



# EduVis

Visualização de padrões provenientes de dados educacionais

## Relatório Técnico

Sandra Gama e Daniel Gonçalves

Agosto de 2014

Projeto Educare - PTDC/EIA-EIA/110058/2009

## Resumo

Uma análise efetiva do percurso e resultados obtidos pelos alunos no ensino tem uma grande relevância por proporcionar ferramentas úteis à comunidade académica na identificação de problemas e padrões. Atualmente através de aplicações de técnicas de *data mining* é possível identificar esses padrões, pontos fraturantes e fazer previsões do possível desempenho que os estudantes possam vir a ter. No entanto compreender esta informação torna-se muito complexo, uma vez que consiste num amplo conjunto de dados, representados sob a forma de texto, o que impossibilita ou dificulta a sua legibilidade. Adicionalmente, a análise destes resultados frequentemente requer conhecimento razoável de algoritmos matemáticos e estatísticos complexos por parte do utilizador, para que este possa entender e avaliar os padrões obtidos. No entanto, tal facto nem sempre se verifica, considerando o vasto espectro de intervenientes dos processos educativos. É necessário, então, procurar uma forma de representar a informação deste domínio de uma forma perceptível para os utilizadores, de maneira a tirar partido do potencial da informação disponibilizada pelos processos de *data mining* educacional. Dado o potencial já demonstrado da visualização de informação para representar grandes quantidades de dados, foi criado um mecanismo interativo de visualização de informação educacional, evidenciando as várias disciplinas curriculares e relações de precedência e simultaneidade entre as mesmas. Para tal foi desenvolvida uma visualização que integra dois mecanismos complementares, uma vista multi-camada e uma vista multi-matricial. Testes com utilizadores demonstraram o potencial desta solução para a visualização de dados educacionais.

# Índice

<b>1. INTRODUÇÃO</b>	<b>4</b>
1.1. Objetivo	4
1.2. Organização do documento	5
<b>2. TRABALHO RELACIONADO</b>	<b>6</b>
2.1. Ensino Online	6
2.2. Ensino Tradicional	7
2.3. Discussão	8
<b>3. VISUALIZAÇÃO DE PADRÕES EDUCACIONAIS</b>	<b>9</b>
3.1. Padrões Educacionais	9
3.2. Solução	10
3.2.1. <i>Representação Temporal</i>	11
3.2.2. <i>Conectores Visuais para Representação de Percursos Curriculares Típicos</i>	14
3.2.3. <i>Utilização da Cor como Veículo de Informação</i>	16
3.2.4. <i>Visualização Global de Inter-relações: Vista Multi-Matricial</i>	17
3.2.5. <i>Exploração Integrada de Padrões</i>	19
3.3. Exemplo de Exploração	21
<b>4. AVALIAÇÃO</b>	<b>25</b>
4.1. Tarefas representativas	25
4.1. Protocolo de teste	26
4.3. Resultados	27
<b>5. CONCLUSÕES</b>	<b>29</b>
<b>REFERÊNCIAS</b>	<b>30</b>

# 1. Introdução

O número de alunos inscritos no ensino tradicional e online tem vindo a aumentar nas últimas décadas. Relativamente ao ensino universitário, a taxa de alunos que o frequentam teve um crescimento global de quase 8% nos últimos 30 anos [Gao10]. 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 o aumento exponencial de estudantes. Com este crescimento popularizou-se o uso de CMS (*Course Management Systems*) e de LMS (*Learning Management*), ferramentas desenhadas para prestar assistência aos alunos e transpor o ensino para a internet [Kay13]. Com este aumento no número de alunos no ensino tradicional e online surge uma grande quantidade de informação em bruto correspondente aos dados curriculares. A análise eficaz desta informação poderá proporcionar à comunidade académica um meio para ganhar mais conhecimento e compreensão sobre os diferentes problemas que se manifestam e que muitas vezes podem passar despercebidos, podendo agir em conformidade.

Para analisar este tipo de informação são utilizadas técnicas de *data mining*, que neste contexto assume a denominação EDM (*Data Mining* Educacional) e que atualmente tem vindo a ganhar terreno, por desenvolver técnicas para exploração e análise de grandes quantidades de dados provenientes do contexto educacional. Contudo, e apesar da relevância dos padrões obtidos, não é possível fazer uma análise direta dos mesmos. Tal deve-se ao seu formato textual extenso, difícil de ler e interpretar, e ao facto de muitas vezes serem necessários conhecimentos complexos de algoritmos matemáticos e estatísticos para analisar os dados. Neste sentido, para a prospecção de dados ser eficaz, é necessário envolver o utilizador no processo da exploração e combinar criatividade, flexibilidade e conhecimento geral [Keim02]. A visualização tem o potencial para cumprir estes requisitos e tem vindo a ganhar destaque na representação de informação. O seu objectivo é ajudar à compreensão de grandes quantidades de dados, auxiliando na descoberta de tendências, padrões e discrepâncias na informação [Ware12]. Assim, uma visualização bem desenhada pode substituir mecanismos cognitivos apenas com deduções perceptuais, tornando possível a descoberta de propriedades anteriormente desconhecidas, mostrando padrões e sub-problemas, que um conjunto de descrições textuais teriam dificuldade em evidenciar. Terão então acesso a informação necessária à tomada de decisões no sentido de encontrar as soluções mais adequadas para otimizar os processos educacionais.

## 1.1. Objetivo

Este trabalho procura usar técnicas de visualização de informação para possibilitar a análise de padrões educacionais obtidos através de EDM. Assim, o objetivo principal é ***permitir e facilitar a aquisição e assimilação de conhecimentos sobre o percurso curricular dos alunos através de um método de visualização que destaque as cadeiras com maior ou menor sucesso e que forneça mecanismos para analisar padrões educacionais como as relações de precedência e simultaneidade entre disciplinas e a comparação relativa entre as mesmas.***

Para tal, procura-se o estudo e criação de um mecanismo que permita uma interpretação eficaz e eficiente das inter-relações entre as diversas disciplinas dos vários semestres do curso, evidenciando o sucesso e o insucesso e as disciplinas envolvidas em mais padrões, ou seja, que possuem mais relações de precedência ou simultaneidade, e fornecendo mecanismos de filtragem para uma exploração mais detalhada dos dados.

## **1.2. Organização do documento**

O presente documento organiza-se em cinco secções. Na secção 2 é apresentado e discutido trabalho relevante na área. Na secção 3 é exposto o contexto do problema, enunciando e detalhando os padrões resultantes do EDM. São descritas as várias fases do processo de criação da visualização, apresentando e discutindo os vários mecanismos interativos criados. Na secção 4 é descrita a avaliação efetuada com utilizadores. É apresentado o protocolo de teste e são discutidos os resultados obtidos. Na secção 5 são enunciadas as conclusões do presente trabalho, discutindo as contribuições do mesmo no contexto da visualização de informação educacional.

## 2. Trabalho relacionado

Com o crescimento do número de alunos no ensino tradicional e online, várias ferramentas foram criadas no âmbito da representação gráfica de informação educacional.

### 2.1. Ensino Online

Tendo em conta a educação online, os CMS permitem a criação de salas de aula virtuais, onde alunos e professores podem compartilhar informação, tornando possível a participação remota em discussões e a gestão de aulas. Este tipo de interação gera grandes quantidades de dados que precisam de ser geridos de uma forma que forneça aos professores informação relevante sobre o desempenho dos alunos.

Para superar esse desafio, foi criado o CourseVis [Mazza05], sendo usado como uma extensão ao CMS que permite a exploração interativa de dados e manipulação através de diferentes mecanismos de visualização. Um desses mecanismos é a representação tridimensional de participantes num fórum, em que os tópicos são representados como esferas com tamanho proporcional ao número de alunos envolvidos, permitindo *zoom* e visualização panorâmica. Outro mecanismo é a Matriz Cognitiva, que consiste numa matriz na qual os nomes dos alunos são representados ao longo de um eixo e os conceitos da disciplina no outro eixo, sendo associada uma escala de cor ao desempenho, variando entre o verde (sucesso) e o vermelho (insucesso). Uma terceira ferramenta, que retrata o comportamento dos alunos através de gráficos e texto ordenados de forma matricial, tem como objetivo mostrar informação como o acesso aos conteúdos, a assiduidade e o progresso escolar. Um estudo com utilizadores mostrou que as representações gráficas do CourseVis permitem obter informação sobre os aspectos cognitivos dos estudantes de forma rápida e precisa. No entanto, os utilizadores tiveram alguns problemas na compreensão da informação devido à sobreposição de elementos visuais, resultando em dificuldade na leitura dos gráficos na representação do comportamento dos estudantes [Iliinsky11].

As limitações do CourseVis levaram à criação de GISMO (Sistema Gráfico Interativo de Monitorização de Estudantes), que representa visualmente dados de LMS, bastante complexos e difíceis de ler e entender. O GISMO está integrado no LMS Moodle mas pode também ser adaptado para outras plataformas de aprendizagem. Focado numa visualização bidimensional de comportamentos, o GISMO oferece uma visualização simples para parâmetros específicos. Permite a exploração interativa de acessos e detalhes de recursos, e fornece os meios para a exploração dos comportamentos dos alunos que tinham sido considerados relevantes no CourseVis [Mazza05]. Tendo sido usado num curso *online*, o GISMO mostrou ser eficaz para a compreensão dos comportamentos individuais, bem como dos métodos de avaliação, permitindo redesenhar o curso de acordo com as necessidades dos alunos.

## 2.2. Ensino Tradicional

Dado que o ensino tradicional resulta de processos de avaliação, tais como notas de avaliação e outras competências, como assiduidade e participação, a informação resultante pode ser difícil de interpretar. Adicionalmente, a descoberta de padrões e troca de informação são quase impraticáveis neste contexto.

Um dos estudos que têm sido feitos no âmbito da visualização de informação para apoiar os processos educacionais é o AVOJ [Xiaohuan13]. De modo a refletir as diferentes capacidades dos alunos, esta ferramenta permite a exploração e visualização de dados sobre o desempenho. Fornece meios para agrupar os alunos de acordo com suas notas e outros aspetos, como os hábitos de estudo. No entanto, faz uso de um esquema de cores com níveis máximos de brilho e saturação, competindo pela atenção visual do utilizador, o que torna mais difícil encontrar padrões [Iliinsky11]. O sistema oferece um mecanismo de visualização adicional para comparar tendências gerais, composto por um gráfico de barras que mostra estatísticas em tempo real, tornando possível compreender alguns aspectos específicos, como a forma como os alunos gerem as suas sessões de estudo. Os autores acreditam que, uma vez fornecendo um amplo conjunto de dados que não estavam disponíveis anteriormente, será possível obter mais informação sobre os processos educacionais e melhorar a eficácia do ensino.

Outro estudo interessante é o de de Xiaoya et al. [Xiaoya09], uma visualização para análise de resultados dos estudantes universitários num curso de Inglês. Seguem uma visualização de coordenadas paralelas, em que N eixos equidistantes são usados para representar as dimensões de um conjunto de dados multidimensional. No entanto, vários modelos adicionais têm sido utilizados: (i) classificação, que permite aos utilizadores dividir dados em conjuntos arbitrários de aulas, (ii) média, o que torna possível, por meio de interações com o gráfico principal, obter os valores médios de cada conjunto de dados ou subconjuntos de dados, (iii) *box plot*, para medir a dispersão dos dados, (iv) permuta de eixos, o que evidencia relações internas entre atributos; (v) correlação, para calcular a correlação entre dois conjuntos de dados ao longo de dois eixos paralelos, (vi) associação, no qual se pode prever a ocorrência de atributo A com base na ocorrência de atributo B, e (vii) *roll-up* e *drill*, permitindo a representação de dados de diferentes níveis hierárquicos. Estes modelos, além de permitirem a representação e manipulação interativa de dados, facultam uma visão geral imediata e mostram um número considerável de variáveis simultaneamente, permitindo uma análise mais eficiente e eficaz.

Trimm et al. [Trimm12] criaram ainda uma visualização em que os alunos são agrupados de acordo com suas notas. Estes grupos podem ser visualizados utilizando composições que mostram as suas características e variações no tempo. A fim de utilizar a composição, a informação sobre o historial de cada aluno é constituída por uma trajetória bidimensional, representada em dois eixos. Esta representação, embora simples, não é muito eficiente para mostrar simultaneamente características e tendências de uma grande quantidade de dados dos estudantes. A visualização utiliza recursos espaciais para representar trajetórias dos alunos ao longo do tempo usando o algoritmo de nível definido. É utilizada uma técnica de mistura de cor para mostrar características como a média e o desvio padrão em relação aos resultados

dos alunos, representando a informação de uma forma natural, e é utilizada a técnica de *color weaving* para mostrar o valor de um determinado atributo em relação a uma trajetória selecionada aleatoriamente. No entanto, o uso de gradientes pode adicionar ruído à imagem e a variação de brilho e saturação pode também causar a ilusão de proximidade, levando a uma interpretação imprecisa dos dados. No entanto, testes com utilizadores mostraram que esta visualização fornece meios para encontrar novos padrões relevantes nos dados.

A fim de entender o abandono escolar de um número significativo de estudantes relativamente a um curso de ciência da computação, foi desenvolvida uma ferramenta que permite a visualização de padrões de repetição relativamente a sucesso ou insucesso [Wortman07]. Devido à grande quantidade de informação disponível, foi usada uma estrutura visual de nós e arestas. Os nós representam eventos, com largura proporcional ao número de estudantes de cada evento, e as arestas representam a trajetória dos estudantes relativamente aos eventos (exames ou trabalhos), sendo a largura representativa do número de alunos a ser descritos neste comportamento. As cores representam o desempenho dos alunos, permitindo diferenciar grupos de estudantes, identificando os que têm comportamento similar. A visualização é interativa, permitindo a seleção de categorias de estudantes, como *alunos que repetem pelo menos uma disciplina* ou *alunos que nunca reprovaram uma disciplina*. Como resultado, foi possível aos professores obter um conjunto de conclusões a respeito de falhas repetitivas ou implicações de uma determinada cadeira no sucesso noutras disciplinas. A estrutura visual tem, contudo, uma importante lacuna no que diz respeito à sobreposição das linhas mais grossas às mais finas, o que pode tornar difícil perceber alguma informação.

### 2.3. Discussão

Todas as abordagens mencionadas focam a visualização de informação relativamente aos processos de ensino, tradicional ou on-line. Apresentam diferentes técnicas e metodologias para visualizar os dados neste contexto em particular. Dado o âmbito do presente estudo e as particularidades dos dados que pretendemos analisar, em que uma visualização que permita interrelacionar disciplinas e mostrar as suas interdependências apresenta maior relevância, destacamos o trabalho de Wortman e Rheingans [Wortman07] e, mais recentemente, o de de Trimm et al. [Trimm12].

Uma visualização interativa que permita mecanismos interativos, como destaque, comparação e filtragem, é de extrema importância no âmbito deste estudo. No entanto, o último trabalho supracitado, de Trimm et al. [Trimm12], oferece interação limitada sobre as particularidades de nosso contexto. Além disso, o estudo de Wortman e Rheingans [Wortman07], apesar de proporcionar mecanismos de interação, não permite comparação entre disciplinas ou destaque de padrões específicos. Além disso, o esquema de cores usado não alivia a confusão visual que está presente quando muitas arestas diferentes são representadas. Tentámos colmatar as lacunas das ferramentas supracitadas através da criação de uma método de visualização interativo que representa as inter-relações entre as disciplinas de um programa de estudos e permite mecanismos de seleção e filtragem.



### 3. Visualização de padrões educacionais

Com o número crescente de ingressos no ensino superior, e no sentido de proporcionar aos estudantes um curso cada vez melhor adaptado às suas necessidades, é necessário que se consigam identificar os vários padrões existentes. Como resultado, os intervenientes nos processos de ensino terão acesso a melhor e maior conhecimento dos problemas e desafios existentes, podendo agir no sentido de procurar melhorar os processos de ensino.

No contexto do projeto Educare foram analisados dados recolhidos sobre os três primeiros anos dos percursos académicos da Licenciatura em Engenharia informática e Computadores (LEIC) do Instituto Superior Técnico. Utilizando técnicas de EDM, foi possível extrair padrões, que fornecem o número de aprovações e reprovações a determinadas disciplinas e as relações de precedência e simultaneidade entre as mesmas. Um exemplo de precedência seria, “90 dos alunos que não conseguiram aprovação à cadeira A num semestre, também não conseguiram aprovação à cadeira B no semestre seguinte”. Uma relação de simultaneidade poderá ser “10 dos alunos que foram aprovados na disciplina C num determinado semestre, também foram aprovados na cadeira D no mesmo semestre”.

Neste contexto, foi estudada a criação de uma visualização que visa mostrar a informação proveniente destes padrões de forma a que permita ter uma visão de conjunto do percurso curricular dos alunos, as cadeiras críticas e as relações de precedência a simultaneidade entre as mesmas. As próximas seções detalham as características da informação a representar e discutem o processo de desenvolvimento da solução criada para visualização e análise de dados educacionais.

#### 3.1. Padrões Educacionais

Neste trabalho foi utilizado o resultado da aplicação de *data mining* sequencial a dados recolhidos durante nove anos no contexto do percurso académicos na Licenciatura em Engenharia informática e Computadores (LEIC) do Instituto Superior Técnico no contexto do projeto Educare.

O objetivo do *data mining* sequencial é, dado um conjunto de sequências e um limiar de suporte mínimo, descobrir o conjunto de sequências que estão contidas em pelo menos  $\sigma$  sequências do conjunto de dados, isto é, o conjunto de sequências frequentes [Agrawal95]. Esta técnica permite a descoberta de padrões sequenciais frequentes, sendo estes consistentes com o conhecimento de fundo existente. Tal conhecimento pode ser representado por uma gramática livre de contexto, que desempenha o papel de uma restrição no processo de *data mining* sequencial. Este método não só reúne padrões esperados com base no conhecimento de fundo mas, com o uso de relaxamentos, permite também a descoberta de padrões que correspondem a desvios ao comportamento esperado, tornando evidentes algumas tendências potencialmente relevantes que antes eram desconhecidas [Antunes08].

Assim, o conhecimento do currículo escolar foi representado como um autômato finito, estabelecendo a ordem das disciplinas que os alunos deviam completar para concluir o curso. O *data mining* sequencial com três diferentes valores de limiar de suporte (50%, 25% e 20%, correspondente à percentagem do número total de alunos aos quais o padrão se deve aplicar, para poder se considerado) foi realizado, resultando em três diferentes conjuntos de padrões. Evidentemente, quanto menor for o valor do limiar de suporte, maior é o número de padrões resultante da aplicação do *data mining* sequencial. Como um resultado da aplicação do método acima referido, foi gerado um conjunto de padrões textuais que são agrupados por semestre. Estes correspondem a semestres de inscrição e não a semestres letivos no currículo dos alunos. Ou seja, nos exemplos abaixo, quando são referidos um “primeiro semestre” e um “segundo semestre”, estamos a referir-nos a semestres consecutivos de inscrição dos alunos, e não a um ano letivo ou de permanência do aluno no curso em particular.

Os padrões observam a seguinte estrutura:

```
padrão i = (semestre 1, ... , semestre2, N total alunos),  
semestre j = disciplinal || (disciplinal, ... , disciplinaN);
```

Exemplos:

```
['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', '~fisical'], '591']:
```

591 alunos concluíram a disciplina fex no primeiro semestre, mas reprovaram a arqc e fisical no segundo semestre. Neste último caso, as reprovações são representadas pelo símbolo '~'.

Embora a informação textual torne difícil entender padrões específicos e forneça pouca informação geral, esta estrutura de padrões disponibiliza dados sobre as inter-relações entre os diferentes semestres do curso que poderão ser evidenciados através de uma visualização eficaz.

### 3.2. Solução

Os padrões a visualizar, além de serem em grande número (especialmente considerando conjuntos de menor suporte) contêm um conjunto de características, tais como semestres, disciplinas, sucesso, insucesso, etc.. Adicionalmente, como descrito anteriormente, os padrões refletem uma natureza temporal: parte dos acontecimentos representados têm lugar simultaneamente, e parte ao longo do



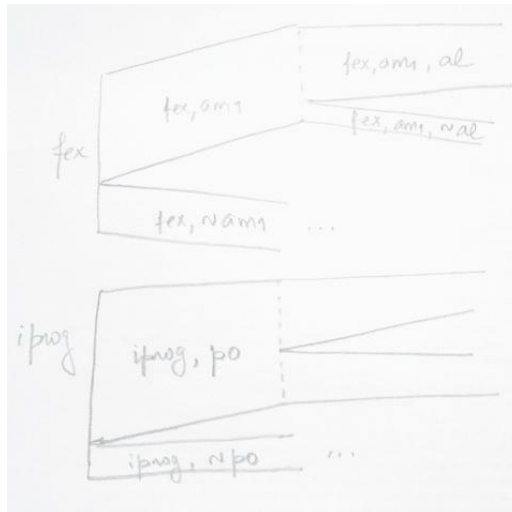


Figura 3.2. Protótipo em papel de uma visualização baseada no Mapa de Minard.

Desta forma, foi implementada uma versão funcional de uma visualização com base nesta abordagem, ilustrada na Figura 3.3.

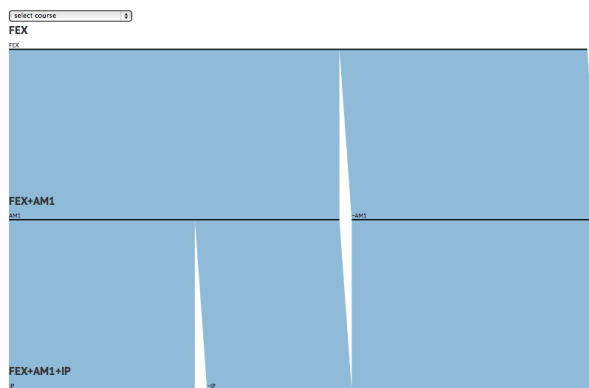


Figura 3.3. Versão funcional de uma visualização baseada no Mapa de Minard com base na abordagem *Parallel Sets* <sup>1</sup>.

No entanto, os primeiros testes com dados reais mostraram que esta solução era insuficiente. De facto, uma das suas maiores vantagens estaria na facilidade de representar ramos convergentes, de padrões que, por caminhos diferentes tivessem levado ao mesmo desfecho. No entanto, essa informação não é passível de ser fielmente retirada dos dados existentes, empobrecendo a visualização. Esta conclusão levou-nos a considerar outras soluções.

Foi feita a exploração de várias abordagens, como a visualização em *Edge Bundling Hierárquico*, *Chord Diagrams* e *Sunburst* (Figura 3.4.). No entanto, todas estas revelaram-se limitadas para a representação das relações temporais de precedência e simultaneidade.

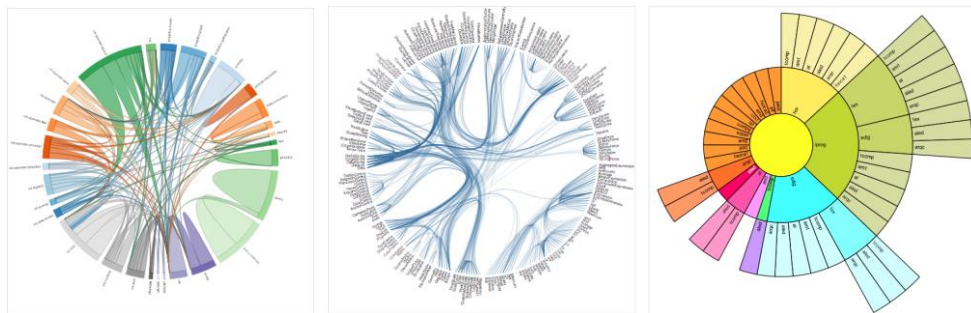


Figura 3.4. *Edge Bundling Hierárquico* (esquerda), *Chord Diagram* (centro) e *Sunburst* (direita)<sup>1</sup>.

Como consequência, optámos por conceber uma nova forma de visualização, que se veio a revelar adequada, que é representada na Figura 3.5.



Figura 3.5. Visualização Multi-Camada (primeira aproximação).

Esta nova abordagem, que denominámos visualização multi-camada, combina informação estática e dinâmica na mesma vista. Cada camada representa um semestre de inscrição dos alunos (e não um semestre dos seus currículos) podendo, portanto, existir disciplinas repetidas entre os vários semestres. Por exemplo, na Figura 3.6. são representados três semestres.



Figura 3.6. Visualização Multi-Camada: representação de semestres e disciplinas.

<sup>1</sup> Baseado em: <http://d3js.org/>

Os nós que representam cada disciplina correspondem a círculos, que se encontram, se necessário, divididos em semicírculos, correspondentes à passagem e reprovação nessa disciplina. O raio dos (semi)círculos é proporcional ao número de alunos envolvidos nos vários padrões referentes a essa disciplina. Este valor tem em conta, para fins de normalização, os valores máximo e mínimo de alunos que aprovaram ou reprovaram uma disciplina, estando sempre situado entre estes limites superior e inferior. Estes limites são mapeados num intervalo de pixéis, para fins de visualização. O valor mínimo deste intervalo corresponde a um valor fixo de quatro pixéis de raio (verificado por experimentação que é o valor mínimo de raio para um círculo de forma a que este seja passível de ser selecionado com algum conforto por parte do utilizador). O valor máximo é calculado dinamicamente e tem em conta a largura da visualização e o número de disciplinas a representar. Assim, o valor do raio de uma determinada disciplina é um mapeamento linear do valor de alunos que se inscreveram à cadeira para este intervalo de valores visuais. Por exemplo, na Figura 3.7. são representadas quatro disciplinas, em que `plf` não dispõe de suporte para o insucesso, enquanto a cadeira de `iar` corresponde a taxas mais baixas de reprovação quando comparada com `am3` e `pest`.



Figura 3.7. Visualização Multi-Camada: representação de sucesso/insucesso.

### 3.2.2. Conectores Visuais para Representação de Percursos Curriculares Típicos

A visualização multi-camada permite a seleção das várias disciplinas realizadas pelos alunos num dado instante de tempo (horizontal), dando origem a restrições que levam à visualização imediata dos vários percursos possíveis seguidos por esses alunos ao longo do curso (vertical). Essa seleção é feita através da passagem do cursor do rato sobre uma das disciplinas, resultando na representação visual dos vários padrões que lhe dizem respeito, sob a forma de arcos que unem as várias disciplinas em cada padrão [Gama14b][Gama14d]. Estes conectores partem do centro do (semi)círculo selecionado e terminam nos (semi)círculos correspondentes às cadeiras com as quais a disciplina atualmente selecionada se inter-relaciona, como ilustrado na Figura 3.8.

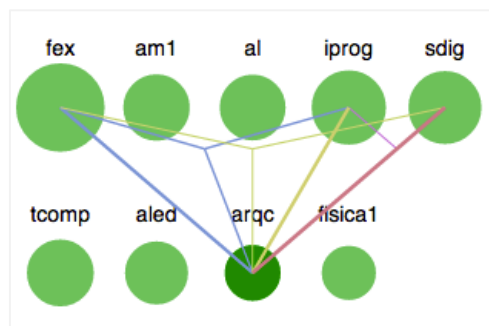


Figura 3.8. Visualização Multi-Camada: representação de percursos curriculares típicos (padrões).

A espessura dos traços é proporcional aos conjuntos de suporte dos padrões e, portanto, ao número de alunos envolvidos. No exemplo da Figura 3.9., o padrão (tcomp, aled) é mais significativo do que o padrão (tcomp, arqc), uma vez que corresponde a um maior número de alunos. O cálculo da espessura dos conectores utiliza uma normalização similar à usada no cálculo do raio dos círculos. Isto é, tem em conta o número máximo e mínimo de alunos presentes nos padrões para restringir o valor da espessura do arco. O cálculo de cada arco, sendo proporcional ao número de alunos é, assim, normalizado. A espessura máxima dos arcos foi escolhida de modo a não obscurecer totalmente as zonas da visualização sobre a qual passam, tanto quanto possível. Assim, a espessura dos conectores é mapeada linearmente, tendo em conta o valor mínimo de um pixel e o valor máximo do diâmetro do círculo mais pequeno representado.



Figura 3.9. Visualização Multi-Camada: representação de padrões mais ou menos significativos.

No entanto, em alguns casos o número de arcos tornava a visualização ilegível, ocultando o nome das disciplinas, como ilustrado na Figura 3.10 (esquerda). Procurámos solucionar esse problema com a implementação de conectores em curvas de Bézier cúbicas, como ilustrado na Figura 3.10 (direita). Estas curvas, partindo do centro da cadeira seleccionada, direcionam-se para o centro das cadeiras inter-relacionadas. Os pontos de controlo localizam-se a uma distância vertical e/ou horizontal da cadeira seleccionada e cada uma das cadeiras inter-relacionadas, como ilustrado na Figura 3.11.

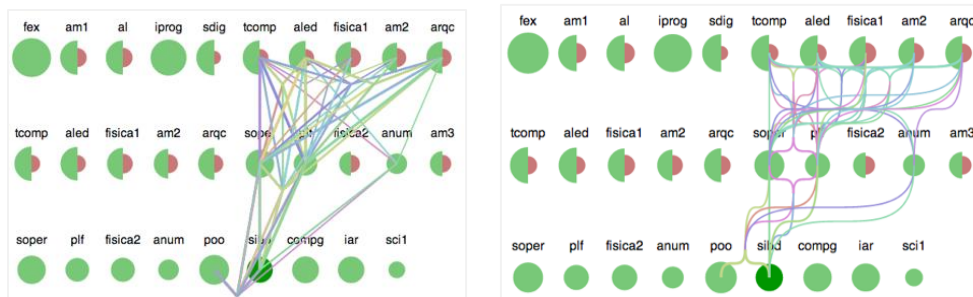


Figura 3.10. Visualização Multi-Camada: conectores lineares (esquerda) e curvos (direita).

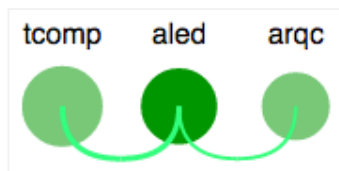


Figura 3.11. Conectores para a cadeira aled: curvas de Bézier com pontos de controlo situados na direção vertical do centro do círculo da cadeira seleccionada e do centro das cadeiras inter-relacionadas. As coordenadas verticais destes pontos de controlo correspondem a metade da distância entre as camadas (semestres).

Implementámos ainda mecanismos de agrupamento (*bundling*) dos arcos: vários conectores que ligam iguais grupos de disciplinas são agrupados num único traço, como representado na Figura 3.12. A espessura do traço resultante corresponde ao valor normalizado para a visualização da soma do número de alunos correspondentes a cada um dos padrões. Esta normalização tem em conta o intervalo de valores considerados para os conectores, ou seja, o valor mínimo de um pixel e o valor máximo do diâmetro do círculo mais pequeno representado.

Testes com utilizadores demonstraram melhorias significativas no desempenho de tarefas de exploração de padrões com a utilização de conectores em curva de Bézier com mecanismos de *bundling* [Gama14e].

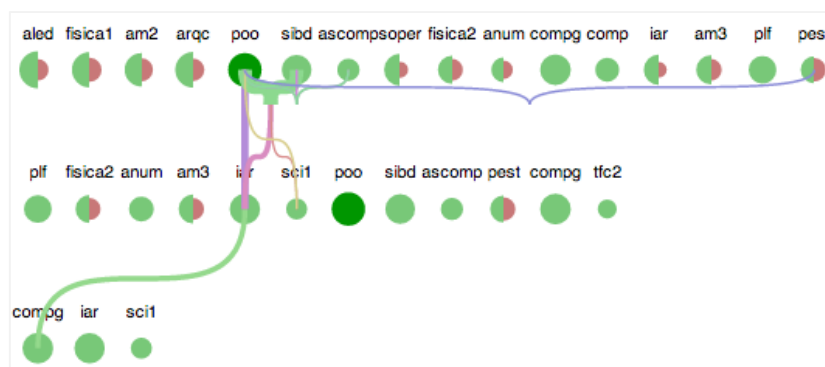


Figura 3.12. Agrupamento (*bundling*) dos conectores curvos.

### 3.2.3. Utilização da Cor como Veículo de Informação

Outro dos problemas com que nos deparámos ao considerar a grande quantidade de arcos, é o facto de estes conterem muita informação. Dado que são representados os padrões associados ao sucesso e à aprovação de uma mesma disciplina, é importante que estes sejam distinguidos.

Decidimos usar a cor para criar esta distinção, atribuindo cores diferentes para arcos referentes a aprovações e reprovações, bem como para oferecer, de forma redundante e complementar, informação sobre o número de alunos envolvidos. Assim, a solução passou por utilizar princípios de mistura de cor para tornar clara a representatividade dos padrões. Em particular, foram utilizados códigos de cor convencionais da sociedade ocidental [Ware12] para representar sucesso e insucesso: cores *quentes* para reprovações e cores *frias* para aprovações.

De forma a definir um conjunto de cores a utilizar, fizemos um estudo com utilizadores que nos permitiu aferir o espaço de cor mais natural para a interpolação. O espaço em mais facilmente os utilizadores conseguiram perceber as cores originais que, quando misturadas, dão origem a uma segunda cor, foi o CIE-LCh, pelas suas características perceptualmente uniformes de cor [Gama14a] [Gama14c]. Do mesmo estudo concluímos que determinados pares de cores produzem melhores resultados perceptuais quando misturados: (vermelho, amarelo) e (azul, verde).



Investigámos, ainda, até que ponto o olho humano consegue distinguir as proporções relativas das cores originais na mistura de cor [Gama14g]. A conjugação dos resultados desta sequência de estudos permitiu decidir como associar cor aos conectores de sucesso e insucesso. Os pares (vermelho, amarelo) e (azul, verde) associados a cores *quentes* e *frias*, respetivamente, foram, assim, mapeados nos conectores de reprovações e aprovações, como ilustrado na Figura 3.13. Assim, quanto maior o número e alunos num padrão correspondente a uma reprovação, mais a cor do conector se aproximará do vermelho (mais *quente*). Por outro lado, quanto maior o número de alunos numa relação correspondente à aprovação de uma disciplina, mais a cor do conector se aproximará do azul (mais *fria*).



Figura 3.13. Códigos de cor para sucesso (esquerda) e insucesso (direita)

### 3.2.4. Visualização Global de Inter-relações: Vista Multi-Matricial

Apesar da visualização acima descrita permitir explorar facilmente as vertentes estáticas e dinâmica dos padrões, não era a mais adequada para uma visualização de conjunto de quais as disciplinas envolvidas em mais padrões (e, portanto, mais críticas no currículo). Tentou-se então colmatar esta lacuna. Desta forma, começou por criar-se um conjunto de protótipos não funcionais em papel para tentar chegar a uma visualização que permitisse representar esta informação, representados na Figura 3.14.

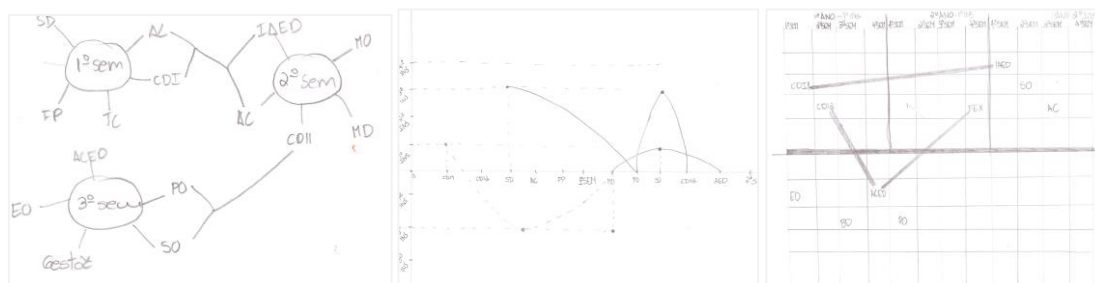


Figura 3.14. Protótipos em papel para a representação global de inter-relações.

Após a discussão dos protótipos, decidimos utilizar uma abordagem multi-matricial, ilustrada na Figura 3.15 [Jordao14b].

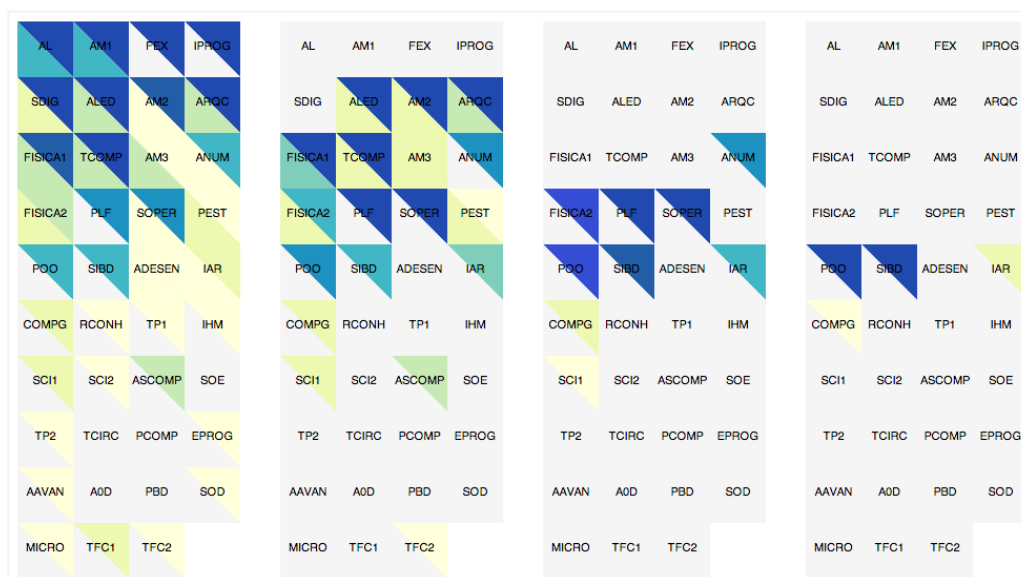


Figura 3.15. Visualização multi-matricial.

Nesta, são apresentadas lado a lado diversas matrizes, correspondendo aos semestres letivos presentes nos padrões. Cada célula da matriz corresponde a uma disciplina, dividida diagonalmente em dois triângulos, em que o triângulo superior representa a aprovação a essa cadeira e o inferior a reprovação. A escolha da divisão das células em triângulos ao invés da utilização de retângulos prendeu-se com a continuidade visual que é mantida pelos primeiros (garantida pelo contacto de pelo menos um ponto entre dois triângulos de células adjacentes), em contraste com a sensação de descontinuidade visual que seria provocada por retângulos. Cada triângulo encontra-se colorido com uma intensidade proporcional ao número de padrões em que se encontra uma referência a passar (ou reprovar) a essa disciplina. Assim é possível, à semelhança dos *heatmaps*, ter uma visão imediata das disciplinas mais importantes (as que figuram em mais padrões). O código de cores utilizado na visualização é consistente com o código utilizado no mecanismo de visualização multi-camada.

Após estar definida a estrutura da visualização multi-matricial, e mesmo antes da implementação estar totalmente concluída, verificou-se uma necessidade de realizar um pré-processamento dos dados para permitir uma resposta em tempo real por parte do sistema. Desta forma, quando a visualização é carregada pela primeira vez, antes de qualquer interação, os padrões são processados, iterados e organizados em estruturas de dados criadas para esse fim, de forma a que aceder à informação sobre determinada disciplina seja fácil e rápido. Esta alteração melhorou substancialmente os tempos de resposta do sistema, melhorando a fluidez da visualização.

### 3.2.5. Exploração Integrada de Padrões

Os dois mecanismos, correspondentes aos dois tipos de vistas (multi-camada e multi-matricial), encontram-se integrados e funcionam de forma coordenada como uma única visualização [Gama14f, Jordao14a]. A exploração de informação relativa a uma determinada disciplina dá-se através da seleção da mesma, ou seja, passando o cursor do rato sobre esta. Uma vez que ambas as vistas (multi-camada e multi-matricial) refletem todas as disciplinas curriculares por semestre, a seleção pode ser feita em qualquer uma das vistas, selecionando um (semi) círculo (vista multi-camada) ou um triângulo (vista multi-matriz). Como resultado, a visualização multi-camada mostra os conectores visuais representativos dos padrões de inter-relação entre cadeiras e a visualização multi-matriz evidencia essas relações através do destaque (*highlight*) das disciplinas envolvidas com a selecionada, enfatizando as dependências (Figura 3.16).

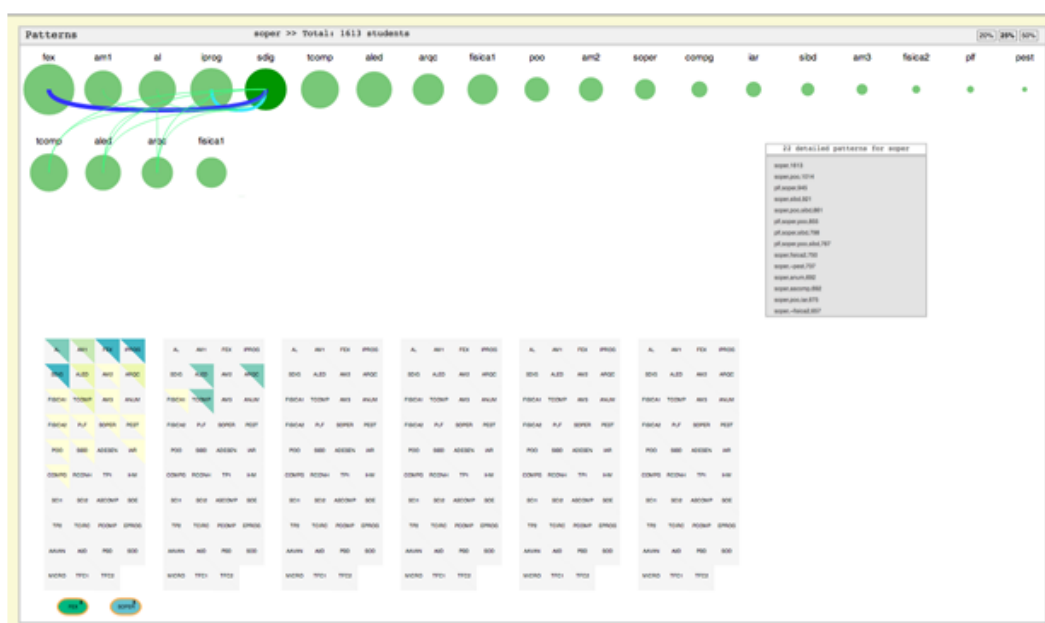


Figura 3.16. Visualização coordenada das vistas multi-camada e multi-matricial.

Pressionando o rato sobre uma disciplina fixa a informação anteriormente mostrada, permitindo comparação simultânea de conjuntos de padrões em simultâneo, como ilustrado na Figura 3.17.



Figura 3.17. Comparação simultânea de conjuntos de padrões.

Pressões subsequentes nas células correspondentes a outras disciplinas já destacadas na vista multi-matricial restringe ainda mais a informação a salientar, permitindo criar filtros para explorar a informação apresentada, que podem ser adicionados e removidos em qualquer fase da interação, como representado na Figura 3.18. Adicionalmente, uma determinada referência pode ter 2 conjuntos de restrições, cada um associado a uma cor diferente. Desta forma, células que pertençam a ambas as restrições vão ser coloridas através do *blending* das cores originais. Ainda, o facto de ser possível remover as restrições por qualquer ordem permite a exploração de vários cenários.

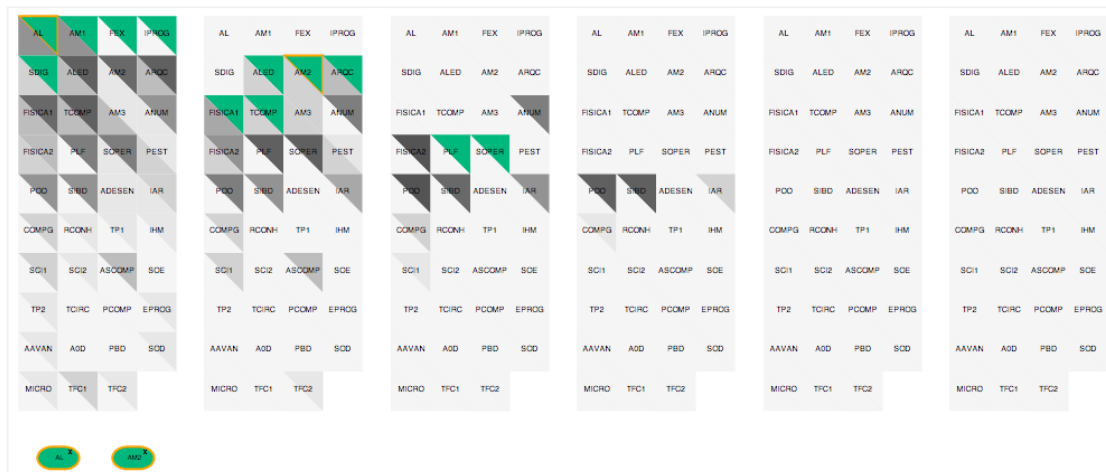


Figura 3.18. Aplicação de filtros (pormenor da visualização multi-matricial).

O protótipo do sistema de visualização está disponível em <http://danielgoncalves.info/research/educare/eduvis/eduvis.html>.



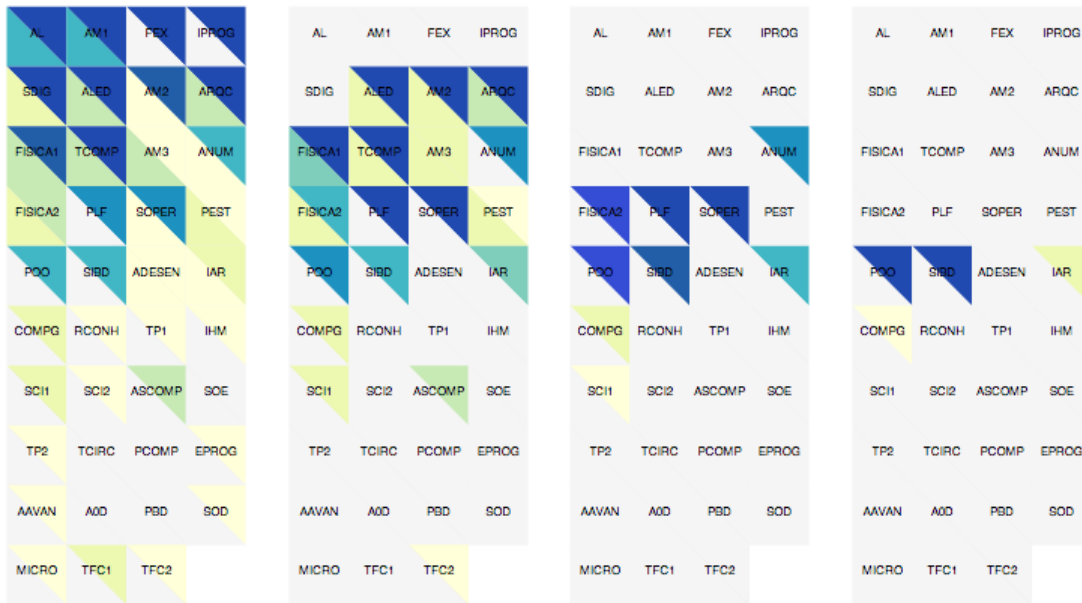


Figura 3.20. Disciplinas com mais padrões associados: vista multi-matricial (pormenor).

De facto, ao seleccionarmos a primeira disciplina (visualizando as suas relações, como ilustrado na Figura 3.21), na vista multi-matricial são destacadas as dependências da mesma (Figura 3.22).

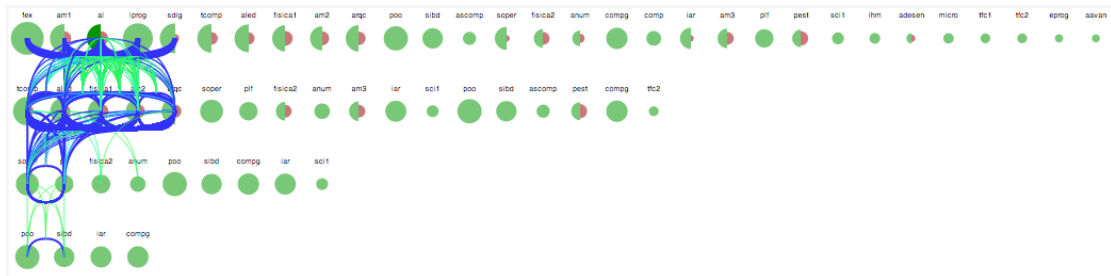


Figura 3.21. Vista multi-camada: padrões associados à disciplina a1.

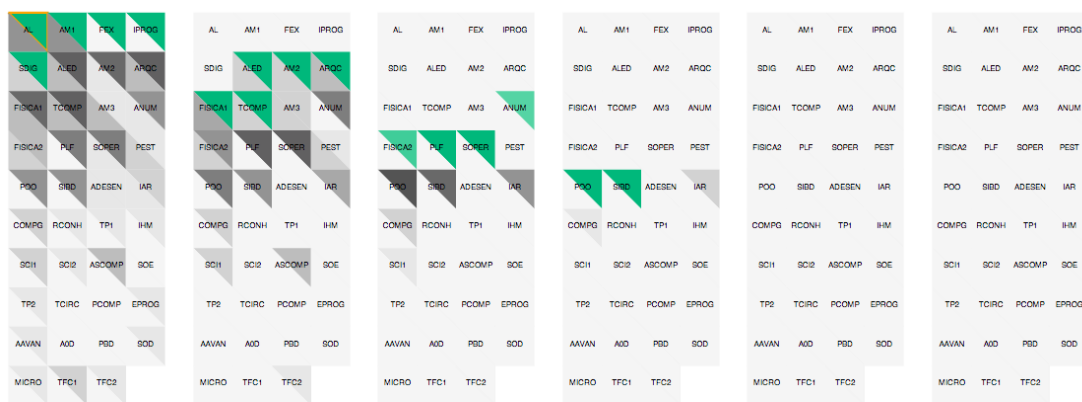


Figura 3.22. Vista multi-matricial: dependências de a1.

Se adicionarmos a restrição am1, criando filtros que incluem as duas disciplinas com mais padrões na vista multi-matricial. São refletidas as alterações na visualização e a vista multi-matricial evidencia as dependências criadas por estas cadeiras face às restantes (Figura 3.23).

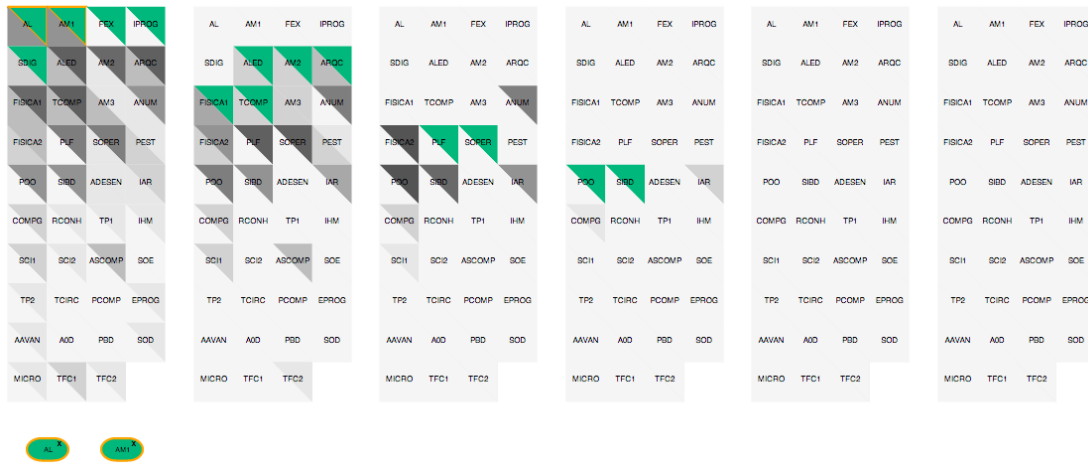


Figura 3.23. Vista multi-matricial: dependências de al e am1.

Se adicionarmos uma terceira restrição, por exemplo *fisical*, verificamos que o número de outras disciplinas condicionado à aprovação nestas três cadeiras é menor face à restrição anterior. No entanto, este número ainda é significativo, mostrando a sua importância no currículo (Figura 3.24).

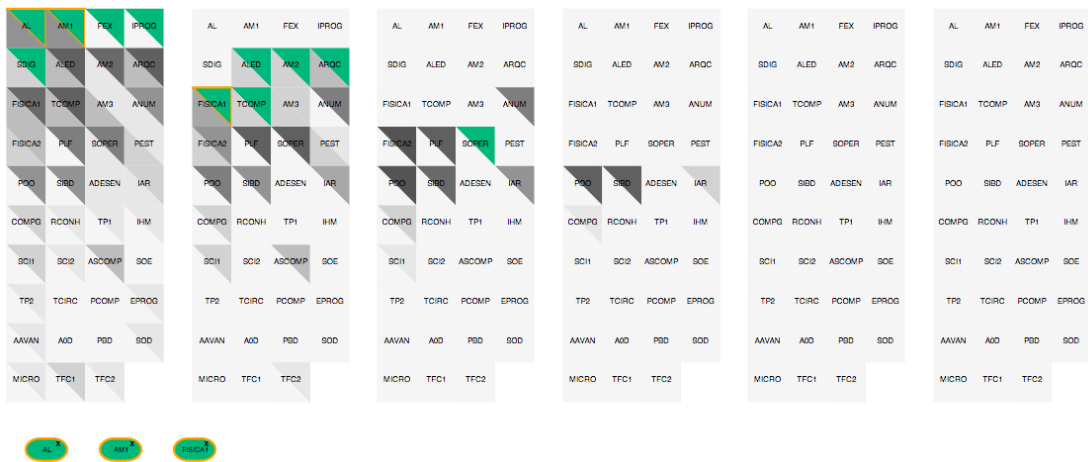


Figura 3.24. Vista multi-matricial: dependências de al, am1 e fisical.

Se continuarmos a adicionar restrições, no entanto, verificamos que o número de dependências vai decrescendo, até conter apenas as disciplinas selecionadas. Podemos remover filtros e acrescentar outros em qualquer altura. Por exemplo, se removermos os filtros am1 e fisical e adicionarmos o filtro fex, o conjunto de restrições será logicamente diferente (Figura 3.25).

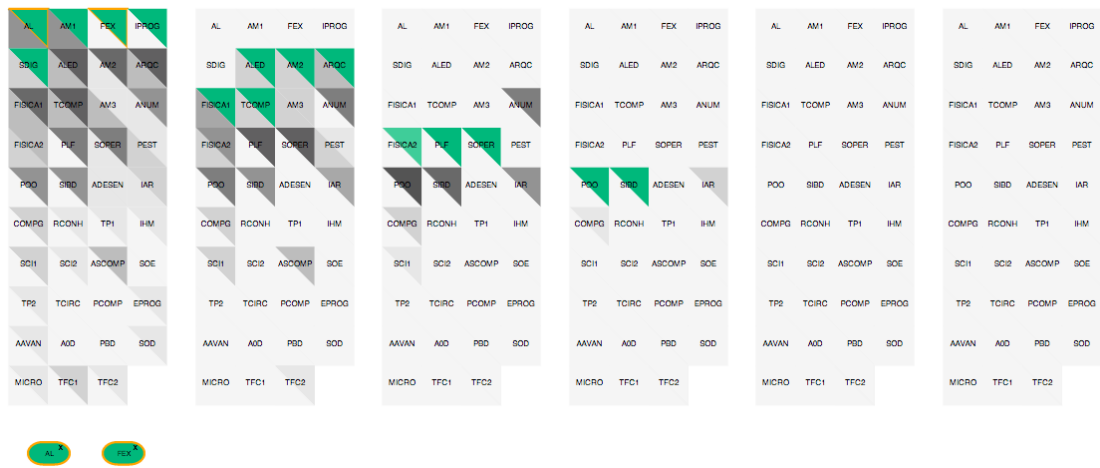


Figura 3.25. Vista multi-matricial: dependências de *al* e *fex*.

Neste caso podemos, por exemplo, verificar que a disciplina *fisica2*, embora dependa de *fex*, não depende necessariamente de *am1*. A conjugação dos mecanismos disponibilizados pela visualização criada permitem, de facto, não só ter uma visão geral dos dados curriculares, mas também analisar a informação disponível de uma forma simples e eficaz, permitindo seleccionar e analisar detalhadamente informação relevante para o contexto.



## 4. Avaliação

Foi realizado um estudo com utilizadores para avaliar a solução criada para a visualização de padrões educacionais. Tendo em conta os objetivos e contexto, pretende-se averiguar, em relação à solução final de visualização:

1. A eficácia e eficiência;
2. A usabilidade, em geral, e a capacidade de aprendizagem, em particular;
3. O grau de satisfação proporcionado no desempenho de tarefas relevantes para o contexto.

### 4.1. Tarefas representativas

Foi escolhido um conjunto de tarefas para testar a eficácia e eficiência da solução final, EduVis, para cumprir os objetivos propostos no que diz respeito à visualização da informação resultante dos processos de EDM. Assim, focam-se na exploração dos vários mecanismos da visualização por forma a obter toda a informação disponível e relevante no presente contexto.

Tendo, assim, em conta a informação representada, foi criado um conjunto de perguntas para avaliar a visualização como um todo e os diversos mecanismos criados, bem como a integração e complementaridade dos mesmos:

1. Quantos semestres são representados?
2. A nível geral, quais as duas cadeiras com mais alunos?
3. Qual o conjunto de cadeiras envolvidas em mais padrões positivos no 2º semestre?
4. Quais as cadeiras que estão relacionadas com a cadeira de *iar* no 4º semestre?
5. Considerando os alunos que fizeram *am1* e *fex* no 1º semestre, quais são as outras cadeiras em que também obtiveram sucesso no 2º semestre?
6. Quais são as cadeiras comuns a quem fez *sdiag* no 1º semestre e obteve aprovação a *aled* no 2º semestre?
7. Considerando a cadeira *poo* no 4º semestre, embora seja a que tem mais padrões associados, é esta a cadeira com mais alunos nesse semestre?

A resposta às perguntas 1, 2 e 3 deverá ser inferida através de uma rápida análise visual da visualização, permitindo uma resposta imediata, sem necessidade de exploração. A questão 1 pode ser respondida utilizando qualquer um dos mecanismos de visualização: o número de camadas na visualização multi-camada ou o número de matrizes coloridas na visualização multi-matricial. A questão 2 pode ser respondida observando, na visualização multi-camada, quais os círculos de maior tamanho e a resposta à questão 3, embora possível de responder na visualização multi-camada, exige exploração, sendo imediata na visualização multi-matricial. É, assim, obtida através da análise da segunda matriz, correspondendo às disciplinas cujo triângulo superior tem maior brilho.

As questões seguintes requerem alguns graus de interação. Na questão 4, para descobrir as relações da cadeira de *iar*, é necessário deslocar o cursor sobre o círculo

da disciplina (visualização multi-camada) ou selecionar o triângulo correspondente à mesma (visualização multi-matricial). 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 selecionadas e consegue observar de imediato os subconjuntos de cada cadeira por estes estarem destacados com uma cor diferente da seleção, ou pode alternativamente clicar no triângulo da cadeira e ficam apenas visíveis as disciplinas que fazem parte do seu subconjunto.

Por fim, a última questão obriga a que o inquirido conjugue informação de ambos os mecanismos, de forma a perceber que uma cadeira, apesar de estar envolvida em muitos padrões (informação extraída do mecanismo multi-matricial), pode não ter muitos alunos inscritos (informação proveniente do mecanismo multi-camada).

#### **4.1. Protocolo de teste**

Para avaliar os participantes foi elaborado um questionário com as tarefas supracitadas. Os testes foram realizados individualmente junto de 20 pessoas, sem a presença de outros inquiridos e com uma duração aproximada de 15 minutos, num ambiente controlado, com condições de luminosidade constante.

Antes de dar início à avaliação, foi dada aos inquiridos uma introdução global, descrevendo o contexto e objetivos do estudo, assim como a visualização, detalhando brevemente os vários artefactos visuais. De seguida foi feita uma pequena demonstração das funcionalidades, sendo posteriormente dada liberdade de exploração a cada utilizador até que o mesmo se sentisse confortável para realizar as tarefas propostas.

Findo este processo inicial, foi entregue aos participantes um questionário com o conjunto de perguntas supracitado, mencionando que a necessidade de que fossem respondidas mediante a realização de tarefas na visualização.

Foi feita uma observação direta das ações 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<sup>2</sup>) 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. Esta última parte do questionário pretende revelar o grau de compreensão dos seguintes aspetos:

1. Número de semestres representados (noção da existência de um espaço temporal, e localização no mesmo).
2. Cadeiras com maior número de aprovações e reprovações;

---

<sup>2</sup> <http://www.measuringusability.com/sus.php>

3. Disciplinas envolvidas em mais padrões, ou seja, com mais inter-relações com outras cadeiras (identificação das cadeiras chave para o sucesso académico dos estudantes);
4. Comparação de padrões de disciplinas distintas.

### 4.3. Resultados

Dos 20 inquiridos, mais de metade são do género masculino. A maioria situa-se no intervalo entre os 25-34 anos de idade e, em termos de escolaridade, 80%, frequenta ou concluiu o ensino superior, como resumido nas tabelas 4.1, 4.2 e 4.3.

Tabela 4.1. Género dos inquiridos

Género	
Feminino	5 (15%)
Masculino	15 (75%)

Tabela 4.2. Idade dos inquiridos

Idade	
18-24	3 (15%)
25-34	10 (50%)
35-44	3 (15%)
45-54	2 (10%)
55-64	1 (5%)
+65	1(5%)

Tabela 4.3. Grau de escolaridade dos inquiridos

Escolaridade	
Básico	2 (10%)
Secundário	2 (10%)
Superior	16 (80%)

A tabela 4.4 apresenta o tempo e número de erros médio de conclusão das tarefas.

Tabela 4.4. Tempo médio e número de erros médio por tarefa

Tarefa	Tempo (s)	N.º de erros
1	3.20	0.35
2	17.60	0.10
3	28.30	0.55
4	23.95	0.30
5	27.70	0.30
6	49.45	0.75
7	45.70	0.30

Tendo em conta o tempo e o número de erros médio relativos ao desempenho de cada uma das tarefas, os valores medidos não permitem uma distinção clara entre as tarefas imediatas (1 a 3) e as tarefas de exploração (4 a 7), com possível exceção das tarefas 6 e 7. Estas duas tarefas que assumem diferenças aparentemente consideráveis a nível de tempo, enquanto a tarefa 6 parece levar a um maior número de erros.

Para aprofundamento destes resultados, foi realizada uma análise estatística. A utilização de um teste *Shapiro-Wilk* levou à verificação de evidências 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. Efetivamente, 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 ( $W = 0$ ,  $cv = 3$ ,  $p < 0.05$ ) 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, calcularam-se os coeficientes de *Pearson*, encontrando uma correlação, embora relativamente fraca, na tarefa 2 ( $r = 0.45$ ,  $p < 0.05$ ) e na tarefa 7 ( $0.49$ ,  $p < 0.05$ ), o que não permite generalizar uma correlação entre um maior tempo para desempenho da tarefa e um maior número de erros. Assim, é possível concluir 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 [Sauro11], foi de 79.47 pontos, mostrando resultados bastante elevados no que diz respeito à usabilidade e capacidade de aprendizagem do sistema. A utilização do mesmo método para calcular as respostas de satisfação contextuais resultou numa pontuação de 92.11, mostrando um grau bastante elevado de satisfação no desempenho das tarefas.

O objetivo estabelecido para este trabalho foi tornar evidentes padrões educacionais. Particularmente, definiu-se relevante destacar o sucesso e insucesso no percurso dos alunos, fornecendo mecanismos de comparação de disciplinas e análise de inter-relações entre as mesmas. Neste estudo com utilizadores foi criada uma lista de tarefas para avaliar a solução criada no que diz respeito ao cumprimento destes objetivos, em termos de eficácia, eficiência, usabilidade e satisfação. Os resultados destes testes validaram a relevância da visualização criada, demonstrando o sucesso da solução quanto ao cumprimento dos objetivos propostos.

## 5. Conclusões

O número de inscrições no ensino tem sofrido um aumento ao longo dos anos. Como consequência, tem vindo a surgir uma grande quantidade de informação em bruto correspondente aos processos educativos. Se analisada efetivamente, esta informação poderá fornecer conhecimento adicional aos intervenientes dos processos de ensino, proporcionando uma forma de diagnosticar e corrigir problemas. Os processos de *data mining* educacional têm vindo a ganhar popularidade, permitindo analisar automaticamente uma grande quantidade de informação neste contexto. No entanto, o formato complexo da informação resultante torna difícil a análise e compreensão da informação, não permitindo tirar total partido da mesma na avaliação e diagnóstico dos processos de ensino.

A visualização de dados tem vindo a ganhar terreno como uma forma de aliviar a carga cognitiva associada à interpretação da informação. Se usada efetivamente, poderá ajudar a descobrir padrões que de outra forma não seriam evidentes, particularmente tendências gerais que seriam impossíveis de obter através de dados textuais.

Este trabalho surgiu no contexto da visualização de dados obtidos através da aplicação de processos de EDM a um conjunto de informação curricular do curso de Engenharia Informática e de Computadores do Instituto Superior Técnico durante 10 anos. Como tal, propusemo-nos utilizar técnicas de visualização de informação para possibilitar a análise dos padrões educacionais existentes.

Como tal, utilizando um método de desenho iterativo e incremental, desenvolvemos uma visualização que integra dois principais mecanismos para representação de dados curriculares orientados por semestre curricular: uma vista multi-camada e uma vista multi-matricial. Estas duas ferramentas funcionam de forma integrada e complementar, permitindo à visualização refletir as disciplinas mais críticas e os padrões educacionais mais relevantes, com foco na representação do sucesso e insucesso. A visualização, além de tornar alguma informação imediatamente perceptível, fornece mecanismos interativos que permitem detalhar a análise da informação, como a comparação simultânea e a filtragem.

Um estudo final com utilizadores permitiu avaliar a solução criada. Estes testes mostraram que a visualização desenvolvida é, de facto, compreensível e útil face aos objetivos que nos propunhamos atingir. O sucesso, não só em termos de eficácia e eficiência, mas também a nível de usabilidade e satisfação, mostrou o potencial da solução criada para visualizar de forma relevante informação educacional.

## Referências

- [Agrawal95] R. Agrawal, R. Srikant. Mining sequential patterns. IEEE International Conference on Data Engineering, 1995.
- [Keim02] D. Keim. Information visualization and visual data mining. IEEE Transactions on Visualization and Computer Graphics, 2002.
- [Mazza05] R. Mazza, V. Dimitrova. Generation of graphical representations of student tracking data in course management systems. International Conference on Information Visualisation, 2005.
- [Wortman07] D. Wortman, P. Rheingans. Visualizing trends in student performance across computer science courses. 38th SIGCSE Technical Symposium on Computer Science Education, 2007.
- [Antunes08] C. Antunes. Acquiring background knowledge for intelligent tutoring systems. International Conference on Educational Data Mining, 2008.
- [Xiaoya09] G. Xiaoya, L. Kan, L. Ping. Visual analysis of college students' scores in english test. 4th International Conference on Computer Science Education, 2009.
- [Gao10] Y. Gao. A study on mass higher education in the world-based on comparative perspectives. International Conference on Education and Management Technology, 2010.
- [Iliinsky11] N. Iliinsky, J. Steele. Designing Data Visualizations, O'Reilly, 2011.
- [Sauro11] J. Sauro. A Practical Guide to the System Usability Scale: Background, Benchmarks and Best Practices, CreateSpace, 2011.
- [Trimm12] D. Trimm, P. Rheingans, M. desJardins. Visualizing student histories using clustering and composition. IEEE Transactions on Visualization and Computer Graphics, 2012.
- [Ware12] C. Ware. Information Visualization: Perception for Design, Elsevier, 2012.
- [Kay13] J. Kay, P. Reimann, E. Diebold, B. Kummerfeld. MOOCs: So many learners, so much potential ... IEEE Intelligent Systems, 2013.
- [Xiaohuan13] W. Xiaohuan, Y. Guodong, W. Huan, H. Wei. Visual exploration for time series data using multivariate analysis method. 8th International Conference on Computer Science Education, 2013.
- [Gama14a] S. Gama, D. Gonçalves, Guidelines for using color blending in data visualization, AVI 2014, International Working Conference on Advanced Visual Interfaces, 2014.

**[Gama14b]** S. Gama, D. Gonçalves, Multi-Level Visualization of Interrelated Data Entities, AVI 2014, International Working Conference on Advanced Visual Interfaces, 2014.

**[Gama14c]** S. Gama, D. Gonçalves, Studying Color Blending Perception for Data Visualization, Eurovis 2014 - Eurographics Conference on Visualization, 2014.

**[Gama14d]** S. Gama, D. Gonçalves, Visualizing Educational Datamining Patterns, Eurovis 2014 - Eurographics Conference on Visualization, 2014.

**[Gama14e]** S. Gama, D. Gonçalves, Visualizing Large Quantities of Educational Datamining Information, iV2014 - 18th International Conference Information Visualisation, 2014.

**[Gama14f]** S. Gama, V. Jordão, D. Gonçalves, EduVis: Visualização Interativa de Dados Educacionais, EPCG, Encontro Português de Computação Gráfica, 2014.

**[Gama14g]** S. Gama, D. Gonçalves, Studying the perception of color components' relative amounts in blended colors, NordiCHI 2014, 8thth Nordic Conference on Human-Computer Interaction, 2014.

**[Jordao14a]** V. Jordão, S. Gama, D. Gonçalves, EduVis: Visualizing Educational Information, NordiCHI 2014, 8th<sup>th</sup> Nordic Conference on Human-Computer Interaction, 2014.

**[Jordao14b]** V. Jordão, EduVis: Visualização Interativa de Dados Educacionais. Tese de Mestrado. Instituto Superior Técnico, 2014.