

INSTITUTO DE CIÊNCIAS MATEMÁTICAS E DE COMPUTAÇÃO
DEPARTAMENTO DE CIÊNCIAS DA COMPUTAÇÃO
BACHARELADO EM CIÊNCIA DE DADOS

ARTHUR HIRATSUKA REZENDE
CAIO ASSUMPÇÃO REZZADORI
GUSTAVO BLOIS
JOÃO PAULO VAN TOL AMARAL GUERRA

ANÁLISE DE ALGORITMOS DE BUSCA NÃO INFORMADA E
INFORMADA

Relatório apresentado como requisito parcial à
obtenção de aprovação na disciplina de Inteligência
Artificial do curso de Bacharelado em Ciência de
Dados do Departamento de Ciência da Computação
da Universidade de São Paulo

Orientadora: Prof^a Dr^a Solange Oliveira Rezende

SÃO CARLOS

2023

1 INTRODUÇÃO

A modelagem de problemas cotidianos pode ser realizada de diferentes maneiras, utilizando as mais diversas estruturas de dados, bem os algoritmos a serem implementados para solucionar o problema escolhido se apresentam em uma vasta gama de possibilidades.

Desta maneira, é necessário compreender tanto qual a melhor estrutura de dados para representar o comportamento a ser modelado quanto quais algoritmos proverão o melhor desempenho para a tarefa.

Nesse contexto, busca-se modelar a malha metroviária da cidade de São Paulo através de um grafo direcionado, com o intuito de desenvolver um programa com o qual, a partir de uma estação de origem e uma estação de destino, seja possível traçar o melhor trajeto (em função do tempo gasto).

Para viabilizar tal projeto, é feito um estudo comparativo de algoritmos de busca (informada e não informada), para então implementar a melhor opção no programa mencionado acima.

2 METODOLOGIA

Em um primeiro momento são implementados os algoritmos de busca informada A* e Best-First e de busca em profundidade (*Depth First Search* - não informada). Além disso, realiza-se o tratamento dos dados para a modelagem do problema na forma de um grafo, utilizando a estrutura de dados implementada na biblioteca *NetworkX*.

No presente trabalho são apresentados dois estudos distintos, sendo que um deles almeja conclusões acerca da eficiência de diferentes variações dos algoritmos de busca. Para isso, utilizam-se os algoritmos implementados pelos autores e também o algoritmo de busca Dijkstra (implementado na biblioteca *NetworkX*). Também são comparados dois tipos de heurística para o algoritmo A*, sendo uma delas a distância em linha reta entre dois pontos e a outra a distância Haversine, que se utiliza da latitude e longitude dos pontos para calcular a distância considerando a curvatura terrestre.

Já o segundo estudo é relacionado à análise do tempo de execução e propriedades dos caminhos encontrados pelos algoritmos de busca implementados pelos autores, a saber busca em profundidade (não informada) e busca A* (informada) com a heurística sendo a distância Haversine.

Para a realização os estudos mencionados, promove-se um experimento que gera 2500 pontos de origem/partida randômicos e, para cada um dos algoritmos, busca-se um caminho entre os nós, sendo as variáveis de interesse o tamanho e a distância total do caminho e o tempo de execução. A fim de minimizar interferências de hardware e software na medição dos tempos, toma-se 5 vezes a medição do tempo de execução, sendo que os resultados são apresentados como a média dos tempos.

3 RESULTADOS E DISCUSSÕES

Em um primeiro momento expõe-se o estudo dos algoritmos de busca, com uma breve descrição dos algoritmos de busca implementados, cujas medidas de tempo de execução e propriedades do caminho são tema de análise.

Em seguida, de forma análoga, procede-se com o estudo comparativo entre busca informada e não informada, sendo feita uma breve descrição das nuances de cada caso. Além disso, utiliza-se o teste Mann-Whitney para obter um resultado que indique se há diferença estatística entre os resultados obtidos.

Como mencionado, os resultados de medida de tempo são apresentados através da média das 5 aferições de tempo, para ambos os casos.

3.1 ESTUDO DE ALGORITMOS DE BUSCA

No presente tópico são compilados os resultados comparativos entre busca informada e não informada e os resultados de comparativos de diferentes algoritmos para cada um dos tipos de busca.

Na Figura 1 são comparados múltiplos algoritmos de busca, tanto informada quanto não informada, com as implementações feitas pelos autores – exceto para o caso do algoritmo Dijkstra.

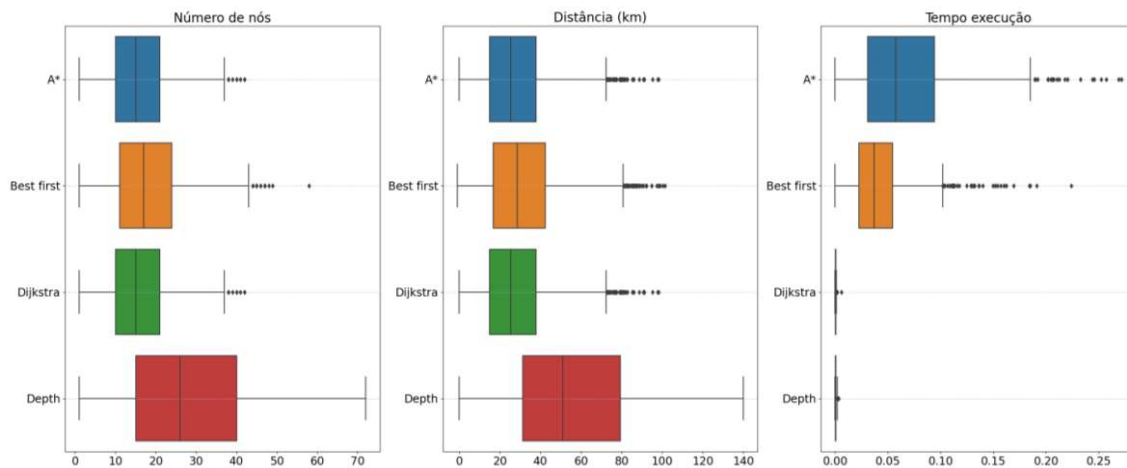


Figura 1 Resultado de Buscas Informadas e Não Informadas

No primeiro caso analisado, é evidente que os algoritmos de busca informada (A^* e Best-First) comportam-se de forma similar se tratando das propriedades dos caminhos (número de nós e distância total), porém, como é esperado, o algoritmo A^* sempre acha os caminhos ótimos, apresentando menores valores de número de nós e distância total.

Observando o tempo de execução, o algoritmo A^* apresentou os piores resultados, e que pode ser parcialmente derivado da falta de otimização na implementação, mas é um resultado esperado, uma vez que a solução é ótima e isso tem um custo atrelado.

Por fim, o algoritmo de busca em profundidade (*Depth*) retornou os caminhos com maiores distâncias e número de nós, porém com tempo de execução baixo comparado com os casos de busca informada.

A seguir é realizada a comparação entre os dois algoritmos de busca não informada (busca em profundidade e Dijkstra). É possível observar, na Figura 2, que a implementação que utiliza a busca com algoritmo Dijkstra mostra resultados melhores na busca de caminhos mais curtos, com casos em que a diferença chega a quase 60 nós entre caminhos encontrados por cada um dos algoritmos.

No caso da distância, a busca em profundidade teve caminhos com a mediana da distância próxima a 60km, contra uma mediana de aproximadamente 25km para os casos de busca Dijkstra. Para ambos os casos o tempo de execução foi similar, de forma que o algoritmo Dijkstra mostra-se a melhor escolha.

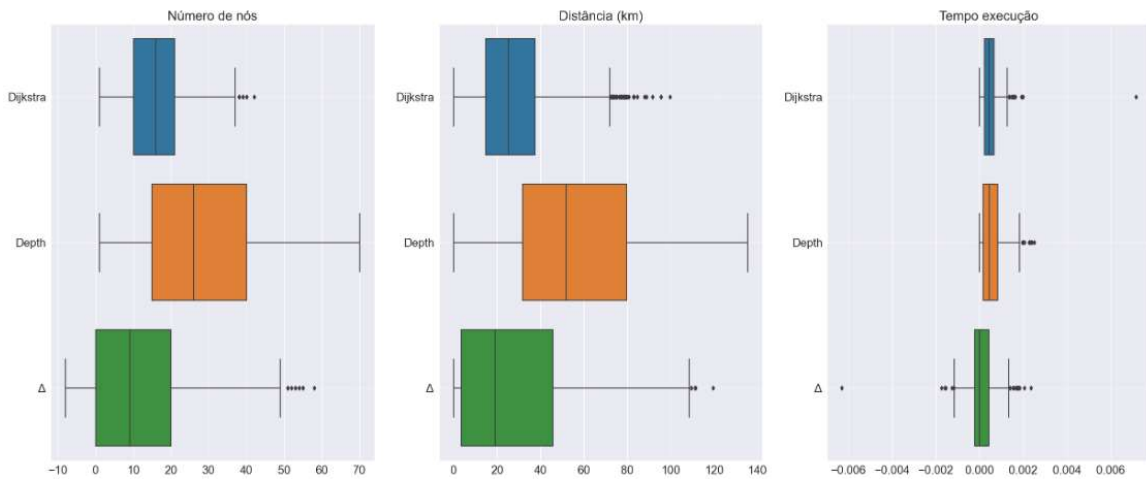


Figura 2 Resultado de Buscas Não Informadas

Na comparação entre os algoritmos de busca informada (Figura 3) são considerados os casos A* e Best-First. No caso do A* são comparadas duas heurísticas: a primeira (Heur0) toma a distância Haversine, que se utiliza da latitude e longitude dos pontos para calcular a distância considerando a curvatura terrestre, e a segunda (Heur1) utiliza a distância em linha reta entre as estações.

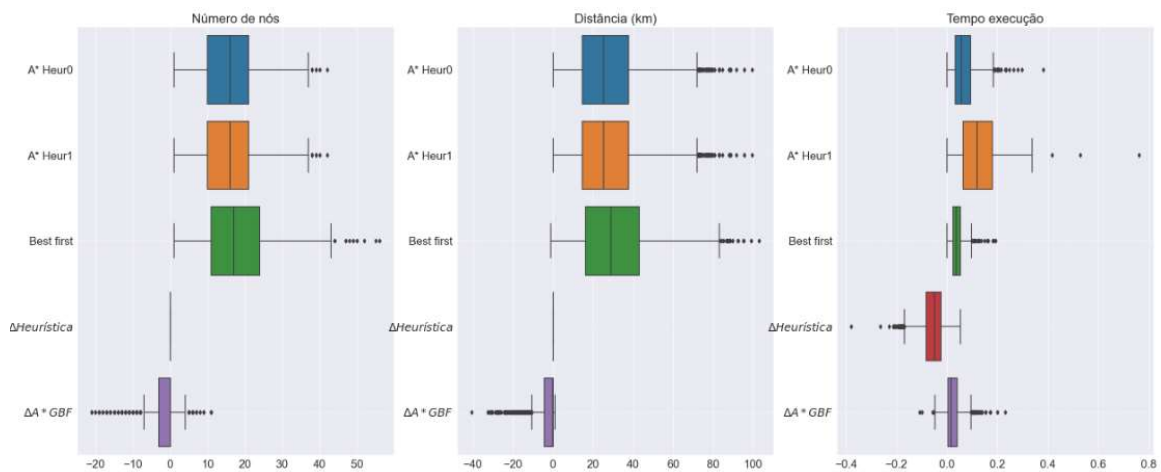


Figura 3 Resultado de Buscas Informadas

É evidente que, no caso do tamanho e distância total dos caminhos, os resultados são similares para as duas heurísticas escolhidas e entre algoritmos. Observando os tempos de execução observa-se que a heurística que tomou a distância Haversine obteve melhores resultados quando comparado com a heurística da distância euclidiana.

Portanto, é evidente que a escolha de uma heurística adequada, além de propiciar a convergência da solução do caminho ótimo mais rapidamente, também exerce influência considerável no tempo de execução do algoritmo.

Comparando os algoritmos A* e Best-First, nota-se que o último apresentou tempo de execução menor, porém ressalta-se que nem sempre apresentou o caminho ótimo, como é evidente ao observar tamanho e distância de caminhos maiores que os obtidos utilizando-se o algoritmo A*.

3.2 COMPARATIVO BUSCA INFORMADA E NÃO INFORMADA

No tópico atual realiza-se o comparativo entre os dois tipos de busca, para isso utilizando o algoritmo de busca informada A* e o algoritmo de busca em profundidade (*Depth*) não informada.

São analisados o tamanho e distância dos caminhos e o tempo de execução, sendo que para cada uma das análises é calculado o p-valor do teste Mann-Whitney, que indica se há diferença estatística entre os resultados obtidos – a hipótese nula é que os resultados são derivados de uma mesma distribuição.

Nas análises abaixo exibe-se a distribuição das observações na plotagem mais à esquerda e as imagens do centro e da direita são derivadas da subtração dos resultados obtidos pelo algoritmo de busca em profundidade dos resultados obtidos pelo algoritmo A*.

Na Figura 4 são apresentados os resultados obtidos relativos ao tamanho dos caminhos. Na imagem à esquerda é observa-se que o p-valor calculado é arredondado para 0, ou seja, existe evidência forte de diferença estatística entre os resultados – o que fica evidente observando as distribuições obtidas.

Na imagem do meio observa-se que aproximadamente 600 caminhos (de 2500) são iguais, não importando o algoritmo utilizado. Porém, observando a figura à esquerda, fica claro que o algoritmo A* conduziu a caminhos menores na maior parte dos casos - conclusão corroborada observando a imagem mais a direita, que explicita que a maior parte da diferença entre caminhos é negativa, ou seja, os caminhos obtidos pela busca em profundidade são maiores que os obtidos na busca A*.

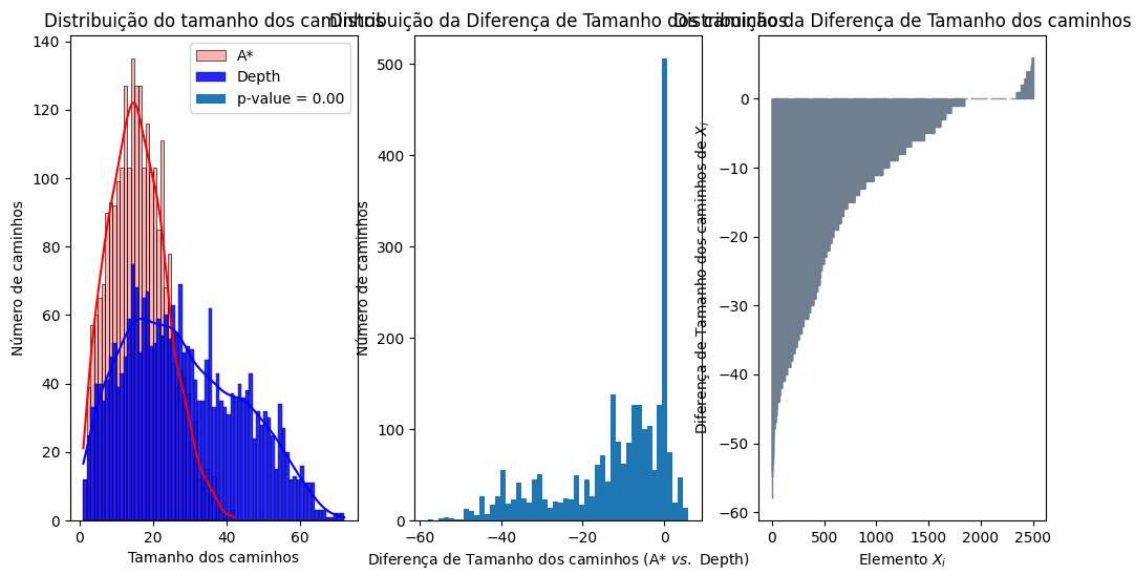


Figura 4 Comparativo buscas – Tamanho dos caminhos

De forma análoga, é conduzida uma análise referente a distância total dos caminhos, com os resultados exibidos na Figura 5. Assim como no caso anterior, o valor p calculado foi aproximado como 0, explicitando diferença entre os resultados apresentados pelos dois tipos de busca.

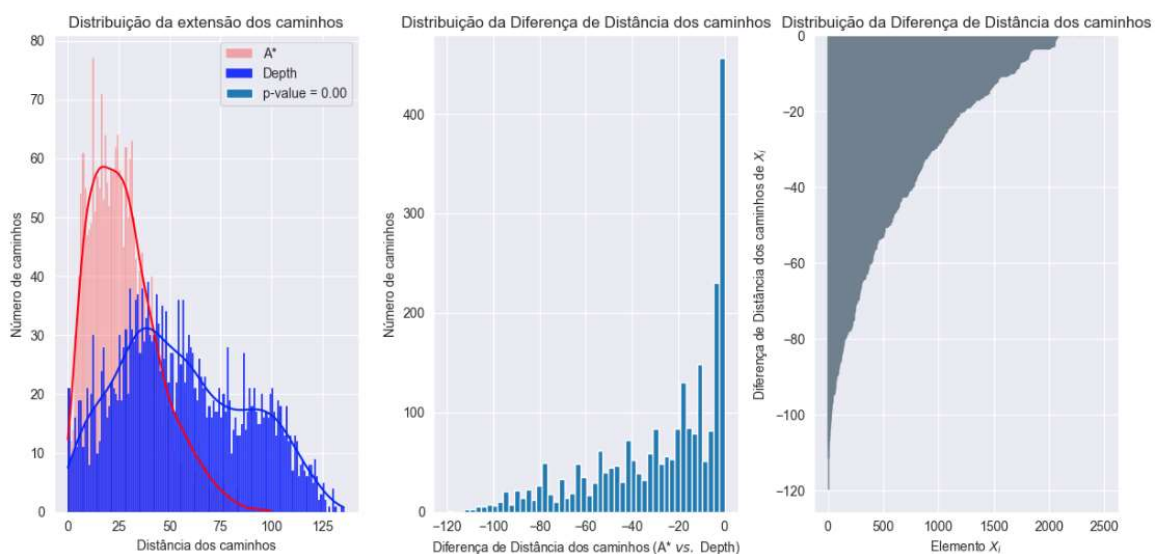


Figura 5 Comparativo buscas – Tamanho dos caminhos

Observando as figuras central e a direita, é evidente que os caminhos obtidos pelo algoritmo A* sempre são ótimos, visto que não houve sequer um caso que a busca não informada retornou um caminho mais curto. Vale ressaltar que nem sempre o caminho mais curto é o que percorre menos nós, de forma que os

resultados dessa análise não invalidam os resultados do comparativo de tamanho total dos caminhos.

Por fim, na Figura 6 são apresentados os resultados obtidos relativos ao tempo de execução. Ressalta-se, na figura à esquerda, que é utilizada a escala log para o eixo x. De forma similar aos casos apresentados, o resultado do teste de Mann-Whitney indicam, com forte evidência, a existência de diferença estatística entre os casos de busca informada e não informada.

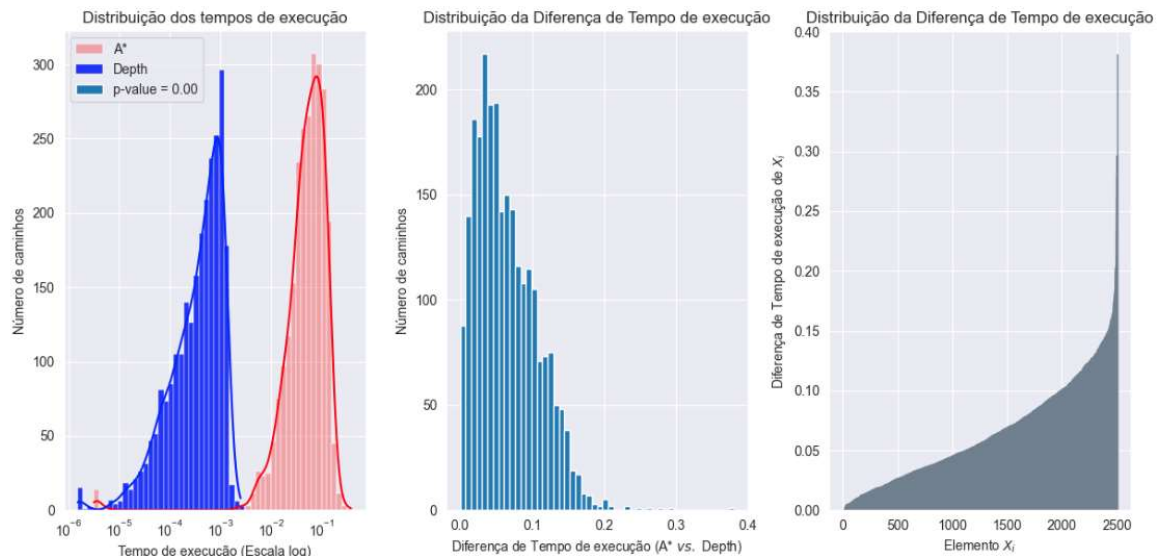


Figura 6 Comparativo buscas – Tamanho dos caminhos

Atentando-se às figuras central e à direita, fica claro que o custo de encontrar um caminho ótimo sempre é refletido no tempo de execução, com o algoritmo A* sendo sempre mais custoso – com diferença de 2 ordens de grandeza, como explicitado na imagem à esquerda.

Na Tabela 1 são compilados os resultados da comparação entre os dois tipos de busca. Observando as proporções (A^* / Depth) dos valores encontrados, é possível observar que os caminhos entre dois nós encontrados pela busca A* resultaram em valores em torno de 68% e 60% menores que os valores apreendidos da busca em profundidade para número de nós e distância, respectivamente, porém com tempo de execução médio 242 vezes maior.

No extremo, o algoritmo A* obteve um caminho de mesma distância que o algoritmo de busca em profundidade, apresentando casos com até 1.8 vezes o número de nós que o caminho encontrado pela busca em profundidade. Esse

resultado explicita que nem sempre o melhor caminho é o que passa pelo menor número de nós. Em relação ao tempo de execução, no pior caso, o A* demorou 1878 vezes mais que o algoritmo de busca em profundidade.

Tabela 1 Resultados comparativos – Busca Informada vs Não Informada

	Caminhos		Tempo de Execução
	Tamanho	Distância	
Proporções (A* / Depth)			
Média	0.678	0.603	242.572
Mínimo	0.052	0.016	0.885
Máximo	1.800	1.000	1878.485
Desvio Padrão	0.295	0.290	228.103
Diferenças em valores absolutos (A* - Depth)			
Média	12.321	17.813	0.066
Máxima	59.000	121.971	0.304

Observando a Figura 7, que detalha a distribuição da proporção dos resultados, fica claro, na imagem esquerda, que grande parte dos resultados são iguais, com aproximadamente 20% dos caminhos com mesmo número de nós e aproximadamente 17% dos caminhos com mesma distância.

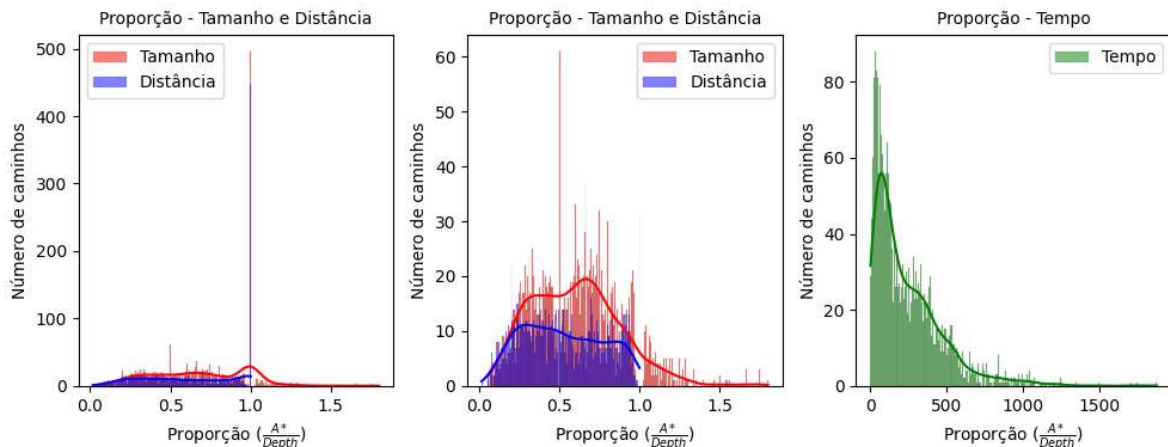


Figura 7 Distribuição da proporção dos resultados

Omitindo os resultados de caminhos iguais obtém-se a imagem central, na qual é possível observar que são poucos os caminhos obtidos com a busca A* que tem mais números de nós do que a busca em profundidade, com uma distribuição relativamente uniforme entre o intervalo [0,1] para as proporções de tamanho e distância dos caminhos.

Em relação à proporção do tempo de execução, retratada na figura direita, é possível observar o fato que 75% dos casos que utilizam a busca A* gastam até 375 vezes mais tempo que a busca em profundidade, com a 50% dos resultados gastando até 171 vezes mais tempo e média de tempo gasto 242 vezes maior.

3.3 IMPLEMENTAÇÃO PARA ROTA DO METRÔ

Uma vez concluído o estudo sobre os algoritmos de busca, é feita a implementação do programa para cálculo de rota. Utiliza-se a estrutura de grafos implementada na biblioteca *NetworkX*, com a busca A*, com os resultados da rota obtida e exibida para o usuário na Figura 8 e o trajeto detalhado que é exibido para o usuário retratado na Figura 9.

A validação foi feita comparando os resultados do programa com os obtidos utilizando a rota calculada pelo Google Maps, com grande parte dos resultados retornando tempo de trajeto bastante similar, e em quase a totalidade de verificações obteve-se a mesma rota.

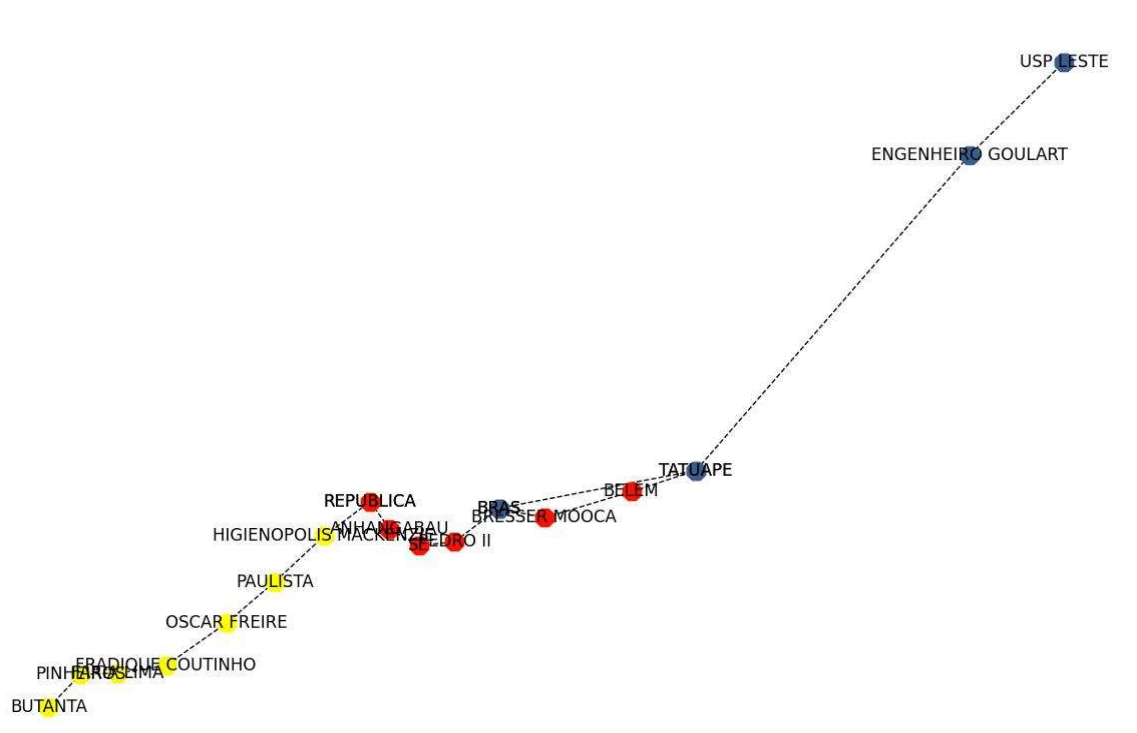


Figura 8 Rota exibida para o usuário

```

Itinerário resumido:
Embarque na linha:      4-AMARELA      Estação:  BUTANTA
Baldeação para linha:  3-VERMELHA     Estação:  REPUBLICA
Baldeação para linha:  12-SAFIRA      Estação:  TATUAPE
Desembarque na linha:  12-SAFIRA      Estação:  USP LESTE
Distância estimada do trajeto: 31.20km

Duração aproximada do trajeto 0 horas 55 minutos 47 segundos

Itinerário
Linha:  4-AMARELA      Estação:  BUTANTA      -> Estação:  PINHEIROS
Linha:  4-AMARELA      Estação:  PINHEIROS    -> Estação:  FARIA LIMA
Linha:  4-AMARELA      Estação:  FARIA LIMA   -> Estação:  FRADIQUE COUTINHO
Linha:  4-AMARELA      Estação:  FRADIQUE COUTINHO -> Estação:  OSCAR FREIRE
Linha:  4-AMARELA      Estação:  OSCAR FREIRE  -> Estação:  PAULISTA
Linha:  4-AMARELA      Estação:  PAULISTA     -> Estação:  HIGIENOPOLIS MACKENZIE
Linha:  4-AMARELA      Estação:  HIGIENOPOLIS MACKENZIE -> Estação:  REPUBLICA

MUDANÇA DE LINHAS NA ESTAÇÃO:  REPUBLICA      Linha de origem: 4-AMARELA ---> Linha destino: 3-VERMELHA
Linha:  3-VERMELHA      Estação:  REPUBLICA      -> Estação:  ANHANGABAU
Linha:  3-VERMELHA      Estação:  ANHANGABAU    -> Estação:  SE
Linha:  3-VERMELHA      Estação:  SE            -> Estação:  PEDRO II
Linha:  3-VERMELHA      Estação:  PEDRO II      -> Estação:  BRAS
Linha:  3-VERMELHA      Estação:  BRAS          -> Estação:  BRESSER MOOCA
Linha:  3-VERMELHA      Estação:  BRESSER MOOCA -> Estação:  BELEM
Linha:  3-VERMELHA      Estação:  BELEM         -> Estação:  TATUAPE

MUDANÇA DE LINHAS NA ESTAÇÃO:  TATUAPE      Linha de origem: 3-VERMELHA ---> Linha destino: 12-SAFIRA
Linha:  12-SAFIRA      Estação:  TATUAPE      -> Estação:  ENGENHEIRO GOULART
Linha:  12-SAFIRA      Estação:  ENGENHEIRO GOULART -> Estação:  USP LESTE

```

Figura 9 Trajeto detalhado exibido para o usuário

4 CONCLUSÕES

Com amplas aplicações, a tarefa de busca se faz presente em praticamente qualquer implementação computacional, fato esse que ressalta a importância do estudo e compreensão do tópico. Desta maneira, o presente estudo busca tanto avaliar variações na implementação da busca, utilizando variações de busca informada e não informada, bem como busca elucidar o impacto da adoção de diferentes heurísticas no caso do algoritmo A*.

Em ambos os estudos apresentados, os resultados são depreendidos de um experimento com 2500 caminhos aleatórios, com tomada de tempo aferida 5 vezes, sendo utilizada a média das aferições, de forma a conferir robustez ao estudo.

No primeiro estudo são comparados diversos algoritmos, sendo evidente que os algoritmos de busca informada obtêm caminhos mais curtos e, portanto, soluções

ótimas. Porém, observando as médias dos tempos de execução, fica claro que o custo de se achar uma solução ótima é refletido no tempo gasto.

Comparando-se somente os casos de busca não informada, o algoritmo implementado na biblioteca *NetworkX* claramente retornou caminhos menores quando comparado com a implementação feita pelos autores da busca em profundidade. Tal fato pode ser por otimização da implementação, mas também se deve à característica intrínseca da busca em profundidade de nem sempre achar o menor caminho.

No caso de busca informada, compara-se duas heurísticas de cálculo de distância, Haversine e Euclidiana, sendo que os resultados apresentados só divergem no tempo de execução, com as propriedades dos caminhos se mantendo. Portanto, conclui-se que a escolha adequada da heurística leva à convergência da solução ótima de forma otimizada.

Nesse mesmo estudo, o algoritmo Best-First apresentou resultados positivos, com tempo de execução menor que as implementações do A*, porém nem sempre retornando soluções ótimas, como fica evidenciado no tamanho dos caminhos e em sua distância total.

No segundo estudo, comparando os algoritmos de A* e busca em profundidade, é evidente que o caso de busca informada conduz a resultados ótimos, com caminhos tão ou mais eficientes em todos os casos. Porém, analisando-se o tempo de execução, observa-se uma diferença de até duas ordens de grandeza, evidenciando que a busca por soluções ótimas em problemas complexos pode ser inviabilizada, de forma que o conhecimento de alternativas – como buscas não informadas - mostra-se extremamente relevante.

A análise das proporções dos resultados demonstra que os caminhos encontrados pela busca A* são sempre ótimos, e mesmo quando passam por mais nós têm menores distâncias. O tempo de execução claramente é bem maior no A*, mas para a aplicação pretendida essa diferença não se faz notória para o usuário, justificando a adoção desse algoritmo como motor de busca.