

Visualização de dados relativos ao MIEET e comparação com cursos congêneres

Rafael Sá (104552) e Luís Laranjeira (81526)

Visualização de Informação, 2020 (Mestrado em Engenharia Informática, Universidade de Aveiro)

Resumo

O presente relatório expõe as visualizações desenvolvidas para apresentar os dados relativos às candidaturas ao concurso nacional de acesso ao ensino superior para o Mestrado Integrado em Engenharia Eletrónica e Telecomunicações da Universidade de Aveiro e para outros cursos congêneres, de forma a que estas possam ser utilizadas pelos professores do Departamento de Eletrónica, Telecomunicações e Informática.

Para as várias visualizações são apresentados os protótipos de baixa fidelidade, os protótipos finais, os resultados das avaliações e testes de usabilidade realizados, e ainda os principais desafios da implementação das visualizações. As soluções de visualização propostas consistem em dois *line charts*, um *grouped bar chart*, um *radar chart*, e um *chord diagram*.

Motivação e Objetivos

O principal objetivo do projeto é a criação de visualizações interativas de dados relativos ao Mestrado Integrado em Engenharia Eletrónica e Telecomunicações (MIEET) e cursos congêneres, de forma a que seja possível estabelecer comparações entre eles. Uma primeira abordagem consiste em converter algumas visualizações do relatório anual com informação do acesso ao ensino superior para os cursos do Departamento de Eletrónica, Telecomunicações e Informática (DETI) para visualizações numa aplicação *web* interativa, sugerindo depois novas visualizações para visualizar outra informação, como a preferência dos alunos ou os concorrentes diretos do MIEET.

Utilizadores e Questões

As visualizações foram desenvolvidas para um conjunto de utilizadores alvo, de forma a auxiliar a análise dos dados e ajudar os utilizadores a responder a um conjunto de questões.

Caracterização dos utilizadores e do contexto de utilização

De uma forma geral, os utilizadores serão todos aqueles que tiverem interesse em ver e analisar informação estatística das candidaturas e colocações no MIEET e estabelecer comparações com cursos congêneres. Porém, as visualizações destinam-se especificamente a pessoal docente ligado ao DETI, em particular ao MIEET, como por exemplo o diretor de curso. Os utilizadores poderão recorrer às visualizações durante a análise anual às candidaturas ao MIEET, ou sempre que quiserem consultar informação estatística ou comparar o MIEET com cursos congêneres.

Questões para responder

As principais questões que as visualizações devem ajudar a responder são:

- Q1. Qual a evolução das notas dos últimos colocados no MIEET?
- Q2. Qual a distribuição das notas de candidatura dos colocados no MIEET?
- Q3. Qual a posição do MIEET nas escolhas dos candidatos?
- Q4. Quais os cursos mais escolhidos pelos candidatos consoante a opção da candidatura?
- Q5. Quais as alternativas dos alunos colocados no MIEET?
- Q6. Quais os principais cursos adversários do MIEET?

De uma forma geral, as visualizações vão também ajudar a perceber como se destaca o MIEET em relação aos cursos congêneres.

Dataset

O *dataset* contém informação relativa às candidaturas ao Concurso Nacional de Acesso ao Ensino Superior para o MIEET e para cursos congêneres. Os cursos congêneres foram selecionados por serem da área de Engenharia

Eletrotécnica, dando preferência a cursos com um elevado número de colocados, notas de acesso superiores ou próximas da do MIEET, proximidade geográfica e também a verificação de que historicamente os candidatos ao MIEET também se candidatam a esses cursos. Para os vários cursos é apresentada uma evolução histórica, com dados entre 2007 e 2016.

Os dados podem ser obtidos a partir de informação disponibilizada no sítio da DGES (<https://www.dges.gov.pt/>), porém os dados utilizados neste projeto foram disponibilizados pelo autor do relatório anual de candidaturas de acesso ao ensino superior para os cursos do DETI, o professor Nuno Lau.

Da informação disponibilizada, foram utilizados os dados relativos às instituições e cursos das várias opções da candidatura, resultado e nota da candidatura, e instituição e curso onde o candidato foi colocado, se aplicável.

Soluções de Visualização

Foram propostas inicialmente cinco visualizações, sendo que duas visualizações foram retiradas do relatório anual de candidaturas de acesso ao ensino superior para os cursos do DETI, uma visualização foi uma proposta para melhorar uma outra já existente no relatório, e duas visualizações foram propostas pelos autores do projeto.

As soluções de visualização propostas neste projeto podem ser agrupadas em duas categorias: visualizações em função da nota de acesso e visualizações em função das opções da candidatura.

De forma a analisar os dados relativos às notas de candidatura para os vários cursos, considerou-se que a informação mais pertinente seria a evolução das notas do último colocado e das notas dos colocados no MIEET. Para representar estes dados foram propostas duas visualizações. A primeira visualização apresenta a evolução temporal das notas de candidatura dos colocados no MIEET e a segunda visualização apresenta a evolução temporal das notas do último colocado no MIEET e cursos congéneres.

Em relação aos dados relativos às opções de escolha das candidaturas, considerou-se que seria pertinente analisar informação referente à percentagem da posição do MIEET nas opções das candidaturas, à relação entre a preferência de um curso consoante a opção, e às alternativas dos candidatos ao MIEET e cursos congéneres. Para representar estes dados foram propostas três visualizações. A primeira visualização apresenta a evolução temporal da posição da opção do MIEET nas preferências dos candidatos, a segunda visualização apresenta as preferências dos candidatos ao MIEET e cursos congéneres consoante a opção de escolha e, por último, a

terceira visualização apresenta as colocações e alternativas dos candidatos ao MIEET e cursos congéneres.

Protótipo de baixa fidelidade e feedback dos utilizadores

Na primeira iteração do projeto foram desenvolvidos protótipos de baixa fidelidade, ilustrativos do modelo conceptual da aplicação, recorrendo ao *software* de prototipagem rápida *Balsamiq*.

Notas dos colocados

Para a visualização da evolução temporal das notas de candidatura dos colocados no MIEET optou-se por um *grouped bar chart*, sendo que no eixo dos xx estão representadas as várias notas, no eixo dos yy é representado o número de alunos colocados, e as várias barras representam a informação para os vários anos. O protótipo de baixa fidelidade desenvolvido é apresentado na figura 1.

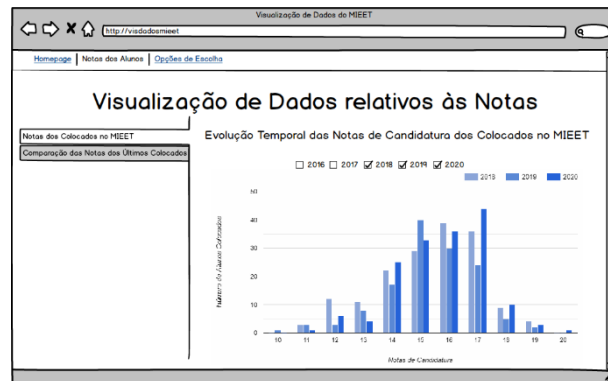


Figura 1: Protótipo de baixa fidelidade para a visualização da evolução temporal das notas dos colocados no MIEET

Esta visualização é uma proposta de melhoria da visualização existente no relatório para estes dados, visto que no teste de usabilidade foi obtido bastante *feedback* a considerar a visualização original confusa e difícil de interpretar. A utilização do *grouped bar chart* é justificada pelo facto de ser um gráfico que permite comparar diferentes grupos de dados [1], neste caso anos. Foi ainda considerada a utilização de um *stacked bar chart*, porém concluiu-se que o *grouped bar chart* era a melhor opção pois, apesar de não ser indicado quando se pretende apresentar muitos grupos de dados [1], permite estabelecer comparações mais facilmente.

Para além de representar esta informação na visualização, foi sugerida a apresentação de métricas, calculadas com base na informação apresentada. Com base nisto, adicionou-se à proposta inicial a apresentação da média dos vários anos para cada nota.

Para esta visualização, a interação proposta inicialmente foi a possibilidade de seleção dos anos para os quais se quer ver os dados, pois com todos

Figura 4: Protótipo de baixa fidelidade para a visualização da relação entre as preferências opção – curso

Considerou-se adequada a utilização de um *radar chart*, pois é uma visualização que permite apresentar dados multivariados na forma de um gráfico bidimensional de três ou mais variáveis quantitativas, e permite identificar *outliers* e semelhanças. Apesar disso, o gráfico possui a desvantagem de ser difícil comparar visualmente os comprimentos dos diferentes raios, uma vez que as distâncias radiais são difíceis de julgar [3].

As interações propostas inicialmente para esta visualização foram a possibilidade de selecionar um ano em específico, e também a possibilidade de filtrar a informação, podendo alterar a visibilidade dos cursos no gráfico. Para além disso, existe a possibilidade de observar uma animação onde é apresentada no gráfico a evolução dos dados ao longo de um intervalo de tempo, sendo que esse intervalo também poderá ser alterado. Durante o teste de usabilidade notou-se alguma dificuldade dos utilizadores na distinção dos cursos apresentados no gráfico, uma vez que as cores ficam sobrepostas. Para solucionar esse problema, foi adicionada à proposta a possibilidade de realçar informação e também uma *tooltip* para indicar qual o curso em questão.

Alternativas dos Candidatos

Para a visualização das colocações e das alternativas dos candidatos ao MIEET e cursos congêneres optou-se por um *chord diagram*. Esta visualização é sugerida pelos autores e é apresentada na figura 5.

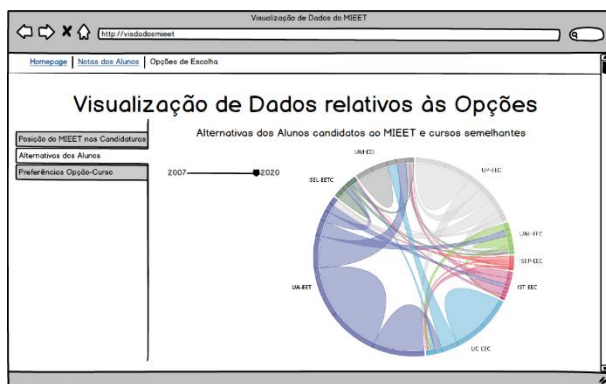


Figura 5: Protótipo de baixa fidelidade para a visualização das colocações e alternativas dos candidatos

Cada curso é representado por um fragmento da parte externa do diagrama, as conexões para o mesmo fragmento de origem representam os alunos colocados no curso, e as conexões para outros fragmentos representam os cursos que eram alternativa dos candidatos.

Considerou-se adequada a utilização do *chord diagram*, uma vez que exibe relações entre entidades e permite comparar semelhanças entre o conjunto de dados. Porém, possui a desvantagem de se tornar difícil de interpretar, algo que pode ser minimizado com a disponibilização de funcionalidades de interação [4].

As interações propostas inicialmente para esta visualização foram a possibilidade de seleção de um ano específico, a possibilidade de realçar apenas as conexões relativas a um curso e a apresentação de uma *tooltip* com os cursos de origem e destino de uma conexão. Através destas interações foi possível simplificar a análise dos dados, uma vez que esta é uma visualização que pode tornar-se difícil de interpretar quando existem muitas conexões representadas. Durante a avaliação da primeira iteração foi sugerida a possibilidade de selecionar apenas os dados relativos a uma opção específica da candidatura, pelo que foi adicionada essa interação à proposta.

Protótipos funcionais

Com base nas visualizações propostas inicialmente, e nas alterações efetuadas com base no *feedback* do teste de usabilidade e da avaliação da primeira iteração, procedeu-se à implementação das visualizações utilizando a ferramenta *d3.js*.

Notas dos colocados

A implementação da visualização da evolução temporal das notas de candidatura dos colocados no MIEET é apresentada na figura 6. Os dados apresentados nesta visualização ajudam os utilizadores a responder à questão Q2 (Qual a distribuição das notas de candidatura dos colocados no MIEET?).

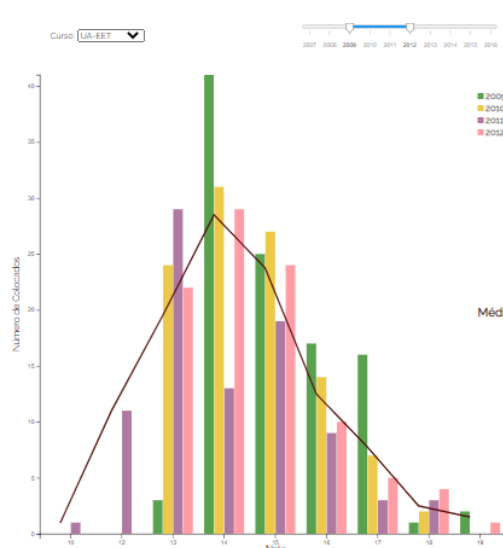


Figura 6: Implementação da visualização da evolução temporal das notas dos colocados no MIEET para responder à questão Q2

A seleção do curso é feita através do *drop-down menu* e a seleção do intervalo temporal é feita através do *slíder* de duas variáveis, ambos apresentados acima do gráfico. Através da legenda, apresentada ao lado do gráfico, é possível mostrar ou ocultar informação do gráfico. A informação relativa à média das notas dos vários anos é representada no gráfico através da linha.

Notas do último colocado

A implementação da visualização da evolução temporal das notas do último colocado no MIEET e cursos congêneres é apresentada na figura 7. Esta visualização ajuda os utilizadores a responder à questão Q1 (Qual a evolução das notas dos últimos colocados no MIEET?).

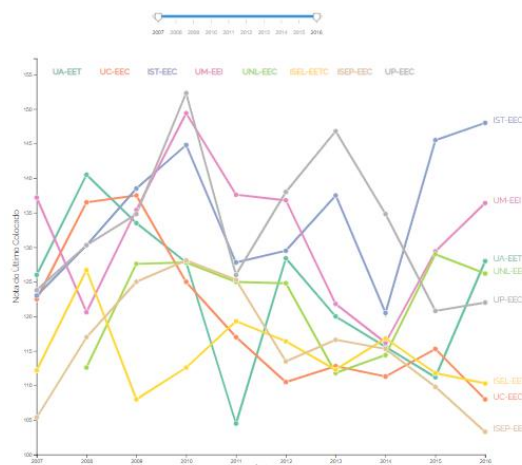


Figura 7: Implementação da visualização da evolução temporal das notas do último colocado para responder à questão Q1

A seleção do intervalo temporal é feita através do *slider* de duas variáveis, e a filtragem dos dados é feita através da legenda, sendo que tanto o *slider* como a legenda são apresentados acima do gráfico. Para além da legenda, é apresentado no final de cada linha o curso correspondente, de forma a facilitar a leitura do gráfico.

Posição do curso nas candidaturas

A implementação da visualização da evolução temporal da posição do MIEET nas preferências dos candidatos é apresentada na figura 8. Esta visualização ajuda os utilizadores a responder à questão Q3 (Qual a posição do MIEET nas escolhas dos candidatos?).

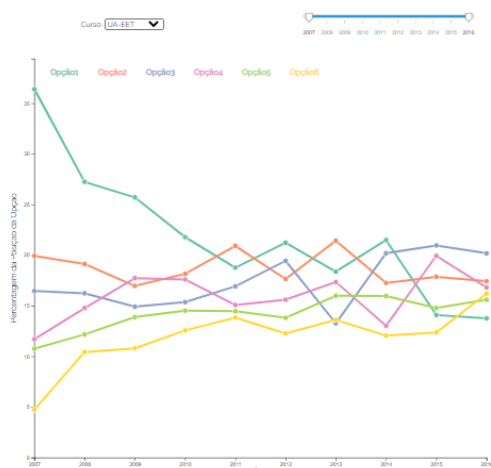


Figura 8: Implementação da visualização da posição de um curso nas candidaturas para responder à questão Q3

A seleção do curso é feita através do *drop-down menu* e a seleção do intervalo temporal é feita através do *slider* de duas variáveis, ambos apresentados acima do gráfico. Através da legenda, apresentada também acima do gráfico, é possível mostrar ou ocultar informação. Para além disso, é apresentada no final de cada linha a opção correspondente.

Preferências Opção – Curso

A implementação da visualização das preferências dos candidatos na escolha do curso consoante a opção é apresentada na figura 9. Esta visualização ajuda os utilizadores a responder à questão Q4 (Quais os cursos mais escolhidos pelos candidatos consoante a opção da candidatura?).

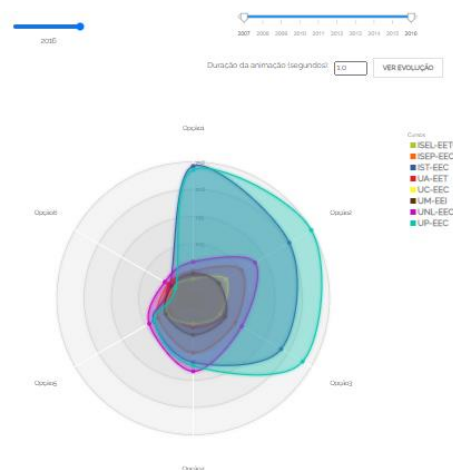


Figura 9: Implementação da visualização das preferências opção - curso para responder à questão Q4

A seleção do ano para o qual se quer visualizar os dados é feita no *slider* de uma variável, apresentado no lado esquerdo acima do gráfico. A filtragem dos dados é feita através da legenda, apresentada ao lado do gráfico. Para definir o intervalo temporal da animação da evolução deve ser utilizado o *slider* de duas variáveis, apresentado no lado direito acima do gráfico, sendo que para iniciar a animação deve ser utilizado o botão apresentado logo abaixo do *slider*.

Alternativas dos Candidatos

A implementação da visualização das colocações e das alternativas dos candidatos é apresentada na figura 10. Esta visualização ajuda os utilizadores a responder às questões Q5 (Quais as alternativas dos alunos colocados no MIEET?) e Q6 (Quais os principais cursos adversários do MIEET?).

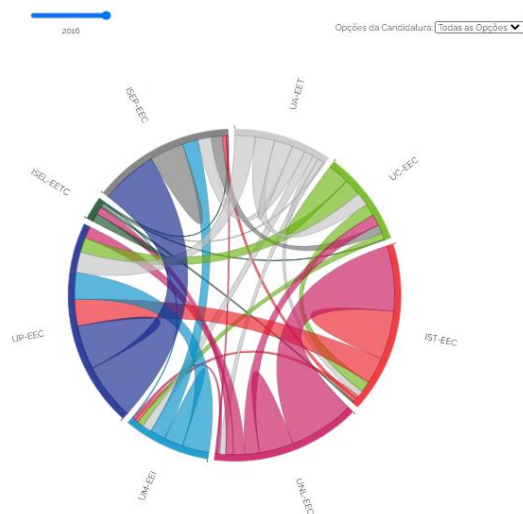


Figura 10: Implementação da visualização das colocações e alternativas dos candidatos para responder às questões Q5 e Q6

A seleção do ano é feita através do *slider* e a seleção das opções que se quer considerar é feita através do *drop-down menu*, ambos apresentados acima do gráfico. Em redor do diagrama é apresentada uma *label* com a indicação do curso ao qual pertence o fragmento.

Desafios de implementação

Para implementar cada uma das visualizações foi utilizado um *template* base, que foi sendo alterado para se adequar às necessidades das visualizações propostas. Os *templates* foram obtidos no sítio da *D3.js Graph Gallery* (<https://www.d3-graph-gallery.com/>).

O principal desafio da implementação foi desenvolver os protótipos das visualizações de uma forma genérica, para que a sua adaptação para os restantes cursos do DETI fosse o mais linear possível. Tendo isso em consideração, as visualizações e a extração dos dados evitam referências explícitas aos cursos analisados neste projeto, existindo apenas referências nas páginas HTML da aplicação. Sendo assim, basta alterar os ficheiros a partir dos quais os dados vão ser extraídos, e referenciar os novos ficheiros de dados no ficheiro *files_list.csv* seguindo o formato já definido. De realçar que os ficheiros de dados necessitam de possuir o cabeçalho na primeira linha. Em relação aos ficheiros HTML, todas as referências explícitas aos cursos analisados neste projeto devem ser alteradas manualmente para os cursos que se quer adaptar a aplicação.

De destacar também o desafio da utilização das funções do D3 para a leitura dos ficheiros de dados. O facto destas funções serem assíncronas obrigou a um aumento da complexidade do código no que diz respeito à extração dos dados e, consequentemente, à criação das visualizações.

Por último, é importante salientar que a implementação dos protótipos não permite que

sejam colocados dados totalmente novos nas visualizações sem que estas sejam refeitas. Esta abordagem mostrou-se necessária para resolver problemas que foram surgindo, como foi o caso da leitura assíncrona de ficheiros. Isto justifica o facto das visualizações serem refeitas aquando da alteração dos dados através dos *sliders*, *drop-down menus*, e da animação da evolução, em vez de ser feita uma transição suave, como era intenção inicial.

Avaliação heurística

Foi efetuada uma avaliação heurística à aplicação, e respetivos protótipos, porém não foi encontrado nenhum problema severo.

A aplicação é consistente ao longo das várias páginas e possui um *design* minimalista. Para além disso, é fornecida ajuda em todas as visualizações, bem como uma introdução aos dados e ao contexto na página inicial. As interações com as visualizações também são intuitivas e consistentes, e deve ser realçado que não é permitida a introdução de *inputs* inválidos por parte do utilizador.

Teste de usabilidade e alterações nos protótipos

Visto que na segunda iteração do projeto os protótipos das visualizações deveriam ser testados com utilizadores tão representativos dos utilizadores alvo quanto possível, com apoio do professor Paulo Dias foi solicitado a três professores do DETI que testassem a aplicação e dessem *feedback*, sendo que não obtida qualquer resposta.

Apesar disso, foram obtidas algumas sugestões do professor Paulo Dias, que também é parte interessada neste projeto, e foram feitas algumas alterações aos protótipos com base nisso.

Nas visualizações onde é utilizado o *line chart*, foi adicionada uma interação, a possibilidade de realçar informação, de forma a facilitar a leitura da informação dos gráficos. Na visualização que apresenta a relação entre opções e cursos foi adicionada a possibilidade de alterar a duração da animação da evolução temporal. Em relação ao *chord diagram*, foi alterada a informação apresentada na *tooltip*, onde para além de apresentar os cursos de ambas as extremidades de uma conexão, é também apresentada a percentagem dos dados da conexão em relação ao total de dados. Para todas as visualizações foi alterada a forma como é apresentada a legenda, colocando a negrito a informação que está visível no gráfico.

Foi criada uma nova visualização, alternativa à visualização onde é apresentada a relação entre opções e cursos com o *radar chart*, na qual é apresentado um conjunto de *small multiples* de *radar charts*, cada um com a representação de apenas um curso (figura 11). Esta visualização

alternativa tem a vantagem de permitir observar com mais precisão as tendências gerais e os padrões de cada curso. Assim, esta divisão em vários *radar charts* produz uma visualização mais eficaz do que tentar agregar todos os dados num único gráfico, já que não existe tanto ruído visual [2].

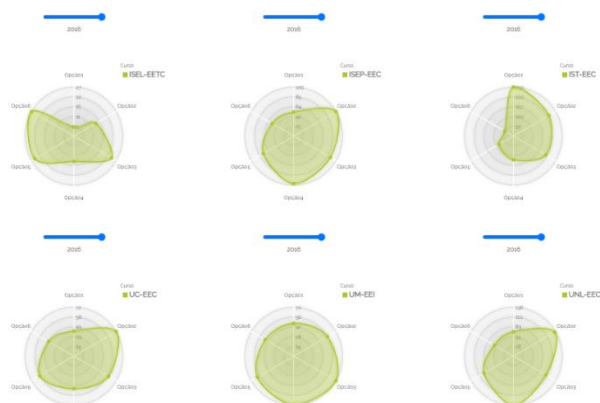


Figura 11: Alternativa à visualização das preferências opção – curso utilizando *small multiples* de *radar charts*

Por último, a aplicação e os protótipos estavam inicialmente em inglês, mas a pedido do professor Paulo Dias foram alterados para português, já que o trabalho realizado servirá de base para uma futura aplicação.

Conclusões e Trabalho Futuro

Concluindo, os objetivos propostos para o projeto foram atingidos, já que as visualizações implementadas permitem ilustrar algumas características relevantes dos dados para os utilizadores alvo.

O processo de desenvolvimento das visualizações exigiu a análise de várias alternativas, de forma a conseguir boas soluções para o problema proposto. De realçar a importância que os testes de usabilidade e as avaliações heurísticas possuem neste processo, já que foram efetuadas alterações pertinentes nas soluções de visualização com base no *feedback* obtido dos utilizadores.

Conclusões da análise aos dados das visualizações

Analisando a visualização correspondente à evolução das notas do último colocado, é possível concluir que tem existido alguma oscilação, sendo que as notas relativas ao MIEET por norma ocupam uma posição intermédia entre os restantes cursos analisados. O ano de 2011 acaba por se destacar entre os restantes, uma vez que existiu uma descida das notas na maioria dos cursos, sendo que no caso do MIEET a descida foi bastante acentuada, colocando-o com a nota de acesso mais baixa.

Em relação à visualização das notas dos colocados, constata-se que tem existido uma tendência para que a média das notas dos colocados no MIEET diminua, sendo que assiste-se a um aumento da afluência de alunos com notas de candidatura inferiores. Nos anos mais recentes analisados, verifica-se que grande parte dos colocados possuem notas de candidatura entre treze e catorze valores.

Analisando a posição do MIEET nas escolhas dos candidatos, é possível observar que os candidatos ao MIEET têm vindo a colocar o curso em opções de menor relevância. Ao longo dos anos, a principal escolha dos candidatos ao MIEET passou de primeira opção para terceira opção, revelando uma clara perda de importância.

Já em relação às preferências na escolha do curso consoante a opção, a conclusão que se destaca é uma clara preferência pelos cursos do Instituto Superior Técnico (IST) e da Universidade do Porto (UP).

Por último, analisando com mais detalhe um exemplo retirado da visualização das alternativas dos candidatos, constata-se que existe um número elevado de alunos candidatos ao MIEET com primeira opção na UP, e poucos alunos candidatos à UP com primeira opção no MIEET. Por outro lado, se se analisar as sextas opções das candidaturas, verifica-se o oposto, existindo um número mais elevado de alunos candidatos à UP com sexta opção no MIEET, e um número muito reduzido de alunos candidatos ao MIEET com sexta opção na UP.

Trabalho futuro

Como trabalho futuro, poderão ser feitos melhoramentos às visualizações existentes. Na visualização que representa as alternativas dos candidatos, onde foi utilizado o *chord diagram*, a seleção da opção da candidatura poderá permitir a seleção de mais do que uma opção em simultâneo, como por exemplo a seleção das opções 1 e 2. Na visualização das notas dos colocados, onde foi utilizado o *grouped bar chart*, poderão ser adicionadas mais métricas ao gráfico, como por exemplo a mediana, desvio padrão, variância, entre outras. Na visualização que representa as preferências opção-curso, onde foi utilizado o *radar chart*, a funcionalidade da animação da evolução temporal poderá permitir a seleção de apenas alguns cursos para estarem visíveis durante a animação, em vez de todos os cursos, e a transição entre os vários anos da animação também poderá ser feita sem refazer o gráfico. A animação também poderá ser disponibilizada na versão da visualização com *small multiples* de *radar charts*. Nas várias visualizações é utilizada a legenda a negrito para indicar a visibilidade no gráfico. Uma alternativa a isso seria a inclusão de uma *checkbox*, o que facilitaria a perceção da informação que está visível ou não.

Poderão também ser criadas novas visualizações com base nas visualizações existentes. Uma alternativa ao *chord diagram* seria integração da técnica *edge bundling* com um mapa, como é apresentado na figura 12.

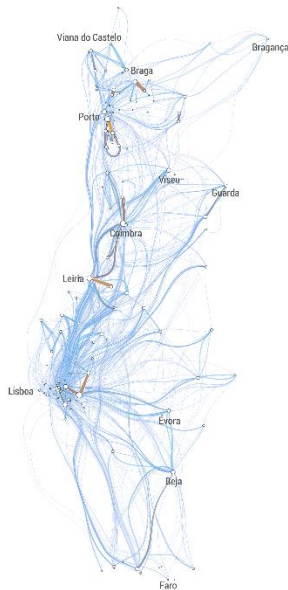


Figura 12: Exemplo de mapa com a técnica *edge bundling* [5]

Para representar a relação entre os cursos e as opções foi utilizado um *radar chart*, porém uma alternativa que também poderia ser interessante para representar essa informação seria um *stacked bar chart*.

Poderá também ser interessante uma visualização que permita compreender a relação entre notas e cursos, de forma a perceber qual o curso que os melhores alunos tendem a escolher. Essa visualização poderia ser desenvolvida tendo como base a visualização das notas dos colocados, onde é utilizado um *grouped bar chart*.

Por último, a aplicação desenvolvida poderá ser adaptada para os restantes cursos do DETI. A aplicação foi desenvolvida tendo em consideração esta finalidade, bastando por isso efetuar as alterações já referidas na subsecção referente aos Desafios da implementação.

Referências

- [1] Chen, C., Härdle, W., Unwin A. (2008). *Handbook of Data Visualization*. (1st ed.). Springer-Verlag Berlin Heidelberg. <https://doi.org/10.1007/978-3-540-33037-0>
- [2] Heer, J., Bostock M., Ogievetsky V. (2010). A tour through the visualization zoo. *Communications of the ACM*, 53(6), 59-67. <https://doi.org/10.1145/1743546.1743567>
- [3] Kim J., Iyer V., Joshi S., Volkin D., Middaugh C. (2012). Improved data visualization techniques for analyzing macromolecule

structural changes. *Protein Science : a Publication of the Protein Society*, 21(10), 1540-1553. <https://doi.org/10.1002/pro.2144>

- [4] Mazza R. (2009). *Introduction to Information Visualization*. (1st ed.). Springer-Verlag London. <https://doi.org/10.1007/978-1-84800-219-7>
- [5] Polisciuc, E., Cruz, P., Maças, C., Amaro, H., Machado, P. (2020). *Transitions of Customers Among Supermarkets*. Computational Design & Visualization Lab. https://cdv.dei.uc.pt/transitions_cutomers/