

Pareamento de casais resolvido por algoritmo de designação

Nome do autor, RA e e-mail:

Israel Campiotti, 155825, i155825@dac.unicamp.br

Nome do autor, RA e e-mail:

Victor Jeronymo, 157490, v157490@dac.unicamp.br

Nome do autor, RA e e-mail:

Desireé Ruttul, 166526, d166526@dac.unicamp.br

Abstract

Problemas de designação em fluxo em redes são caracterizados pelo pareamento um a um de nós de um grafo bipartido. Pensando nisso, utilizamos esse conhecimento para indicar a pessoas que responderam a um questionário seus possíveis pares ideais. Para isso utilizamos como método o algoritmo húngaro, verificando sua eficácia nesse tipo de problema uma vez que obtivemos um sucesso em quase 100% dos participantes da pesquisa

1 Introdução

Fluxo em redes é uma área da Pesquisa Operacional que trabalha com grafos, a cada arco do grafo associamos um fluxo, custo ou capacidade. Diversas são as variações e aplicações deste campo de estudo, desde problema como do Caixeiro Viajante a problemas de transporte. Este trabalho visa mostrar o funcionamento de uma destas áreas, os chamados problemas de designação.

Um problema de designação é caracterizado por um pareamento um a um entre nós de um grafo orientado bi-partido, onde o fluxo se inicia em um dos lados e é relacionado somente a um destino do lado oposto. Isto significa, por exemplo, atrelar professores às disciplinas, pessoas às tarefas ou formação de casais monogâmicos. Este último se tornou muito popular graças a sites e aplicativos de relacionamento, transformando-se numa indústria lucrativa.

Utilizamos justamente este último exemplo para exemplificar a aplicabilidade deste problema no mundo real. Na seção 2, descrevemos o problema que nos propomos a resolver, mostrando como os dados foram obtidos, como os custos foram construídos, além de uma visão geral dos dados coletados. Na seção 3 detalhamos a formulação matemática de um problema de designação ao destacar a função objetivo e as restrições. Seção 4 desenha um breve resumo de como o problema proposto foi solucionado, bem como informações importantes obtidas. A seção 5 contém as considerações finais, comentando aspectos interessantes do problema.

2 Descrição do problema

Buscamos resolver o problema de pareamento de casais usando dados reais, coletados através de um formulário que disponibilizamos em uma página do Facebook acessada por diversos estudantes da Unicamp. Com as repostas em mãos, calculamos uma matriz de custo relativa a afinidade entre as pessoas, ou seja, quanto menor o custo, maior a compatibilidade entre as duas pessoas. As perguntas estão contidas no Apêndice, mas dependendo da pergunta devemos acrescentar um valor diferente ao custo.

2.1 Criando os custos

A principal pergunta é relativa ao sexo dos participantes. Permitimos que estes se caracterizem como pertencentes ao gênero masculino ou feminino, porém possam buscar somente um ou ambos sexos. Assim, se a pessoa i , do sexo masculino, procura alguém do sexo feminino, mas a pessoa j , do sexo feminino procura alguém do mesmo sexo que ela, então o custo entre as pessoas i e j é muito grande (big M). Agora, se j procura alguém do sexo masculino ou de ambos sexos, o custo entre os dois é zero. O mesmo raciocínio é aplicado para casos homo afetivos.

Outra importante questão condiz ao tipo de relacionamento que o participante procura, permitimos que sejam selecionada uma de três opções: sério, casual ou disponível a ambas possibilidades. Ou seja, suponha que a pessoa i procura algo sério, enquanto a pessoa j procura algo diferente. Se j procura algo estritamente casual, então o custo entre os dois é acrescido em big M, porém se j estiver disponível para ambos casos, então nenhum custo é acrescido. Nenhum custo é adicionado se ambos procuram pelo mesmo tipo de relacionamento.

A próxima categoria de perguntas diz a respeito da altura, idade, curso, signo e opção alimentar. Assim como no caso do gênero, definimos estas questões como perguntas de dois sentidos, ou seja, ambos o que a pessoa i busca e o que a pessoa j procura para criar o custo entre os dois, porém estas perguntas apesar de serem importantes, não devem excluir totalmente a possibilidade de i e j serem pareados juntos.

Por exemplo, na opção alimentar permitimos que seja escolhida uma entre três opções: vegetariano, vegano e nenhum dos dois e que o participante busque alguém vegetariano, vegano ou não se importe com esta opção. Logo, se a pessoa i é indiferente a opção alimentar do parceiro, mas é vegana, então o custo é acrescido somente pela preferência da pessoa j . Se j procurar somente veganos ou vegetarianos acrescentamos um custo pre-determinado, caso a opção também não importe a j então o custo não é acrescido. O mesmo é feito para o signo.

O caso da idade, altura e curso é ligeiramente diferente, onde o custo a ser acrescido depende do quanto diferente as respostas são. Por exemplo, se a pessoa i procura alguém entre 22 e 24 anos, o custo entre ela e alguém de 18 a 21 ou 24 a 27 anos é menor do que entre alguém entre 28 a 30, que é menor do que alguém entre 31 a 35 e assim sucessivamente. No caso do curso, separamos por similaridades,

sendo exatas mais próximo de biológicas, depois educação física, humanas, letras e artes. Lembrando que o custo entre i e j é acrescido levando em consideração tanto as preferências de i quanto de j .

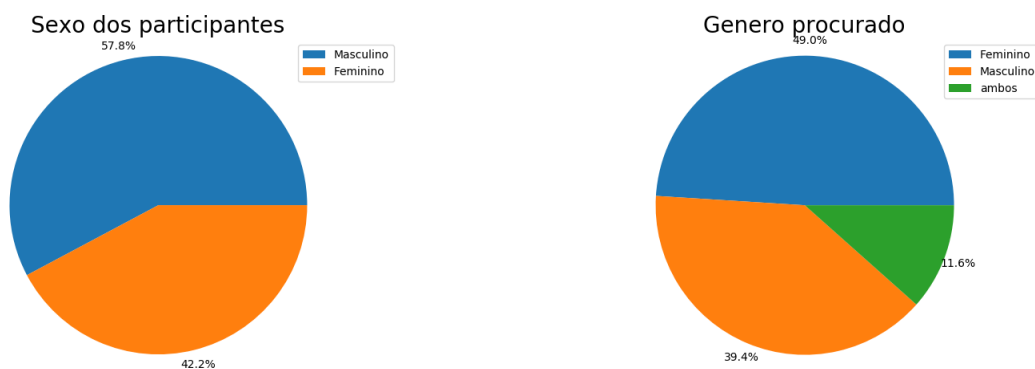
Quanto as perguntas relacionadas a gostos musicais, de livro, filme e como gosta de passar as noites de sexta-feira, realizamos o seguinte tratamento. Criamos arbitrariamente uma lista de proximidade para as respostas, por exemplo, se a pessoa i gosta mais de musica clássica, então ela provavelmente gosta mais de jazz do que de rock, depois indie, hip-hop/rap, pop, sertanejo e por fim funk. Logo, as alternativas mais próximas crescem custo menores do que alternativas mais distantes.

Por fim, para pergunta referente as séries de TV, se a pessoa i e a pessoa j gostam da mesma série diminuimos um pouco o valor do custo entre as duas.

2.2 Dados coletados

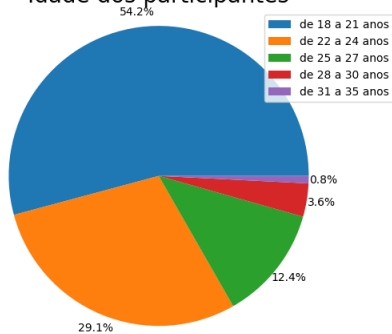
Os dados coletados foram aqueles apresentados pelos participantes do formulário a seguir apresentamos um resumo dos resultados para cada pergunta fazendo comentários onde se mostra necessário. Mostramos somente as opções que foram selecionadas pelo menos uma vez.

1. Sexo dos participantes e por qual procuram. Como vemos no gráfico abaixo a partição do sexo masculino foi superior a participação do sexo feminino, o que poderia gerar um grande desbalanceamento no pareamento se não possibilitássemos procurar alguém do mesmo sexo ou de ambos.

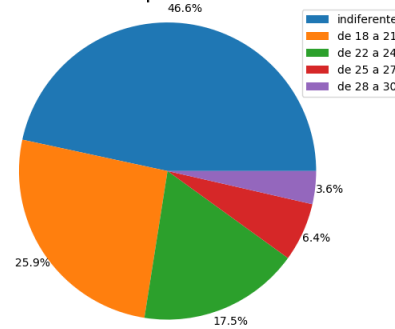


2. Idade dos participantes e por qual procuram. A maioria dos participantes possui entre 18 a 21 anos, porém a maioria das pessoas não se importa pela idade do parceiro.

Idade dos participantes

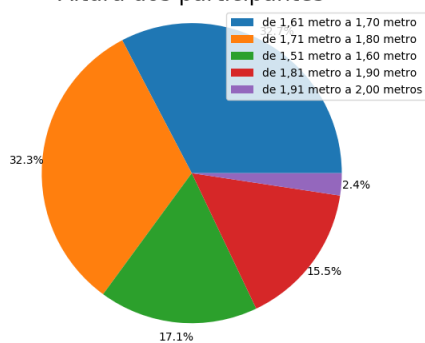


Idade procurada

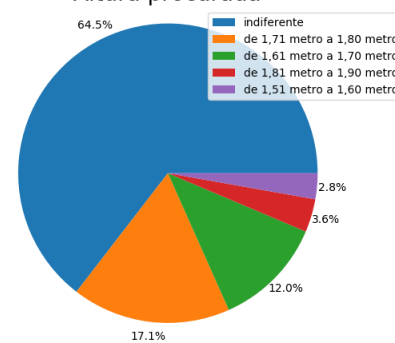


3. Altura dos participantes e altura procurada.

Altura dos participantes

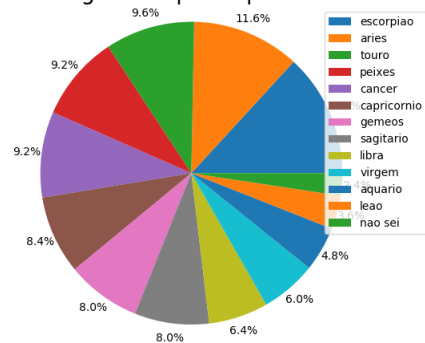


Altura procurada

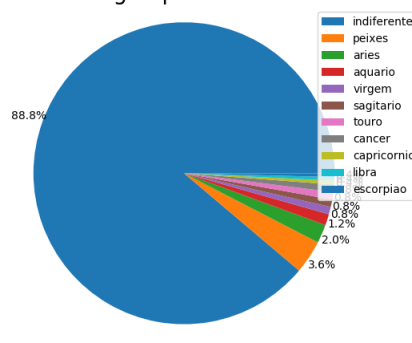


4. Signo dos participantes e signo procurado.

Signo dos participantes

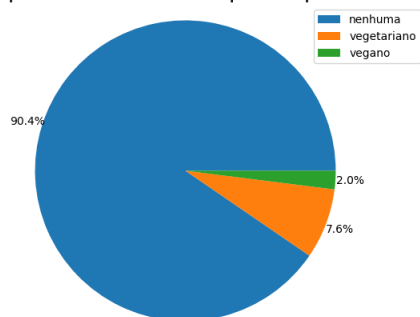


Signo procurado

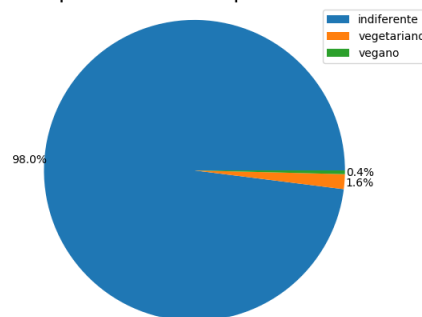


5. Opção alimentar dos participantes e opção procurada. Embora esta pergunta acrescentasse um custo relativamente alto, podemos ver que foi decisiva em poucos casos, já que a maioria não se importa com a opção alimentar do parceiro.

Opcao alimentar dos participantes

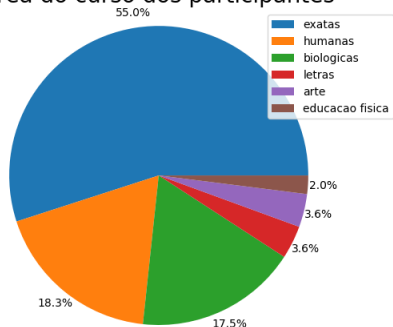


Opcao alimentar procurado

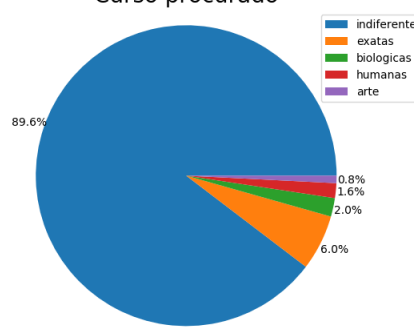


6. Curso dos participantes e curso procurado.

Area do curso dos participantes

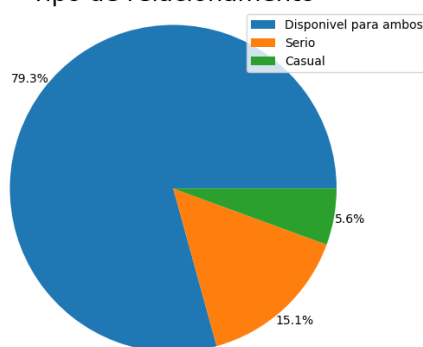


Curso procurado



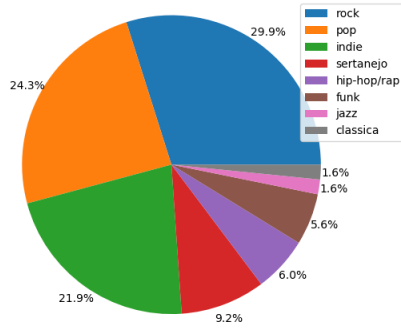
7. Tipo de relacionamentos procurado pelos participantes. Esta pergunta também era muito importante no valor do custo, porem foi irrelevante na maioria dos casos, contudo serviu para diferenciar alguns dos participantes.

Tipo de relacionamento

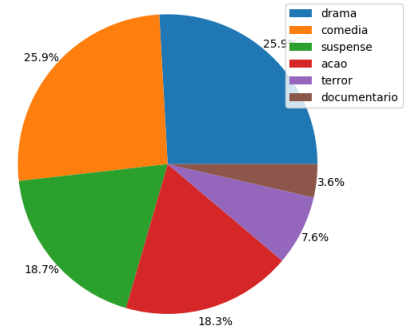


8. Demais perguntas.

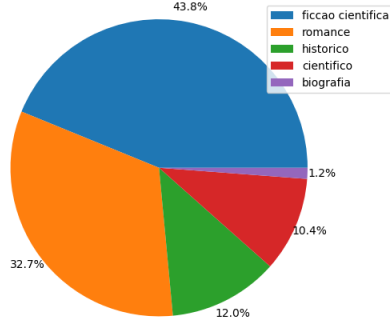
Musica favorita dos participantes



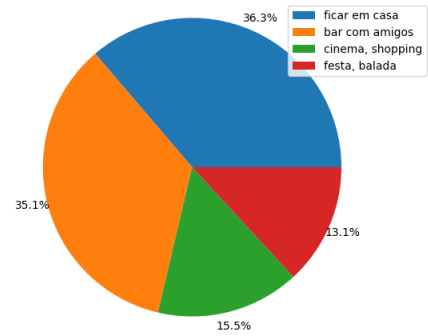
Filme favorito dos participantes



Livro favorito dos participantes



Sexta a noite dos participantes



3 Modelagem matemática

Um problema de designação pode ser modelado da seguinte forma[4]:

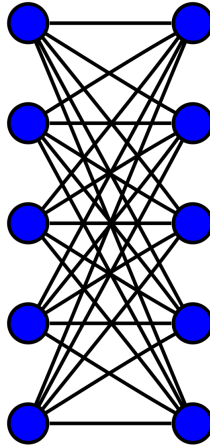
$$\begin{aligned}
 & \text{Min} \sum_{i=1}^m \sum_{j=1}^m c_{ij} x_{ij} \\
 & s.a. : \sum_{j=1}^m x_{ij} = 1, \forall i = 1, \dots, m \\
 & \sum_{i=1}^m x_{ij} = 1, j = 1, \dots, m \\
 & x_{ij} \in \{0, 1\}
 \end{aligned} \tag{1}$$

onde c_{ij} é o custo entre a pessoa i , do lado esquerdo, e a pessoa j , do lado. x_{ij} é o fluxo do entre i e j , se $x_{ij} = 1$ então a pessoa i é designada a pessoa j . Note que as restrições, impedem que a pessoa i e a pessoa j sejam pareados com mais de uma pessoa.

Agora, se o numero de nós do lado esquerdo for n e o do lado direito for m , com $n < m$, então o problema e formulado da seguinte forma:

$$\begin{aligned}
& \text{Min} \sum_{i=1}^m \sum_{j=1}^m c_{ij} x_{ij} \\
& s.a. : \sum_{j=1}^m x_{ij} = 1, \forall i = 1, \dots, n \\
& \sum_{i=1}^n x_{ij} \leq 1, j = 1, \dots, m \\
& x_{ij} \in \{0, 1\}
\end{aligned} \tag{2}$$

Esta formulação permite que nós do lado direito não sejam designados a nenhum no do lado esquerdo. A seguir mostramos como um grafo de designação é caracterizado. Lembrando que apesar do grafo ser totalmente conectado, cada nó do lado esquerdo é designado a somente um nó do lado direito. Além disso a formulação 2 permite que o lado esquerdo possua menos nós que o direito.



4 Resolvendo o problema

Para resolver o problema utilizamos custos obtidos para escrever uma matriz de custo, onde a entrada c_{ij} representa o custo entre o participante i e j . O tratamento dos dados foi feito utilizando códigos em Python[1], primeiramente transformamos as respostas as perguntas em números, tornando mais fácil sua utilização na criação do custo descrita anteriormente. No caso que supomos uma ordem de proximidade a reposta tivemos o cuidado de que os números representasse isto bem, ou seja, possuísem uma ordenação, no caso da pergunta sobre como passar sexta-feira a noite, fizemos:

- Ficar em casa $\rightarrow 0$
- Ir ao cinema ou shopping $\rightarrow 1$
- Sair com amigos ao bar $\rightarrow 2$
- Ir para festas $\rightarrow 3$

Com estes dados e com a matriz de custos resolvemos o problema de designação em MATLAB[2], onde utilizamos uma implementação do algoritmo de Húngaro[3]. Resolvemos o problema para três casos.

O primeiro caso foi feito utilizando todos os participantes no lado esquerdo do grafo assim como no lado direito, ou seja, utilizamos a formulação 1 do problema de designação. Note que como o custo de $c_{ij} = c_{ji}$, então obtivemos $x_{ij} = x_{ji} = 1$ caso a pessoa i fosse designada ao participante j . Como o grupo de amostragem não é simétrico, alguns participantes foram designados a si mesmos. Denotamos na tabela abaixo a quantidade destes casos.

O segundo caso foi feito colocando somente as mulheres no lado esquerdo do grafo, utilizamos portanto a formulação 2 do problema de designação, permitindo que pessoas no lado direito do grafo terminassem sem par. Esta formulação da preferência as mulheres, uma vez que elimina a competição com o sexo masculino a participantes no lado direito do grafo que busquem parceiros de ambos gêneros.

O terceiro caso foi feito utilizando somente homens no lado esquerdo do grafo, sendo então semelhante ao caso anterior, porem eliminando a competição com o sexo feminino. Ambos os segundo e terceiro caso permitem que $x_{ij} = x_{ji} = 1$, no caso em que alguma mulher seja designada a outra, ou algum homem a outro.

Em seguida apresentamos a tabela com o resumo dos resultados para cada caso. O tamanho nos informa quantos nós do lado esquerdo para a quantidade no lado direito do grafo, os custos são calculados sem considerar as designacoes que terminaram com bigM, e finalmente bigMs, denota a quantidade de vezes que esta designação aconteceu.

| Preferência | Tamanho | Custo Total | Menor Custo | Maior Custo | Média dos Custos | bigMs |
|-------------|--------------|-------------|-------------|-------------|------------------|-------|
| Todos | 251 para 251 | 1765 | -1 | 37 | 7.57 | 18 |
| Mulheres | 105 para 251 | 717 | -1 | 37 | 6.82 | 0 |
| Homens | 146 para 251 | 985 | -1 | 37 | 7.69 | 18 |

5 Comentários

Como utilizamos dados reais (as repostas do questionário), não obtivemos uma simetria, logo, algumas pessoas foram designadas a si mesmas, pois não havia par disponível para ele. Vemos que isto ocorreu somente a homens que procuravam mulheres, pois o número de participantes do sexo masculino superou o do sexo feminino.

Além disso, perguntas que julgamos importantes e que adicionavam custos altos, tiveram uma grande porcentagem de participantes as julgando indiferentes, como opção alimentar, altura, idade ou signo, porém serviram para refinar melhor os parceiros daquelas pessoas que as julgaram importantes. Podemos concluir então que as principais adições aos custos foram feitas pelas perguntas de gostos

pessoais, estilo musica, filme, livro e como gostam de passar o tempo livre no fim de semana.

Observamos também que houveram pares que tiveram seus custos reduzidos devido a gostos similares em séries de TV, pois somente assim poderíamos obter um custo negativo, como apresentado na tabela anterior.

Foram consideradas um total de 251 repostas, o que nos dá um número razoável de pessoas, contudo as perguntas podem não ter sido as mais precisas ou não serem suficientes para definir com precisão a afinidade entre duas pessoas, mas atingiram o objetivo proposto, mostrando como podemos modelar e solucionar situações reais de problemas de designação.

5.1 Comentários sobre anexos

Em anexo possuímos alguns arquivos, a seguir explicamos o que cada um contém:

- *formulario.csv*: Contem as informações como recebidas pela formulário do Google aos quais os participantes responderam. A única alteração feita foi a retirada de acentos para facilitar a implementação dos códigos. Note também que este arquivo contem algumas pessoas duplicada, uma vez que elas responderam mais de uma vez o formulário, porém somente a ultima participação foi considerada.
- *labels.csv*: Contem as respostas ao formulário, mas já transformadas em números e sem repetições de participantes.
- *custos.csv*: Contem os valores de custo entre os participantes, matriz de custo utilizada no algoritmo de Húngaro.
- *designacoes.csv*: Contem os participantes bem como suas respectivas designações em cada caso. C1 representa o resultado do primeiro caso e assim sucessivamente. Quando nada esta presente na designação significa que a pessoa foi designada a si mesma.
- *perguntas.pdf*: Contem as perguntas como apresentadas aos participantes.

References

- [1] Python 3.6. Users manual and reference manual.
- [2] MATLAB R2018. Documentation.
- [3] "Munkres' Assignment Algorithm, Modified for Rectangular Matrices", <http://csclab.murraystate.edu/bob.pilgrim/445/munkres.html>
- [4] Bazaraa, M.S. and Jarvis, J.J. and Sherali, H.D. *Linear Programming and Network Flows*. 2010.

- [5] S. C. Poltroniere, K. C. Poldi, F. M. B. Toledo and M. N. Arenales. A coupling cutting stock-lot sizing problem in the paper industry, *Annals of Operations Research*, 157:91–104, 2008.