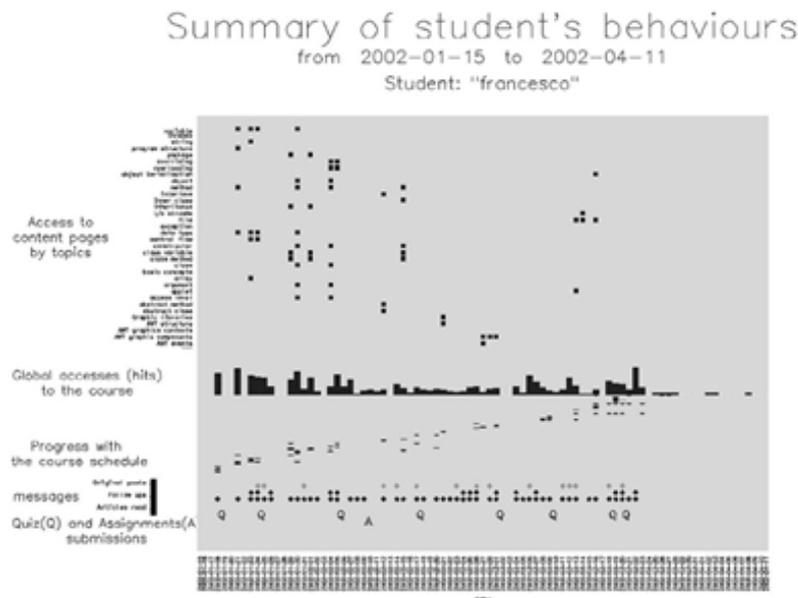


Figura 2.4: Gráfico de comportamento

teractive Student Monitoring System), uma ferramenta de visualização para monitorização de estudantes, com o objectivo de responder a algumas limitações reveladas na avaliação do CourseVis tais como a falta de conexão entre as representações gráficas e a informação correspondente proveniente do CMS.

2.1.2 Monitoring an Online Course with the GISMO Tool: A Case Study

Um dos problemas com que os estudantes do ensino à distância se deparam é com a falta de suporte por parte dos professores . Devido à natureza da comunicação mediada por computador, os alunos tendem a estudar sozinhos com pouca ou nenhuma interacção com outros estudantes. Nestes casos o papel do professor é muito importante pois tem de monitorizar as actividades dos estudantes, providenciar apoio a quem necessita e facilitar o processo de aprendizagem. Questões que o tutor tem de fazer tais como "Estão os estudantes a participar em discussões?" , "Têm estudado a matéria dada no curso?" e "Têm obtido um bom desempenho nos questionários?" são importantes para o sucesso de ensino à distância, mas este tipo de monitorização não é fácil de alcançar com LMS comuns, pois proporcionam pouca ajuda aos docentes de forma a que estes consigam ter conhecimento e compreensão dos processos cognitivos e sociais de uma classe à distância.

Foi então desenvolvida a ferramenta gráfica GISMO para monitorizar um curso à distância[8], que representa visualmente dados recolhidos por LMS e que podem ser utilizados por professores para ganhar conhecimento dos seus alunos e estar a par do que acontece em turmas à distância. Os dados provenientes de LMS são complexos e estão organizados em formatos tabulares, o que na maioria dos casos os torna difíceis de ler e inapropriados para as necessidades dos intrutores. O GISMO adopta o paradigma da visualização de informação, em que a informação é apresentada visualmente e sustenta-se nas habilidades perceptuais da visão humana para a sua interpretação, implementando algumas das visualizações indicadas úteis pelos professores, baseando-se na experiência anterior com a ferramenta CourseVis [1].

Esta nova ferramenta está integrada no LMS Moodle ² mas também é adaptável a outras plataformas de aprendizagem. Após a utilização do GISMO num curso online, este provou ser uma ferramenta poderosa para pelo menos três actividades, monitorizar turmas e comportamentos individuais, avaliar participações em fóruns e redesenhar o curso de acordo com as necessidades dos estudantes.

Na Figura 2.5 está representado um gráfico com os acessos ao curso, em que cada quadrado representa pelo menos um acesso ao curso feito por um estudante numa determinada data. O histograma na parte de baixo da imagem mostra o número global de acessos por todos os estudantes numa determinada data.

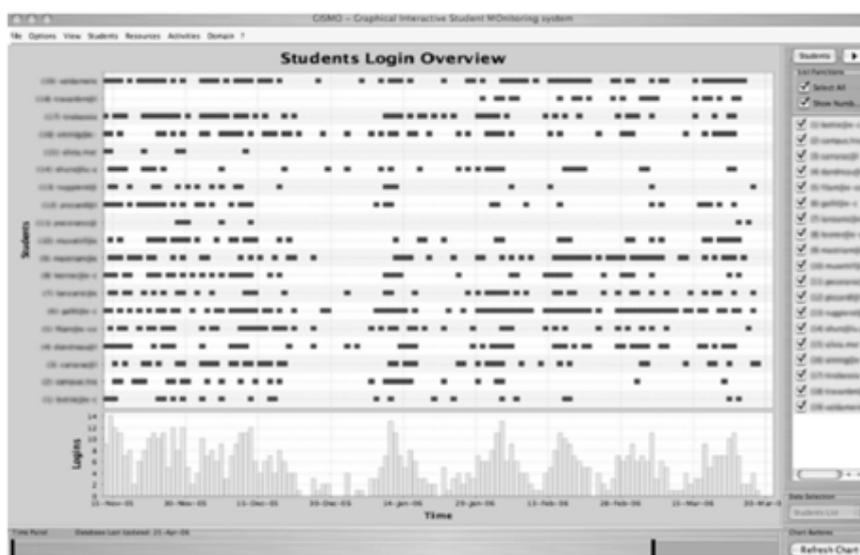


Figura 2.5: Acessos dos estudantes ao curso

A Figura 2.6 representa o número global de acessos por um estudante a todos os recursos do curso. Se o utilizador clicar com o botão direito do rato em uma das barras do histograma e seleccionar a opção "Details" pode ver os detalhes dos acessos para aquele estudante. Na Figura 2.7 são visíveis esses detalhes, tais como, em que dias o estudante visitou um recurso específico e quantos acessos globais foram feitos a todos os recursos em cada dia do curso.

A Figura 2.8 mostra os acessos de todos os estudantes aos recursos do curso.

O GISMO acaba por oferecer uma visualização simples de parâmetros específicos (autenticação e acesso a recursos) tornando esta informação muito mais legível que as longas listas de registos disponíveis no Moodle e em quase todos os LMS. Por outro lado permite ao professor comparar diferentes aspectos do comportamento de uma classe, e refinar certos tipos de informação, que de outra forma seria genérica, tal como o facto de haver muitas autenticações por semana, quais foram mesmo os alunos que se autenticaram e o que fizeram depois disso. Verificar estes dados com o GISMO demora apenas alguns minutos enquanto que fazê-lo sem qualquer tipo de representação gráfica seria muito mais difícil e moroso.

A análise dos dados acima mencionados permitiu identificar quais os recursos menos usados, que podem ser provavelmente pouco adequados ao curso ou até mesmo não ter nenhuma utilidade, quais os recursos que só são procurados durante alturas específicas e que então podem ser escondidos de forma a simplificar a interface do curso, recursos acedidos por todos mas só algumas vezes e recursos acedidos por todos de uma forma regular. Desta

² <http://www.moodle.org> (2013)

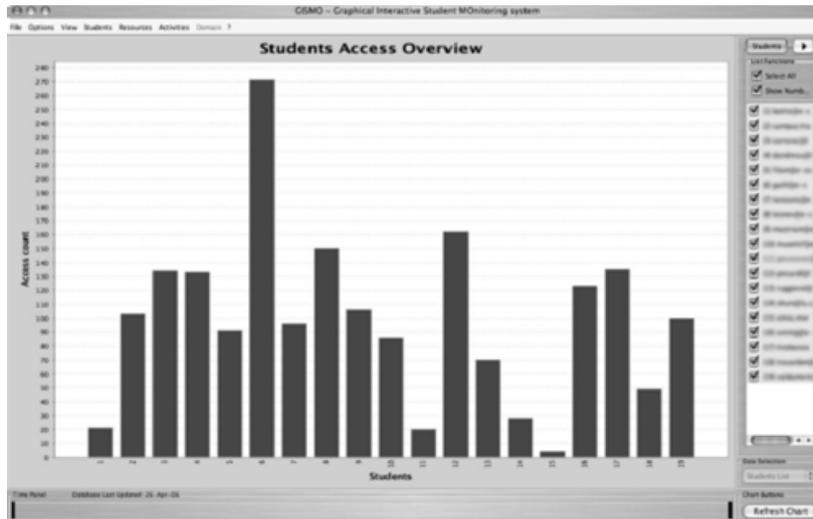


Figura 2.6: Número de acessos globais feitos por um estudante aos recursos do curso

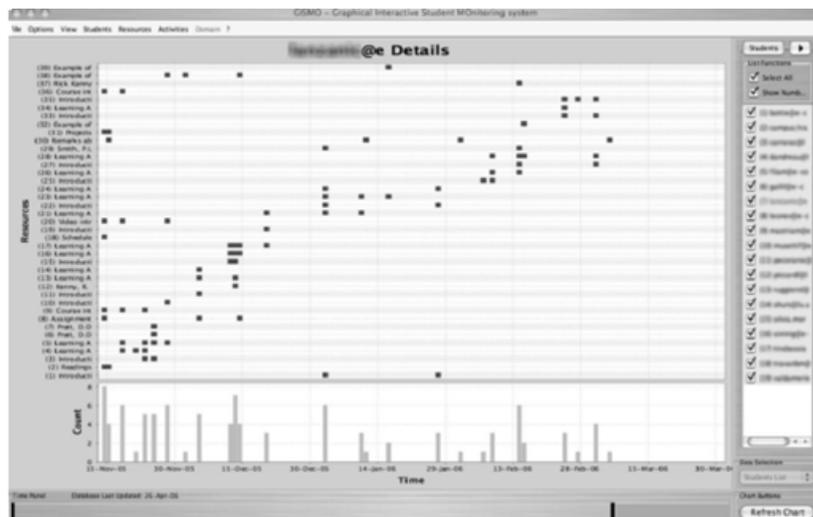


Figura 2.7: Detalhes dos acessos feitos por um aluno aos recursos do curso

forma esta ferramenta pretende permitir que os professores consigam estar a par do que acontece em turmas de ensino à distância e que possam melhorar os cursos, tornando-os cada vez mais adequados às necessidades dos alunos, de modo a que estes consigam obter o maior sucesso possível.

2.1.3 Work In Progress - Curriculum Visualization

Os catálogos de cursos são uma referência e fonte de informação sobre os cursos para os alunos. Estes catálogos levantam alguns problemas por fazerem apenas uma descrição textual, fazendo com que novos alunos se sintam mais confusos com as opções existentes e tenham dificuldade em tomar uma decisão mais consciente.

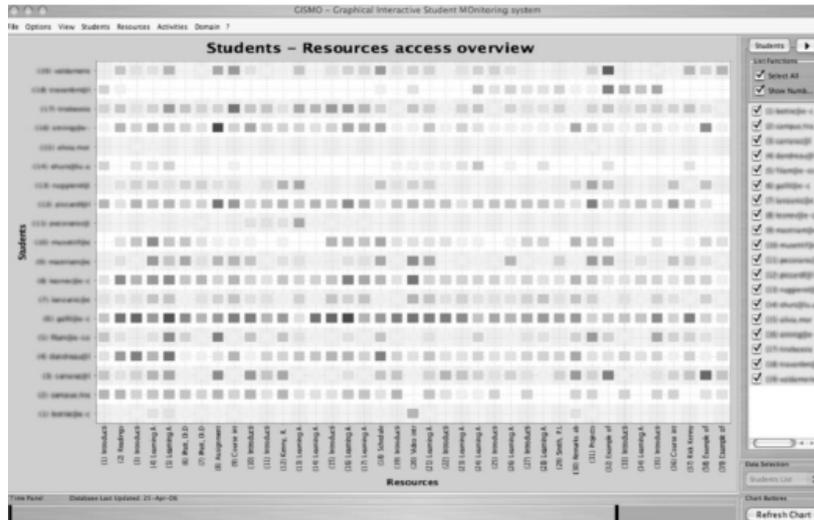


Figura 2.8: Gráfico dos acessos de todos os estudantes a todos os recursos do curso

Uma solução para para este tipo de problema é a visualização de informação, que permite a melhor compreensão e interiorização da mesma. Foi desenvolvida uma aplicação, Curriculum visualization [3], que permite visualizar a informação curricular gerando um grafo direccionado, em que um nó sombreado representa uma cadeira opcional e um nó em branco representa uma cadeira obrigatória, e as arestas representam a relação entre as cadeiras, Figura 2.9.

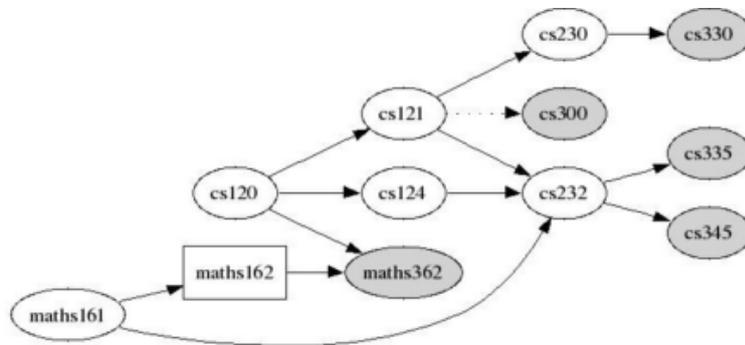


Figura 2.9: Visualização do currículo do curso de ciências da computação

Espera-se que desta forma seja possível que várias instituições educacionais e de pesquisa possam melhorar a sua forma de comunicar e que os alunos estejam melhor informados e possam tomar mais conscientemente as suas decisões sobre as escolhas de cursos. A aplicação encontra-se em desenvolvimento e conseguiu provar-se que através da visualização os alunos conseguiram de facto ficar mais esclarecidos sobre os percursos que podem fazer.

Existem algumas melhorias em vista para a aplicação tais como incorporar mais informação para a visualização dos dados, pois existem algumas lacunas e os diagramas gerados são estáticos. O autor do artigo menciona que irá existir uma nova versão para colmatar

estas falhas de forma a melhorar os resultados obtidos, mas até à data não existe notícia de que qualquer nova versão desta aplicação tenha sido desenvolvida.

2.2 TÉCNICAS E MÉTODOS DE VISUALIZAÇÃO DE INFORMAÇÃO APLICADOS A DADOS EDUCACIONAIS

Neste ponto são abordadas algumas técnicas e métodos de visualização para representar dados académicos.

2.2.1 *Visual Analysis of College Student's Scores in English Test*

De forma a conseguir representar a informação da disciplina de Inglês de uma maneira que torne mais eficaz a sua leitura e compreensão, foi utilizada a visualização de informação para representação de dados. Foi escolhida como técnica de representação o método de coordenadas paralelas em que são usados N eixos paralelos e equidistantes para representar dimensões de um conjunto de dados multidimensionais, mas por ser insuficiente para descobrir mais detalhes da informação são utilizados alguns modelos interativos que permitem reduzir os resultados desinteressantes obtidos pelos algoritmos de Data Mining e aumentar a fiabilidade dos resultados [17]. Os modelos utilizados, visíveis na Figura 2.10, são a classificação, que permite aos utilizadores dividirem os dados em conjuntos arbitrários de classes, a média, que possibilita através de interações com o gráfico, obter a média total ou de partes de dados, *boxplot*, que permite medir a dispersão dos dados, a troca de eixos, que permite aos utilizadores descobrirem relações internas entre dois atributos, a correlação, que calcula a correlação entre dois conjuntos de dados em dois eixos paralelos, a regra da associação, em que o utilizador pode prever a ocorrência do atributo A baseada na ocorrência do atributo B, e *Roll-up* e *Drill-Down*, permitem apresentar os dados em diferentes níveis hierárquicos.

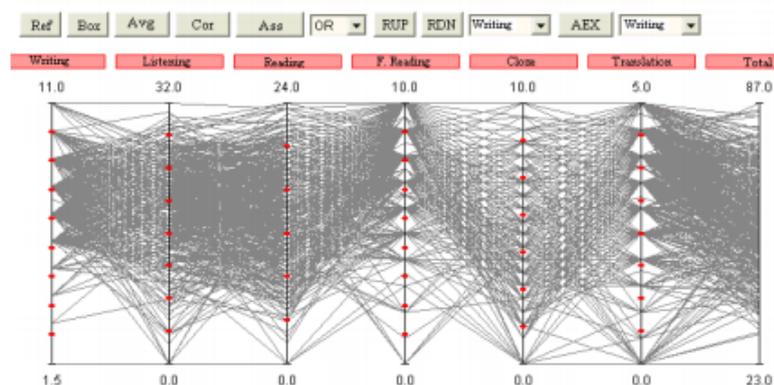


Figura 2.10: Dados em coordenadas paralelas

Na Figura 2.11 é utilizado o modelo interativo da classificação, onde os utilizadores podem agrupar os dados conforme necessitam, trocando a ordem dos eixos, facilitando assim a análise dos mesmos, observando-se por exemplo que os resultados na escrita são pobres independentemente dos alunos terem no seu total notas altas ou baixas.

Na Figura 2.12 são usados os modelos *boxplot* e média, sendo possível concluir que a maioria dos estudantes consegue ler e usar correctamente o vocabulário mas têm mais

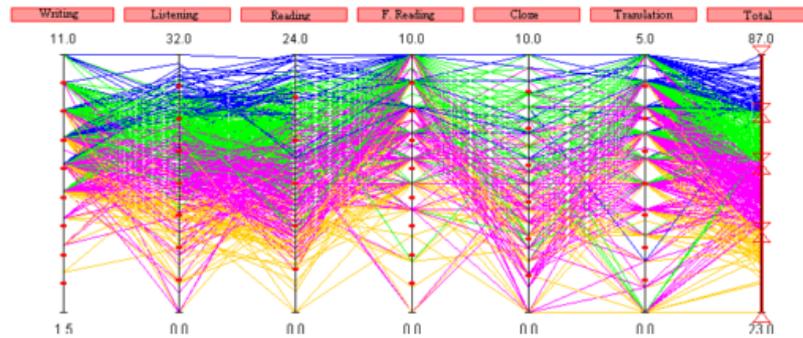


Figura 2.11: Classificações dos alunos

dificuldade na escrita, permitindo assim aos professores tomarem medidas apropriadas para melhorarem os pontos mais fracos. Foi também calculada a correlação, usando o modelo respectivo, entre a leitura e a escrita, mostrando que estas não são independentes, pelo que é necessário que as suas capacidades sejam desenvolvidas como um todo, como mostra a Figura 2.13.

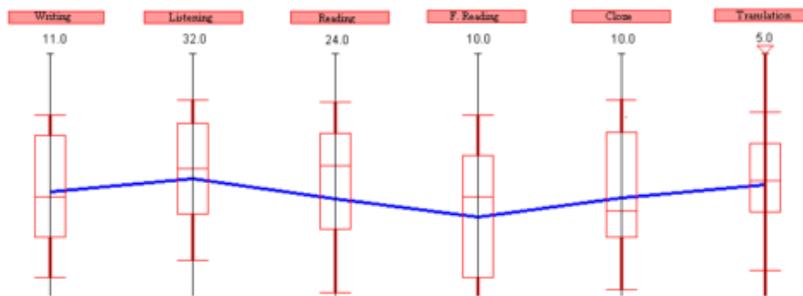


Figura 2.12: Comparação das médias e Boxplot

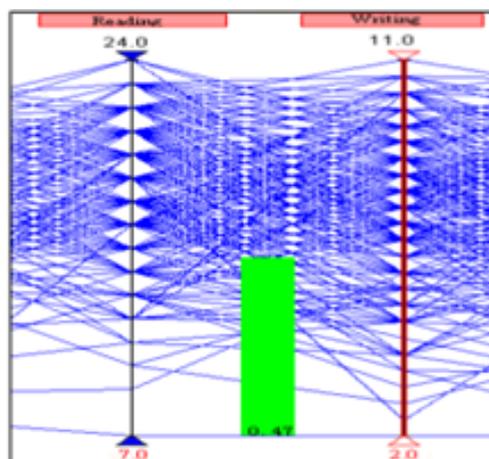


Figura 2.13: Correlação entre leitura e escrita

Por fim é utilizada a regra associativa que permite descobrir relações entre dados com atributos diferentes. Na Figura 2.14 é visível a utilização deste método e o seu resultado neste caso indica que apenas 53% dos estudantes que tiveram 15 valores, tiveram nota superior a 9 na componente escrita.

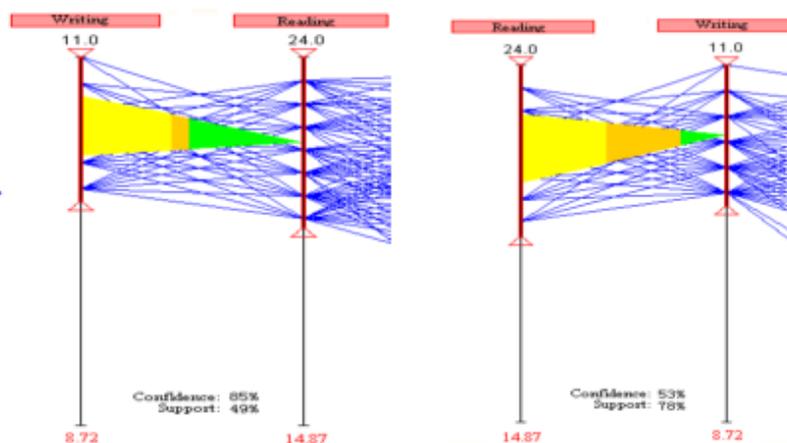


Figura 2.14: Regra da associação entre leitura e escrita

Através da utilização destes modelos interactivos que permitem visualizar e interagir com a informação, e que têm como vantagem permitir uma rápida visão geral dos dados [14] e mostrar muitas variáveis simultaneamente[5], concluiu-se que as capacidades de leitura, escrita e fala dos estudantes não podem ser encaradas como independentes e que uma capacidade não pode substituir a outra. Futuramente espera-se conseguir estender este estudo a várias universidades e as técnicas de visualização serão aplicadas de forma mais detalhada.

2.2.2 Visual Exploration for Time Series Data Using Multivariate Analysis Method

De modo a reflectir as diferentes habilidades dos estudantes é desenvolvido um sistema que permite a exploração e visualização de dados sobre o desempenho dos alunos, chamado AVOJ (Analysis and Visualization for Online Judge of Teaching Data) [4]. A partir deste sistema é possível fazer uma análise visual que mostra as tendências mais gerais nos processos de ensino, agrupar os estudantes de acordo com as suas capacidades, analisar o seu progresso, e monitorizar a informação em tempo real. A utilização da visualização dos dados deve-se ao facto de evitar que os leitores se sintam confusos e não consigam interpretar resultados, é útil também para ajudar a descobrir novos padrões e trocar informação. Os estudantes foram agrupados de acordo com as suas notas e capacidades. Para representar a informação e tornar fácil a sua leitura foi utilizado um esquema de cores com o objectivo de se conseguir distinguir os dados com mais clareza e assim facilitar a sua análise, Figura 2.15.

Na Figura 2.16 são utilizados múltiplos eixos para monitorizar e mostrar estatísticas em tempo real.

Como conclusão deste estudo, é possível perceber a quantidade de estudantes com boas notas e que têm como capacidades serem inteligentes e trabalhadores, os estudantes com más notas e pouco trabalhadores e os estudantes que são eficientes e inteligentes mas que demonstram interesse apenas durante o período necessário para completarem a sua tarefa. Espera-se que com a ajuda do sistema AVOJ seja possível melhorar os resultados educacionais e que este promova o desenvolvimento da educação.

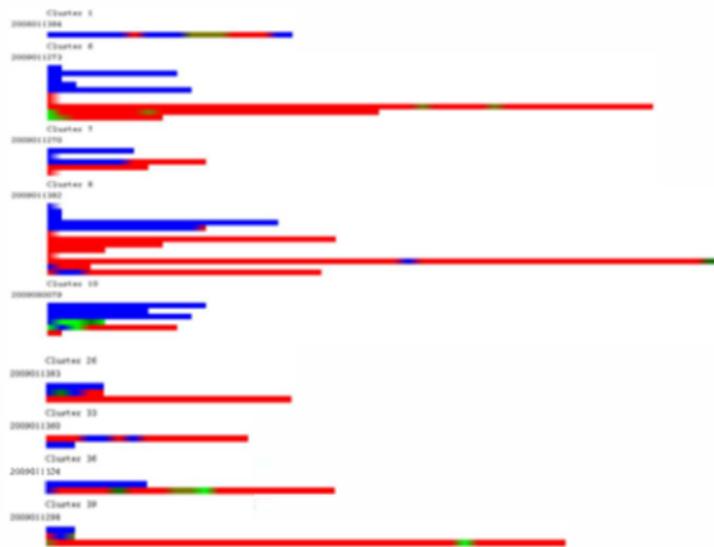


Figura 2.15: Agrupamento de estudantes

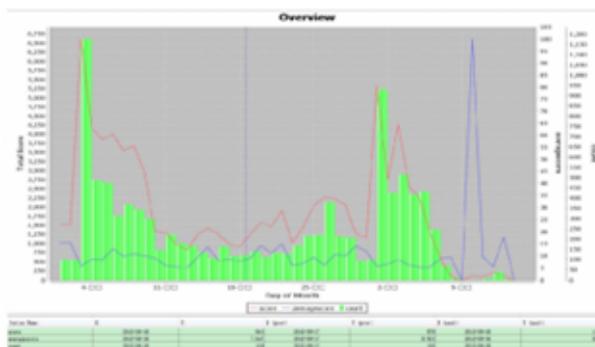


Figura 2.16: Comparação de vários parâmetros

A visualização apresentada tem apenas um inconveniente presente na Figura 2.15. As cores escolhidas por terem o mesmo nível de saturação e brilho e por serem cores que estando juntas resultam num visual mais chocante acabam por competir pela atenção visual do leitor aumentando a dificuldade em descobrir padrões ou dados subjacentes. Seria por isso preferível ter sido utilizada uma paleta cujas cores estivessem mais próximas na roda de cores criando um visual mais agradável e variando o nível de brilho[12]. Relativamente à Figura 16, utilizar este tipo de representação nesta situação é uma boa opção pois permite reduzir o ruído visual e a quantidade de texto, e é uma solução eficaz para representar dados qualitativamente devido aos eixos permitirem definir áreas ou agrupamentos [11], e assim os utilizadores podem inquirir e observar de imediato as tendências mais gerais e perceber melhor a situação geral.

2.2.3 Visualizing Student Histories Using Clustering and Composition

A recolha de dados do percurso académico dos estudantes, depois de analisada, pode ser útil na descoberta de informação importante para melhorar o desempenho dos alunos, melhorar

aspectos do curso tais como a carga de trabalho, e a nível individual permite prever o desempenho futuro de um estudante.

Neste estudo os estudantes foram agrupados de acordo com as suas notas, os grupos formados podem ser visualizados usando composições que mostram as características dos grupos e variações no seu histórico [13]. Para se conseguir utilizar as técnicas de composição, a informação do histórico de cada estudante é uma trajectória de duas dimensões. Para tal utiliza-se uma representação por eixos que permite uma representação mais fiel das características temporais dos dados, Figura 2.17. Esta representação, apesar de simples torna-se pouco eficaz quando se deseja ver as características e tendências dos dados de muitos estudantes de uma só vez. Na Figura 2.18 também é utilizada uma representação bi-dimensional com uma escala de cores vermelho-amarelo-verde, por ser intuitiva e fácil de associar as cores aos resultados dos alunos, verde significa muito bons resultados e vermelho significa maus resultados. Este esquema de cores pode torna-se problemático para daltónicos devido às cores verde e vermelho. O facto de se utilizar também este tipo de representação faz com que haja sobreposição de valores, para fazer uma distinção dos mesmos, o autor sugere a utilização da opacidade.

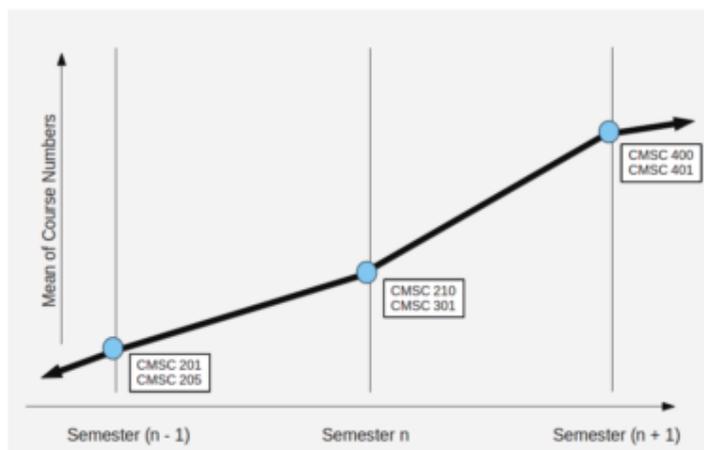


Figura 2.17: Representação bi-dimensional do histórico dos estudantes

De forma a poder mostrar as características espaciais das trajectórias dos alunos, que permitem ao utilizador ver a densidade, forma e compactação dos conjuntos de trajectórias é utilizado o algoritmo *Level Set*, que modela a expansão bi-dimensional de uma fronteira ao longo do tempo a partir de uma condição inicial e consegue calcular como a fronteira se expande e quando irá alcançar uma determinada localização, como mostra a Figura 19. A parte brilhante representa a expansão da fronteira ao longo do tempo e o afastamento da média da trajectória e a parte escura representa o tempo inicial. De modo a conseguir guardar a distância à média da trajectória, o algoritmo foi modificado de modo a reter o tempo paramétrico da média da trajectória, permitindo assim que as várias trajectórias sejam transformadas e compostas juntas, para a renderização final.

A fase final da composição consiste em atribuir cores a cada píxel na renderização final. É necessário atribuir um esquema de cores adequado para que o leitor consiga visualizar de forma correcta a estrutura espacial dos dados e consiga fazer uma interpretação correcta dos mesmos. Para tal é utilizada a abordagem *color blending* e o método *weaving*, Figura 2.20.

A abordagem *color blending* permite mostrar a densidade, média e o desvio padrão enquanto que o segundo método mostra o valor de um atributo de uma trajectória seleccionada aleatoriamente.

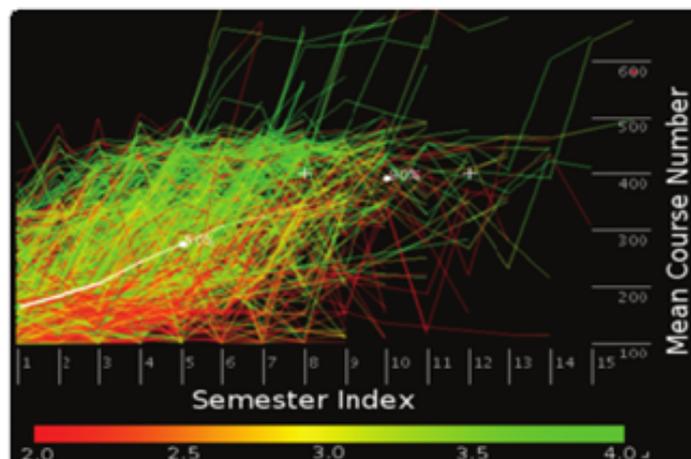


Figura 2.18: Representação bi-dimensional a cores do histórico dos estudantes



Figura 2.19: Exemplo do algoritmo *Level Set* quando a média da trajetória é usada como condição inicial.

O esquema de cores escolhido para a primeira abordagem continua a ser vermelho-amarelo-verde, alterando a sua saturação de acordo com o desvio padrão e variando o brilho de acordo com a densidade. A segunda abordagem utiliza o modelo de cores HSB, em que é seleccionado o *hue* apropriado para cada *weaving cell* do atributo da trajetória escolhida. A abordagem *color blending* tem como vantagem permitir mostrar grandes quantidades de informação de forma intuitiva e o método *woven coloring* obriga o leitor a avaliar a variabilidade e valor geral para uma dada região da composição. Na Figura 2.21 é usado um esquema alternativo de cores, usando uma escala de cores *Brewer*, nos tons vermelho-laranja-amarelo, que funciona melhor no método *weaving*, ao invés do modelo *blended* por esta escala conter apenas índices discretos de cores.

Ao concluir este estudo confirmaram-se várias tendências de que já havia conhecimento, enquanto que novas relações foram descobertas. Por exemplo na Figura 2.22 são visíveis as trajetórias femininas e masculinas. A imagem da esquerda utiliza a técnica da composição, onde os resultados parecem muito semelhantes mas na imagem da direita são visíveis variações nas notas entre os dois sexos, concluindo-se que o sexo feminino teve melhores notas mas a sua progressão foi um pouco mais lenta.

Descobriu-se que é muito importante que um estudante consiga um bom resultado na primeira cadeira de ciências da computação, pois só os alunos com melhores notas nesta cadeira tiveram sucesso na continuidade do curso, permitindo assim identificar alunos que necessitem de ajuda por parte dos professores.

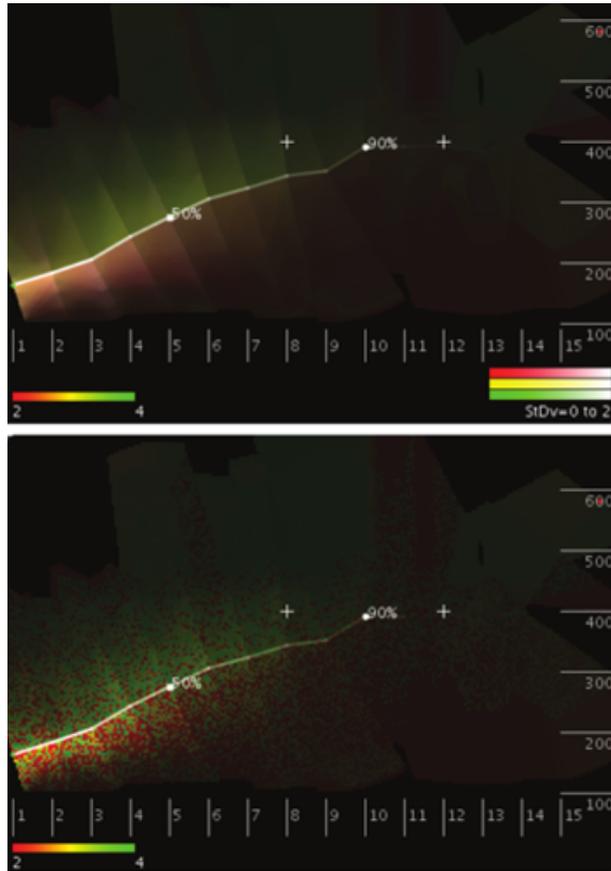


Figura 2.20: color blending e o método weaving

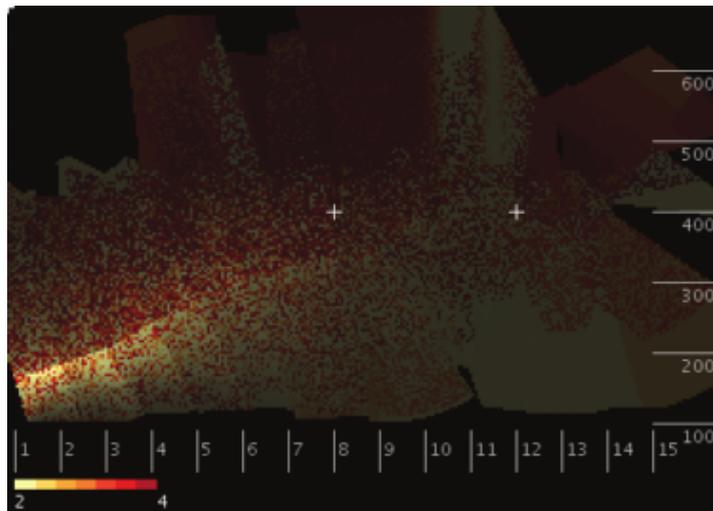


Figura 2.21: Método weaving com uma escala de cores Brewer

Detectaram-se também alguns problemas com esta abordagem tais como a incapacidade de descobrir a causa das tendências acima descobertas.

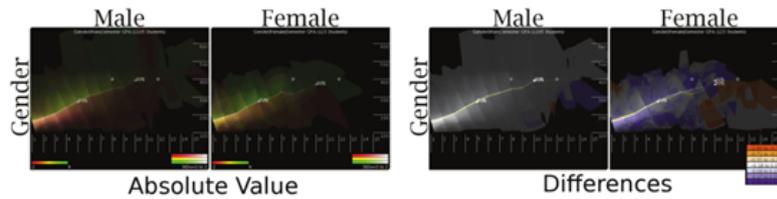


Figura 2.22: Diferenças nas notas entre os dois géneros.

Como conclusão, a nível de visualização, na Figura 2.19 a utilização de gradientes em geral é arriscada pois este efeito acrescenta ruído à imagem e adiciona uma sensação de luminosidade e sombra que leva o cérebro a assumir a região mais clara como mais próxima e a mais escura como estando mais longe [12], provocando exactamente o efeito contrário ao pretendido neste caso, já que, como acima mencionado, a parte brilhante(mais clara) representa o distanciamento.

2.2.4 Visualizing Trends in Student Performance across Computer Science Courses

Numa tentativa de isolar a razão porque muitos estudantes abandonam o curso na área de ciências da computação, foi elaborado um estudo que permite visualizar que padrões acontecem e se repetem sobre os chumbos e sucessos dos alunos enquanto frequentam o curso[16]. Após a recolha dos dados e por serem em grande quantidade, é essencial representar a informação de modo a poder visualizá-la, tornando-se muito mais fácil para os leitores conseguirem percebê-la e analisá-la. Para tal foi utilizada uma estrutura de nós e arestas com cores, em que as cores representam o desempenho dos alunos permitindo assim diferenciar melhor grupos de estudantes e identificar grupos com comportamentos semelhantes. Os nós representam eventos e a sua largura representa o número de estudantes em cada evento e as arestas representam os alunos deslocando-se entre os eventos e a sua largura também é referente ao número de alunos que se está a deslocar entre cada nó. Na Figura 2.23 estão representados percursos de vários estudantes enquanto que na Figura 2.24 se encontra o percurso de um só estudante.

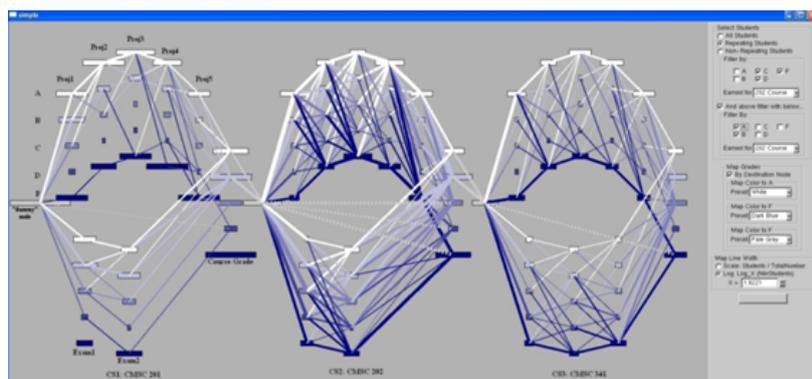


Figura 2.23: Representação em grafos do percurso dos estudantes.

É possível interagir com a aplicação, seleccionando categorias de estudantes tais como alunos que repetiram pelo menos uma cadeira e alunos que nunca chumbaram a nenhuma cadeira. Os estudantes que não façam parte do resultado da pesquisa não são apresentados, mas a sua presença é representada pela largura dos nós, que não é alterada. O objectivo deste



Figura 2.24: Percurso de um estudante.

estudo é ajudar os professores a identificar mais rapidamente problemas e tendências de modo a aumentar o sucesso dos alunos nesta área. Ao representar os dados obtidos facilitando a sua leitura foi possível tirar algumas conclusões sobre como por exemplo estudantes que chumbam a determinada disciplina, provavelmente irão também reprovar noutra, e a situação contrária, mostrando que alunos que tiram determinada nota numa cadeira, têm maior probabilidade de passar a outra cadeira. Foi também possível concluir que quando se chumba a uma disciplina, voltar depois a repeti-la não implica que se obtenha aprovação.

O estudo refere ainda que futuramente tenciona adicionar técnicas que permitam o reconhecimento automático de padrões na visualização dos dados, permitindo assim tomar medidas mais adequadas para ajudar os alunos a terem maior sucesso académico. A estrutura de nós e arestas utilizada para representar os dados, neste caso acaba por dificultar um pouco a visualização por haver sobreposição de arestas, principalmente quando arestas mais grossas se sobrepõem a arestas mais finas, ocultando-as, tornando-se mais complicado para o leitor perceber se existem alunos a deslocarem-se entre nós e se são em grande ou pequena quantidade.

2.2.5 *A Multidimensional Data Visualization Method Based On Parallel Coordinates and Enhanced Ring*

Têm sido apresentadas várias propostas de técnicas de visualização de informação, mas a grande maioria foca-se apenas nas características dos dados, tais como relações hierárquicas, mas sem as quantificar. Mostrar mais detalhes sobre a distribuição e relação dos dados durante o processo de visualização é muito importante mas encontra alguns obstáculos tais como, quando se lida com grandes quantidades de dados, torna-se difícil quantificar a relação devido ao próprio método e ao tamanho do ecrã e a sua visualização em massa acaba por levar à desorganização. É então apresentado um método de visualização baseado em coordenadas paralelas e *enhanced ring* (PCER) que possibilita uma visualização eficaz de dados em massa a nível qualitativo e quantitativo por proporcionar aos utilizadores uma melhor percepção da informação relativamente a outros métodos usados [9]. O método das coordenadas paralelas com esquemas interactivos permite aos utilizadores interagirem directamente com as parcelas das coordenadas e possibilita uma grande flexibilidade em explorar e descobrir padrões subjacentes. Este método melhora os efeitos de visualização mas os utilizadores não conseguem ainda quantificar tendências ou relações num conjunto de dados. O método *enhanced ring chart* utiliza um gráfico em anel para representar os dados, de modo a que cada anel é determinado por uma propriedade dos dados, e é estabelecida uma relação entre

os diferentes anéis do gráfico. Assim, se os utilizadores quiserem obter informação sobre as notas dos estudantes em cada disciplina, então esta vai ser uma propriedade, representada por um anel. A vantagem deste método sobre o *ring chart*, é que o primeiro permite adicionar mais propriedades de modo a dividir os dados, e os dados de cada anel são diferentes. Neste exemplo, no *ring chart*, Figura 2.25, o conjunto de dados dos três anéis é o mesmo, os estudantes que tiveram uma classificação acima de 79 na disciplina de física, enquanto que no *enhanced ring chart*, Figura 2.26, os conjuntos de dados dos três anéis são diferentes, por exemplo o conjunto de dados do segundo anel é sobre estudantes que tiveram acima de 79 na disciplina de física, 83 na disciplina de Chinês e 89 na disciplina de Inglês.

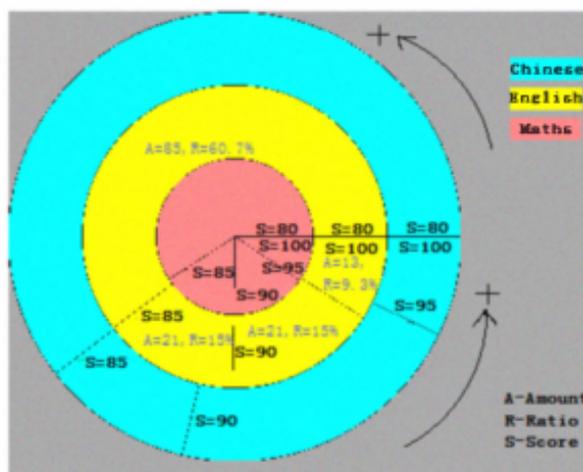


Figura 2.25: Gráfico em anel normal, usa o mesmo conjunto de dados para os três anéis.

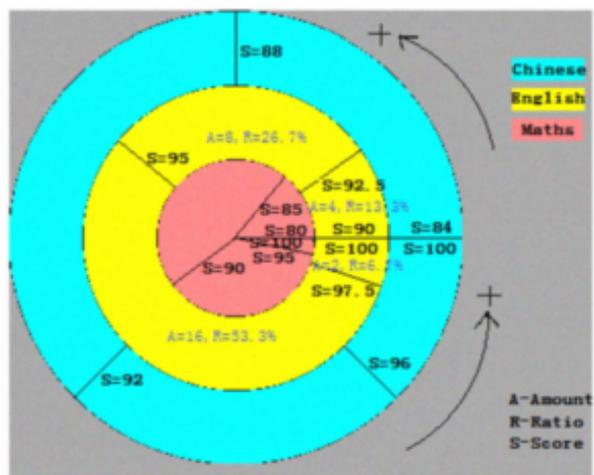


Figura 2.26: Enhanced Ring Chart.

Utilizando o método PCER, que combina as duas técnicas acima descritas, os utilizadores podem seleccionar a dimensão e o alcance dos dados onde ocorre o desalinho, de modo a poderem ver a proporção dos dados e ajustar a escala de acordo com as suas preferências de forma a obterem um resultado mais detalhado. Quando é seleccionada uma dimensão nas

coordenadas paralelas, esta dimensão é adicionada ao gráfico em anel como uma propriedade ou como condição. As coordenadas paralelas redesenham a informação de acordo com o conjunto de dados seleccionados, Figura 2.27. Nesta imagem são visíveis cinco áreas representadas por letras. A área A mostra estatísticas preliminares, em B são especificadas as condições de escala, em C o utilizador define a condição, propriedades e a proporção para o *ring chart*. As áreas D e E são utilizadas para visualização.

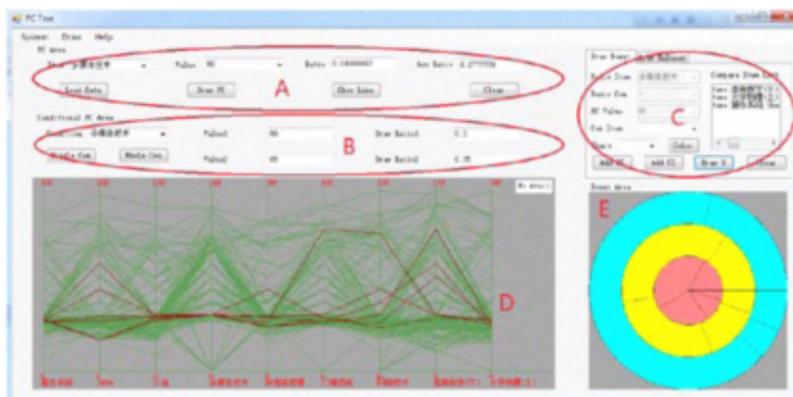


Figura 2.27: Visualização dos dados sobre as notas dos alunos usando PCER

Ao utilizar esta técnica de visualização é então possível visualizar dados multidimensionais e ter uma visão qualitativa e quantitativa mais clara dos mesmos. Não obstante as vantagens que este tipo de visualização pode ter, é necessário ter em conta que as pessoas têm muito mais facilidade em estimar e comparar áreas rectangulares que áreas circulares e o tamanho das fatias do gráfico em anel, como tal o seu uso é mais indicado quando a precisão não é a palavra de ordem. Então nesta situação, Figura 2.27, teria sido preferível ter-se utilizado um gráfico de barras normal por permitir uma melhor comparação dos dados [12].

2.2.6 Discussão

Os trabalhos acima descritos apresentam variadas técnicas e metodologias para a visualização de dados, obtendo alguns mais sucesso que outros. Segue-se uma análise a cada linha da Tabela 1 que resume as características principais de cada artigo, para facilitar a sua exploração.

Na primeira linha é indicado qual o tipo de visualização usada em cada artigo. Existem alguns que se evidenciam pela sua relevância para o tema deste trabalho tal como, *Visualizing Student Histories Using Clustering and Composition*. Este mostra as trajectórias do percurso dos alunos, o que permite depois avaliar as dependências entre as diferentes cadeiras do curso de ciências da computação, muito à semelhança do que se pretende fazer neste tema. Os restantes artigos, apesar dos tipos de dados apresentados serem diferentes dos dados que irão ser utilizados nesta tese, são relevantes por mostrarem algumas das várias vertentes em que a visualização pode ser útil no campo educacional, tais como as relações que pode haver entre as diferentes capacidades dos alunos e suas dependências, a forma como os resultados dos estudantes são influenciados pelas suas capacidades, descobrir padrões de reprovações e estabelecer relações entre cadeiras.

Outro aspecto importante, relevante nos artigos apresentados, que se encontra representado na segunda linha da Tabela 1, é a eventualidade de o utilizador interagir com a visualização. A possibilidade de interactuar com os dados leva a que os leitores se sintam

Tabela 1: Sumarização das características dos diferentes trabalhos.

	Generation of Graphical Representations of Student Tracking Data in Course Management Systems	Monitoring an Online Course with the GISMO Tool: A Case Study	Work In Progress - Curriculum Visualization	Visual Analysis of College Student's Scores in English Test	Visual Exploration for Time Series Data Using Multivariate Analysis Method	Visualizing Student Histories Using Clustering and Composition	Visualizing Trends in Student Performance across Computer Science Courses	A Multidimensional Data Visualization Method Based On Parallel Coordinates and Enhanced Ring
Método de Visualização	Cognitive Matrix e Single-axis composition	Ferramenta gráfica GISMO	Grafos direccionados	Coordenadas Paralelas	Método de análise multivariada	Composição e Clustering	Grafos não direccionados	Coordenadas Paralelas e <i>Enhanced Ring Chart</i>
Interação	Sim	Não	Não	Sim	Não	Não	Sim	Sim
Cor	Sim, apesar dos tons escolhidos serem desadequados em alguns gráficos	Utilização de imagens a preto e branco	Sim	Sim	Sim	Sim, utilização de gradientes	Sim	sim
Adequação do Método de visualização	Desadequado em algumas situações	Adequado	Adequado	Adequado	Adequado	Adequado	Apresenta algumas desvantagens	Desadequado para o fim a que se destina
Principal desvantagem?	Sobreposição de dados e Escolha de cores desadequada	Não permitir interação	Não permitir interação		Utilização de cores desadequada e Não permitir interação	Utilização de gradientes com efeito contrário ao pretendido e Não permitir interação	Sobreposição de dados	Escolha de gráfico pouco adequada

mais capacitados e motivados a explorar a visualização por si próprios. Parte dos trabalhos mencionam e indicam essas mesmas vantagens.

Os restantes artigos por serem estáticos, têm como inconveniente o facto de só poderem oferecer algumas vistas dos dados, pelo que são necessárias várias vistas estáticas para mostrar todas as perspectivas da mesma informação. Também ocorre a situação de que o número de dimensões para representar os dados é limitado quando se quer apresentar todos os elementos ao mesmo tempo na mesma imagem [10].

Na linha seguinte é observável quais os artigos que usam cores nas suas visualizações. Assim, destacam-se os trabalhos *Generation of Graphical Representations of Student Tracking Data in Course Management Systems*, *Visual Exploration for Time Series Data Using Multi-Variate Analysis Method* e *Visualizing Student Histories Using Clustering and Composition*.

Os primeiros dois artigos mencionados apresentam algumas fragilidades ao nível da forma como usam as cores para representar os dados, por trabalharem com métodos de visualização que usam eixos para representar a informação e por estes deixarem subjacente que existe uma ordem. A cor no cérebro humano não tem um alinhamento natural, por exemplo, não existe uma regra que dite que o azul é seguido do verde tal como existe para a ordem numérica, em que o 1 é seguido do 2, como tal, deveriam seguir outro tipo de abordagem em termos de cor. O terceiro artigo peca pela utilização de gradientes para mostrar os resultados do algoritmo *Level Set* acabando por provocar o efeito contrário ao pretendido.

Na penúltima linha são mencionados quais os artigos que utilizam métodos desadequados para demonstrar os objectivos pretendidos. Neste aspecto, destaca-se a utilização de gráficos em anel no artigo *A Multidimensional Data Visualization Method Based On Parallel Coordinates and Enhanced Ring*, em que as desvantagens deste tipo de gráficos se encontram descritas no respectivo ponto.

Também são de referir os gráficos usados no artigo *Visualizing Trends in Student Performance across Computer Science Courses*, que acabam por não ser uma boa escolha por dificultarem a visualização de dados, devido à sobreposição de arestas. O trabalho *Genera-*

tion of Graphical Representations of Student Tracking Data in Course Management Systems, também tem alguns pontos que deveriam ser melhorados a nível das visualizações usadas, que não se relacionam com o método usado, mas sim com a escolha de cores desadequada, já indicado anteriormente.

Por fim, a última linha da Tabela 1, tendo em conta as dimensões anteriores, resume quais as principais desvantagens de cada artigo. A falta de interacção por parte de alguns dos trabalhos analisados destaca-se como sendo o principal problema apontado, e pelas razões acima descritas, é um requisito fundamental para a visualização desenvolvida nesta tese.

Outro problema que se repetiu em dois trabalhos, foi a utilização desadequada de cores, em parte pela utilização do espectro do arco-íris, com o objectivo de gerar sequencialidade, que é uma propriedade que este mapa não possui [21].

A sua utilização com o objectivo de mostrar continuidade ou máximos e mínimos deve ser feita em aplicações que estejam enraizadas na cultura dos utilizadores, como por exemplo, cartas meteorológicas, caso contrário o utilizador não saberá qual a ordem que está a ser considerada [21].

Tendo em conta estas características, para a implementação da visualização consideraram-se apenas variações no brilho da cor, para mostrar sequências de valores mais ou menos elevados.

Por fim, outra situação também relevante, é a sobreposição de artefactos visuais. Ao haver sobreposição de dados perde-se informação, uma vez que esta fica muito mais difícil de perceber ou mesmo ilegível. Assim, ao desenvolver a visualização proposta, teve-se em conta este problema, usando métodos que o evitassem.

Como conclusão, retiraram-se aprendizagens importantes deste capítulo, sobre os erros mais comuns, e que são de evitar, bem como aspectos positivos sobre por exemplo os métodos de visualização mais adequados para representar certos tipos de dados.

Esta análise foi também relevante para saber o que se tem feito na área da visualização de informação educacional, e como os seus resultados são usados, servindo de guia para o desenvolvimento de novas funcionalidades neste trabalho. Assim este distingui-se relativamente aos artigos mencionados, por facultar uma visão geral das relações entre as várias cadeiras de um curso, dando a perceber quais as disciplinas mais importantes e mais críticas para que se tenha sucesso nas cadeiras dos semestres seguintes. Possui também a característica de filtrar informação, que não se encontra em nenhum dos trabalhos descritos, permitindo saber relativamente a uma dada disciplina, qual a consequência de ter tido aprovação ou reprovação a outra cadeira simultaneamente ou em semestres separados. Procura-se assim, criar uma ferramenta inovadora e útil para a comunidade académica para a análise de informação de padrões educacionais.

PROJETO EDUCARE

O projecto Educare consiste na visualização e modelação de comportamentos em educação. O seu objectivo é providenciar a comunidade educativa com um sistema de apoio à decisão tendo como um dos principais objectivos, um conjunto de técnicas de descoberta e visualização de informação, inovadoras e adaptadas à realidade do contexto educacional.

O projecto desenrolou-se em quatro fases distintas, o desenho do sistema, a definição de técnicas de descoberta de comportamentos e antecipação de falhas e o desenho de visualização de comportamentos. Esta tese está inserida na última tarefa, que tem como principal objectivo estudar uma forma de representar padrões educacionais.

Nesse contexto, foi primeiramente desenvolvida uma primeira visualização, pela Eng. Sandra Gama, designada por visualização multi-camada, Figura 3.1, que combina informação estática e dinâmica na mesma vista.

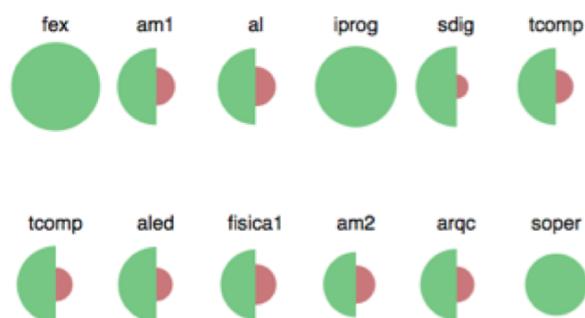


Figura 3.1: Visualização Multi-Matricial

Nesta visualização, cada cada camada corresponde a um semestre de um programa curricular, representando as disciplinas desse semestre que têm algum tipo de relação com as outras. Assim, as cadeiras são geralmente representadas como círculos verdes com tamanho proporcional ao número total de estudantes que as completaram. Quando existem dados sobre o insucesso, o círculo da cadeira é sub-dividido em dois semi-círculos, mostrando informação sobre a aprovação e a reprovação através do código convencional de cor [21]. O verde à esquerda, representa o número de alunos aprovados, e o vermelho, à direita, mostra o número de estudantes que reprovaram. Este modelo permite explorar com facilidade a vertente temporal, que corresponde à escolha de várias disciplinas realizadas pelos alunos num dado instante de tempo, dando origem a restrições que levam à visualização imediata dos vários percursos seguidos pelos alunos ao longo do curso, como se pode ver na Figura 3.2.

Apesar destas funcionalidades, é possível representar mais tipos de informação transmitida pelos dos dados utilizados. Possíveis exemplos são, uma visão do conjunto de quais as disciplinas envolvidas em mais padrões, e portanto mais críticas no plano curricular, e

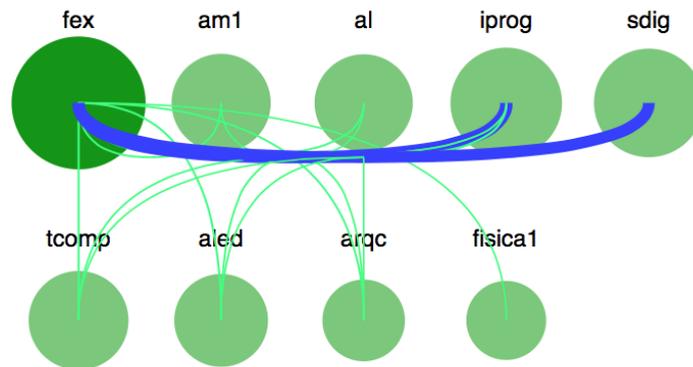


Figura 3.2: Exploração das vertentes temporal e dinâmica

comparação de padrões e filtragem dos mesmos. Nesse sentido, a visualização anterior não é a mais indicada para demonstrar tais características, pelo que surgiram duas possibilidades: (1) realizar alterações à visualização já desenvolvida ou (2) desenvolver um novo mecanismo que mostrasse os novos tipos de informação.

A primeira opção foi rapidamente descartada pois dificultaria a tarefa de mostrar as várias vistas obtidas, através do cruzamento das variáveis existentes. A sua usabilidade seria também posta em causa, pois ou ficaria muito mais limitada ou então correria o risco de se tornar menos efectiva por ter um maior grau de complexidade, e de levar o utilizador a focar maioritariamente a sua atenção na manipulação da ferramenta, ao invés de o fazer na informação transmitida.

Justificou-se portanto o desenvolvimento de uma nova visualização, para complementar a anterior, de modo a preencher as lacunas mencionadas.

Esta denomina-se visualização multi-matricial, e é a resposta ao desafio proposto no Capítulo 1, construída com base nas aprendizagens feitas no Capítulo 2 sobre o trabalho relacionado, estando integrada no mecanismo anterior, funcionando os dois de forma fluida e coordenada. A sua concepção, foco deste trabalho, será descrita no Capítulo 4

3.1 PADRÕES EDUCACIONAIS

Neste trabalho foi utilizado o resultado da aplicação de *data mining* sequencial, a dados recolhidos durante nove anos, sobre os três primeiros anos dos percursos académicos da Licenciatura em Engenharia informática e Computadores (LEIC) pela equipa da Prof.a Dra Cláudia Antunes, do Instituto Superior Técnico.

O *data mining* sequencial tem como objectivo, dado um conjunto de sequências e um limiar de suporte mínimo, descobrir o conjunto de sequências que estão contidas em pelo menos δ sequências do conjunto de dados [22]. Esta técnica permite a descoberta de padrões sequenciais frequentes. Este método reúne os padrões esperados com base no conhecimento de fundo e também permite com o uso de relaxamentos, para a descoberta de padrões que correspondem a desvios ao comportamento esperado, tornando evidentes algumas tendências relevantes que antes eram desconhecidas.

Assim, para a obtenção dos dados finais, foram utilizados três valores de limiar de suporte (50%, 25% e 20%), que representam a percentagem de alunos, de entre todos, que satisfazem o padrão. Neste caso resultaram três conjuntos diferentes de padrões. Quanto

menor for o valor do limiar de suporte maior é o número de padrões resultante da aplicação do *data mining* sequencial. Os padrões gerados estão no formato textual e têm a seguinte estrutura:

*Padrão i = (semestre 1,...,semestre2, N total alunos),
Semestre j = disciplina1 OU (disciplina1,..., disciplinaN);*

Exemplo:

['am2', '1683'] : 1683 alunos obtiveram aprovação à disciplina am2.

['fex', ['tcomp', 'aled'], '1168'] : 1168 alunos tiveram aprovação a fex no primeiro semestre, e a tcomp e aled no segundo semestre.

['fex', ['~arqc', '~fisica1'], '591'] : 591 alunos concluíram a disciplina fex no primeiro semestre, mas reprovaram a arqc e fisica1 no segundo semestre. Neste último caso, as re-provações são representadas pelo símbolo "~".

Quando é referido que os alunos tiveram ou não aprovação a um determinado semestre, este não corresponde a um semestre lectivo mas sim ao semestre de inscrição do aluno. Não se considera por exemplo, primeiro semestre e segundo semestre de determinado ano lectivo. Neste caso o que é tido em conta é a sequencialidade das disciplinas a que o aluno foi aprovado ou reprovado. Assim, como exemplo, um estudante pode ter feito a cadeira *am1* no quarto semestre e *am2* no quinto e outro aluno, no mesmo padrão, pode ter estado inscrito a *am1* no segundo semestre e a *am2* no terceiro.

Por essa razão existem disciplinas repetidas na visualização. Desta forma esta estrutura informa quais as relações entre os diferentes semestres de inscrição e das várias cadeiras da licenciatura do curso em questão.

Mais uma vez, dado o formato textual dos padrões em causa, a sua análise é complexa, e perde-se a visão global do aproveitamento dos alunos e das dependências das várias disciplinas. Tais problemas podem ser contornados com uma visualização eficaz, que tornará estas inter-relações evidentes.

SOLUÇÃO PROPOSTA

Para chegar à visualização final foi tida em conta a pesquisa feita no trabalho relacionado, e as aprendizagens daí retiradas. Através da sua análise e consequente discussão, viram-se os erros cometidos e formas de os evitar futuramente. Houve também a necessidade de ultrapassar alguns desafios principais. Segue-se a descrição dos requisitos relacionados com a usabilidade da visualização:

- i Chegar a uma representação visual que facilite a compreensão e interpretação dos dados por parte dos utilizadores;
- ii Fornecer mecanismos de interacção, para que seja possível interagir e manipular a visualização de forma fácil, tendo em conta que o utilizador deve focar a sua atenção na informação que está a manipular e não no uso da ferramenta [14];
- iii Integrar as duas visualizações num só mecanismo, sem haver sobrecarga visual;
- iv Escolher cores adequadas para a representação dos diferentes tipos de informação acessível através da análise dos padrões.

Relativamente à funcionalidade do mecanismo desenvolvido, definiram-se os seguintes requisitos:

- i Dar a noção de sequencialidade, entre semestres. O utilizador deve perceber as relações entre as cadeiras num padrão, ao longo dos vários semestres representados.
- ii Ver quais os padrões que correspondem a cada disciplina
- iii Distinguir a quantidade de padrões em que uma disciplina está envolvida, através do brilho das cores
- iv Compreender quais as cadeiras a que é importante ter aprovação, para se ter sucesso a disciplinas futuras.
- v Filtrar padrões. Quando se tem um padrão seleccionado em que entram várias cadeiras, se o utilizador quiser observar dentro desse mesmo padrão, as relações só de algumas cadeiras, pode seleccioná-las e observar as ligações que se criam.
- vi Ter dois padrões seleccionados simultaneamente, assim é possível comparar as relações de duas cadeiras com padrões distintos.
- vii Fazer o *blend* de cores, quando existem cadeiras que se sobrepõem em padrões diferentes.

- viii Que as alterações provenientes da interacção com a visualização multi-matricial se reflitam na visualização multi-camada e vice-versa. Assim o utilizador consegue ver toda a informação disponibilizada de uma vez só, sem ter de interagir separadamente com os dois mecanismos.

Nas secções seguintes segue-se a descrição do caminho percorrido ao longo do desenvolvimento da visualização. É explicado o processo evolutivo do desenvolvimento da visualização, desde os primeiros protótipos até ao mecanismo final, descrevendo as ideias subjacentes, as diferentes abordagens tomadas e o porquê de terem sido escolhidas para implementar os requisitos definidos. São também expostos os problemas encontrados e a forma como foram solucionados.

4.1 PROTÓTIPOS DE BAIXA FIDELIDADE

Tendo em conta os diferentes tipos de informação que é possível extrair dos padrões educacionais, e de criar uma visualização que os mostre adequadamente, optou-se pelo desenho iterativo e incremental com forte ênfase na prototipagem.

Nesse sentido, foram criados vários protótipos de baixa fidelidade, de forma a explorar diferentes ideias e visualizações. Estes consistem em esboços simples desenhados à mão, utilizando lápis e borracha, e que registaram graficamente várias ideias instantâneas, de possíveis soluções para o problema proposto.

Estes esboços permitiram testar várias visões e ajudaram a decidir qual seria o melhor ponto de partida para começar o desenvolvimento do sistema. Para criar os vários protótipos é necessário ter presente que existem três dimensões neste tipo de dados:

- i Disciplinas
- ii Número de semestres
- iii Aprovações e Reprovações.

Para que a visualização seja eficaz é essencial que as relações entre estas variáveis sejam claras, e que a lista de requisitos acima definida seja cumprida para que o sistema satisfaça os objectivos a que se propõem. Seguem-se então alguns protótipos, as suas descrições, e os motivos porque foram rejeitados, até chegar ao esboço seleccionado.

Na Figura 4.1, está desenhado um grafo direccionado de forças, cujo objectivo seria unir as cadeiras dos diferentes semestres, caso estas tivessem algum tipo de relação. Este esboço tem como desvantagens não ter em conta a representação de aprovações e chumbos, dificuldade em perceber a sequencialidade dos semestres e por fim ao tentar representar todas as cadeiras existentes nos padrões, a sua legibilidade ficaria comprometida.

A Figura 4.2 demonstra uma tentativa da utilização de coordenadas paralelas. Neste caso o objectivo seria representar as várias dimensões sem a utilização de eixos cartesianos. As linhas verticais indicam os semestres, e os números representam cadeiras. Este protótipo também se revelou pouco adequado devido à dificuldade em mostrar o aproveitamento dos alunos, e ao grande número de padrões existentes. Nesta técnica, o número de dados visualizados ao mesmo tempo é limitado a aproximadamente 1000 itens, devido à sobreposição de linhas [18], mas nos dados fornecidos, principalmente nos suportes menores, este número é grandemente ultrapassado, chegando aos 16.000 itens.

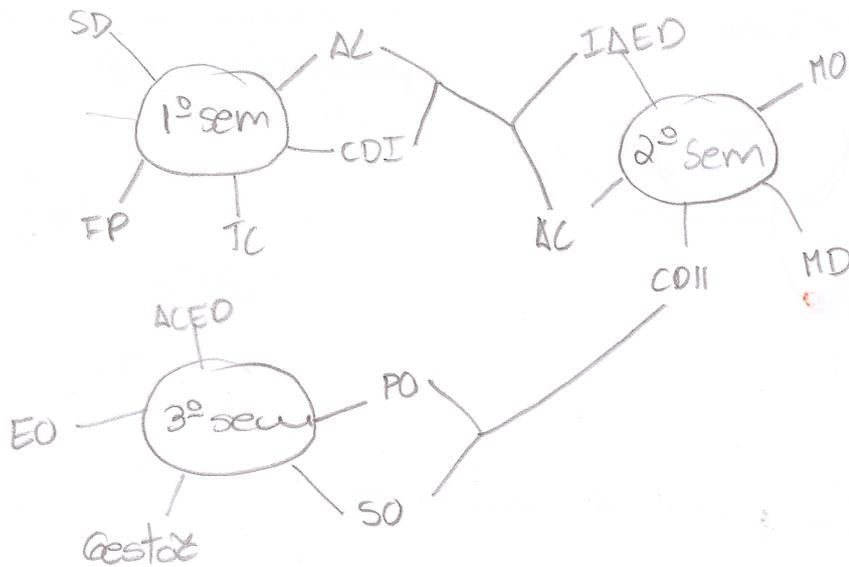


Figura 4.1: Pbf : grafo de forças

A imagem seguinte, Figura 4.3, representa uma tentativa de utilização de eixos cartesianos. Por transmitirem sequencialidade e ordenação, e por estes serem dois atributos necessários para compreender os padrões usados, foi feito um esboço da sua possível utilização. No eixo vertical, estão presentes o número de semestres. Da origem para a parte superior estão representados os semestres com dados positivos, e da origem para a parte inferior, estão representados os semestres com dados negativos. Ao longo do eixo horizontal estão as diferentes disciplinas. Este protótipo tenta mostrar de forma mais perceptível o aproveitamento dos alunos, relativamente aos anteriores. Assim cada linha representaria um padrão, e bastaria segui-la para ver o percurso dos alunos ao longo das diferentes cadeiras. Apesar de neste caso se conseguir representar as várias dimensões, há o problema do avultado número de padrões existentes, levando a uma grande sobreposição de linhas, que neste caso acabariam mais uma vez por comprometer a legibilidade do sistema. Por esta razão este esquema também não foi considerado.

Por fim, a Figura 4.4 apresenta o protótipo que após posteriores iterações levou à visualização final. A escolha de matrizes para a representação dos dados teve em conta a facilidade e clareza, relativamente aos outros esboços, de como é possível demonstrar as relações entre as três dimensões existentes, bem como a sequencialidade que é possível retirar dos padrões. Neste caso, os semestres são separados por colunas verticais, no centro da matriz, existe uma divisória horizontal que separa os padrões referentes aos chumbos a às aprovações. As linhas mostram as relações entre as diferentes cadeiras, unindo-as e a sua espessura demonstra o número de relações envolvidas.

4.2 EDUVIS

Após se ter decidido como seria a estrutura da visualização, através da análise dos protótipos de baixa fidelidade, deu-se início à sua implementação. Esta nova visualização após estar

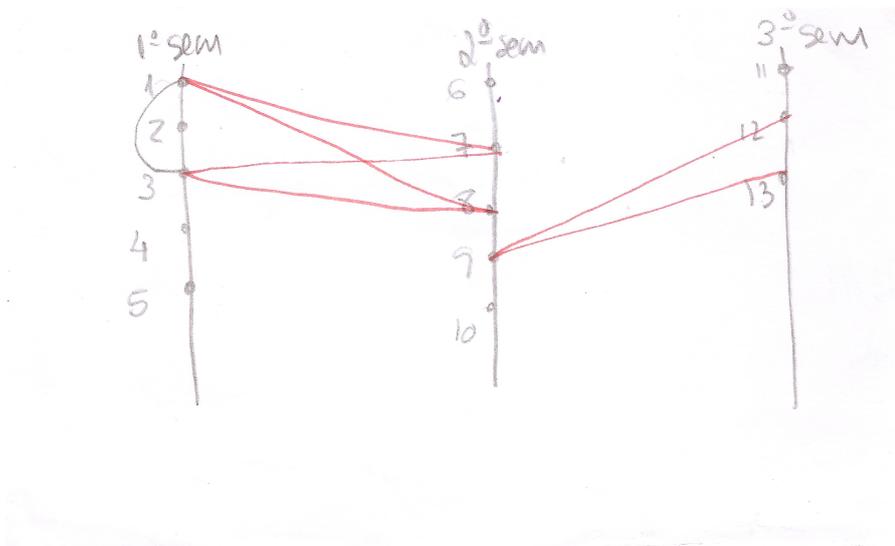


Figura 4.2: Pbf : coordenadas paralelas

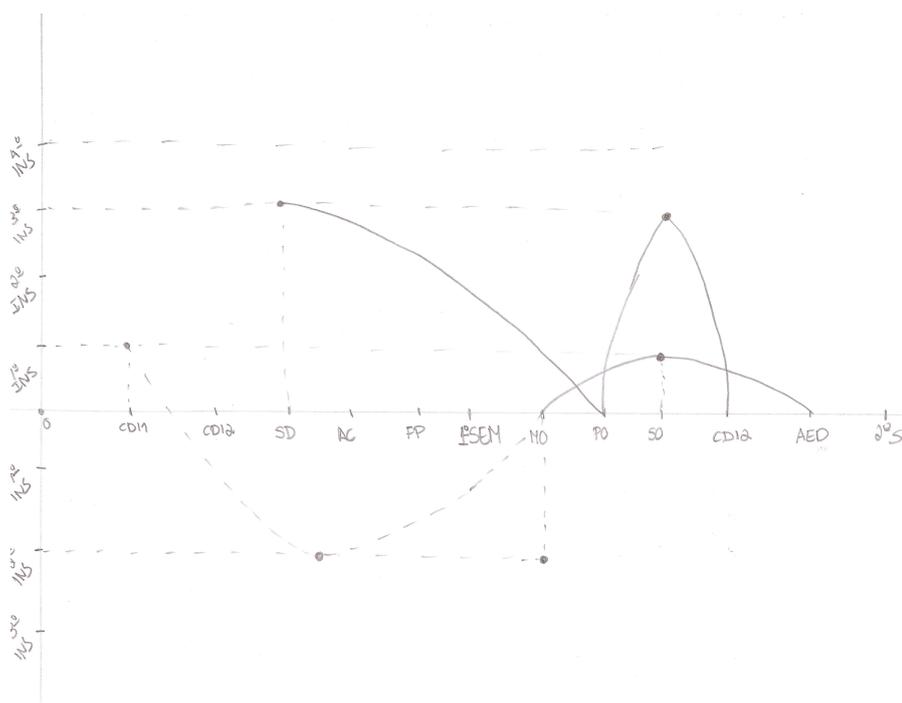


Figura 4.3: Pbf : eixos cartesianos

concluída, foi então integrada no mecanismo que já tinha sido previamente desenvolvido, visualização multi-camada, no contexto do projecto Educare.

Assim, surgiu o EduVis, que corresponde a uma visualização coordenada, composta por dois mecanismos integrados, e que tira partido da sua conjugação. O primeiro método, desenvolvido pela Eng. Sandra Gama, constitui uma representação multi-camada das disciplinas do programa curricular, inter-relacionadas através de conectores visuais. O

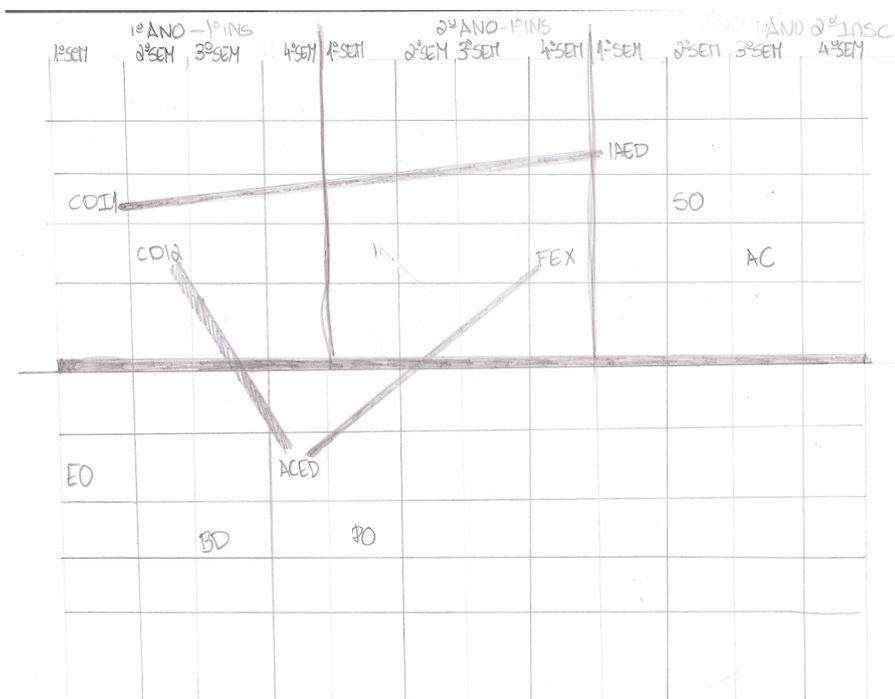


Figura 4.4: Pbf matriz

segundo método, objectivo desta tese e que constitui a solução ao problema enunciado, é uma representação multi-matricial em que são apresentadas as disciplinas, sendo os respectivos padrões representados através de mapas de cor.

As duas ferramentas possuem características distintas, e têm como finalidade complementarem-se na informação que retiram e transmitem dos padrões analisados.

4.3 VISUALIZAÇÃO MULTI-MATRICIAL

A visualização multi-matricial, é constituída por 6 matrizes em que cada matriz representa um semestre. O número de matrizes deve-se ao facto de os padrões disponibilizarem dados até ao sexto semestre, valor esse que é extensível, caso fossem usados dados com um maior intervalo temporal.

Cada matriz tem as disciplinas dos primeiros três anos da licenciatura pré-Bolonha em Engenharia Informática do Instituto Superior Técnico. As disciplinas, consideradas a unidade de avaliação, são representadas por quadrados. Cada quadrado é dividido diagonalmente em dois triângulos, de forma a dar informação sobre as aprovações de cada cadeira. Assim, o triângulo superior refere-se aos sucessos e o inferior aos chumbos de uma determinada disciplina.

4.4 ESTADO INICIAL

Antes de haver qualquer interacção por parte do utilizador com a visualização, esta encontra-se no seu estado inicial, onde é apresentado um *heatmap*. Neste caso, são

visíveis as cadeiras sobre as quais existem informação a nível de aprovações e chumbos, através da utilização de cores que variam entre o amarelo e o azul escuro. As disciplinas sem dados são representadas a cinzento, por ser uma cor associada à neutralidade e por transmitir a sensação de retrocesso, permitindo que os outros tons se evidenciem [12]. A luminância representa a quantidade de padrões em que a cadeira está envolvida: quanto menor o seu valor (cor mais "escura"), maior é o número de relações em que essa cadeira tem com outras disciplinas. Se a cadeira tiver poucas ou nenhuma dependências, então fica associada a um valor elevado(cor mais "clara"). Na Figura 4.5 é visível como a cor varia consoante o o número de padrões de cada cadeira ao longo dos semestres, a Figura ?? mostra em pormenor, os diferentes tons da variação da cor.

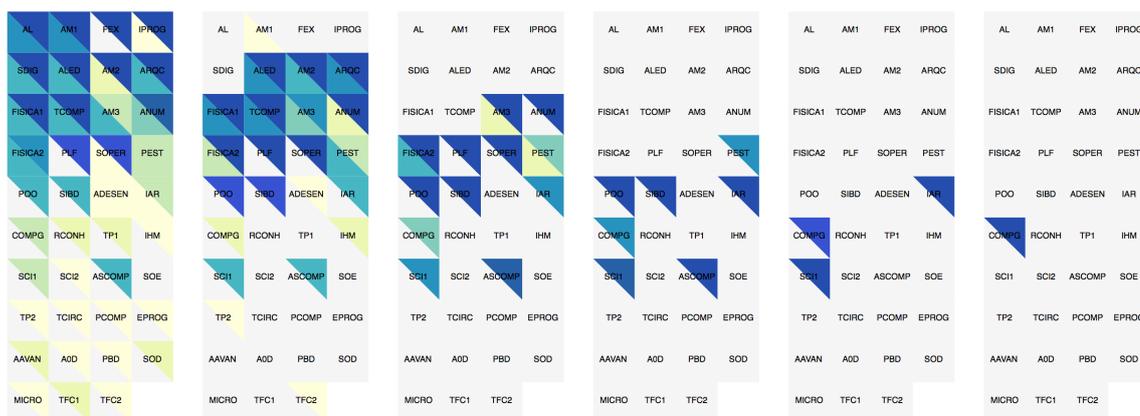


Figura 4.5: *Heatmap*, estado inicial

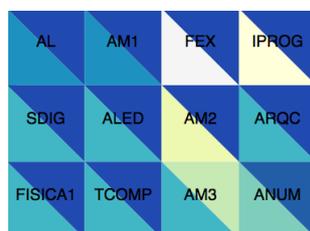


Figura 4.6: Variação da cor

4.5 INTERACÇÃO

O mecanismo multi-matricial permite alguns tipos de interacção, de forma a motivar o utilizador a explorar a visualização por si próprio e tornando mais efectiva a interpretação dos resultados [14]. Assim, é possível simplesmente movimentar o cursor sobre um dos triângulos de uma disciplina para ver as relações de cada uma ou então pode seleccioná-las e comparar resultados, e ainda efectuar alguns filtros para que seja possível comparar relações de cadeiras do mesmo padrão.

4.5.1 *Mouseover*

Este tipo de interacção destaca todas as dependências de uma determinada cadeira. Desta forma, para observar quais os padrões associados ao sucesso de uma disciplina, coloca-se o cursor sobre o triângulo superior da cadeira em questão, ficando visíveis todas as suas relações a nível de aprovações e chumbos. De forma análoga, para visualizar os padrões associados à sua reprovação, coloca-se o cursor sobre o seu triângulo inferior.

Para demonstrar estas dependências são usadas duas cores, azul e vermelho, que representam respectivamente, aprovações e chumbos de uma cadeira. A escolha de vermelho para as reprovações deveu-se à associação desta cor com perigo ou situações negativas, enquanto que a utilização de azul para as aprovações acontece por o azul estar associado a situações positivas. A luminância destas duas cores reflete a sua quantidade de relações/dependências. Assim o brilho diminui com o crescimento de relações e aumenta para a situação inversa. Não são utilizadas cores totalmente saturadas a fim de evitar que estas compitam pela atenção do utilizador.

Por exemplo, ao colocar o cursor no triângulo superior da disciplina *am1*, no primeiro semestre, ficam realçadas todas as cadeiras com que esta está relacionada em tons de azul e vermelho, a nível de aprovações e reprovações, Figura 4.7. Com qualquer tipo de interacção, o *heatmap* para as restantes cadeiras que não estão seleccionadas mantém-se, usando novamente a cor cinzenta, pelas mesmas razões acima descritas. O seu brilho varia consoante o número de dados associados, e é sempre mais escuro que as restantes cadeiras sem padrões para que não se confundam.

Quando o cursor sai do triângulo, se não houver nenhuma disciplina seleccionada, a visualização volta ao seu estado inicial, como mostra a Figura 4.5.

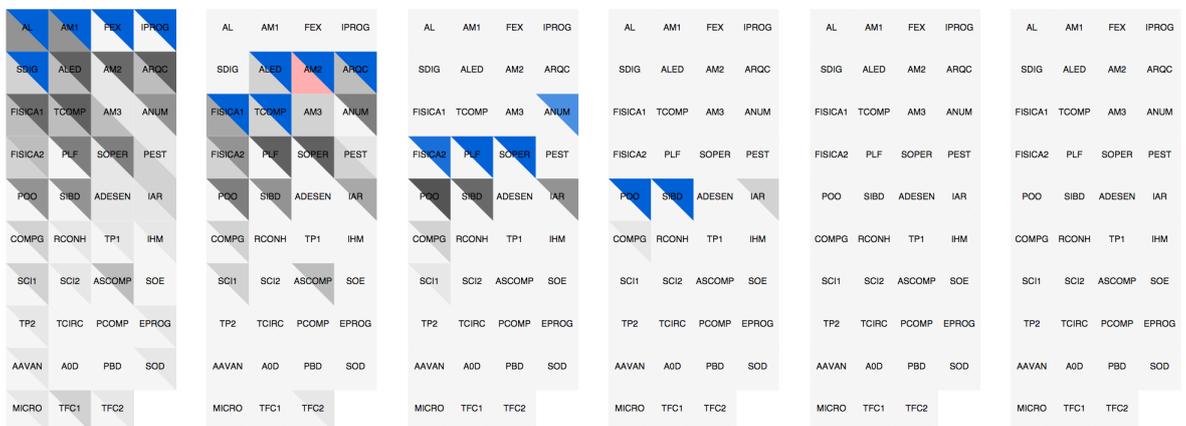


Figura 4.7: Relações da disciplina *am1*

4.5.2 *Seleção de Disciplinas*

Relativamente à selecção de disciplinas, é possível que o utilizador queira fazer uma comparação simultânea das relações de disciplinas distintas, ou de aprovações e reprovações da mesma disciplina, podendo observar de que modo passar ou reprovar a esta pode influenciar o percurso académico dos alunos. Para esse caso, é necessário clicar em um

dos triângulos de duas cadeiras diferentes ou da mesma cadeira. Os padrões em que a disciplina está envolvida ficam visíveis novamente, mas neste caso ficam fixos e não desaparecem com o deslocar do cursor, como Não é possível ter mais que duas selecções, por acabar por dificultar de forma significativa a percepção e interpretação dos dados por parte dos utilizadores, provocando sobrecarga visual. As cores usadas para estas selecções são o verde e o azul, por fazerem parte do espaço de cores mais natural para a mistura de cores, provocando melhores resultados perceptuais quando houver cadeiras que se sobrepõem [19] [20], ver Figura ??.

Os triângulos das cadeiras que foram seleccionadas pelo utilizador ficam destacadas por um contorno laranja, como é mostrado na Figura 4.9.



Figura 4.8: Dupla selecção



Figura 4.9: Contorno de uma disciplina seleccionada

Sempre que o utilizador selecciona uma cadeira, são desenhadas pequenas elipses, ver Figura ??, com o nome da disciplina e com a sua cor. Assim o utilizador, pode saber a ordem pela qual foi fazendo as suas selecções. Neste caso, primeiro foi seleccionada *sdig*, depois num padrão diferente, perceptível por serem cores distintas, *tcomp*, e uma restrição ao padrão de *tcomp*, por terem a mesma cor, *am2*. Também é possível remover as cadeiras seleccionadas, clicando na cruz que existe dentro de cada elipse.



Figura 4.10: Elipses com o nome das disciplinas seleccionadas

Dentro das relações de cada disciplina seleccionada, é possível aplicar algumas restrições. Ao passar o cursor sobre os padrões de uma cadeira previamente seleccionada, são visíveis

em tons de lima, quais as relações comuns entre a cadeira inicial e a cadeira onde se encontra o cursor actualmente. Escolheu-se esta cor por ser uma das cores intermédias entre o azul e o verde e por não provocar grande contraste, de modo a evitar adicionar mais ruído visual, mas que ao mesmo tempo se destaca destes dois tons, tornando perceptível o resultado da interacção do utilizador.

Ao clicar num desses triângulos, apenas ficam visíveis os padrões comuns às duas cadeiras seleccionadas. Fez-se assim uma restrição ao conjunto de relações iniciais da primeira disciplina que o utilizador escolheu seleccionar. Na Figura 4.11, é visível uma possível restrição dos padrões da cadeira *fex*. Na Figura 4.12, o utilizador restringiu os padrões positivos de *Fex* no primeiro semestre ao seleccionar os padrões negativos de *física1* no segundo semestre, ficando só visíveis as relações comuns às duas cadeiras. É possível restrições indefinidamente, até chegar a um ponto limite em que todas as cadeiras fazem parte do mesmo conjunto de relações.



Figura 4.11: Uma possível restrição dos padrões da disciplina *fex*

Por fim, pode dar-se o caso de duas cadeiras de padrões distintos ficarem sobrepostas. Tal sucede-se quando há cadeiras que são comuns a dois padrões diferentes. Nesse caso, é feito o *blend* com as cores usadas inicialmente, azul e verde, para preencher o quadrado das disciplinas. A Figura 4.13 mostra uma visão global dos padrões seleccionados e o seu *blend*

A Figura 4.14 mostra em detalhe as cadeiras cuja cor foi alterada para o *blend* de cores.

4.6 IMPLEMENTAÇÃO

Nesta secção serão descritos quais os algoritmos mais importantes usados no desenvolvimento da visualização multi-matricial, bem como as principais estruturas de dados utilizadas. É explicada a sua evolução ao longo do tempo, e quais as decisões que foram tomadas. São também descritos os principais desafios, e questões que foram ocorrendo durante o desenvolvimento, as hipóteses que foram consideradas e a forma como foram ultrapassados.

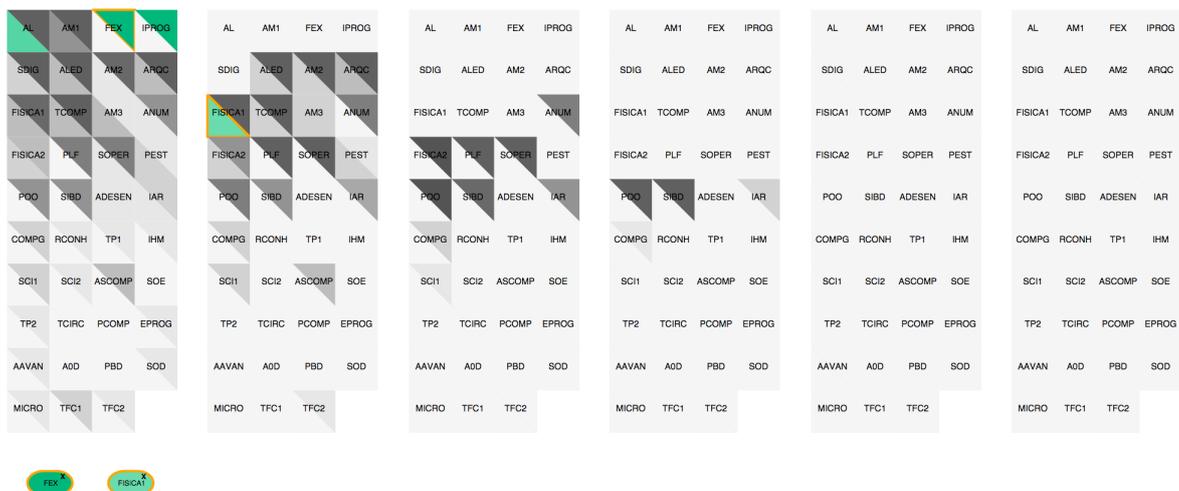


Figura 4.12: Restrição dos padrões comuns às cadeiras de *fex* e *fisica1*

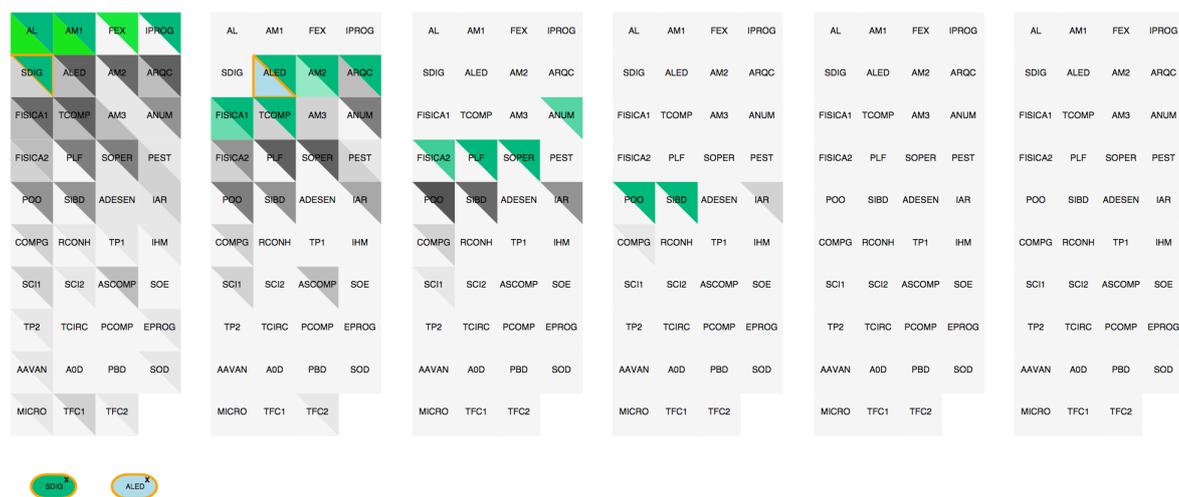


Figura 4.13: Visão de células sobrepostas nos padrões das disciplinas de *sdig* e *aled*

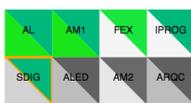


Figura 4.14: *Blend* das disciplinas *al*, *am1* e *fex*

4.6.1 Processamento de padrões Educacionais

Após estar definida a estrutura da visualização, e mesmo a sua implementação não estando totalmente concluída, foi necessário ler e processar os padrões fornecidos, de forma a fazer o seu mapeamento visual, para que se pudesse observar o comportamento da visualização, nos seus estágios iniciais.

Esta necessidade surgiu devido ao facto de os ficheiros disponibilizados serem muito extensos, nomeadamente os que contêm os padrões com a utilização de limiar de suporte de 20% e 25%. Nestes dois casos, a quantidade de padrões gerados é naturalmente muito significativa. Na Figura 4.15 é visível o número total de de padrões de cada suporte.

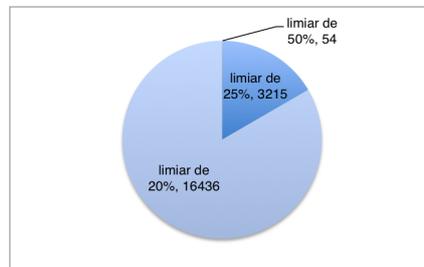


Figura 4.15: Percentagem de padrões em cada suporte

Por esta razão, desde o início do desenvolvimento, foi fundamental ter sempre presente a necessidade de fazer um tratamento eficiente dos dados, para que a visualização respondesse de uma forma fluida, sem grande atraso, à interacção por parte do utilizador. Inicialmente, a leitura dos padrões era feita no momento em que o utilizador realizava algum tipo de acção. A utilização desta abordagem funcionou apenas durante o início do desenvolvimento da visualização, quando a sua estrutura ainda estava a ser construída, e havia apenas uma funcionalidade, mostrar as relações entre disciplinas.

O seu uso revelou-se um problema, pois ao fazer o processamento dos padrões, sem ter uma estrutura de dados previamente montada, levou a que o atraso entre qualquer tipo de interacção e o seu resultado fosse notório, interferindo na habilidade do utilizador conseguir completar as suas tarefas de forma eficiente, tornando o sistema oneroso e impossibilitando a realização de objectivos, criando um problema de usabilidade.

A solução tomada para contornar esta situação, passou por realizar um pré-processamento dos dados. Estes são carregados para o cliente web no seu estado inicial, e quando a visualização é carregada pela primeira vez, antes de qualquer interacção, os padrões são lidos, iterados e organizados numa estrutura de dados, segue-se a sua descrição:

Matriz $i = [matriz1 [Célula1, \dots, célula39], \dots, matriz6 [Célula1, \dots, célula39]]$, *Célula* $l =$
Objecto

Objecto : N^o padrões,

Disciplina,

Padrão1, \dots, Padrão n

A estrutura escolhida é composta por *arrays* de *arrays*. O primeiro conjunto de vectores, refere-se às matrizes existentes, existindo 6 vectores, um por cada matriz. O segundo conjunto, guarda 39 vectores, um por cada célula, que por sua vez se refere a uma disciplina.

Por fim cada célula é um objecto composto pelo o nome da disciplina em questão, número de padrões em que a disciplina entra, e por um vector que contém os seus padrões. Optou-se por guardar estes campos, por conterem a informação sobre o que se pretende representar e para facilitar o seu acesso pelas várias funcionalidades da visualização.

Esta alteração melhorou substancialmente os tempos de resposta do sistema, mas no padrão de menor suporte os resultados continuavam a não ser satisfatórios. Tal deve-se

ao facto de uma disciplina ao ter vários padrões associados, tem na maioria das vezes muitos valores repetidos:

$$fex = [fex,am1], [fex,am1,iprogr]$$

Neste exemplo, é visível que a disciplina *am1* está presente mais do que uma vez, gerando iterações desnecessárias e atraso nos tempos de resposta, pois basta que apareça uma vez, para informar que pertence ao padrão, neste caso, de *fex*.

Por esta razão foi necessário estruturar os dados de outra forma. A estrutura anterior manteve-se, e foi criada outra somente para a leitura de dados. Neste caso, cada disciplina passou a ter um array associado, que contém todas as disciplinas com que esta se relaciona:

$$Cadeira\ I = Cadeira\ 1 \times Cadeira\ 2 \times Cadeira\ 3$$

Deixou de se ter em conta os padrões de que a disciplina faz parte, para se passar a considerar apenas as cadeiras que fazem parte do seu universo de padrões. Exemplo:

$$fex = [fex,am1,iprogr]$$

Limitou-se desta forma o número de elementos do array de uma disciplina a um máximo de 38, que é o número total de cadeiras existentes, resultando numa diferença muito elevada ao tempo de resposta da visualização por parte de algum tipo de interacção.

Considerando a disciplina de *iprogr*, no suporte de 20%, esta tem associados 7652 padrões. Com a implementação da primeira estrutura de dados, o tempo de resposta ao deslocar o cursor sobre a sua célula e ver um resultado é de 47,68 segundos, enquanto que na nova estrutura o tempo é de 0,22 segundos, ver Figura ??.

	1º Estrutura	2º Estrutura
Mouseover	47,69 s	0,11 s
Seleccção	49 s	0,13 s

Tabela 2: Tempos de resposta da Visualização ao deslocamento do cursor e à selecção de uma célula

4.6.2 Evolução da Visualização

Na fase inicial do desenvolvimento, existiam duas matrizes. Cada matriz tinha presente todos os semestres, mas a primeira era referente aos sucessos dos alunos, e a segunda aos seus chumbos. As relações entre as várias disciplinas, seriam feitas através de conectores visuais, linhas que uniam as várias cadeiras, representando as suas dependências. Esta solução acabou por ser descartada, por ser mais complexa. Quando o utilizador está a analisar uma visualização, tem uma parte limitada do seu raciocínio para dedicar à avaliação do problema. Uma parte da sua atenção é dedicada à descodificação do sistema enquanto que a outra parte é dedicada ao entendimento do mesmo [14]. Ao ter duas matrizes que juntavam todos os semestres em si, e que iriam utilizar linhas que se transpunham de uma matriz para outra, a sobrecarga visual seria maior, o utilizador não perceberia de imediato qual o número de semestres existentes, nem quais as cadeiras que

possuem padrões relativamente a aprovações e chumbos ao mesmo tempo por exemplo. Perderia assim, mais tempo a interpretar a visualização para alcançar a mensagem que esta pretende transmitir.

Dadas as desvantagens acima mencionadas, passou a ser preferível representar as aprovações e chumbos na mesma matriz, e para que o utilizador conseguisse perceber logo numa primeira análise qual a dimensão temporal que está a ser usada, definiu-se que uma matriz representaria um semestre de inscrição.

A utilização de linhas como conectores, para demonstrar as relações entre as várias cadeiras, também foi posta de parte. Dado o elevado número de dependências que algumas disciplinas têm, principalmente nos suportes de 20% e 25%, acabaria por se tornar imperceptível a quantidade de relações em que cada cadeira está envolvida, bem como, quais as outras cadeiras com que se relaciona, devido à sobreposição de um grande número de linhas. Em alternativa, passaram a ser utilizados mapas de calor com diferentes esquemas de cores. Esta solução foi adoptada por estes serem representações gráficas de dados que permitem expressar valores e relacioná-los em grandes conjuntos de dados [14], sem dificultar a legibilidade dos mesmos.

Regressando à representação de sucessos e chumbos na mesma matriz, a forma de o fazer passou por dividir ao meio verticalmente, cada quadrado referente a uma disciplina. Assim, a parte esquerda representaria as aprovações e a parte direita as reprovações, Figura 4.16.

	AL	AM1	FEX	I PROG
SDIG	ALED	AM2	ARQC	
FISICA1	TCOMP	AM3	ANUM	

Figura 4.16: Divisão vertical de células

Esta opção levava à existência de linhas verticais a meio dos quadrados, acentuando a ideia de haver uma divisória, mas sem continuidade, que é o oposto do que se pretende, visto ser a mesma cadeira. Por essa razão a divisão vertical foi rejeitada, e passou a ser feita uma divisão diagonal, em que já não existem linhas visíveis que atravessam o quadrado, e é transmitida à mesma uma noção de divisão, mas contínua, ver Figura 4.17.

Relativamente ao *mouseover* e a selecção de padrões, estas eram as duas funcionalidades mais prejudicadas em termos de tempo de resposta e para o bom funcionamento da visualização, era essencial contornar este problema, ver secção Processamento de padrões Educacionais. O aspecto visual de ambas as funcionalidades, manteve-se ao longo da implementação, mas a forma como os dados são acedidos foi mudada.

Para ser mais fácil de identificar a cadeira em questão, cada célula possui uma identificação única, então para o *mouseover*, é feita uma procura pelo identificador onde o utilizador colocou o cursor, no *array* que guarda a informação dos padrões de cada dis-

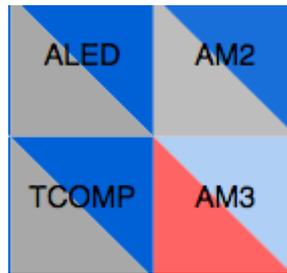


Figura 4.17: Divisão diagonal de células

ciplina. Assim obtêm-se todas as cadeiras que fazem parte do padrão da disciplina em questão, pintado metade dos seus quadrados de azul ou vermelho, consoante seja uma aprovação ou reprovação.

Quando uma disciplina é seleccionada, é guardada num vector, juntamente com os padrões com que se relaciona. Ao ser desseleccionada, estes dados são removidos do vector.

Se o utilizador quiser fazer uma restrição a um padrão que esteja seleccionado, é feita uma procura pelas cadeiras comuns existentes nos padrões das cadeiras seleccionadas. Por exemplo, se o utilizador clicar em *fex* na primeira matriz e em *tcomp* na segunda matriz, é feita uma busca pelos padrões de *fex* que contêm obrigatoriamente estas duas cadeiras. A ordem pela qual as restrições são feitas e são removidas é irrelevante. Caso se faça uma restrição pela ordem *fex*, *tcomp*, *iprog* e *sdig*, é indiferente se primeiro é removida a disciplina de *iprog* ou *iprog*. Os quadrados pintados vão ser sempre o resultado dos padrões comuns às cadeiras que ficaram seleccionadas.

Sobre as elipses, estas vão buscar a informação sobre qual foi a cadeira escolhida, através do vector que guarda as cadeiras que foram seleccionadas. Também permite desseleccionar disciplinas, removendo-as do vector em questão.

As cores usadas tanto no quadrados como nas elipses, são obtidas através do número de padrões em que cada disciplina entra. Essa informação encontra-se disponível numa das duas principais estruturas de dados definidas, que estão acima descritas. Dessa forma é calculada a variação da cor, passando o número de padrões como argumento, através de uma função fornecida pela biblioteca D3.js¹, que é usada para realizar esta visualização. Existe um caso particular da utilização de cores, que acontece quando existe sobreposição de padrões. Nesse caso os quadrados são coloridos de acordo com o valor obtido a partir da mistura das cores das duas cadeiras seleccionadas.

4.7 INTEGRAÇÃO NO EDUCARE

A integração das duas visualizações foi feita de forma a que funcionassem de forma fluida e coordenada, passando a constituir um só sistema. Sempre que o utilizador efectua uma acção num dos mecanismos, esta reflecte-se automaticamente no outro, sendo que as duas visualizações se complementam, transmitindo de forma completa toda a informação que é possível extrair do conjunto de dados usados.

Realizar a sua integração foi uma tarefa complexa, pois estes modelos inicialmente foram desenvolvidas de forma independente, pelo que cada um tem as suas funções e estruturas.

¹ <http://d3js.org>

Desse modo, foi necessário compreender e analisar o código que produz a visualização multi-camada, para depreender como funcionam as funções principais, e saber que estruturas de dados invocar.

Para efectuar as mudanças necessárias e conseguir realizar a integração, do lado da visualização multi-matricial trabalhou-se com os vectores que guardam as células seleccionadas. Assim sempre que se realiza uma acção num círculo, o seu valor, que corresponde a uma disciplina, é passado como argumento às funções do mecanismo multi-matricial, que vão tratar de adicionar ou remover padrões dos vectores que guardam esse tipo de informação. O mesmo método foi usado para a situação inversa. Sempre que se interage com uma matriz, o resultado dessa interacção, que será o nome da disciplina que o utilizador escolheu, é passado como argumento a funções da visualização multi-camada, que tratam de desenhar os conectores visuais para a cadeira em questão.

Assim, sempre que o utilizador desloca o cursor sobre um círculo de uma disciplina e se acendem todas as suas ligações, também nas matrizes se iluminam todos os quadrados dessa cadeira, e vice-versa.

O mesmo acontece com a selecção de disciplinas, se uma disciplina estiver seleccionada no primeiro mecanismo, também vai o vai ficar no segundo mecanismo, ou a situação inversa, caso se tenha seleccionado uma célula, também o círculo correspondente fica seleccionado.

Para reflectir a restrição de padrões efectuada nas matrizes, na visualização multi-camada, foi necessário realizar algumas alterações mais profundas ao seu código. Com estas mudanças, adicionou-se uma nova funcionalidade ao mecanismo multi-camada, que passou a permitir a restrição de padrões a partir dos seus círculos.

Foi necessário criar algumas funções extras, que fizessem a conversão dos argumentos que as funções das duas visualizações recebem para os formatos que estas recebe,

Existem algumas situações que não têm repercussões dadas as características de cada visualização, como por exemplo, se houver duas selecções no mecanismo multi-matricial, só o círculo correspondente à primeira cadeira seleccionada vai estar assinalado.

Este foi essencialmente o principal desafio desta integração e pelas razões acima descritas foi um processo moroso e trabalhoso. Na Figura 5.1 é visível a integração dos dois mecanismos, que corresponde à fase final do EduVis. Este caso ilustra a situação em que o utilizador seleccionou a disciplina de *fex* no mecanismo multi-camada e essa acção repercutiu-se na visualização multi-matricial, ficando também o triângulo correspondente seleccionado.

Segue-se uma tabela, que resume os dados que são possíveis extrair de cada visualização.

	Taxas de Aprovação	Comparação visual de sucessos e insucessos	Relações	Cadeiras Base	Quantidade de padrões envolvidos	Comparação de padrões	Filtrar padrões	Blend de cores
Multi-camada	Sim	Sim	Sim	Não	Não	Não	*	Não
Multi-matricial	Não	Não	Sim	Sim	Sim	Sim	Sim	Sim

Tabela 3: Sumarização das características da visualização multi-camada e multi-matricia

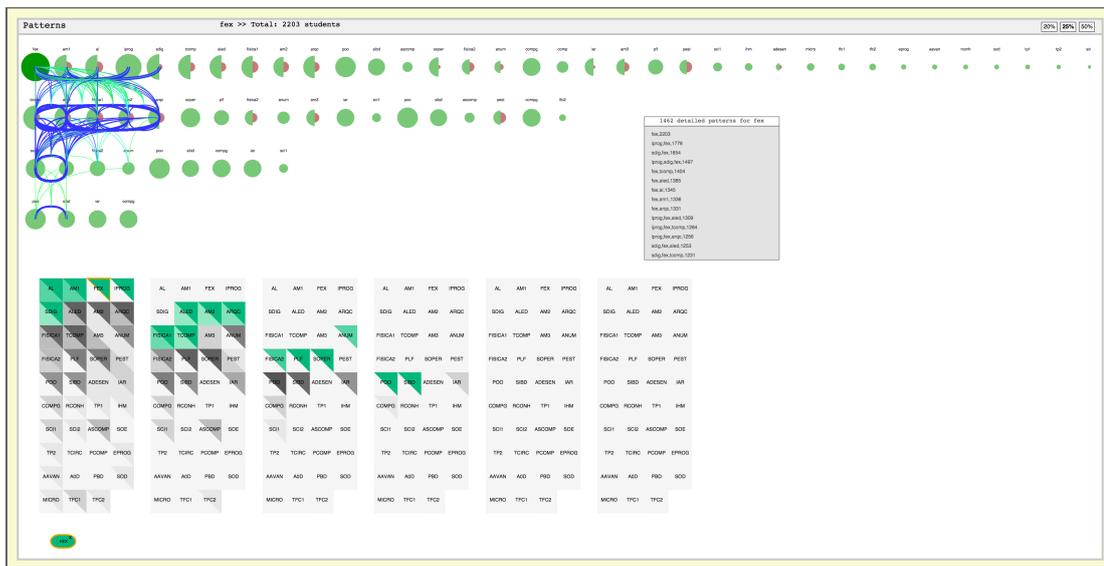


Figura 4.18: EduVis:Versão final, caso demonstrativo da selecção de uma disciplina e a sua repercussão nas duas visualizações

Analisando a primeira coluna da Tabela 3, esta refere-se às taxas de aprovação de cada disciplina. Estes dados só são visíveis na primeira visualização, através do raio dos círculos, que é proporcional ao número de alunos aprovados ou reprovados. Consequentemente também só nesta visualização é possível fazer a comparação entre sucessos e insucessos, coluna 2, permitindo uma observação imediata desses resultados, sem necessidade de exploração adicional.

Relativamente às relações de cada disciplina, estas são visíveis nos dois mecanismos, embora com abordagens diferentes. A primeira visualização utiliza linhas como conectores visuais para demonstrar as dependências entre as várias cadeiras, sendo que, em disciplinas que tenham muitos padrões por vezes é difícil percepcionar quais as cadeiras envolvidas, principalmente se os seus círculos tiverem um raio pequeno, podendo ficar quase ocultados pelos conectores.

O mecanismo multi-matricial utiliza um *heatmap* para fazer essa representação, não havendo sobreposição de artefactos visuais. Dessa forma complementa esta característica, por ajudar a clarificar os relacionamentos existentes, para que não haja perda de informação.

Inferir quais são as cadeiras com mais relações/dependências, ou seja, em que é mais importante ter aprovação para ter sucesso a disciplinas futuras é possível na segunda visualização. É também possível observar o caminho percorrido pelos estudantes até terem aprovação a determinada cadeira. Esta informação é transmitida pelo *heatmap* inicial, ou então se o utilizador decidir explorar as relações de alguma disciplina. Observar a quantidade de padrões envolvidos em cada disciplina, também um dos objectivos da visualização multi-matricial, é possível através do *heatmap*, e da leitura do número de padrões na caixa informativa, no lado direito das visualizações. É também possível compará-los, por essa visualização permitir a selecção de duas cadeiras ao mesmo tempo.

Relativamente a restringir padrões, o segundo mecanismo permite filtrá-los, a fim de poder comparar cadeiras dentro do mesmo padrão. Após a integração, no âmbito desta

tese, foi adicionada essa funcionalidade ao mecanismo multi-camada, de modo a tornar a visualização final mais coerente.

Por último, a visualização multi-matricial, permite ver uma mistura das duas cores usadas na selecção de padrões, o verde e o azul, sempre que existem cadeiras comuns em dois padrões distintos e que estejam nesse momento seleccionados. Assim, está descrito o que cada visualização permite efectivamente observar, e o que é comum a ambas.

AValiação

Este capítulo apresenta e discute as avaliações efectuadas na utilização do EduVis. De forma a validar a solução proposta, foram realizados testes com utilizadores, observando o seu comportamento, e a forma como interagem com a visualização. Na secção 4.1 está descrito qual o protocolo utilizado, bem como o objectivo que se pretende atingir com a realização destes testes. Na secção 4.2 é feita a discussão dos resultados obtidos e na secção 4.3 é feita uma análise e posterior discussão dos mesmos.

5.1 DESCRIÇÃO DO PROTOCOLO

Após o desenvolvimento da solução proposta, pretende-se avaliar o sistema, em várias vertentes, de modo a concluir se este cumpre os objectivos propostos na secção 1.1. Desta forma, avalia-se a eficiência da sua utilização, verificando se o mecanismo desenvolvido produz os resultados esperados, a sua usabilidade, como a facilidade de aprendizagem por parte dos inquiridos, a rapidez no desenvolvimento das tarefas propostas e a quantidade de erros cometidos, e por fim o grau de satisfação demonstrado.

Para avaliar os participantes, foi elaborado um questionário com 7 tarefas representativas das principais funcionalidades da visualização, a 20 utilizadores, e que vão ao encontro dos requisitos definidos. Os testes foram realizados individualmente, sem a presença de outros inquiridos, num ambiente controlado, com boas condições de luz e têm a duração aproximada de 15 min.

Antes de iniciar o questionário, foi dada uma explicação sobre os objectivos do sistema, explicando o significado dos vários artefactos visuais. De seguida foi feita uma pequena demonstração das funcionalidades da visualização, e depois foi dada liberdade a cada utilizador para que a explorasse, até se sentir confortável para realizar as tarefas propostas.

Findo este processo inicial, foi entregue aos participantes um questionário com um conjunto de tarefas, que têm ser respondidas mediante a realização de acções na visualização.

Como método de avaliação, foi feita uma observação directa das acções dos inquiridos, solicitando que pensassem em voz alta, para que fosse possível tomar notas dos raciocínios feitos. Durante a fase de testes, foi medido o tempo que duração de cada tarefa e foram contabilizados os erros cometidos. Por fim, foi pedido aos participantes que respondessem a um inquérito online de satisfação, dividido em duas partes: a primeira parte corresponde ao System Usability Scale (SUS)¹ e a segunda parte diz respeito a um pequeno conjunto de perguntas que têm como objectivo avaliar também, usando uma escala de Likert de 5 pontos, o grau de dificuldade sentido pelos participantes ao realizar cada tarefa.

¹ <http://www.measuringusability.com/sus.php>

Esta última parte do questionário pretende revelar o grau de compreensão dos seguintes aspectos:

- i Número de semestres representados. É importante para saber se o utilizador tem noção da existência de um espaço temporal, e se se consegue situar no mesmo.
- ii Quais as cadeiras com maior número de aprovações e reprovações.
- iii Quais as cadeiras com maior número de precedências. Estes dois últimos pontos permitem saber se é facilmente perceptível a identificação das cadeiras chave para o sucesso académico dos estudantes.
- iv Quais as cadeiras envolvidas em mais padrões.
- v Comparação de padrões de disciplinas distintas.

5.2 TAREFAS

Relativamente às tarefas seleccionadas para avaliar a visualização, são as seguintes:

- i Quantos semestres estão representados?
- ii A nível geral, quais as duas cadeiras com mais alunos?
- iii Qual o conjunto de cadeiras envolvidas em mais padrões positivos no 2º semestre?
- iv Quais as cadeiras que estão relacionadas com a cadeira de iar no 4º semestre?
- v Considerando os alunos que fizeram am1 e fex no 1º semestre, quais são as outras cadeiras em que também obtiveram sucesso no 2º semestre?
- vi Quais são as cadeiras comuns a quem fez sdig no 1º semestre e obteve reprovação a alled no 2º semestre?
- vii Considerando a cadeira compg no 4º semestre, embora seja uma das disciplinas com menos padrões associados, é esta a cadeira com menos alunos nesse semestre?

As perguntas de 1, 2 e 3 são respondidas utilizando o suporte maior, de 50%, por ser o suporte com menor quantidade de padrões e por serem perguntas mais genéricas e simples acabando por funcionar como uma introdução às restantes tarefas, de maior complexidade. São perguntas cuja resposta deve ser inferida apenas pela análise visual da visualização, permitindo uma resposta imediata, sem necessidade de exploração. A questão 1 pode ser respondida contando o número de matrizes com cores, representativas do número de padrões, ou então contando o número de camadas de círculos existentes. A questão 2 pode ser respondida observando quais os círculos de maior tamanho e por fim a solução da questão 3 é obtida através da análise da segunda matriz, percepcionando quais são as disciplinas cujos o triângulo superior têm maior brilho.

As questões seguintes requerem que o utilizador interaja com a visualização, explorando-a. Para responder à questão 4, é necessário deslocar o cursor para descobrir as relações da cadeira de iar. As questões 5 e 6 requerem a aplicação de filtros, mas na questão 6, por haver sobreposição de cadeiras, existe uma mistura de cores que o utilizador deveria conseguir distinguir. A visualização permite que a questão 5 possa ser respondida de duas formas. O utilizador pode deslocar o rato sobre as cadeiras que estão seleccionadas e consegue observar de imediato os subconjuntos de cada cadeira por estes estarem destacados com uma cor diferente da selecção, ou então pode efectivamente clicar no triângulo da cadeira e apenas ficam visíveis as disciplinas que fazem parte do seu subconjunto.

Por fim, a última questão obriga a que o inquirido conjugue a informação de ambos os mecanismos. O seu objectivo é mostrar a importância da interacção de ambos os mecanismos, em que o utilizador pode retirar informação distinta porém complementar. Assim a tarefa demonstra que uma cadeira apesar de estar envolvida em poucos padrões, pode ter mais alunos inscritos que outras cadeiras com maior número de relações.

5.3 DESCRIÇÃO DOS RESULTADOS

Dos 20 utilizadores inquiridos, mais de metade são do género masculino, a média de idades situa-se no intervalo dos 25-34 anos, e a maioria, 80%, frequenta ou concluiu o ensino superior, ver Tabelas 4,5 e 6 respectivamente.

Géneros	Números	Percentagem
Feminino	5	15%
Masculino	15	75%

Tabela 4: Género dos inquiridos

Idades	Números	Percentagem
18-24	3	15%
25-34	10	50%
35-44	3	15%
45-54	2	10%
55-64	1	5%
+65	1	5%

Tabela 5: Idade dos inquiridos

Grau de escolaridade	Números	Percentagem
Ensino Básico	2	10%
Ensino Secundário	2	10%
Ensino Superior	16	80%

Tabela 6: Grau de escolaridade

Na Tabela 7 estão sumarizados o tempo médio de conclusão de cada tarefa e o número médio de erros cometidos.

Tarefa	Tempo(segundos)	Desvio Padrão Tempo	Número de erros	Desvio Padrão Erros
1	3.20	0,477	0.35	1,249
2	17.60	0,3	0.10	12,560
3	28.30	0,590	0.55	15,802
4	23.95	0,458	0.30	13,109
5	27.70	0,458	0.30	9,8138
6	49.45	0,622	0.75	28,251
7	45.70	0,714	0.30	25,277

Tabela 7: Tempo e número de erros médio

Nas Figura 5.1 e Figura 5.2 estão visualmente representadas a distribuição dos tempos de resposta dos utilizadores e da média dos erros cometidos respectivamente, às perguntas efectuadas. Na Figura 5.1 é possível ver que a maioria dos tempos de resposta se encontram no intervalo entre 20 e 40 segundos, enquanto que na Figura 5.2 a média dos erros cometidos situa-se entre os valores 0,3 e 0,45.

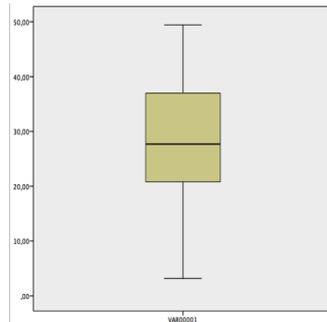


Figura 5.1: Distribuição da média dos tempos de resposta

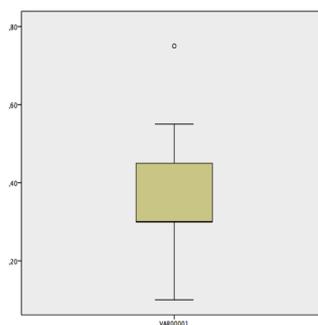


Figura 5.2: Distribuição da média dos erros cometidos

Sobre os tempos e número de erros médio, relativos ao desempenho de cada uma das tarefas, embora as tarefas 1, 2 e 3 sejam imediatas, sem necessidade de exploração adicional, os valores medidos não permitem criar uma generalização acerca do tempo e número de erros para desempenho destas tarefas face às tarefas de exploração, 4 a 7.

Para um maior aprofundamento de resultados, foi realizada uma análise estatística, aplicando um teste *Shapiro-Wilk*. Este mostrou evidência contra uma distribuição normal na maior parte dos casos ($p < 0.05$), pelo que foi aplicado um teste não paramétrico, o teste Wilcoxon signed-rank para descobrir as diferenças significativas entre amostras. Efectivamente, a nível de tempo a tarefa 1 é significativamente mais rápida que as restantes, ($z_{1-2} = -3.72, z_{1-3} = -3.92, z_{1-4} = -3.92, z_{1-5} = -3.92, z_{1-6} = -3.92, z_{1-7} = -3.92, p < 0.05$), enquanto que a tarefa 2 é significativamente mais rápida que as tarefas 3,5,6 e 7, ($z_{2-3} = -2.39, z_{2-5} = -2.54, z_{2-6} = -3.32, z_{2-7} = -3.25, p < 0.05$).

Por outro lado, as tarefas 3,4 e 5 são significativamente mais rápidas que as tarefas 6 e 7 ($z_{3-6} = -2.31, z_{3-7} = -2.17, z_{4-6} = -3.45, z_{4-7} = -2.76, z_{5-6} = -3.36, z_{5-7} = -2.63, p < 0.05$). No entanto, no que diz respeito ao número de erros, existem diferenças significativas apenas entre a tarefa 6 e as que têm menos erros, a tarefa 2 ($W = 6, cv = 17, p < 0.05$) e a tarefa 4 ($W = 0, cv = 3, p < 0.05$), e entre as tarefas 2 e 3 ($W = 5, cv = 8, p < 0.05$).

Numa tentativa de encontrar uma correlação entre o tempo e o número de erros, calculou-se os coeficientes de *Pearson*, encontrando uma correlação, embora relativamente fraca, na tarefa 2 ($r = 0.45, p < 0.05$) e na tarefa 7 ($r = 0.49, p < 0.05$), o que não permite generalizar uma correlação entre um maior tempo para desempenho da tarefa e um maior número de erros. É possível concluir assim, que a complexidade temporal de uma tarefa no EduVis, não leva a um maior número de erros no desempenho da mesma.

Tendo em conta o questionário de satisfação, a pontuação relativa ao SUS segundo os correspondentes parâmetros de cálculo [24], foi de 79.47 pontos, mostrando resultados bastante elevados no que diz respeito à usabilidade e capacidade de aprendizagem do sistema. Usando o mesmo método para calcular as respostas de satisfação contextuais, obteve-se um resultado de 92.11 pontos, mostrando que os objectivos propostos foram alcançados no que diz respeito à satisfação no desempenho das tarefas.

5.4 DISCUSSÃO DE RESULTADOS

O estudo dos testes realizados, juntamente com observação directa e a utilização de medidas de desempenho, permitiram descobrir alguns problemas com grau de severidade baixo e tecer algumas considerações acerca do EduVis, visando futuras melhorias. As tarefas realizadas dividem-se em dois tipos, (1) as que permitem uma análise imediata e (2) as tarefas em que é necessário explorar a visualização, interagindo com ela. Assim a secção 5.4.1 é referente ao primeiro tipo, que contempla as três primeiras tarefas, e a secção 5.4.2 é referente ao segundo tipo, que contém as restantes questões.

5.4.1 *Análise Imediata*

Aquando da realização das várias tarefas por parte dos utilizadores, foram observados os diferentes passos que os inquiridos executaram, bem como os raciocínios que realizaram, permitindo assim perceber a razão dos erros cometidos.

Relativamente à primeira questão, foi verificado que 6 participantes (30%) cometeram o mesmo tipo de erro, enquanto que 70% não cometeu erro nenhum. Nas perguntas 1,2 e 3 é utilizado o suporte de 50%, em que apenas existem dados nos dois primeiros semestres, sendo portanto essa a resposta certa. Os participantes que não responderam correctamente a esta questão, não conseguiram identificar o número real de semestres representados, demonstrando alguma confusão por observarem na visualização multi-camada duas linhas de círculos, em que cada linha representa um semestre e na visualização multi-matricial 6 matrizes. Neste caso, apesar de apenas as duas primeiras matrizes apresentarem alguma informação sobre os padrões usados, optou-se por continuar a representar as 6 matrizes de forma a dar conhecimento do número total de semestres existentes, apesar de apenas no suporte de 20% existirem dados até ao sexto semestre. A maioria dos restantes inquiridos responderam de forma quase imediata a esta questão, tornado-a a nível de tempo, como demonstra o teste *Wilcoxon signed-ran* na secção 4.2, a tarefa de mais rápida execução. Tal deve-se à sua baixa complexidade e de corresponder a um aspecto geral, imediatamente perceptível na visualização.

Na avaliação da questão 2, houve dois erros diferentes por parte de dois utilizadores. Era pedido que fosse indicado quais as duas cadeiras com mais alunos, bastando para isso verificar quais os círculos de maior tamanho, a resposta correcta seria *fex* e *ipro*. Um dos participantes, compreendeu que a primeira disciplina com mais alunos era efectivamente

fex, devido ao tamanho do círculo desta cadeira, facilmente se destacar em relação aos restantes, mas ao mencionar a segunda disciplina, cometeu um erro e ao invés de indicar *ipro*, indicou *sdig*. Tal deve-se ao facto de os dois círculos terem quase o mesmo tamanho, por o número total de alunos ser muito semelhante, e por o utilizador não ter colocado o cursor por cima das disciplinas em questão, verificando qual o número de alunos que lhe corresponde. No segundo erro, o inquirido ao observar os círculos, inicialmente pensou que estes estariam ordenados pelo número de alunos, devido ao primeiro círculo ser o maior e de haver uma diminuição de tamanho dos mesmos até chegar ao último círculo, transmitindo uma falsa ideia de sequencialidade, levando a que desse como resposta *fex* e *am1*. Após uma análise mais cuidadosa compreendeu que tinha cometido um engano, por verificar a existência de alguns círculos maiores e menores fora de ordem, alterando a sua resposta para *fex* e *ipro*. Esta tarefa foi a segunda mais rápida a ser realizada, por também ser de percepção quase imediata. Apenas o tamanho similar dos círculos com mais disciplinas, fez com que os utilizadores demorassem mais tempo a chegar a uma conclusão. Passando para a tarefa seguinte, 50% dos utilizadores cometeu erros até chegar à solução correcta. Nesta questão, a resposta certa seria *aled*, *arq* e *tcomp*, como as cadeiras envolvidas em mais padrões no segundo semestre. Para inferir tal conclusão era necessário observar o mecanismo multi-matricial e ver quais seriam as disciplinas com cores mais brilhantes.

Foram cometidos 3 tipos diferentes de erros por parte dos inquiridos, que consistem em (1) não associar o brilho das cores ao número de padrões de cada disciplina, (2) não compreender que o número de padrões é visível logo no estado inicial de cada matriz e (3) não ser capaz de distinguir entre os diferentes níveis de brilho. O primeiro erro foi cometido por 3 utilizadores, que não relacionaram de todo a variância do brilho com o número de padrões, mostrando-se indecisos nas suas escolhas e um pouco confusos acabando por dar respostas erradas ou incompletas, havendo mesmo participantes que basearam as suas respostas no tamanho dos círculos.

O segundo erro foi comum a 4 utilizadores, e deveu-se a estes não terem compreendido de imediato que bastava olhar apenas para o brilho das cores, e por essa razão terem perdido algum tempo a deslizar o cursor sobre várias disciplinas, ou clicando nos seus triângulos, mas acabaram eventualmente por chegar à solução correcta.

A última falha foi cometida por um inquirido, que revelou ser daltónico. Como tal teve dificuldade em distinguir a variação de cores usada no mapa de calor inicial, acabando por não conseguir dar a resposta correcta.

5.4.2 Exploração da Visualização

Relativamente às restantes tarefas, a tarefa 4 implicava a mudança de suporte para 25%, por só a partir desse valor existirem padrões referentes a reprovações. Para o utilizador conseguir responder correctamente a esta questão, tinha de deslocar o cursor até à disciplina de *iar* na quarta matriz, e observar quais os triângulos que ficavam azuis e/ou vermelhos. A resposta certa seria *aled*, *tcomp*, *plf*, *soper* e *poo*. Houve 5 participantes que não deram a resposta completa e um participante que deu uma resposta errada. Os 5 inquiridos que não mencionaram todas as cadeiras, tiveram dificuldade em distinguir correctamente o azul mais escuro que preenche os triângulos, do cinzento mais escuro dos triângulos circundantes, permitindo concluir que não repararam que estes possuíam a cor azul. O sexto participante que não conseguiu responder, teve muita dificuldade em distinguir a cor azul da cinzenta, independentemente do brilho. Continuando a

análise de resultados, segue-se a questão 5, cuja resposta seria *aled*, *am2*, *arqc*, *fisica1* e *tcomp*. A maioria os utilizadores que responderam acertadamente a esta questão, fizeram-no seleccionando a disciplina de *am1* e depois passando o cursor sobre *fex* na visualização multi-matricial, ou na visualização multi-camada, observando as cadeiras que ficaram destacadas com outra cor. Como mencionado na secção 5.2, poderiam também ter seleccionado a cadeira de *fex* em qualquer um dos sistemas, em que só ficariam visíveis na matriz, o subconjunto de cadeiras comuns a estas duas. Existiram também 6 utilizadores que cometeram erros. O erro mais comum por parte dos inquiridos, foi que mesmo realizando bem a tarefa, não souberam interpretar os resultados, mostrando-se confusos e acabando por escrever todas as cadeiras com resultados na segunda matriz, não analisando as cores diferentes. Um dos utilizadores por sofrer de daltonismo, não se apercebeu de todo da diferença de cores. Relativamente à questão 6, a resposta correcta é, *al*, *am1* e *fex*. Cerca de 65% dos utilizadores cometeram erros, e foi a tarefa que levaram mais tempo a concretizar. Através da observação dos comentários e acções realizadas, foram visíveis duas situações que levaram ao elevado número de erros: (1) dificuldade por parte dos inquiridos em identificar a mistura das cores das cadeiras sobrepostas e (2) dificuldade em seleccionar um padrão negativo de uma cadeira, que já possui um padrão positivo seleccionado. No primeiro caso, não foi óbvio para muitos dos utilizadores a mistura de cores, sendo que percorreram durante algum tempo as matrizes até chegarem a uma conclusão, sendo que por vezes cometeram erros nas respostas que deram. Na segunda situação, alguns utilizadores ao verem um quadrado com um triângulo seleccionado, concluíam que já não podiam fazer mais nenhuma acção sobre a disciplina em questão, pelo que, por vezes desseleccionavam o triângulo positivo da cadeira de *aled*, ou então só ao fim de algumas tentativas percebiam que podiam seleccionar o triângulo inferior. Houve também episódios em que alguns participantes não cometiam os erros anteriores mas depois apenas escolhiam como resultado a disciplina de *fex*, por ter apenas um triângulo seleccionado. Por fim, a última tarefa, obteve melhores resultados relativamente à quantidade de erros cometidos em relação à anterior, apesar de os tempos de realização desta tarefa e da tarefa 6 serem muito próximos. Para esta responder a esta questão, o utilizador teria de observar o brilho da célula de *compg* da quarta matriz. Após o participante fazer esta análise, teria de olhar para a camada de círculos correspondentes ao quarto semestre e ver qual seria o círculo de menor tamanho e então comparar o nome da disciplina dita no enunciado com a disciplina do círculo mais pequeno. A resposta correcta a esta questão é, "não", demonstrando que não existe uma relação directa entre o número de padrões de uma cadeira e o número de alunos inscritos na mesma. A maioria dos utilizadores respondeu acertadamente, e conseguiu chegar a esta conclusão, demonstrando que relacionaram correctamente a informação dos dois sistemas. Os erros cometidos prendem-se com alguns utilizadores não terem identificado correctamente o quarto semestre na visualização multi-camada, ou então deram como resposta todas as cadeiras desse semestre, indicando que não interpretaram bem a pergunta. Na Tabela 8 encontram-se sumarizados os principais erros cometidos em cada tarefa.

De seguida é também analisado um conjunto de *Boxplots* com o tempo que cada inquirido levou a realizar uma tarefa, e o número de erros que cometeu.

Começando por analisar as Figuras relativas ao número de erros por utilizador, é notório que este valor é menor nas tarefas 1, 2 e 4, ver Figuras 5.3, 5.5 e 5.11 respectivamente, não ultrapassando o rácio 1 erro por inquirido. É de notar que estes valores representam informação diferente dos valores apresentados na Tabela 7, que se refere à média de erros cometidos em cada tarefa. Uma tarefa pode ter uma média de erros baixa, por poucos

	Erro1	Erro2
Tarefa 1	Dificuldade em distinguir semestres	
Tarefa 2	Dificuldade em identificar número de alunos	
Tarefa 3	Dificuldade em associar o brilho da cor ao número de padrões	Dificuldade em distinguir cores
Tarefa 4	Dificuldade em distinguir cores	
Tarefa 5	Dificuldade em distinguir cores	Não compreender o conceito de restrição de padrões
Tarefa 6	Dificuldade em distinguir mistura de cores	Dificuldade em compreender selecção negativa e positiva de cadeiras
Tarefa 7	Dificuldade em distinguir semestres	

Tabela 8: Sumarização dos erros cometidos em cada tarefa

inquiridos terem cometido falhas nessa questão, mas no entanto ter havido um utilizador que cometeu 4 falhas nessa tarefa.

Neste sentido, destacam-se as tarefas 3,6 e 7, ver Figuras 5.9, 5.13 e 5.15 respectivamente, por serem a que tem valores mais elevados de erros, em que a tarefa 7 possui 3 erros por utilizador.

Relativamente ao tempo demorado a concluir cada tarefa, é notório que as tarefas 3, 6 e 7, Figuras 5.10, 5.14 e 5.16 foram as mais demoradas.

Assim, observando o resultado dos vários Boxplots e da Tabela 7 que contém a média dos resultados obtidos, é possível concluir que a tarefa 6 é a mais complexa, com o maior número médio de erros e com o tempo de concretização mais elevado, no entanto é a tarefa 7 que tem o valor mais elevado de erros por utilizador. Verifica-se também que nem sempre existe uma correspondência entre o número de erros e o tempo. Esta tendência dá-se em todas as tarefas com excepção da 6

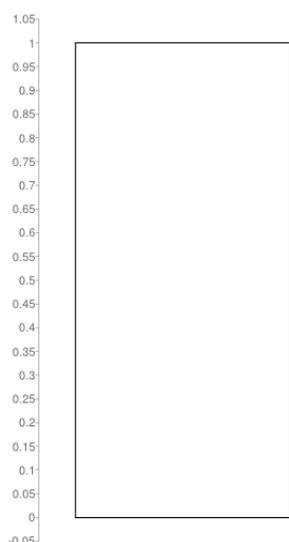


Figura 5.3: Boxplot com o numero de erros cometidos na tarefa 1

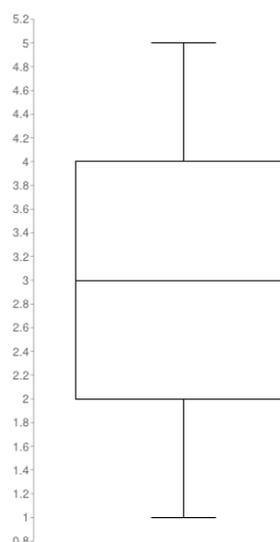


Figura 5.4: Boxplot do tempo demorado para realizar a tarefa 1

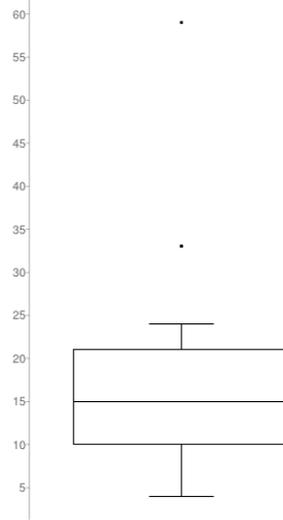
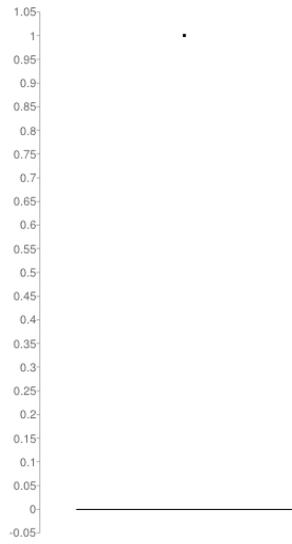


Figura 5.5: Boxplot com o numero de erros cometidos na tarefa 2

Figura 5.6: Boxplot do tempo demorado para realizar a tarefa 2

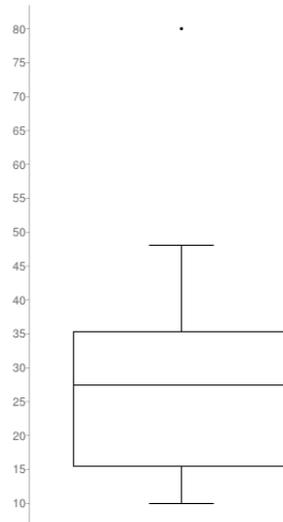
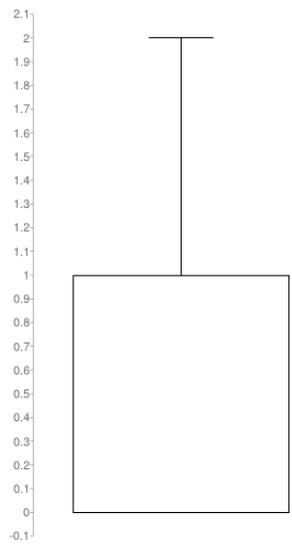


Figura 5.7: Boxplot com o numero de erros cometidos na tarefa 3

Figura 5.8: Boxplot do tempo demorado para realizar a tarefa 3

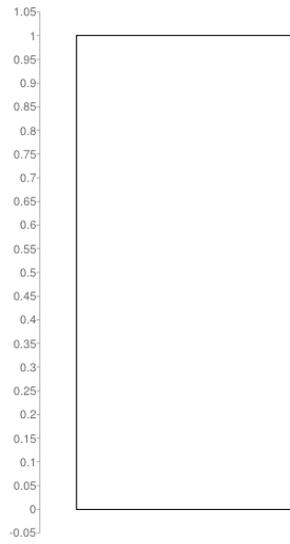


Figura 5.9: Boxplot com o numero de erros cometidos na tarefa 4



Figura 5.10: Boxplot do tempo demorado para realizar a tarefa 4

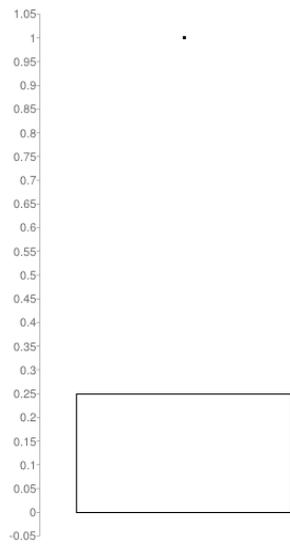


Figura 5.11: Boxplot com o numero de erros cometidos na tarefa 5

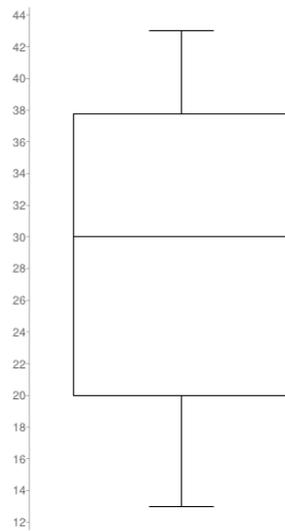


Figura 5.12: Boxplot do tempo demorado para realizar a tarefa 5

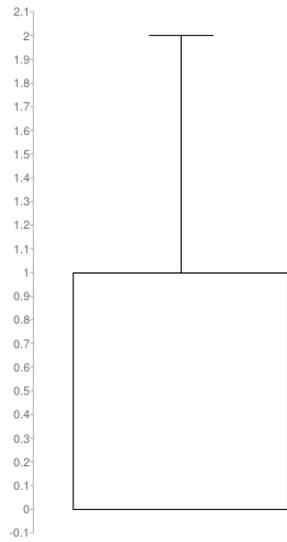


Figura 5.13: Boxplot com o numero de erros cometidos na tarefa 6

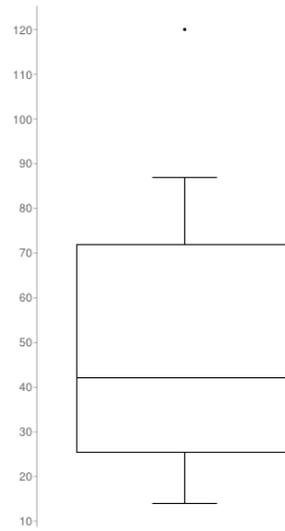


Figura 5.14: Boxplot do tempo demorado para realizar a tarefa 6

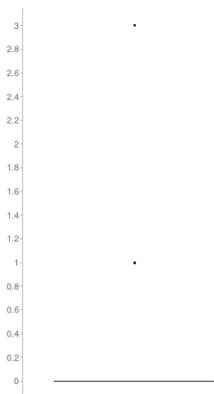


Figura 5.15: Boxplot com o numero de erros cometidos na tarefa 7

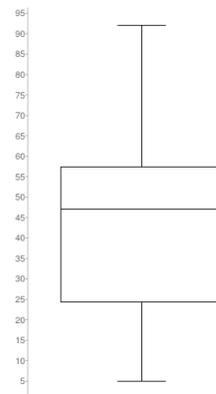


Figura 5.16: Boxplot do tempo demorado para realizar a tarefa 7

5.4.3 Discussão dos Objectivos Alcançados

Face ao objectivo deste trabalho, que pretende ser uma ferramenta de análise do percurso académico dos alunos, e aos requisitos definidos, pode considerar-se que ambos foram cumpridos, tendo em conta o resultado dos testes realizados com utilizadores.

Considerando o requisito da sequencialidade, pode considerar-se que a visualização o transmite com sucesso. Observando o número de erros cometidos na questão 1, bem como o tempo demorado para concluir a tarefa, que são relativamente baixos, conclui-se que a maioria dos inquiridos conseguiu responder correctamente a esta questão.

A nível da distinção da quantidade de padrões em que uma disciplina está envolvida, bem como quais são as cadeiras base num semestre através do brilho da cor, considera-se que o requisito é cumprido com algum sucesso. Observando o resultado da questão 3, que testa estas condições, verifica-se que o número médio de falhas é de 55%, pelo que existem alguns erros a considerar, e que indicam o caminho para possíveis melhorias.

Relativamente aos padrões que correspondem a cada disciplina, a visualização cumpre com sucesso esta condição. A taxa de erros da tarefa 4, que se refere a este requisito, é baixa, 30%, mostrando que a maioria dos inquiridos conseguiu realizá-la com sucesso. As falhas cometidas mostram que houve dificuldade em alguns casos, na distinção dos tons mais escuros de azul e cinzento, enquanto que nos tons mais claros, não houve problemas.

Passando para o requisito de filtragem de padrões, observando os resultados da questão 5, que lhe é correspondente, verifica-se que a taxa de sucesso é de 70%, pelo que se pode também considerar que este requisito foi atingido.

Por último, segue-se o requisito da utilização da mistura de duas cores, quando existem cadeiras sobrepostas. Tendo em conta o valor médio da percentagem de erros da questão 6, que se refere a este requisito, ser de 75%, este objectivo destaca-se por ser o mais problemático e o que mais confusão gerou entre os inquiridos, considerando-se que não foi atingido com sucesso, por a taxa média de erros ser tão elevada. Existem duas situações que levaram a este valor, estas são a dificuldade em identificar a mistura de cores que preenche um triângulo de uma disciplina, e a dificuldade em perceber que é possível seleccionar os dois triângulos do mesmo quadrado.

Assim, pode concluir-se que, os objectivos definidos foram cumpridos, existindo algumas melhorias que podem ser realizadas, na tentativa de diminuir a taxa média de erros por cada tarefa. A única excepção, refere-se ao requisito sobre a percepção da mistura de cores. Esta funcionalidade deve ser trabalhada e melhorada, desenvolvendo uma solução para os tipos de erros cometidos e desse modo, baixando a taxa média de erros.

As melhorias sugeridas à implementação destes requisitos encontram-se descritas na secção 6.1 Trabalho Futuro.

CONCLUSÕES FINAIS E DISCUSSÃO

A principal motivação para este estudo prendeu-se essencialmente com o desenvolvimento de um método de visualização para a representação de grandes quantidades de dados, de modo a permitir a sua análise. Nesta dissertação, os dados usados são padrões emergentes das actividades educacionais dos alunos do curso de Engenharia Informática e de Computadores, obtidos através de *data mining* sequencial.

Neste contexto, as técnicas de *data mining* têm mostrado ser relevantes para encontrar padrões nos dados, mas o resultado da aplicação dos mesmos leva muitas vezes a conjuntos de dados difíceis de ler, interpretar e analisar. Face a este desafio, surgiu o EduVis, uma visualização que representa informação e fornece mecanismos de exploração e filtragem, que permitem a comparação simultânea de disciplinas, bem como realizar restrições aos dados observados. Retira também partido da cor, de forma a diminuir a sobrecarga que se poderia gerar dados os vários tipos de artefactos com propriedades semelhantes presentes, enfatizando a informação relevante.

De forma a chegar ao estado final da visualização apresentada, foi percorrido um longo caminho. O primeiro passo foi dado com o desenho de protótipos de baixa fidelidade, onde foram desenhadas várias tentativas, até se chegar a uma opção viável, que fizesse o mapeamento adequadamente das variáveis existentes, e que facilita-se a compreensão e a interpretação dos dados por parte dos utilizadores. De seguida, com um esboço de uma visualização já delineado, começou o desenvolvimento da mesma. Este foi um processo iterativo e incremental, com vista a ter uma visualização o mais eficaz e fluida possível, possibilitando maneiras satisfatórias de manipular e explorar a informação representada. Findo o seu desenvolvimento, realizou-se a sua integração, na visualização previamente existente, com o objectivo de que estes dois mecanismos trabalhassem coordenadamente e de forma natural.

Posteriormente foram efectuados testes com utilizadores, de forma a avaliar a usabilidade da visualização, e que permitiram a obtenção de métricas referentes ao desempenho do classificador utilizado.

Os testes realizados com utilizadores demonstraram que a visualização torna imediatamente evidente um conjunto relevante de informação geral e permite uma fácil recolha de dados através de mecanismos de exploração e comparação interactiva. Os participantes perceberam imediatamente aspectos como o número de semestres representados, bem como as disciplinas com maior ou menor sucesso utilizando os mecanismos disponíveis para comparar, filtrar e recolher informação e mostraram satisfação no desempenho das tarefas.

6.1 TRABALHO FUTURO

Apesar de se considerar que a visualização é bem sucedida no cumprimento dos seus objectivos, existem vários aspectos que carecem de melhorias futuras, e que se tornaram evidentes através dos erros cometidos pelos inquiridos. Começando por analisar a falha cometida nas questões 1 e 7, que se refere à dificuldade em distinguir o número de semestres representados, seria de considerar eliminar as matrizes que não possuem dados na camada multi-matricial. O facto de haver sempre 6 matrizes, faz com este mecanismo acabe por não ficar consistente com a visualização multi-camada, que apenas representa os semestres que contêm informação, e acabou por induzir em erro alguns utilizadores.

Outro problema levantado por vários utilizadores, é a questão de por vezes, estes se esquecerem do significado das cores, ou então, não associarem a cada mecanismo o seu significado, como por exemplo, o mecanismo multi-matricial foca-se principalmente na visualização do número de relações e de padrões de cada cadeira, enquanto que o mecanismo multi-camada demonstra o número de alunos aprovados ou reprovados a cada cadeira. Assim, posteriormente seria desejável adicionar uma legenda à visualização, que desse uma breve informação sobre o objectivo de cada modelo. Seria também de considerar uma possível alternativa para casos de daltonismo, de modo a proporcionar a utilização do EduVis nestas situações. Estas soluções serviriam para evitar os erros cometidos na questão 3.

Outro erro comum, foi a dificuldade em distinguir as cores usadas, quando se ilumina um padrão, e são pintados os triângulos azuis. Por haver cadeiras, em que o azul é muito escuro, se o seu triângulo estiver adjacente a um quadrado que esteja cinzento, e a sua luminância for igualmente baixa, pode haver alguma confusão na distinção das duas cores, e o utilizador perder alguma informação, como aconteceu na questão 4.

Apesar das alterações sugeridas, é possível concluir, com base nos resultados obtidos que o Eduvis fornece os meios para representar os resultados de informação educacional de uma forma que proporciona à comunidade educativa perceber um conjunto de padrões que não seriam evidentes de outra forma.

Ao permitir diagnosticar problemas como reprovações comuns a determinadas cadeiras, ou informando sobre dependências entre estas, o sistema será útil para que os coordenadores de programas de estudo, tal como professores, tenham os meios para encontrar soluções para as limitações existentes, promovendo o sucesso na educação.

BIBLIOGRAFIA

- [1] R. Mazza and V. Dimitrova, Generation of Graphical Representations of Student Tracking Data. In *Course Management System 9th International Conference on Information Visualization*, 2005
- [2] Y. Gao, A study on Mass Higher Education in the world Based on Comparative Perspective. In *Education Institute of Xiamen University*, 2010
- [3] P. Gestwicki, Work In Progress. In *38th ASEE/IEEE Course Management System 9th International Conference on Information Visualization*, 2008
- [4] Y. Guodong and W. Huan, Visual Exploration for Time Series Data Using Multivariate Analysis Method. In *Computer Science and Education (ICCSE)*, 2013
- [5] J. Heer and M. Bostock and V. Ogievetsky, A Tour Through the Visualization Zoo. In *Communications of the ACM*, 2010
- [6] J. Kay and P. Reimann and E. Diebold and B. Kummerfeld, MOOCs: So Many Learners, So Much Potential. In *University of Sydney*, 2013
- [7] D. Keim, Information Visualization and Visual Data Mining, Visualization and Computer Graphics. In *IEEE Transactions on (Volume:8 , Issue: 1)*, 2002
- [8] R. Mazza and L. Botturi, Monitoring an Online Course with the GISMO Tool: A Case Study. In *Journal of Interactive Learning Research*, Volume 18, Issue 2, 2007
- [9] Y. Chen and C. Xianchen and C. Hongqian, A Multidimensional Data Visualization Method Based On Parallel Coordinates and Enhanced Ring. In *International Conference on Computer Science and Network Technology*, 2011
- [10] S. Murray, *Interactive Data Visualization*, 2013, O'Reilly Media, Gravenstein Highway North
- [11] J. Steele and N. Iliinsky, *Beautiful Visualization.*, 2010, O'Reilly Media, Gravenstein Highway North
- [12] J. Steele and N. Iliinsky, *Designing data visualizations*, 2011, O'Reilly Media, Gravenstein Highway North
- [13] D. Trimm and P. Rheingans and M. DesJardins, Visualizing Student Histories Using Clustering and Composition. In *IEEE Transactions on Visualization and Computer Graphics*, 2012
- [14] E. Valiti, Avaliação de Usabilidade de Técnicas de Visualização de Informações Multidimensionais. In *Universidade Federal do Rio Grande do Sul, Instituto de informática, Programa de Pós-Graduação em Computação*, 2008
- [15] C. Ware, *Information Visualization, Perception for Design, Second Edition*, 2004, Morgan Kaufmann Publishers
- [16] D. Wortman and P. Rheingans, Visualizing Trends in Student Performance across Computer Science Courses Categories and Subject Descriptors. In *SIGCSE '07 Proceedings of the 38th SIGCSE technical symposium on Computer science education*, 2007

- [17] G. Xiaoya and L. Kan and L. Ping, Visual Analysis of College Students Scores. In *English Test, I. Conference and C. Science*, 2009
- [18] KEIM, D.A. Visualization Techniques for Mining Large Database: A Comparison. In *IEEE Knowledge and Data Engineering, IEEE Transactions on (Volume:8 , Issue: 6)* 923-936, Dec. 1996.
- [19] S. Gama, D. Gonçalves, Studying Color Blending Perception for Data Visualization. In *Eurovis 2014 - Eurographics Conference on Visualization*, 2014.
- [20] S. Gama, D. Gonçalves, Studying the perception of color components relative amounts in blended colors. In *NordiCHI 2014, 8th Nordic Conference on Human-Computer Interaction*, 2014
- [21] C. Ware, *Information Visualization, perception for design, Third Edition*, 2013, Morgan Kaufmann Publications, San Francisco
- [22] R. Agrawal, R. Srikant, Mining sequential patterns. In *Proceedings of the IEEE International Conference on Data Engineering*, pag. 3-14, 1995.
- [23] Hodge,V., Austin, J., textitA survey of outlier detection methodologies, *Artificial Intelligence Review*, 22(2), 2004, (pp. 85-126).
- [24] J. Sauro. *A Practical Guide to the System Usability Scale: Background, Benchmarks and Best Practices*. CreateSpace, 2011.

APÊNDICE A

QUESTIONÁRIO DE SATISFAÇÃO

Esta avaliação está enquadrada na tese de mestrado "EduVis: Visualização Interactiva de Dados Educacionais"

1. Gostaria de usar o sistema frequentemente

- 1
- 2
- 3
- 4
- 5

2. Achei o sistema desnecessariamente complexo

- 1
- 2
- 3
- 4
- 5

3. Achei o sistema fácil de usar

- 1
- 2
- 3
- 4
- 5

4. Penso que precisaria da ajuda técnica para usar este sistema

- 1
- 2
- 3
- 4
- 5

5. Penso que as funcionalidades do sistema estão bem integradas

- 1
- 2
- 3

- 4
- 5

6. Achei o sistema inconsistente

- 1
- 2
- 3
- 4
- 5

7. Penso que a maior parte das pessoas aprenderia rapidamente a usar este sistema

- 1
- 2
- 3
- 4
- 5

8. Achei o sistema muito complicado de usar

- 1
- 2
- 3
- 4
- 5

9. Senti-me confiante ao usar o sistema

- 1
- 2
- 3
- 4
- 5

10. Precisei de aprender bastante antes de começar a usar este sistema facilmente

- 1
- 2
- 3
- 4
- 5

11. Consegui facilmente perceber o número de semestres representado

- 1
- 2
- 3

- 4
- 5

12. Consegui perceber quais as cadeiras com mais sucesso e insucesso

- 1
- 2
- 3
- 4
- 5

13. Consegui facilmente perceber quais as cadeiras que estão envolvidas em mais padrões

- 1
- 2
- 3
- 4
- 5

14. Consegui perceber facilmente a precedência entre cadeiras

- 1
- 2
- 3
- 4
- 5

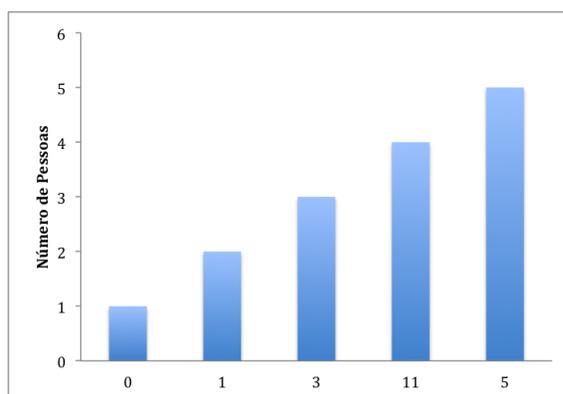
15. Consegui facilmente visualizar simultaneamente os padrões de duas cadeiras diferentes

- 1
- 2
- 3
- 4
- 5

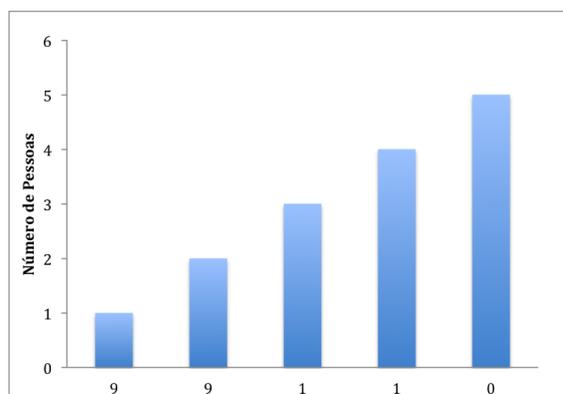
APÊNDICE B

RESULTADOS DO QUESTIONÁRIO DE SATISFAÇÃO

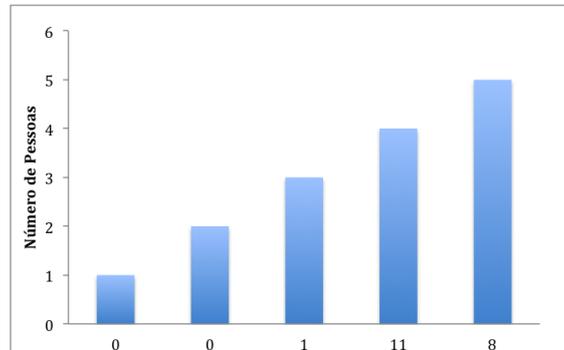
Questão 1 - Gostaria de usar o sistema frequentemente?



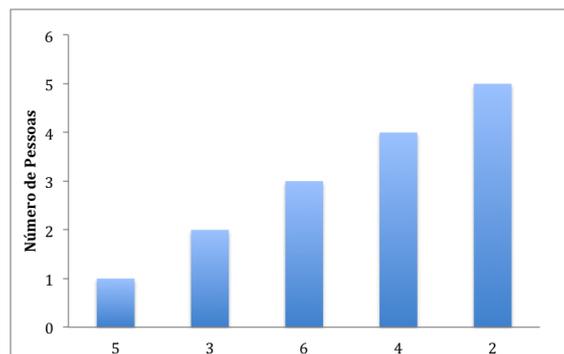
Questão 2 - Achei o sistema desnecessariamente complexo



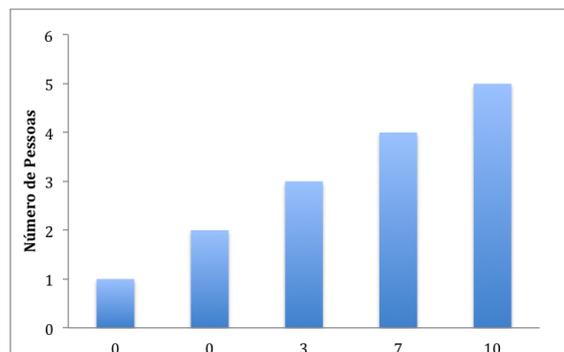
Questão 3 - Achei o sistema fácil de usar



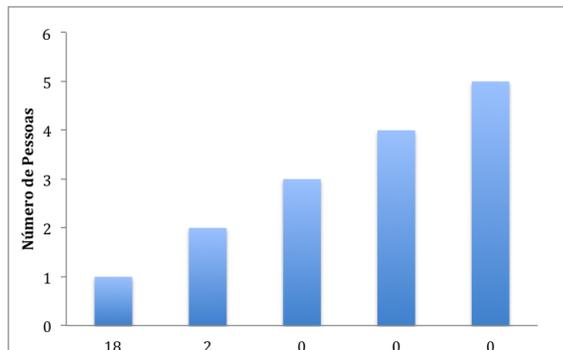
Questão 4 - Penso que precisaria da ajuda técnica para usar este sistema



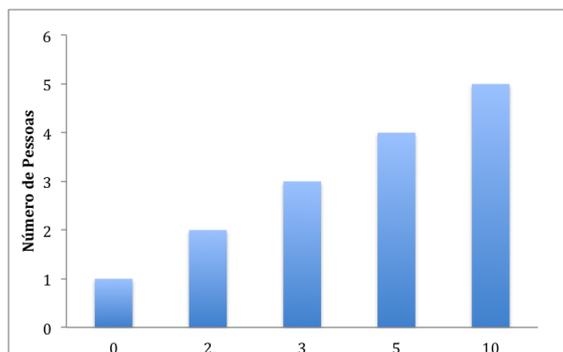
Questão 5 - Penso que as funcionalidades do sistema estão bem integradas



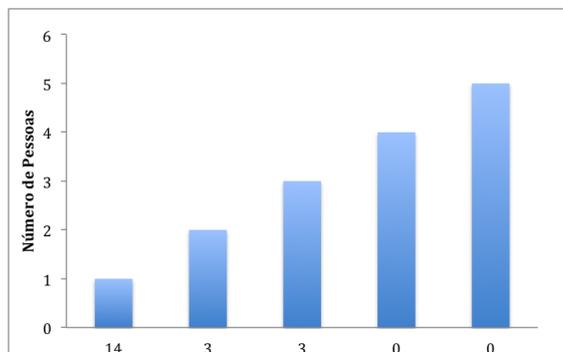
Questão 6 - Penso que as funcionalidades do sistema estão bem integradas



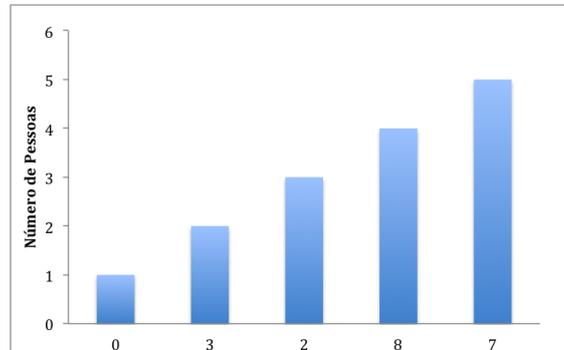
Questão 7 - Penso que a maior parte das pessoas aprenderia rapidamente a usar este sistema



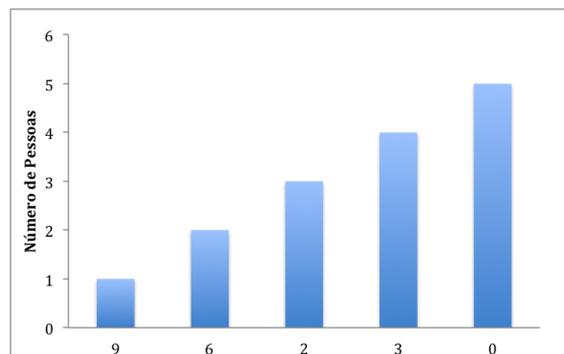
Achei o sistema muito complicado de usar



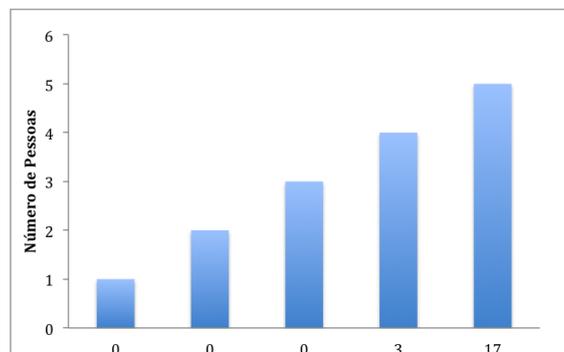
Questão 9 - Senti-me confiante ao usar o sistema



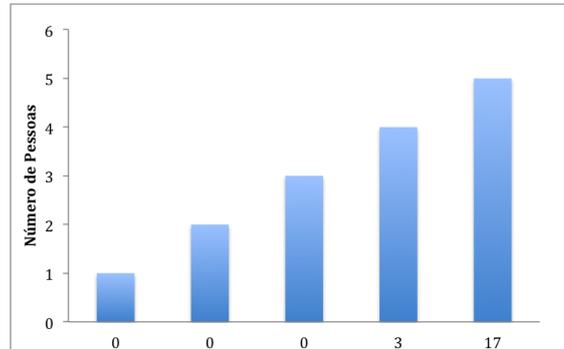
Questão 10 - Precisei de aprender bastante antes de começar a usar este sistema facilmente



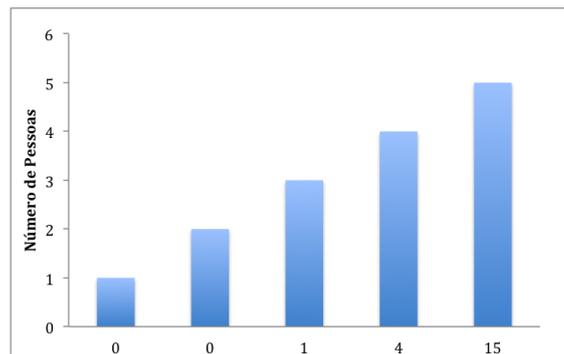
Questão 11 - Consegui facilmente perceber o número de semestres representado



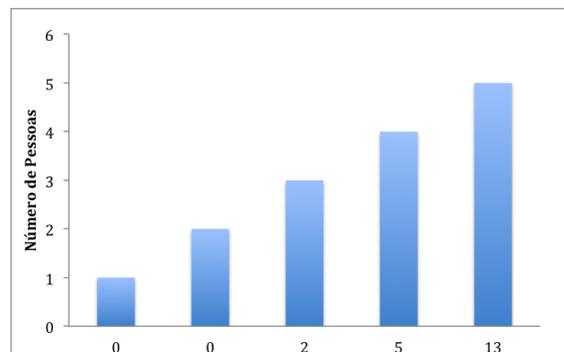
Questão 12 - Consegui perceber quais as cadeiras com mais sucesso e insucesso



Questão 13 - Consegui facilmente perceber quais as cadeiras que estão envolvidas em mais padrões



Questão 14 - Consegui perceber facilmente a precedência entre cadeiras



Questão 15 - Consegui facilmente visualizar simultaneamente os padrões de duas cadeiras diferentes

