

# Visualização de Fluxos Migratórios: Os Resultados das Candidaturas ao Ensino Superior Público

Tiago Brito  
DETI  
Universidade de Aveiro  
Aveiro  
tlfbrito@ua.pt

Joaquim Madeira  
DETI / IEETA  
Universidade de Aveiro  
Aveiro  
jmadeira@ua.pt

Beatriz Sousa Santos  
DETI / IEETA  
Universidade de Aveiro  
Aveiro  
bss@ua.pt

## Resumo

Os candidatos ao Ensino Superior Público são colocados nas diferentes instituições através do Concurso Nacional de Acesso (CNA), o que pode originar a sua migração para outro distrito ou região autónoma. São propostas três visualizações interativas — representação matricial, diagrama cordal e representação georreferenciada — que permitem a apresentação, exploração e análise dos resultados do CNA, sendo usados os dados relativos aos candidatos colocados na 1ª fase dos concursos de 2012, 2013 e 2014. Estas visualizações foram avaliadas por três processos: avaliação heurística, testes de usabilidade supervisionados e recorrendo à utilização do Amazon Mechanical Turk (AMT).

## Palavras-Chave

Visualização, fluxos migratórios.

## 1. INTRODUÇÃO

Os jovens que concluem o Ensino Secundário e se candidatam ao Ensino Superior Público português são colocados, nas diferentes instituições de ensino universitário ou politécnico, através do Concurso Nacional de Acesso (CNA). Embora muitos dos candidatos venham a frequentar instituições do seu distrito ou região autónoma (DRA) de origem, outros são colocados em cursos e instituições que obrigam à sua *migração* no interior do território nacional.

No âmbito da dissertação de mestrado do primeiro autor [Brito 15], foram desenvolvidas e avaliadas três visualizações interativas — representação matricial, diagrama cordal e representação georreferenciada — que permitem apresentar, explorar e analisar os fluxos migratórios originados pelas colocações do CNA. Essas visualizações, e os dados relativos aos cerca de 115.000 candidatos colocados na 1ª fase dos concursos nacionais de 2012, 2013 e 2014, estão disponíveis numa plataforma *Web*.<sup>1</sup>

Nas duas próximas secções são referidos alguns trabalhos na área de Visualização de Fluxos Migratórios e descritos os dados do CNA que foram disponibilizados pela Direção-Geral do Ensino Superior (DGES). Na secção seguinte são apresentadas as três visualizações propostas, e as suas características mais importantes, bem como a plataforma criada para as disponibilizar ao público. De seguida são descritos os processos de avaliação da plataforma e das

visualizações desenvolvidas — avaliação heurística, testes de usabilidade supervisionados e recorrendo à utilização do *Amazon Mechanical Turk (AMT)* —, e apresentados alguns dos seus resultados. Na última secção são apresentadas algumas conclusões e ideias para trabalho futuro, bem como exemplificados alguns resultados da exploração dos resultados do CNA usando as visualizações desenvolvidas.

## 2. VISUALIZAÇÃO DE FLUXOS MIGRATÓRIOS

A visualização de fluxos migratórios é uma das áreas de aplicação da Visualização de Dados e de Informação que, tal como outras, tem as suas raízes em representações surgidas no século XIX.

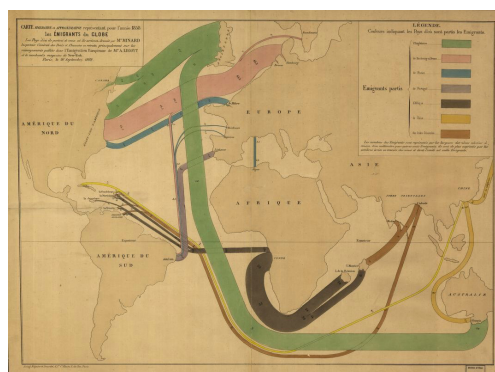


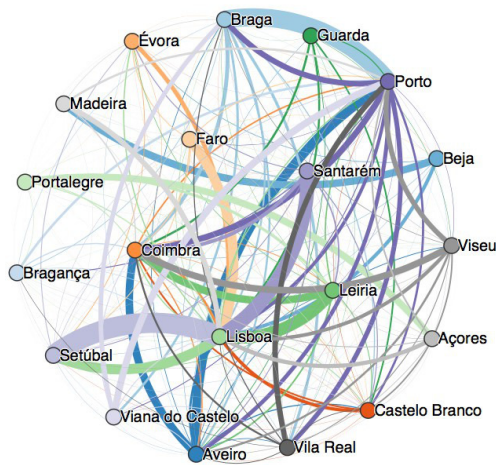
Figura 1. Fluxos de emigração em 1858.<sup>2</sup>

<sup>1</sup>[www.portugal-migration.info](http://www.portugal-migration.info)

<sup>2</sup>C. Minard, 1862, [www.datavis.ca/gallery/minbib.php](http://www.datavis.ca/gallery/minbib.php)

Uma dessas primeiras visualizações parece ser o mapa-mundo de Charles Minard apresentando os fluxos de emigração para o ano de 1858, em que a espessura das faixas representando os fluxos é proporcional ao número de emigrantes e é usada cor para melhor identificar a origem de cada fluxo migratório (ver Fig. 1).

As primeiras representações de fluxos migratórios sobre mapas, desenvolvidas com recurso a computadores, aparecem nas décadas de 1960 e 1970, surgindo mais tarde as primeiras ferramentas de software de suporte à sua criação. Uma dessas ferramentas mais referenciadas é o *Flow Mapper*, cuja primeira versão foi desenvolvida por W. Tobler na década de 1980.

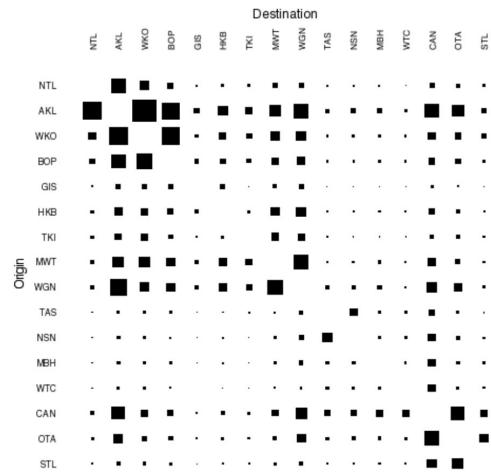


**Figura 2. Representação ingénua, usando um grafo, dos candidatos colocados no Ensino Superior em 2012.**

A representação de fluxos migratórios usando grafos orientados é uma alternativa simples e intuitiva à representação georreferenciada: os nós do grafo representam a origem e o destino de cada fluxo; os atributos de cada fluxo são, por exemplo, mapeados na espessura e na cor do correspondente arco. No entanto, se o grafo for (quase) completo, e mesmo que o número de nós seja pequeno, este tipo de representação direta torna-se inútil (ver Fig. 2), a menos que se use alguma estratégia para posicionamento dos nós e agrupamento de arcos.

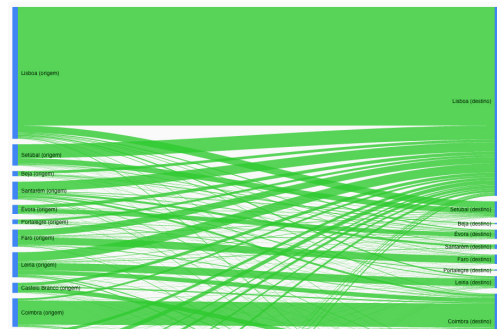
Atendendo a este problema, John Bryant propôs a visualização matricial da Figura 3: o grafo orientado é representado pela sua matriz de adjacências e o tamanho do quadrado associado a cada elemento da matriz é proporcional ao valor do correspondente fluxo migratório [Bryant 11]. Trata-se de um tipo de representação familiar, por semelhança com as representações em tabela, que permite representar um grande número de fluxos sem sobreposições, mas pode não ser fácil avaliar e comparar o valor dos vários fluxos migratórios.

Fluxos de diferentes tipos são habitualmente representados por Diagramas de Sankey. No caso dos fluxos migratórios é usado um diagrama vertical simples: as origens estão re-



**Figura 3. Visualização das migrações internas na Nova Zelândia [Bryant 11].**

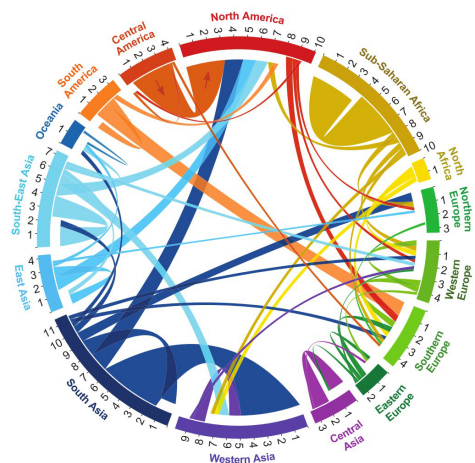
presentadas do lado esquerdo e os destinos do lado direito; a espessura dos arcos unindo origens e destinos é proporcional ao número de migrantes. No entanto, estes diagramas verticais podem apresentar problemas de usabilidade devido ao seu tamanho: o diagrama pode não ser totalmente representável no ecrã e ser difícil seguir a maioria dos fluxos, mesmo efetuando deslocamentos verticais. Já depois do início do trabalho descrito nesta comunicação, a DGES inaugurou um *site* que apresenta estatísticas relativas aos candidatos ao Ensino Superior em 2014, disponibilizando o Diagrama de Sankey da Figura 4.



**Figura 4. Vista parcial do Diagrama de Sankey representando as colocações dos candidatos ao Ensino Superior em 2014.<sup>4</sup>**

A representação em diagrama cordal apresenta um *layout* alternativo às visualizações anteriores: as origens e destinos dos fluxos são dispostos radialmente ao longo de uma circunferência; os fluxos entre origens e destinos são representados por cordas (i.e., arcos) de diferentes cores e de espessura proporcional ao número de migrantes — nalgumas destas visualizações uma única corda de espessura variável representa, para cada par (origem, destino), os flu-

<sup>4</sup>DGES, [www.dges.mec.pt/estatisticasacesso/](http://www.dges.mec.pt/estatisticasacesso/)



**Figura 5. Diagrama cordal representando os fluxos migratórios entre várias regiões do globo [Abel 14].**

xos migratórios em ambos os sentidos. O grupo de N. Sander vem desenvolvendo visualizações intuitivas de fluxos migratórios com base em diagramas cordais (ver Fig. 5) [Abel 14].

### 3. DADOS DO CONCURSO NACIONAL DE ACESSO

O Concurso Nacional de Acesso (CNA) ao Ensino Superior Público é, em termos do número de candidatos, visibilidade e impacto social, um dos mais importantes concursos anuais realizados pela Administração Pública. Para os últimos três anos, o número de candidatos e de colocados na primeira fase do CNA é apresentado na tabela seguinte (Tab. 1).

Ano	Candidatos	Colocados	%
2012	45.078	40.415	90%
2013	40.419	37.415	93%
2014	42.408	37.778	89%

**Tabela 1. Concurso Nacional de Acesso: número de candidatos e de colocados na 1ª fase (2012 a 2014).**

A DGES disponibilizou aos autores dados relativos aos cerca de 115.000 candidatos colocados na 1ª fase dos concursos nacionais de 2012, 2013 e 2014. Esses dados foram fornecidos num ficheiro Excel com 6 folhas de trabalho: 2012, 2013, 2014, GAES<sup>5</sup>, Estabelecimento e Distrito.

Nas três primeiras folhas, e para o respetivo ano, estão listados os atributos de cada candidato colocado (ver Tab. 2); as restantes folhas contêm informação complementar relativas às colunas (i.e., atributos) codificados. A folha GAES associa ao código de cada GAES a sua localização, que é usada para identificar o distrito ou região autónoma de ori-

<sup>5</sup>Gabinete de Acesso ao Ensino Superior.

Atributo	Descrição
Ano	Ano de candidatura
IDAluno	Identificador único
FaseCand	Fase de candidatura (apenas 1ª fase)
GAES	Código do GAES de origem
EstabelCol	Código da instituição de colocação
NotaCand	Nota de candidatura
AnoNasc	Ano de nascimento
Sexo	Género

**Tabela 2. Atributos associados aos candidatos colocados.**

gem dos colocados. A folha Estabelecimento permite associar ao código de cada estabelecimento de ensino a correspondente instituição em que cada candidato é colocado (universidade, faculdade, instituto politécnico, etc.).

Os dados recebidos da DGES foram processados para (i) verificar a sua integridade, (ii) remover informação considerada desnecessária (p.ex., a indicação da fase de colocação), (iii) agregar-lhes a população de cada distrito ou região autónoma, e (iv) optimizá-los para leitura, exportando-os para ficheiros CSV.

Para facilitar a análise dos dados foram definidos os seguintes indicadores numéricos, em que  $c(i, j)$  representa o número de candidatos oriundos de  $i$  e colocados em  $j$ , sendo  $i$  e  $j$  distritos ou regiões autónomas:

- Total de candidatos colocados em  $d$ :  
 $Tcol(d) = \sum_i c(i, d)$
- Total de candidatos oriundos de  $d$ :  
 $Tori(d) = \sum_j c(d, j)$
- Saldo de colocados para  $d$ :  
 $S(d) = Tcol(d) - Tori(d)$
- Rácio de colocados para  $d$ :  
 $R(d) = [Tcol(d)/Tori(d)] \times 100$
- Saldo de colocados entre  $u$  e  $v$ :  
 $S(u, v) = c(u, v) - c(v, u)$
- Percentagem de candidatos *migrando* de  $u$  para  $v$ :  
 $M(u, v) = [c(u, v)/Tori(u)] \times 100$

Os valores destes indicadores são usados em legendas e tabelas auxiliares, ou mapeados em atributos visuais (cor, espessura e/ou tamanho) nas três visualizações propostas.

### 4. VISUALIZAÇÕES DESENVOLVIDAS

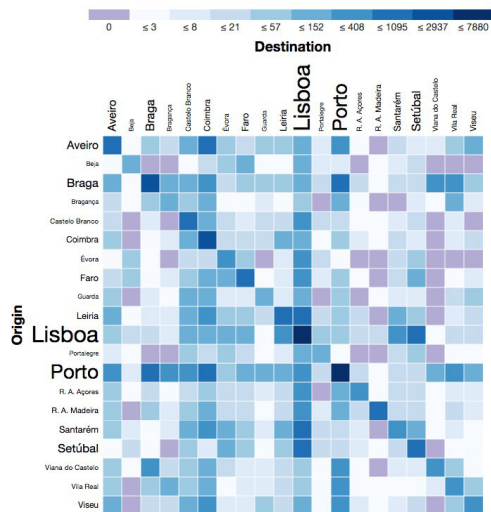
A colocação dos candidatos ao Ensino Superior nas diferentes instituições pode ser encarada como uma *migração*. Assim, podem ser desenvolvidas/adaptadas técnicas de representação de fluxos migratórios para criar visualizações interativas que possibilitam a representação, exploração e análise dos resultados anuais do CNA.

Descrevem-se de seguida as três visualizações desenvolvidas — representação matricial, diagrama cordal e

representação georreferenciada —; a biblioteca D3.js [Bostock 11] foi usada como base da implementação.

### 4.1. Representação Matricial

Esta visualização (ver Fig. 6) tem por base a matriz de adjacências do grafo orientado associado à colocação dos candidatos, de modo semelhante ao trabalho de J. Bryant. Há, no entanto, diferenças significativas que serão explicadas de seguida.



**Figura 6. Visualização matricial, representando o número de candidatos colocados, com a correspondente escala de cores.**

Como se pode ver na Figura 6, as linhas e colunas da matriz correspondem aos distritos e regiões autónomas (DRA) de origem dos candidatos e em que os candidatos são colocados. Para destacar os distritos mais populosos, o tamanho das letras da correspondente designação é proporcional à população de cada DRA. Acima da matriz é colocada a escala de cores correspondente aos dados representados.

Para cada conjunto de dados selecionado, e para cada par (origem, destino) — i.e., para cada fluxo de candidatos colocados —, é possível visualizar dois conjuntos de valores:  $c(i, j)$ , o número de colocados, ou  $S(i, j)$ , o saldo das colocações.

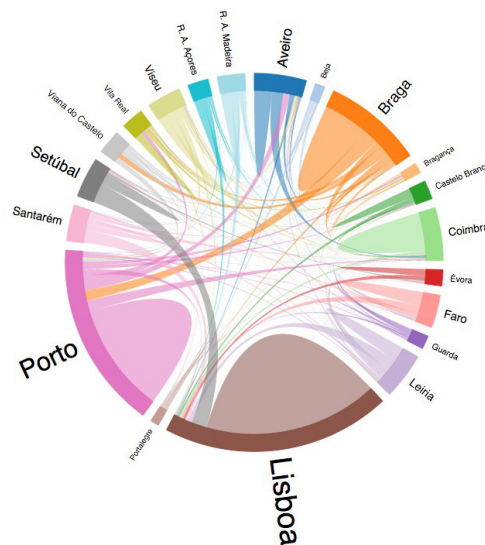
Para evitar enganos, são usadas duas escalas de cores: o número de candidatos colocados é mapeado de modo logarítmico numa escala de cores com 9 tonalidades de azul; o saldo positivo, nulo ou negativo de colocados é representado pelas cores verde, cinzento e encarnado, respetivamente. No caso de não existirem candidatos colocados, a correspondente célula da matriz é preenchida a roxo, ficando claramente destacada das restantes células. A escolha do conjunto de valores a representar desencadeia um conjunto de transições animadas de forma a adaptar progressivamente a visualização à nova escala de cores.

Uma das características particulares desta visualização é a capacidade de ordenar os distritos e regiões autónomas (i.e., as linhas e colunas da matriz) segundo quatro

critérios: designação, número total de candidatos, número total de colocados ou população. Este tipo funcionalidade é especialmente útil para comparar diferenças e destacar padrões, tal como sugerido por [Haroz 12].

### 4.2. Diagrama Cordal

Esta visualização (ver Fig. 7) é a única desenvolvida com base num *layout* D3.js predefinido (*Chord Layout*). No entanto, esta é uma das representações mais difíceis de gerar, pois necessita de estruturas de dados particulares para armazenar os dados e os outros elementos adicionais (texto, cores, etc.).



**Figura 7. Visualização usando o diagrama cordal.**

Tal como é habitual nos diagramas cordais, na periferia estão representados os vários DRA e estes encontram-se repartidos de acordo com a proporção dos candidatos colocados de cada DRA. Tal como na representação matricial, o tamanho da designação de cada DRA é proporcional à respetiva população.

Embora seja uma representação semelhante à referida na Secção 2, foram tomadas algumas opções particulares:

- Cada fluxo de colocados é representado, a menos que seja nulo, não se estabelecendo um (habitual) limite mínimo de colocados para a sua representação.
- A cada DRA é associada uma cor diferente, escolhida de um conjunto de cores de 20 tons divergentes.
- Entre cada par (origem, destino) é apenas desenhado um arco (i.e., uma corda) de espessura variável, que representa os dois fluxos com sentidos opostos, o que diminui o número de arcos representados.
- A espessura em cada extremidade de um arco é proporcional ao número de candidatos colocados no correspondente DRA.

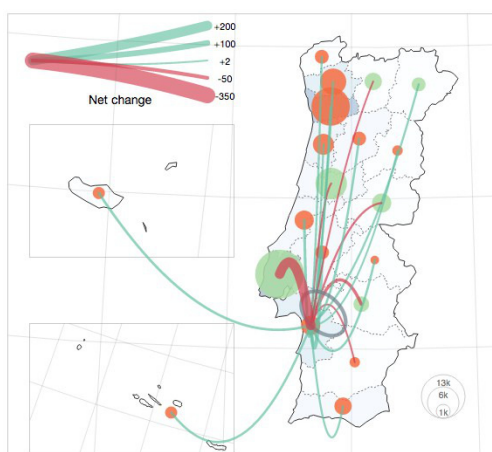
- A cor de cada arco é a cor da sua extremidade (i.e., DRA) que originou maior número de candidatos colocados.
- O arco que representa o conjunto de candidatos de um DRA colocados nesse mesmo DRA é representado como uma *bossa*, com a cor do DRA associado.
- Os arcos associados a cada DRA estão ordenados, em sentido horário, pelo número dos correspondentes candidatos colocados.<sup>6</sup>

A seleção, por filtragem, de um novo conjunto de dados a ser representado origina uma interpolação do estado atual para o estado final, sendo, por exemplo, removidos os arcos desnecessários, adicionados eventuais novos arcos e reposicionados todos os componentes do diagrama de acordo com os dados recém-selecionados. Assim, surge um efeito de sucessão do estado anterior para o subsequente, que não existe numa implementação mais tradicional em que o diagrama é subitamente redesenhado, e o utilizador pode mais facilmente perceber diferenças entre os dois estados.

### 4.3. Representação Georreferenciada

A representação georreferenciada desenvolvida (ver Fig. 8) permite visualizar os fluxos de candidatos colocados sobre o mapa de Portugal e das suas duas Regiões Autónomas.

Nesta visualização é associada (por omissão) uma cor a cada distrito ou região autónoma (DRA), usando 20 tons divergentes, permitindo uma clara associação entre cada cor e um DRA, o que ajuda o utilizador a identificar cada distrito e a sua fronteira. É possível alterar este mapeamento para uma escala de tons de azul representando a população de cada DRA: quanto mais escuro for o tom, maior o número de habitantes.



**Figura 8. Visualização georreferenciada.**

A cada DRA está associado um círculo de raio variável, posicionado no seu centróide, que permite a fácil seleção

<sup>6</sup>Assim, é fácil identificar os casos em que o maior número de candidatos é colocado ou não no mesmo DRA.

do respetivo DRA: pormenor importante no caso das regiões autónomas e de distritos de menor área; evita-se também mapear qualquer atributo usando a forma/área de cada DRA. O círculo selecionado é realçado por uma linha negra ao longo do seu perímetro, para facilitar a perceção da sua seleção, uma vez que os fluxos de candidatos colocados serão representados com origem no círculo selecionado.

O raio de cada um destes círculos é proporcional ao valor do indicador numérico selecionado para a visualização: número total de colocados ( $Tcol(d)$ ), número total de candidatos ( $Tori(d)$ ) ou rácio de colocados ( $R(d)$ ). Já a cor de cada círculo é definida pelo valor de outro indicador, o saldo de colocados ( $S(d)$ ): encarnado para um saldo negativo, verde para um saldo positivo e cinzento para um saldo nulo.

Os arcos que representam os fluxos de candidatos colocados são desenhados entre os círculos associados aos DRA, sendo a cor e espessura do arco ( $u, v$ ), e com destino em  $v$ , definidas do seguinte modo:

- A espessura do arco é proporcional ao valor absoluto do saldo de colocados,  $S(u, v)$ .
- A cor do arco será verde-claro se  $S(u, v) > 0$ , encarnado se  $S(u, v) < 0$ , ou cinzento se  $S(u, v) = 0$ .
- O arco com origem e destino em  $u$  (lacete), representando os candidatos oriundos de  $u$  e aí colocados, é processado de modo particular, sendo-lhe atribuída a cor cinzento.

Enquanto que cada lacete é desenhado como um arco de elipse, os outros arcos são desenhados como curvas quadráticas de Bézier, usando a linguagem SVG.

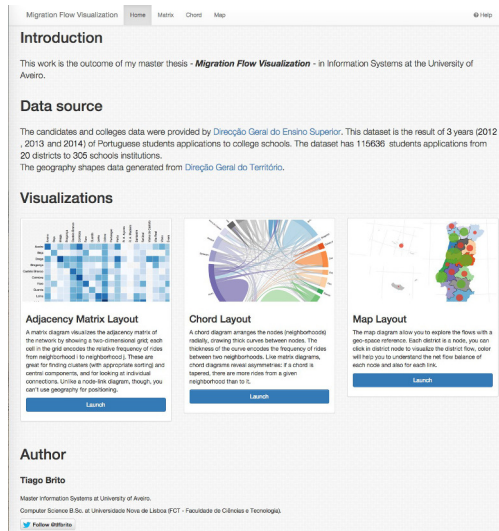
O modo como o valor do saldo de candidatos colocados é mapeado nos vários arcos permite distinguir, de forma clara, as situações em que esse saldo é positivo ou negativo, bem como ter uma ideia do seu valor. Um efeito interessante é conseguido desenhado os arcos com uma animação simples, após a seleção de um DRA: os arcos associados a saldos positivos são desenhados com destino no DRA selecionado, enquanto que os arcos associados a saldos negativos são desenhados com origem no DRA selecionado.

Devido à diversidade de elementos visuais e variáveis mapeadas em simultâneo, esta representação contém duas legendas de forma a auxiliar o utilizador na interpretação da informação visual existente.

### 4.4. Plataforma Web

As visualizações desenvolvidas foram alojadas numa plataforma *Web* ([www.portugal-migration.info](http://www.portugal-migration.info)), tendo como objetivos principais facilitar o acesso às visualizações, contextualizar o seu âmbito e descrever as várias funcionalidades num mesmo local, auxiliando os utilizadores na compreensão e exploração dos dados disponibilizados. No sentido de internacionalizar a utilização da plataforma, foi escolhida a língua inglesa para a sua

apresentação. A página de entrada contém uma introdução a cada uma das visualizações, acompanhada de uma imagem ilustrativa, visível na Figura 9.



**Figura 9. Página de entrada da plataforma desenvolvida.**

Cada uma das visualizações é acessada através do seu botão de ação e acompanhada por um painel de filtros (comuns às três visualizações) e controlos que permitem filtrar a informação existente (por género dos candidatos, ano de candidatura e tipo de instituição) e interagir com cada visualização.

Estão também disponíveis de forma transversal às três visualizações: uma *tooltip* que apresenta os indicadores numéricos respetivos quando é selecionado um distrito ou um fluxo de colocados; uma tabela que apresenta a lista de instituições e o número de colocados em cada uma delas, para um dado fluxo de colocados selecionado.

## 5. AVALIAÇÃO

As características e funcionalidades da plataforma *Web* e das visualizações interativas desenvolvidas foram avaliadas através de métodos de avaliação heurística, testes de usabilidade supervisionados e recorrendo à utilização do *Amazon Mechanical Turk (AMT)*, para detetar eventuais limitações e avaliar a sua eficácia.

### 5.1. Avaliação Heurística

Esta avaliação foi efetuada de modo independente por dois avaliadores<sup>7</sup>, com base nos modelos de Forsell, de Nielsen e de Zuk [Forsell 10, Nielsen 95, Zuk 06], de modo a avaliar a gravidade dos problemas de usabilidade detetados através de duas métricas: aplicabilidade da heurística (escala de 0 a 5) e gravidade do problema (escala de 0 a 4).

Os problemas mais importantes detetados, e as soluções encontradas, foram os seguintes:

<sup>7</sup>Alunos do curso de Mestrado em Sistemas de Informação da U. Aveiro, sem relação com o primeiro autor.

- As escalas de cores originais não eram corretamente distinguíveis por daltónicos, originando interpretações enviesadas e perda de informação — os tons das escalas de cores foram alterados para minorar essa perda de informação.
- O estado dos elementos do painel de filtros e controlos não era apresentado de modo uniforme, nomeadamente para a seleção do género dos colocados — foi feita uma correção para garantir a uniformização desejada.
- Ajuda de contexto insuficiente, dado que, além de ser usada a língua inglesa, ocorre um conjunto de termos que não são intuitivos para o público em geral — foi introduzido um mecanismo de ajuda contextual guiada, transversal a todas as visualizações, que apresenta uma descrição de todos os filtros e controlos e destaca a funcionalidade em contexto.

O resultado global das duas avaliações heurísticas foi considerado positivo e construtivo, confirmando a consistência existente entre as características das três visualizações, e o modo como são apresentadas e podem ser usadas na plataforma. Foi também destacada a intenção subentendida a cada visualização, de que cada uma delas melhor se adequa a realizar algumas tarefas específicas.

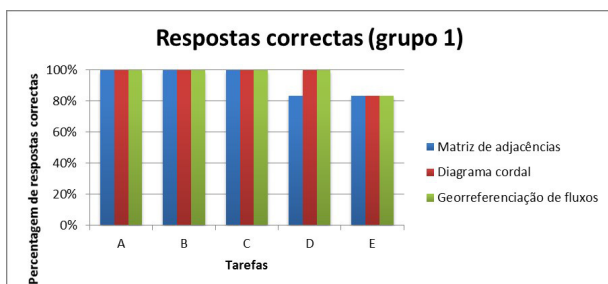
### 5.2. Teste de Usabilidade

O teste de usabilidade compreendeu a realização, por um conjunto de 12 utilizadores, de uma sequência de tarefas com propósito bem definido, e seguiu a metodologia habitual: numa primeira fase foi efetuada uma apresentação da plataforma e das suas funcionalidades, e dada uma explicação dos procedimentos do teste; de seguida, os utilizadores realizaram as tarefas predefinidas e avaliaram o grau de dificuldade de cada uma, usando uma escala de Likert (5 níveis). A prestação de cada utilizador foi acompanhada por um observador que, entre outros pormenores, registou o tempo necessário à conclusão de cada tarefa e o modo como foi executada.

As tarefas propostas foram divididas em dois grupos:

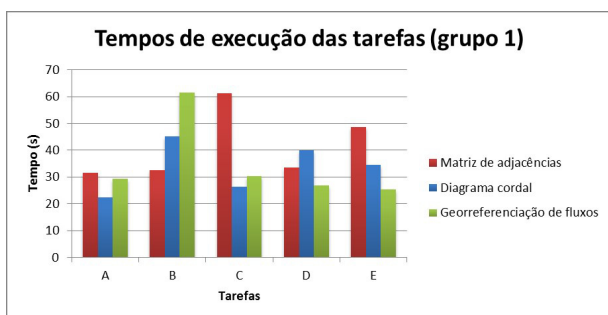
1. Para cada uma das visualizações, foram solicitadas tarefas distintas, mas equivalentes na sua dificuldade e no seu objetivo, de modo a introduzir e ambientar os utilizadores às visualizações e exibir as suas diferenças principais.
2. Sem que fosse indicado o tipo de visualização a usar, foi novamente solicitada a realização de tarefas, devendo o utilizador escolher a visualização que julgasse mais apropriada. A resolução de algumas tarefas era mais fácil através de uma escolha adequada, ou mesmo impossível sem a escolha da única visualização apropriada.

Cada um dos 12 utilizadores realizou um total de 21 tarefas: o primeiro grupo continha 5 tarefas para cada visualização e segundo grupo 6 tarefas adicionais.



**Figura 10. Percentagem de tarefas concluídas corretamente para o primeiro grupo.**

Para o grupo 1, os resultados mostram que os participantes conseguiram concluir todas as tarefas; no entanto, a tarefa E<sup>8</sup> foi errada por 20% dos participantes, para todas as visualizações (ver Fig. 10), o que pode indicar um eventual problema de interpretação do seu enunciado.



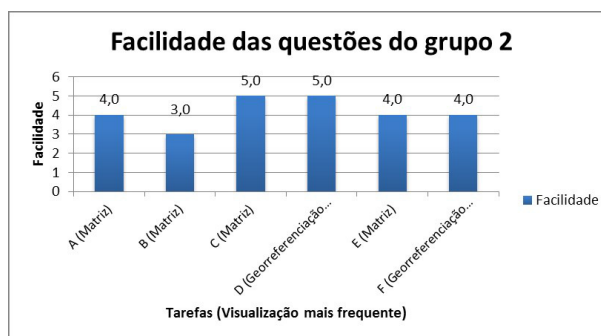
**Figura 11. Tempo médio de conclusão das tarefas do primeiro grupo.**

O gráfico da Figura 11 mostra que, apesar de todos os participantes terem concluído as tarefas do primeiro grupo, a sua resolução foi realizada com tempos bastante distintos. Destacam-se, sobretudo, os tempos médios mais longos para a tarefa B, utilizando a representação georreferenciada, e para a tarefa C, usando a representação matricial.

Para as tarefas do segundo grupo, o gráfico da Figura 12 mostra uma categorização relativamente consistente quanto à facilidade de execução das tarefas propostas. Fica também claro que grande parte dos participantes optou por utilizar a representação matricial, exceto nas tarefas D e F, onde a resolução era apenas possível através da representação georreferenciada. Isto indica que, apesar dos participantes preferirem utilizar a representação matricial, conseguem identificar claramente a visualização mais correta para resolver tarefas de características georreferenciadas.

As críticas e os comentários apontados durante o teste de usabilidade são também uma contribuição valiosa dos participantes. Destacam-se os julgados mais importantes:

<sup>8</sup>Identificar o distrito que recebe o maior número de colocados oriundos de um dado distrito X, excluindo X.



**Figura 12. Mediana do nível de facilidade atribuído e visualização mais frequente escolhida, para as tarefas do segundo grupo.**

- Dificuldade em encontrar os totais dos DRA na representação matricial.
- Dificuldade na análise das visualizações cordal e georreferenciada, devido ao número de arcos presentes.
- Na visualização georreferenciada, os lacetes não têm o destaque necessário e são difíceis de descobrir.

### 5.3. Amazon Mechanical Turk

Apesar dos testes de usabilidade supervisionados serem proveitosos em termos dos dados e comentários recolhidos, a sua preparação e realização são trabalhosas, sendo difícil angariar participantes. Por este motivo, nos últimos anos tem assumido importância a avaliação das características de visualizações usando *crowdsourcing*, em que os participantes nas avaliações são recrutados através de serviços especializados, como o *Amazon Mechanical Turk (AMT)* [Kittur 08].

Após uma análise exploratória do *AMT*, foi decidido efetuar uma avaliação das visualizações desenvolvidas usando essa ferramenta. Para isso foram desenvolvidos três guiões — um para cada visualização a avaliar —, equivalentes nos seus objetivos e na dificuldade das suas questões/tarefas. Cada guião tinha 10 questões: 3 questões demográficas, 1 teste de daltonismo, 2 questões de validação do perfil do participante e 4 tarefas específicas para a visualização a avaliar. O processo de avaliação usando o *AMT* decorreu num período de 24 horas, com 90 participantes (30 para cada guião), com origem em diferentes países.

Entre outros resultados, foram registados o tempo de conclusão de cada uma das tarefas e o número de respostas corretas de cada participante. Uma análise preliminar revelou que os valores do tempo médio de conclusão de todas as tarefas são próximos quando são usados a representação matricial e o diagrama cordal (237 e 220 segundos), com um valor médio mais elevado quando é usada a representação georreferenciada (264 segundos). Esta última representação também originou um menor número médio de respostas corretas (64%), número que é mais elevado para a utilização do diagrama cordal (76%) e

da representação matricial (84%). A menor familiaridade com a geografia de Portugal, por parte dos participantes, talvez possa explicar, em parte, estes resultados.

## 6. CONCLUSÃO

Foram desenvolvidas, avaliadas e disponibilizadas na *Web* três visualizações interativas para a representação, exploração e análise dos fluxos de colocação dos candidatos ao Ensino Superior Público, a partir dos distritos e regiões autónomas de origem (DRA) e sendo usados os dados (2012 a 2014) da 1ª fase do Concurso Nacional de Acesso.

De notar que se trata de um tema com interesse para o público em geral, e que tem associado um conjunto de dados de tamanho expressivo, com um número razoável de atributos associados a cada candidato colocado

Apresentam-se de seguida alguns resultados globais da exploração dos dados, que exemplificam a utilidade das visualizações desenvolvidas:

- O mais importante fluxo de colocados oriundos de um DRA corresponde sempre aos colocados nesse DRA, exceto para o distrito de Setúbal.
- Para os candidatos oriundos dos distritos de Lisboa, Porto, Castelo Branco e Coimbra, a percentagem de colocados no seu próprio distrito é superior a 50%.
- Cada um dos outros DRA coloca mais candidatos no distrito de Lisboa do que o inverso.
- Setúbal é o único DRA de onde o mais importante fluxo de colocados não só não é para o próprio distrito, como também este é superior a 50% (para o distrito de Lisboa).
- Évora e Beja são os distritos cujos os candidatos são colocados num menor número de DRA — Évora: 7 distritos de colocação em 2012 e 2013; Beja: 8 distritos de colocação em 2014.
- A Região Autónoma da Madeira (2013 e 2014) e o distrito de Viana do Castelo (2012 a 2014) são os DRA que recebem colocados de um menor conjunto de DRA: um total de 10 DRA; de outros 10 não recebem qualquer colocado.

Como trabalho futuro referem-se: a vantagem em implementar mais sugestões resultantes da avaliação heurística e do teste de usabilidade supervisionado; o desenvolvimento de funcionalidades adicionais com base noutros atributos de cada colocado, como a nota de candidatura; e o interesse em manter a plataforma atualizada, acrescentando anualmente os dados das colocações, que irão sendo solicitados à DGES. Seria também muito interessante obter atributos adicionais associados a cada candidato colocado e dados adicionais de cada instituição. Por exemplo, permitindo analisar os diversos cursos em que os candidatos são colocados, o número de vagas existentes e sobranes, etc.

Terá também de ser considerada a aplicação das visualizações desenvolvidas a outro tipo de cenários, no sentido de avaliar a sua eficácia e escalabilidade, permitindo uma visualização hierárquica e/ou seletiva dos dados, quando o número de pares (origem, destino) aumentar significativamente.

## 7. AGRADECIMENTOS

Agradece-se à Direção-Geral do Ensino Superior a disponibilização dos dados utilizados neste trabalho.

Agradece-se, também, a todos os que colaboraram na avaliação das visualizações desenvolvidas quer realizando as avaliações heurísticas quer participando nos testes de usabilidade.

## Referências

- [Abel 14] G. J. Abel e N. Sander. Quantifying global international migration flows. *Science*, 343(6178):1520–1522, 2014.
- [Bostock 11] M. Bostock, V. Ogievetsky, e J. Heer. D3 data-driven documents. *IEEE Trans. Vis. Comp. Graphics*, 17(12):2301–2309, Dec 2011.
- [Brito 15] T. Brito. Visualização de fluxos migratórios: Os candidatos ao ensino superior público (2012 a 2014). Dissertação de Mestrado em Sistemas de Informação, Universidade de Aveiro, 2015.
- [Bryant 11] J. Bryant. Visualising internal migration flows. *New Zealand Population Review*, 37:159–171, 2011.
- [Forsell 10] C. Forsell e J. Johansson. An heuristic set for evaluation in information visualization. Em *Proc. Int. Conf. Advanced Visual Interfaces, AVI '10*, páginas 199–206. ACM, 2010.
- [Haroz 12] S. Haroz e D. Whitney. How capacity limits of attention influence information visualization effectiveness. *IEEE Trans. Vis. Comp. Graphics*, 18(12):2402–2410, Dec 2012.
- [Kittur 08] A. Kittur, E. H. Chi, e B. Suh. Crowdsourcing user studies with mechanical turk. Em *Proc. SIGCHI Conf. Human Factors in Computing Systems, CHI '08*, páginas 453–456. ACM, 2008.
- [Nielsen 95] J. Nielsen. How to conduct a heuristic evaluation, 1995.
- [Zuk 06] T. Zuk, L. Schlesier, P. Neumann, M. S. Hancock, e S. Carpendale. Heuristics for information visualization evaluation. Em *Proc. AVI Workshop BEyond Time and Errors: Novel Evaluation Methods for Information Visualization, BELIV '06*, páginas 1–6. ACM, 2006.