

Sistemas Ubíquos - Relatório Intermédio

DEI - Mestrado em Engenharia Informática

Francisco Fernandes 2018278239

João Marcelino 2018279700

5 de maio de 2022

1 Introdução

Este relatório tem como objetivo descrever o trabalho realizado ao longo do projeto no âmbito da disciplina de Sistemas Ubíquos sobre a acessibilidade das habitações aos transportes públicos e às diferentes zonas da cidade de Coimbra. Com este projeto pretende-se decompor a cidade em diferentes zonas industriais/comércio/lazer e calcular a acessibilidade de cada casa a todas essas zonas podendo, assim, identificar zonas de baixa acessibilidade e recomendar potenciais novas linhas de transportes públicos.

Primeiramente, serão abordadas as características dos dados recebidos inicialmente pelo grupo assim como o pré-processamento realizado sobre eles. De seguida será descrito o processo de clustering realizado e as decisões feitas acerca dos algoritmos e parâmetros utilizados. Por fim, será feita uma pequena introdução ao algoritmo de pontuação de acessibilidade desenvolvido até à data, seguida de uma conclusão.

2 Dados

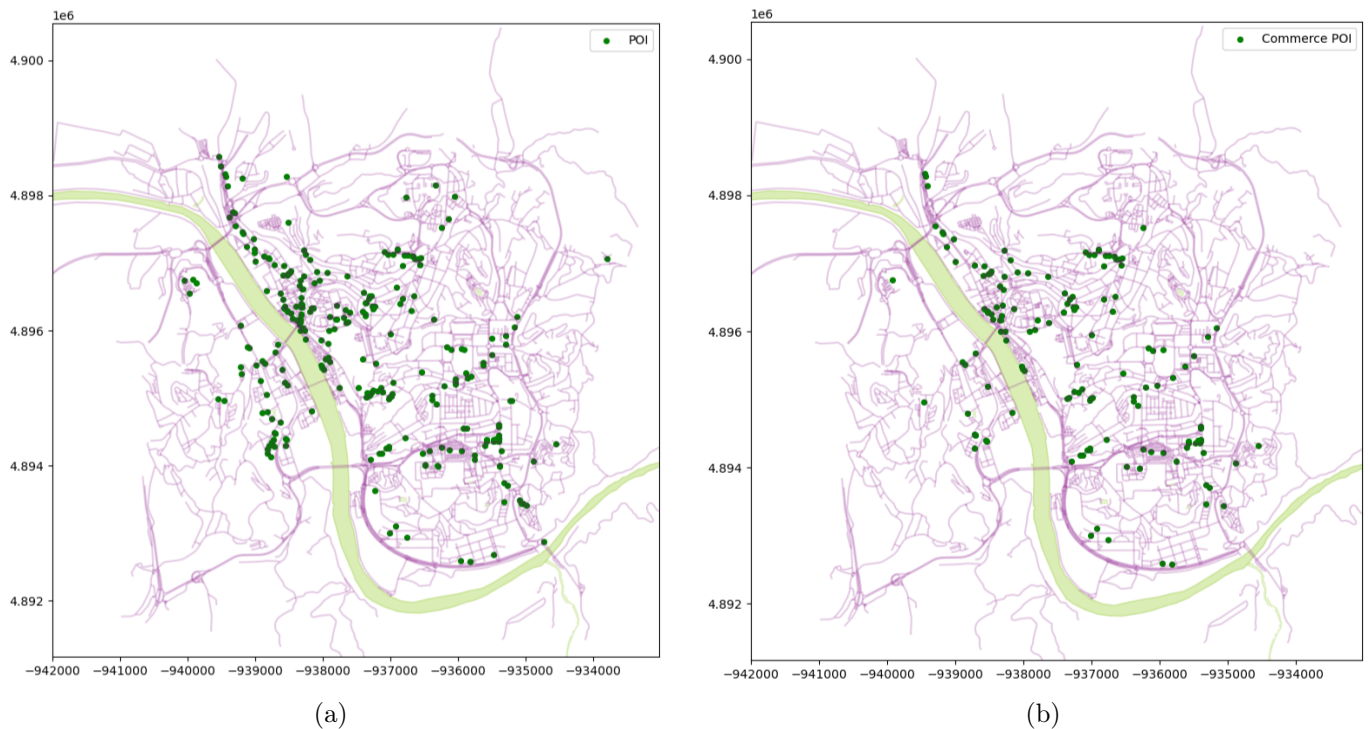


Figura 1: (a) Pontos de interesse de todas as categorias. (b) Pontos de interesse da categoria de comércio. De acordo com a taxonomia definida, estes compõem a maioria do dataset.

Categoria	Nº de pontos	Subcategorias
Comércio	145	<i>restaurant, cafe, pub, supermarket, fast_food, atm, bakery</i>
Educação	6	<i>school, library, kindergarten</i>
Infraestrutura	108	<i>police, post_box, recycling, post_office, parking, telephone, bank, fuel, information, toilet</i>
Lazer	7	<i>theatre, museum, cinema, attraction, nightclub</i>
Saúde	14	<i>pharmacy, hospital</i>

Tabela 1: Taxonomia e número de pontos de interesse que compõe cada categoria.

Para este projeto foi fornecido um dataset com 280 pontos de interesse localizados na cidade de Coimbra. Cada ponto encontrasse identificado com o nome, localização geográfica e tipo de ponto. A partir da informação acerca do tipo de cada ponto, foi possível agrupar o dataset em 5 grupos de classes diferentes e obter uma distribuição como aquela visível na tabela 1.

Para além do agrupamento realizado sobre os pontos de interesse, foi também feita a exclusão de pontos que se encontravam fora o centro da cidade de Coimbra, acabando então por obter uma disposição de pontos igual à da figura 1.

Apesar de o foco deste projeto estar na cidade de Coimbra, o objetivo final será tornar os processos e algoritmos utilizados na averiguação das zonas e acessibilidade da cidade modulares o suficiente para que possam ser aplicados a outras cidades (ex.: Lisboa, Porto, etc).

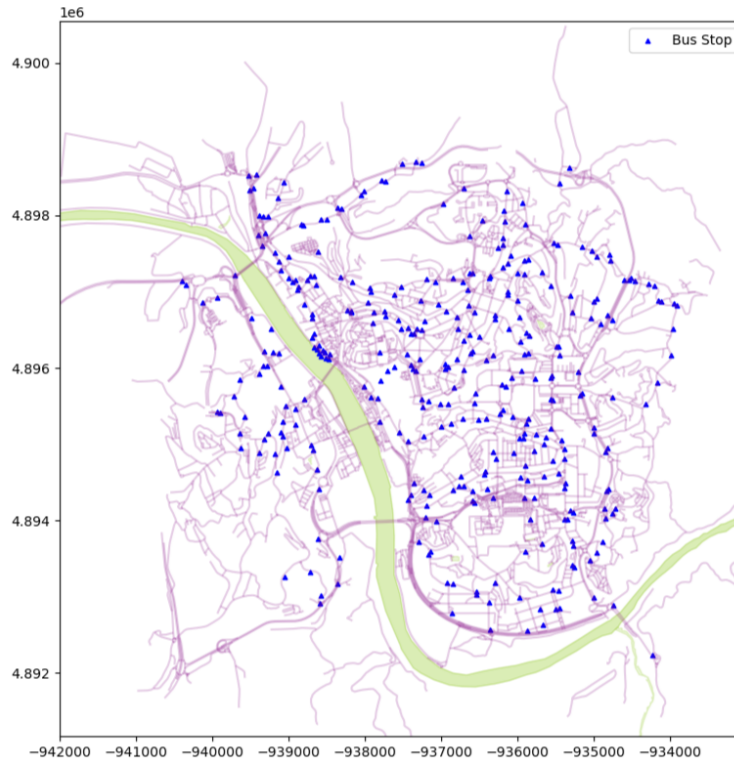


Figura 2: Paragens de autocarros da cidade de Coimbra.

Foram também recolhidas manualmente as localizações de paragens de autocarros através da informação disponibilizada pelo OpenStreetMap [1]. Este dataset conta com a localização de 352 e, apesar de não conter horários ou rotas de autocarros, permitiu ao grupo prototipar o algoritmo que calcula o nível de acessibilidade de cada zona da cidade.

3 Clustering

Depois da recolha de todos os dados necessários, procedeu-se à identificação de zonas de comércio da cidade. Ao invés de calcular zonas de todas as categorias, preferiu-se desenvolver e estudar os algoritmos de clustering focando-se apenas em POIs de comércio. Desta forma conseguiu-se reduzir

a complexidade desta tarefa e, no final, obter um algoritmo aplicável e eficaz para qualquer tipo de zona.

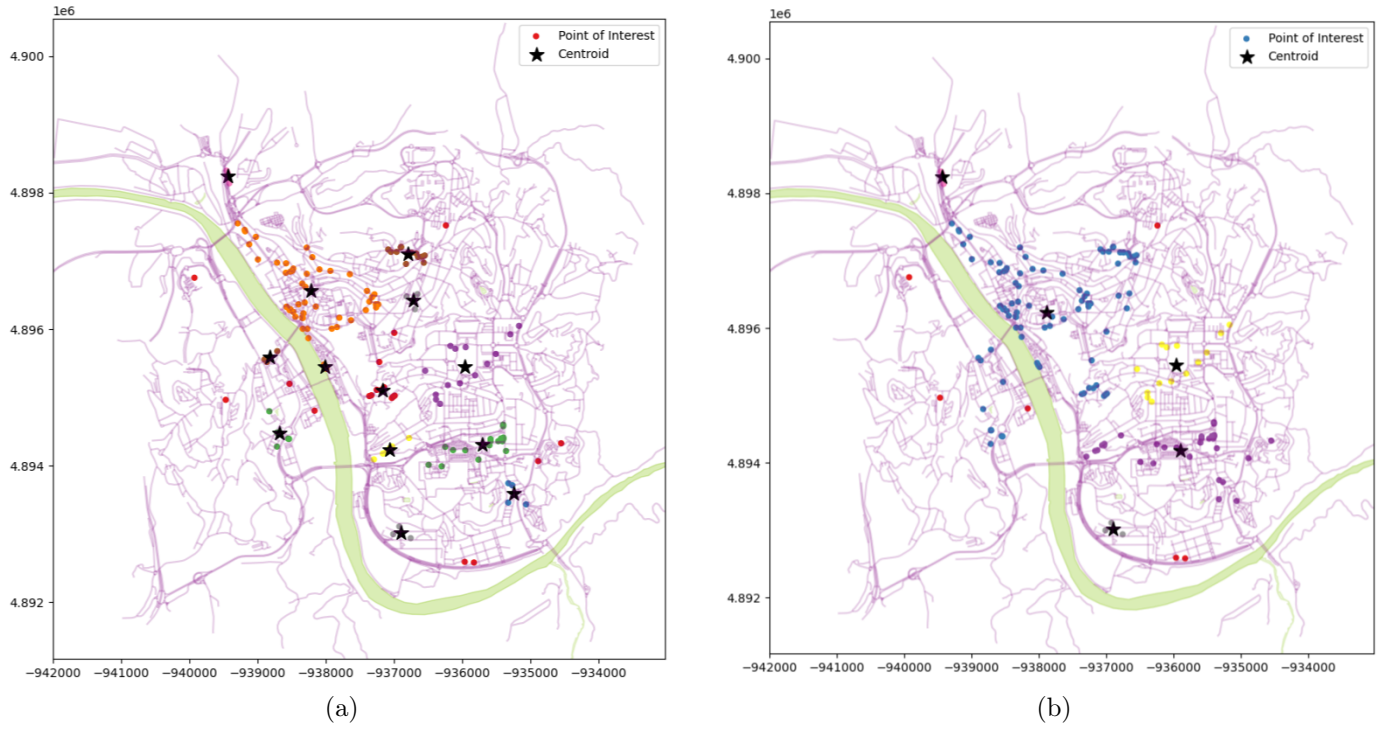


Figura 3: (a)Resultado do DBSCAN com $\epsilon = 400m$ e $n = 3$. (b)Resultado do DBSCAN com $\epsilon = 500m$ e $n = 3$. O cluster a que pertence cada ponto encontra-se sinalizado pela cor e os centroides são sinalizados pelos marcadores em estrela.

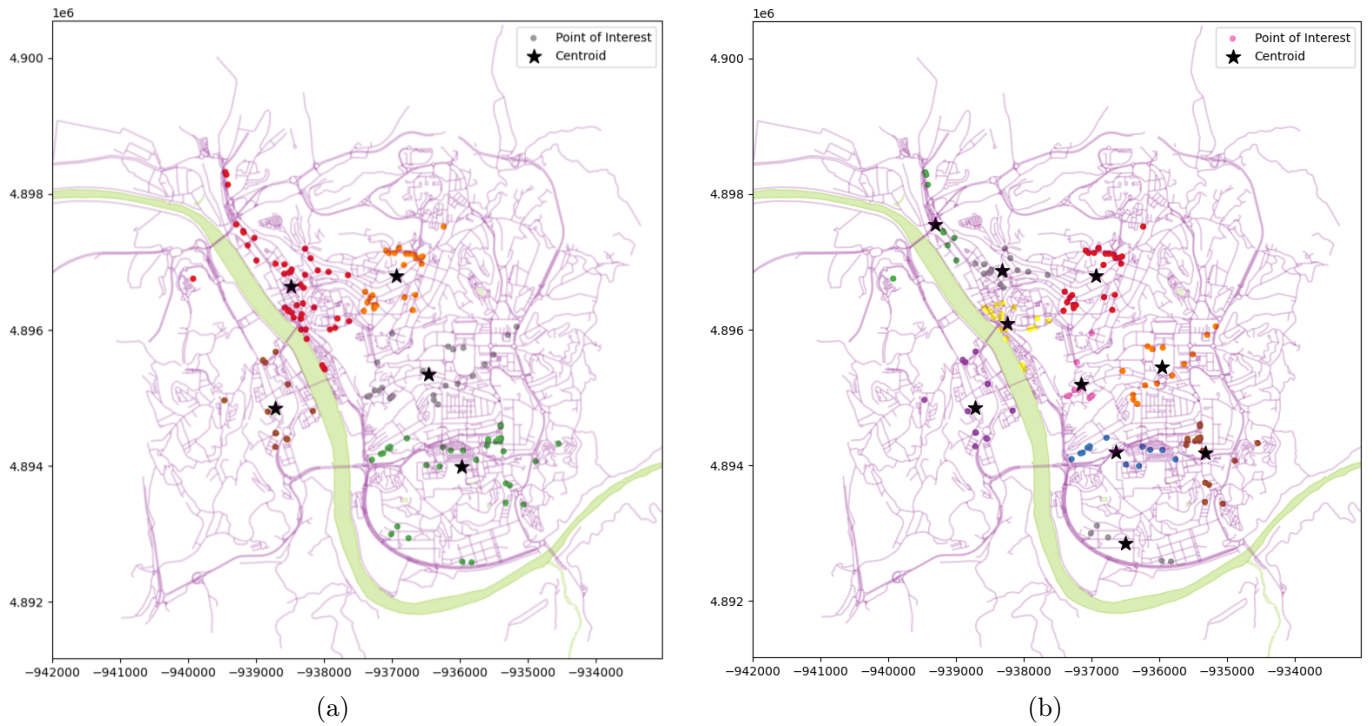


Figura 4: (a)Resultado do AGNES com $K = 5$. (b)Resultado do AGNES com $K = 10$.

Para o clustering dos POIs foram considerados três algoritmos diferentes: AGNES, K-Means Clustering e DBSCAN. Enquanto que os dois primeiros necessitam do n^o de clusters à priori

como hiperparâmetros, o DBSCAN requer um raio de clustering e um n^o mínimo de pontos que constituem um aglomerado.

Como é possível observar na figura 4, os algoritmos AGNES e K-Means por vezes consideram dois conjuntos de pontos bastante distintos como sendo um único aglomerado, posicionando o *centroid* no ponto médio entre esses dois conjuntos. Este efeito é indesejável visto que idealmente todos os pontos que constituem um aglomerado deveriam encontrar-se a distâncias que se possam percorrer a pé. Outra desvantagem destes dois algoritmos é o facto de não suportarem a deteção de anomalias que podem influenciar em muito o cálculo da localização de cada aglomerado.

Por outro lado, o DBSCAN requer uma parametrização cuidadosa no que toca ao raio de vizinhança. Uma das formas abordadas para a validação dos resultados produzidos pelo DBSCAN foi o estabelecimento de uma distância máxima entre quais quer dois pontos do mesmo *cluster* (ex.: 500m) e uma distância mínima obrigatória entre *centroids* de aglomerados diferentes. Através desta abordagem pode-se considerar a aglomeração apresentada na figura 3(b) como não ideal.

Pelas razões referidas acima, o algoritmo que se mostrou mais apropriado para o contexto deste projeto foi o DBSCAN com $\epsilon = 400m$ e $n = 3$.

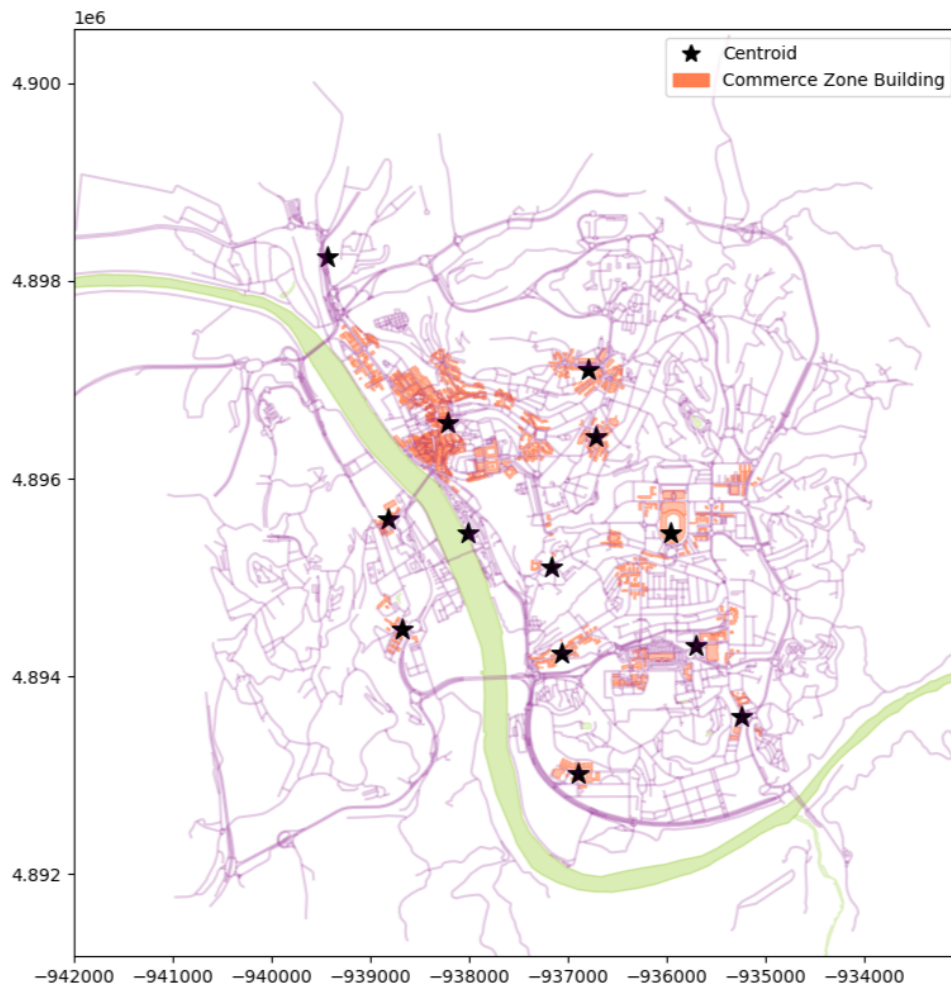


Figura 5: Interseção dos clusters apresentados na figura 3(a) com um dataset de edifícios.

Uma das possíveis aplicações da aglomeração de pontos de interesse é a visualização das varias zonas de comércio que compõem a cidade tal como é mostrado na figura 5. Este mapa foi construído através da interseção de cada ponto de interesse (pertencente a um *cluster*) com *buffer* de 100m e um dataset de edifícios disponibilizados pelo OpenStreetMap [1].

4 Travel Time e Pontuação

Após a escolha de algoritmo de cluster, foi implementado um algoritmo preliminar com o objetivo de calcular a distância de qualquer ponto da cidade até ao cluster mais próximo. Este algoritmo consiste em dividir a cidade numa rede de quadrados idênticos em tamanho e calcular duas distâncias: distância de um dado ponto da grelha à paragem de autocarro mais próxima; distância da paragem mais próxima ao centro do *cluster* mais próximo da paragem.

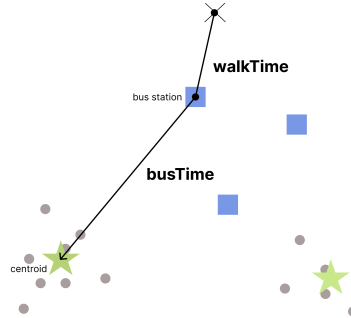


Figura 6: Exemplificação do algoritmo. O tempo total é constituído pela soma de tempo de deslocamento a pé e tempo de viagem de autocarro. Foi considerada uma velocidade de caminhada de 5km/h e velocidade de autocarro de 30km/h .

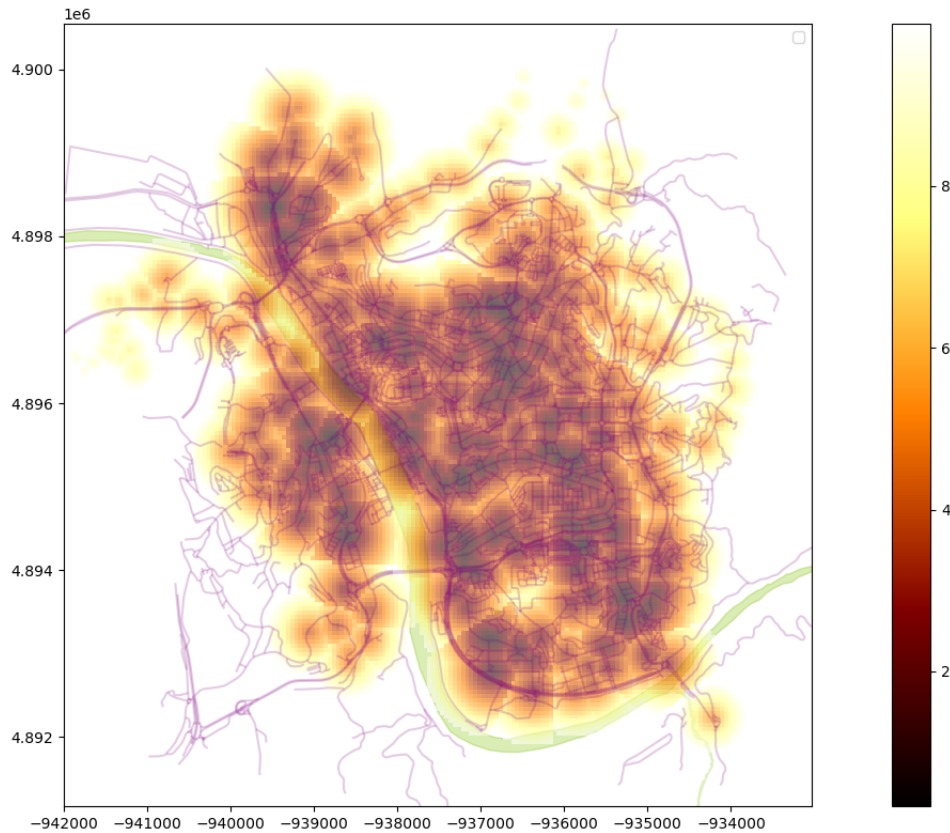


Figura 7: Heatmap dos tempo de viagem a zonas de comércio inferiores a 10 minutos.

Ao invés de contar com trajetórias em linha reta, os algoritmos desenvolvidos futuramente irão contar com trajetórias e tempos de viagem fornecidos pelas APIs *Distance Matrix* e *Directions*

disponibilizadas pela Google. De modo a diminuir os cálculos e futuros pedidos à API escolhida para descobrir a distância mais curta entre pontos, foi decidido pelo grupo diminuir a resolução da grelha de 50m para 100m. Apesar de esta poder provocar piores resultados para a continuação do trabalho, foi uma escolha necessária para controlar os custos.

4.1 Função de Pontuação

De modo a avaliar quais os clusters mais interessantes por quadrado, foi utilizado o algoritmo anterior para descobrir quais os clusters mais próximos do centro do quadrado, e é escolhido o melhor cluster de acordo com a função de score que é escolhida previamente. Até ao momento, foram testadas as seguintes funções de pontuação:

- Inverso do tempo de viagem até ao cluster
- Inverso do tempo de viagem até ao cluster * número de pontos de interesse no cluster

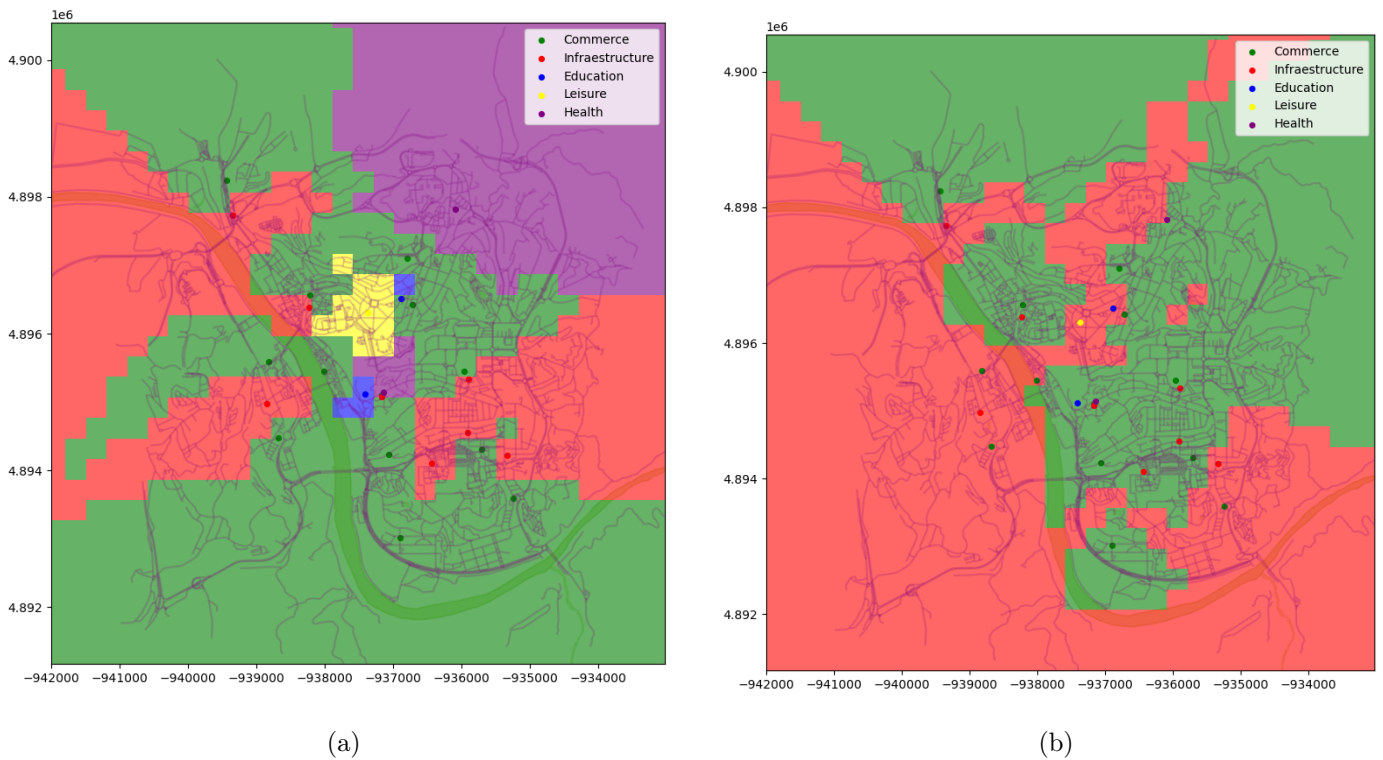


Figura 8: (a) Resultado da função de pontuação inverso do tempo de viagem. (b) Resultado da função de pontuação inverso do tempo de viagem * número de pontos de interesse

5 Trabalho Futuro

Como trabalho futuro, o nosso grupo irá utilizar a Google API Maps Platform para melhorar o cálculo de distâncias para os clusters, utilizando diferentes métodos de viagem como transportes públicos e/ou particulares. Outro foco seria implementar novas funções de pontuação e melhorar as trabalhadas até ao momento, de modo a melhorar a qualidade de escolha para cada quadrado.

Para finalizar iremos alterar o foco do projeto para a zona metropolitana de Lisboa, com o objetivo de termos pontos de interesse mais diversificados, mais clusters e mais meios de transporte, e assim, mais dados para testar as nossas hipóteses.

Referências

- [1] “Openstreetmap - <https://www.openstreetmap.org/>,” 2022.