

# Análise e Comparação Comportamental de traces de Redes Veiculares(VANETs)

**Túlio Nunes Polido Lopes<sup>1</sup>, Felipe Domingos da Cunha<sup>1</sup>**

<sup>1</sup>Instituto de Ciências Exatas e Informática

Pontifícia Universidade Católica de Minas Gerais (PUC-MG)

Av: Dom José Gaspar, 500 Coração Eucarístico - Belo Horizonte - MG

tulionp.lopes@gmail.com, felipe@pucminas.br

**Resumo.** Este artigo trata de uma análise estatística da movimentação de veículos em grandes cidades. O objetivo da análise é contribuir no entendimento do padrão de movimentação veicular nas metrópoles. Este estudo foi feito utilizando duas bases de dados de cidades diferentes. A primeira de Roma, cujos objetos de estudo são táxis, e a segunda, de Dublin, cujos objetos de estudo são ônibus. O objetivo deste estudo é uma investigação da movimentação veicular, com intuito de entender melhor a dinâmica de veículos em cidades para uso em decisões estratégicas de controle de trânsito e otimização de construção de rotas de aplicativos de transporte ou veículos autônomos.

## 1. Introdução

Desde a popularização de veículos automotores no século XX, diversas cidades, em especial as grandes metrópoles, vêm sofrendo com a crescente demanda por controle de trânsito. No Brasil, a cidade de São Paulo é o ápice desse problema, e, mesmo utilizando frequentes rodízios de veículos e alterações nas vias, não há mudança significativa Cintra [2013]. É possível encontrar na literatura trabalhos cujo objetivo é a solução a longo prazo desse obstáculo. E com o advento de veículos autônomos, é plausível que serão utilizadas redes veiculares para automação e controle do trânsito.

Redes Veiculares consistem numa rede computacional de veículos que trafegam por uma área (normalmente considera-se uma cidade) cujos automóveis se comunicam através de pontos de acesso distribuídos de forma a cobrirem toda região Liang et al. [2015]. O uso dessa tecnologia para controle automatizado permite uma melhora no tempo de resposta dos veículos, gerando um progresso na sua circulação.

Para a implementação de um sistema para controle e comunicação de movimentação de automóveis, o ideal é avaliar a melhor forma de desenvolvimento dos protocolos dessa rede, posto que o estudo prévio da interação veicular otimizaria a seleção de suas necessidades primordiais. Esses protocolos, para um funcionamento eficaz, precisam levar em conta informações como velocidade dos veículos, tipos de veículos, pontos de maior entrada/saída, entre outras. Mais um ponto fundamental é a diferença de tráfego entre cidades e quais implicações essas mudanças comportamentais podem gerar no funcionamento de uma rede veicular. Sabendo dessa meta, são de grande valia estudos comportamentais da movimentação veicular, pois esses possibilitariam uma visão mais assertiva de como construir um mecanismo de comando dos automóveis que permita um fluxo mais eficiente.

Portanto, o presente estudo, com o intuito de entender melhor como se dá a movimentação de diferentes tipos de veículos em grandes metrópoles, buscou desenvolver uma metodologia para exploração de *traces* veiculares. Foram investigados dados de GPS para estabelecer quais padrões de comportamento podem ser identificados para cada base considerando as métricas selecionadas. Essa investigação foi feita utilizando-se uma base referente à cidade de Roma (Itália) e outra à cidade de Dublin (Irlanda), cujos conjuntos de automóveis são compostos por táxis e ônibus, respectivamente. Essas bases foram escolhidas devido à sua facilidade de acesso e manuseio.

Este trabalho se encontra dividido da seguinte forma: na Seção 2, os trabalhos relacionados são descritos de acordo com sua importância e influência para o artigo desenvolvido. Em seguida, na Seção 3, é descrita a metodologia utilizada para analisar as bases, onde é retratado a forma de disposição dos dados de cada base, as etapas desenvolvidas da análise, a descrição de algoritmos para secção de *traces* e as métricas avaliadas. Na Seção 4 são descritos e analisados os resultados obtidos das análises individuais de cada métrica, assim como da análise espaço-temporal. Na Seção 5 são tiradas conclusões com base nos dados obtidos e descritos possíveis caminhos a serem tomados em trabalhos futuros.

## 2. Trabalhos Relacionados

Nesta Seção são descritos os trabalhos que serviram de base para o desenvolvimento deste estudo, bem como sua importância para o mesmo. Eles tratam de tanto de algoritmos desenvolvidos para a extração de *traces* individuais em dados de GPS contínuos, como é o caso da base de Roma utilizada neste artigo, quanto métricas de avaliação de dados de GPS de redes veiculares. De uma forma geral, estes trabalhos são referências metodológicas para extração de informações do comportamento veicular descrito por suas base de dados.

Dentre os artigos utilizados para obtenção de informações de parâmetros e métricas de análises de *traces* estão: Basta et al. [2016], Celes et al. [2019] e Gainaru et al. [2009]. O primeiro desenvolve uma proposta de um modelo de mobilidade social para ser utilizados em simulações de redes veiculares. Ele se baseia em esferas sociais, onde cada proprietário de veículo se conecta a outros proprietários por uma rede familiar, laboral ou de amizades. O segundo estuda a mobilidade de ônibus intramunicipais e seu impacto na construção de redes veiculares em uma cidade. No último é feita uma proposta de um modelo de mobilidade social, considerando a relevância de simulação microscópica, isto é, simulação de informações como perfil do motorista.

O artigo Gong et al. [2015], foi utilizado como referência para o algoritmo de secção de viagens individuais de *traces* contínuos. Ele estabelece parâmetros e métodos para seccionar dados de GPS de *smartphones* que não estejam divididos em viagens. Também importante para o desenvolvimento do algoritmo de secção de *traces* é o artigo Li et al. [2008] que busca analisar o movimento de pessoas utilizando dados de GPS de celulares com a finalidade de definir quais os pontos de parada durante o trajeto, por exemplo, *shoppings*, lojas, edifícios etc.

Kong et al. [2018] possui informações importantes para a implementação da matriz de origem/destino desenvolvida na atual pesquisa. Nele é proposto um método de simulação de dados de veículos particulares a fim de facilitar a obtenção de *datasets* para

estudos futuros.

Os seguintes artigos serviram de introdução ao tema investigado: Ning et al. [2017] e Cintra [2013]. O primeiro dá uma breve explanação sobre o que são as *Vehicular Social Networks* (VSN) comparando-as com redes sociais e com redes ad-hoc veiculares (VANET). Já o segundo descreve dados sobre o problema do trânsito de São Paulo ainda em 2013, onde há quilômetros de engarrafamentos que leva condutores a perderem até 2 horas por dia parados.

Com as informações obtidas nos estudo citados, o trabalho corrente busca fazer uma análise do comportamento veicular em uma rede, com o objetivo de definir e entender padrões de comportamento dos veículos em uma grande cidade. O conhecimento alcançado pode ser utilizado futuramente de forma auxiliar na construção de uma rede veicular inteligente, que precisará estabelecer o comportamento padrão dos veículos autônomos, de acordo com protocolos, atentando antecipadamente aos movimentos comuns para o trânsito de uma metrópole.

### 3. Metodologia

Esta Seção busca descrever a metodologia utilizada na pesquisa. Ela é dividida em 3 partes, sendo: Base de dados, que descreve os dois *traces* utilizados, com suas características particulares que levaram a abordagens específicas; Ferramentas, que cita e explica quais ferramentas foram utilizadas para o desenvolvimento do estudo, linguagem, bibliotecas, *softwares* e técnicas de programação; Etapas do processo de estudo das bases, que explica como foi feita a limpeza das bases, o funcionamento do algoritmo de divisão de *traces* e o desenvolvimento da Matriz OD, utilizada como fonte de informações importantes acerca dos dados;

#### 3.1. Base de dados

Dois *traces* foram usados durante as análises. Um *trace* de Roma que contém dados de táxis (de 04/02/2014) e um *trace* de Dublin que contém dados de ônibus (de 01/01/2013). No *trace* de Roma cada linha representa um ponto de GPS gravado, possuindo, necessariamente nesta ordem: Um inteiro para representar o ID do veículo; Um *String* contendo a data e hora em que o ponto foi gravado; Um *float* para a longitude; Um *float* para a latitude; Um *boolean* que define se o ponto sofreu calibração ou não. Os veículos do *trace* de Roma são todos táxis da região.

**Quadro 1 - Exemplo dos Dados de GPS de Roma**

ID	time	long-x	lat-y	is-calibrated
101	2014-02-04 05:00:01	12.48884	41.90304	1
101	2014-02-04 05:00:03	12.48875	41.90301	0

Fonte: Do autor.

O *trace* de Dublin contém dados de GPS de ônibus municipais e possui uma maior quantidade de atributos a disposição. Dentre esses novos atributos tem-se: *at stop*, que define se o ônibus está em um ponto de parada, *stop ID*, que identifica a parada atual do veículo, *journey ID*, que identifica o ID da viagem atual, e *line ID*, que identifica a linha do veículo.

**Quadro 2 - Exemplo dos Dados de GPS de Dublin**

Timestamp	line ID	journey ID	long	lat	stop ID	vehicle ID	at stop
1356998403	747	3493	-6236852	53425327	7411	40040	0
1356998405	27	3883	-6233417	53342232	395	33521	0

Fonte: Do autor.

### 3.2. Ferramentas

Neste trabalho foram utilizadas diversas ferramentas com foco em análise de dados. Essas ferramentas, de forma geral, são bibliotecas da linguagem Python<sup>1</sup>, ou ferramentas de apoio externo cuja função é auxiliar no entendimento dos resultados obtidos. Primeiro, serão descritas as bibliotecas utilizadas.

A linguagem escolhida para a extração de informações da base foi Python. Isso se deve ao fato de ser uma linguagem com sintaxe simples e grande apoio comunitário no desenvolvimento de ferramentas de Ciência de Dados. Neste trabalho, utilizou-se quatro bibliotecas, sendo elas: NumPy<sup>2</sup>, Pandas<sup>3</sup>, Matplotlib<sup>4</sup> e Seaborn<sup>5</sup>. A primeira é útil para a aceleração de cálculos matemáticos envolvendo grandes volumes de dados. Seu funcionamento leva em conta a tipagem dinâmica da linguagem que leva a um custo de processamento. A segunda biblioteca, Pandas, é utilizada no carregamento do conjunto de dados. Assim como Numpy, ela promove uma otimização de processamento ao carregar os dados de uma base em *Data Frames*. Além disso, ela possui funções para automatizar a limpeza e balanceamento dos dados, garantindo resultados mais promissores e condizentes com a realidade. As bibliotecas Matplotlib e Seaborn são utilizadas na plotagem de gráficos com os resultados obtidos dos cálculos feitos anteriormente.

Das ferramentas de apoio, foi utilizada OpenStreetMap<sup>6</sup>, uma base de dados contendo catálogos de mapas mundiais para uso gratuito. A outra ferramenta é o leitor de planilhas, que foi utilizado para acelerar pesquisas simples de instâncias e informações existentes nos dados. Também foi utilizada a técnica de Matriz de Origem e Destino. Uma abordagem computacional que nos permite obter mais informações de redes veiculares, focando nos pontos iniciais e finais de viagens. Ela é melhor descrita na Seção 3.3.4, que explica sua lógica e funcionamento.

### 3.3. Etapas de estudo da base

O trabalho proposto se divide em 5 etapas descritas na Figura 1, sendo elas: A limpeza da Base, Análise Inicial, Secção de Traces, Matriz O/D, Análise das Métricas.

O primeiro passo é realizar uma limpeza da base, onde são excluídos os atributos não utilizados e as instâncias problemáticas, como as duplicadas ou com atributos incompletos. A segunda etapa consiste de uma análise inicial onde são utilizadas métricas básicas de estatística como gráficos de densidade e mapas de calor para avaliação do comportamento do trânsito de Roma. Na terceira parte, foi desenvolvido um algoritmo

<sup>1</sup><https://www.python.org/>

<sup>2</sup><https://numpy.org/>

<sup>3</sup><https://pandas.pydata.org/>

<sup>4</sup><https://matplotlib.org/>

<sup>5</sup><https://seaborn.pydata.org/>

<sup>6</sup><https://www.openstreetmap.org/about>

para secção do *trace* de Roma, por não estar dividido em viagens individuais. Seu desenvolvimento é importante para uma melhor avaliação do *trace* de Roma por viabilizar a implementação de uma matriz de origem e destino (Matriz OD). Essa matriz foi desenvolvida na quarta etapa, para uma análise mais bem fundamentada das informações. Com ela é possível fazer a exploração mais aprofundada dos dados, o que possibilita definir as áreas com maior saída e maior entrada de veículos em determinado intervalo de tempo. Ela foi construída com a finalidade de dividir as cidades de Roma e Dublin em áreas quadradas que representassem quartéis, possibilitando uma exploração quantitativa da base através de diversas métricas.

As métricas foram avaliadas na última etapa, onde há uma avaliação estatística dos resultados obtidos e interpretação dos mesmos.

**Figura 1. Fluxograma das Etapas do Trabalho.**



Fonte: Do autor.

### 3.3.1. Limpeza da base

A limpeza da base de dados é feita utilizando basicamente a biblioteca pandas, do Python, citada na Seção 3.2. Essa biblioteca permite a carga completa da base de dados na memória, já possuindo algumas funções para automatizar o processamento de limpeza. Primeiro foram excluídas todas instâncias duplicadas, para evitar que os cálculos feitos posteriormente possuissem um enviesamento. Também foram removidas instâncias com atributos nulos, evitando qualquer problema de cálculo durante o processamento dos dados. Esses erros em instâncias específicas nas bases de dados normalmente ocorrem durante a leitura, como quando um sensor falha e gera uma instância problemática que deve ser removida.

### 3.3.2. Algoritmo para secção do *trace* de Roma

O passo seguinte deste projeto engloba a divisão do *trace* de Roma, que possui uma gravação contínua de dados, por não possuir separação das trajetórias individuais. Ela é feita com a investigação de intervalos de tempo e distância para encontrar de forma aproximada os momentos onde provavelmente houve o fim/início de uma trajetória de táxi. O algoritmo é descrito a seguir.

No Algoritmo 1, o *for* externo (linha 3) percorre a lista de IDs, criando uma nova lista de paradas para cada veículo na lista e um novo limite para o *for* interno. O ciclo interno (linha 6) percorre a lista de pontos do ID representado pelo atributo *key*. Primeiro é calculado a distância entre o ponto *i* e o ponto *i+1*. Caso essa distância ultrapasse o limite, cuja escolha é explicada na Seção 3.3.3, o próximo passo é o cálculo do tempo entre os dois pontos. Caso o limite de tempo também seja ultrapassado, adicionamos o

índice do ponto  $i$  na lista de separações (linha 11). Ao fim de ambos os ciclos, a lista é retornada contendo os índices respectivos a cada ID.

---

**Algorithm 1** Algoritmo de Secção de viagens

---

```

1: Entrada: Um vetor dos IDs ( $Keys$ ) e um dictionary contendo cada ID e seus respectivos
   pontos ( $P$ ).
2: Saída: Uma lista contendo os índices onde devem ocorrer as separações das viagens.
3: for  $key$  in  $Keys$  do
4:    $separator[key] = \text{nova lista}$ 
5:    $numPoints = |P[key]|$ 
6:   for  $i=0$  to  $numPoints - 1$  do
7:      $dist = \text{distancia}(P_i, P_{i+1})$ 
8:     if  $dist > limiteDist$  then
9:        $temp = \text{tempo}(P_i, P_{i+1})$ 
10:      if  $temp > limiteTempo$  then
11:         $separator[key].append(i)$ 
12:      end if
13:    end if
14:   end for
15: end for
16: Retorna  $separator$ ; FIM

```

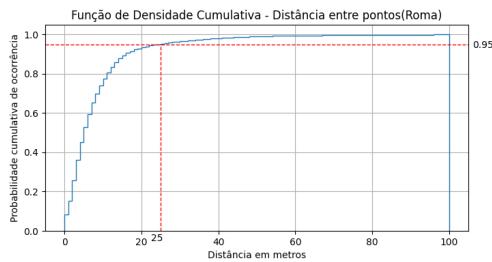
---

### 3.3.3. Cálculo dos limites utilizados no algoritmo de secção de traces

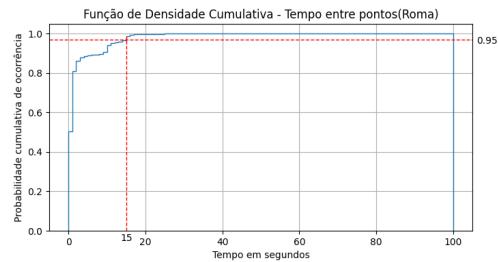
Os limites utilizados no algoritmo de separação de viagens foram escolhidos com base nas Funções de Densidade Cumulativa (CDF) apresentadas nas Figuras 2 e 3, retiradas da base de dados de Roma.

Para os testes, foram selecionados valores a serem utilizados como limites no algoritmo de secção de *traces* levando em conta os dados apresentados nas CDFs. No caso da distância, foi escolhido o limite de 25 metros, representando 95% dos valores. Em relação ao tempo entre dois pontos, foi escolhido o limite de 15 segundos, representando 95% dos dados.

**Figura 2. Distância entre pontos (Roma)**



**Figura 3. Tempo entre pontos (Roma)**



### 3.3.4. Matriz Origem-Destino (OD)

O quarto passo das cinco etapas em que o trabalho está dividido é o desenvolvimento de uma matriz OD para cada base. Essa matriz possui os valores de altura e largura de

cada célula fixos em 100m, formando um quadrado de 100mx100m. Por conta disso, as matrizes formadas não possuem o mesmo número de linhas e colunas, já que as áreas de cada base de dados não formam necessariamente um quadrado. Esses valores foram escolhidos para simular um quarteirão, o que agrupa os pontos de interesse por quarteirões nas cidades.

Utilizando a biblioteca Pandas, são filtrados os pontos iniciais e finais de uma base pelo valor Data/Hora, e os mesmos são inseridos em cada respectivo quarteirão de acordo com sua posição geográfica. Dessa forma, obtém-se uma matriz com células (quarteirões) mais cheias e outras mais vazias, segundo a quantidade de pontos nelas presentes.

Essa abordagem permite observar as áreas de maior concentração de início de viagens, bem como as áreas de maior concentração de fim de viagens. Essas informações são visualizadas através de mapas de calor ou histogramas 2D, e através da busca manual da localização exata das células em destaque. Essas informações são discutidas na Seção 4

### **3.3.5. Métricas Avaliadas**

A quinta parte, e final, da avaliação das bases exige a escolha de algumas métricas para comparação de ambas as bases com objetivo de extrair informações a respeito das diferenças observadas em cada métrica. Neste trabalho foram selecionadas quatro métricas de estudo para ambas as bases, sendo: viagens por veículo, que nos indicam quão movimentada é uma cidade diariamente, e quanto a população dessa mesma cidade depende de deslocamentos veiculares para seus afazeres diárias; Tamanho das trajetórias, que demonstra o nível de movimento pendular da metrópole e a intensidade do movimento horizontalizado; Tempo das trajetórias, que aponta o período diário em que indivíduos estão se deslocando naquele município; e Velocidade média das trajetórias, nos indicando quão estruturada está a cidade para movimentação veicular em massa diária.

Com a implementação de uma Matriz OD para cada base, foi analisado o padrão de movimentação dos veículos nas cidades. Utilizando as informações obtidas, foram construídos gráficos para uma análise Espaço-Temporal das bases. Eles informam a tendência de movimentação da cidade, qual sentido e direção a maior parte do fluxo se dispõe e as principais localidades com predisposição de se tornarem pontos iniciais/finais de uma viagem.

## **4. Resultados**

Na presente Seção, são discutidos os resultados obtidos a partir das técnicas utilizadas. Para isso, ela foi dividida em duas Subseções. A primeira Subseção (4.1) discute os resultados obtidos a partir das métricas de descrição da base, sendo elas: quantidade de viagens por veículo (em 24 horas), tamanho das trajetórias, duração das trajetórias e velocidade média dos veículos. Na segunda Subseção 4.2, foram observados e avaliados dados extraídos com o uso da Matriz OD. Nela são discutidos informações obtidas a partir da movimentação de veículos para regiões específicas das respectivas cidades, definindo intervalos de horário para comparação.

## 4.1. Avaliação das métricas

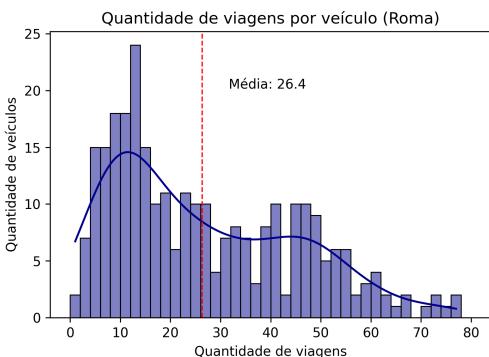
Foram estudadas 4 métricas de descrição das bases para melhor entendimento das diferenças e semelhanças entre os dois conjuntos de dados examinados. A primeira delas é a quantidade de viagens que cada veículo faz num período de 24 horas. A segunda métrica é o tamanho das trajetórias e a terceira, sua duração. A última métrica é a velocidade média dos veículos durante uma trip. Elas permitem uma comparação comportamental dos dois tipos de veículos, táxis, no caso de Roma, e ônibus, no caso de Dublin.

Também foram avaliadas histogramas 2D dos pontos de origem e dos pontos de destino das viagens feitas pelos veículos. O objetivo dessa análise é entender melhor o padrão comportamental da movimentação desses veículos no interior da cidade. Esses histogramas permitem o entendimento de como a cidade se movimenta ao longo do dia, quais pontos são críticos, quais são as tendências de movimentação de acordo com o horário, etc.

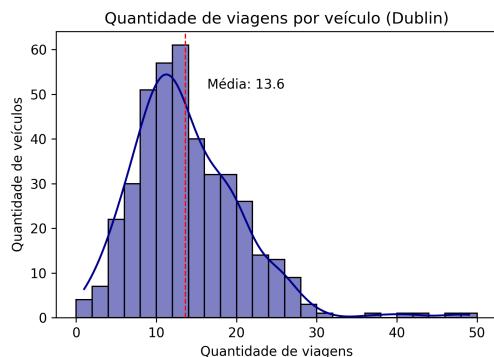
### 4.1.1. Viagens por veículo

Os gráficos de viagens por veículo mostram quantos veículos possuem determinada quantidade de viagens. Nesse caso, as viagens dos ônibus são consideradas apenas pontos iniciais aos finais, desconsiderando paradas. As Figuras 4 e 5 ilustram um histograma com a quantidade de viagens por veículo para Roma e Dublin, respectivamente.

**Figura 4. Quantidade de viagens por veículo (Roma)**



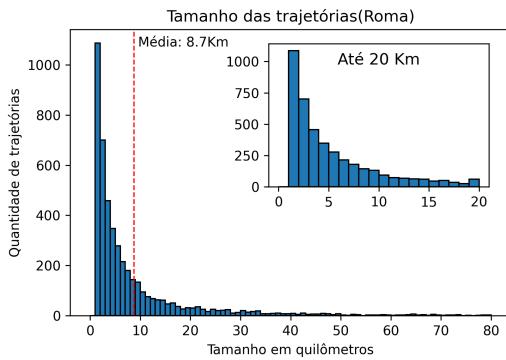
**Figura 5. Quantidade de viagens por veículo (Dublin)**



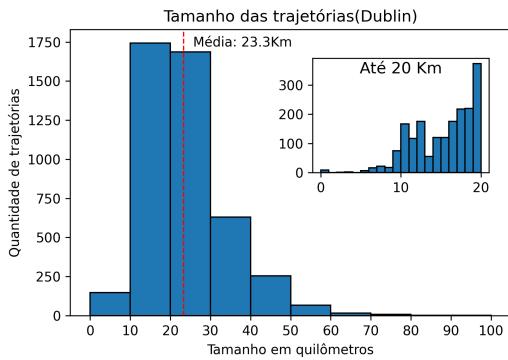
Vê-se uma quantidade média menor de viagens para os veículos de Dublin, com uma concentração grande de veículos com até 30 viagens por dia. Roma mostra uma distribuição mais espaçada, contendo veículos com até 80 viagens e uma média de 26 viagens por veículo, praticamente o dobro da média de Dublin. Como ônibus são veículos com trajetórias normalmente longas, isso explicaria a quantidade menor de viagens diárias. Outro detalhe importante é que a distribuição de Dublin pode ser identificada como uma distribuição Poisson, com a formação de um sino seguido de uma cauda.

#### 4.1.2. Comprimento das trajetórias

**Figura 6. Tamanho das trajetórias (Roma)**



**Figura 7. Tamanho das trajetórias (Dublin)**



A segunda métrica analisada foi o tamanho das trajetórias presentes nas bases de Roma e Dublin. As Figuras 6 e 7 ilustram histogramas do tamanho das trajetórias para Roma e Dublin, respectivamente. No caso de Roma, percebe-se um comportamento de cauda longa, com grande quantidade de trajetórias pequenas, a maioria menor que 10Km, e uma média de 8.7Km. Em Dublin percebe-se uma formação de sino com média de 23.3Km e poucas trajetórias abaixo de 10Km.

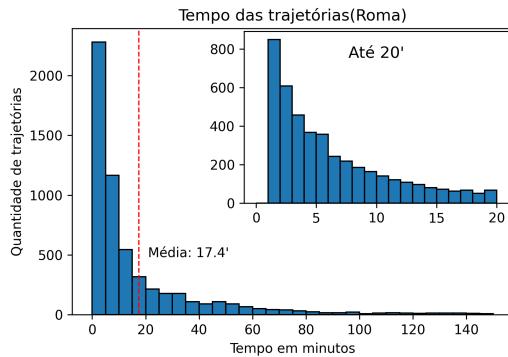
Isolando-se as trajetórias até 20Km, é percebido que o conjunto de dados de Dublin possui um crescimento do tamanho, possuindo trajetórias entre 10Km e 20Km. De forma diferente, Roma continua com o comportamento de cauda longa, com poucas instâncias maiores que 10Km. Isso pode ser explicado pela diferença comportamental dos veículos. Ônibus possuem trajetórias já traçadas e planejadas, sendo incomum viagens pequenas pois seriam ineficientes. Táxis são comumente usados por pessoas que precisam de deslocamento rápido, o que seria mais apto a ser escolhido para trajetórias curtas ou longas que necessitem de deslocamento rápido.

#### 4.1.3. Duração das trajetórias

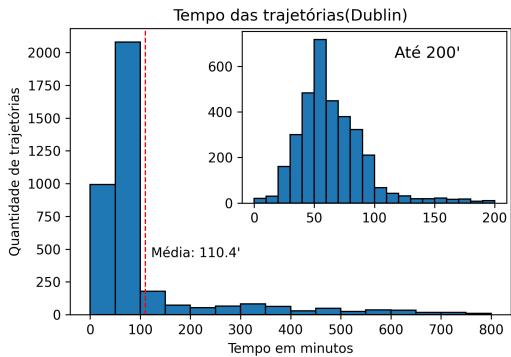
Os histogramas de tempo das trajetórias representam quantas trajetórias estão incluídas em determinado intervalo de tempo. As Figuras 8 e 9 ilustram esses histogramas para as cidades de Roma e Dublin, respectivamente. A diferença de resultados das duas bases é muito alta. Portanto, no caso de Roma a escala escolhida vai de 0 a 140 minutos, sendo o foco do gráfico de 0 a 20 minutos. Em Dublin a escala vai de 0 a 800 minutos, sendo o foco de 0 a 200 minutos.

Roma mostra possuir trajetórias muito curtas com média de 17.4 minutos, onde é observável o comportamento de cauda longa quando estabelecemos o limite do gráfico de até 140 minutos. Dublin possui uma média de tempo muito maior, sendo 110.4 minutos, e mantém o comportamento de formação de um sino entre 0 e 150 minutos. É um padrão de comportamento esperado para ônibus que tendem a ter trajetórias longas.

**Figura 8. Duração das trajetórias (Roma)**



**Figura 9. Duração das trajetórias (Dublin)**

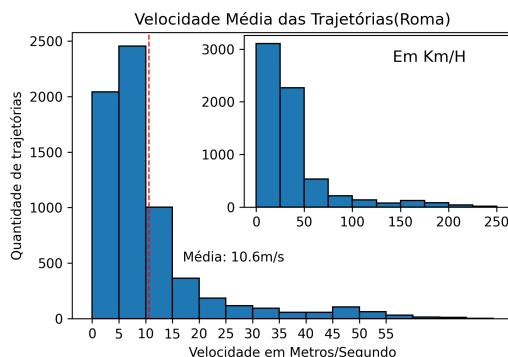


Nos focos dos gráficos, ou gráficos internos, percebe-se um comportamento de sino mais clara para os valores de Dublin. Já Roma demonstra o mesmo comportamento de cauda longa, onde a quantidade de trajetórias existentes diminui conforme a duração cresce.

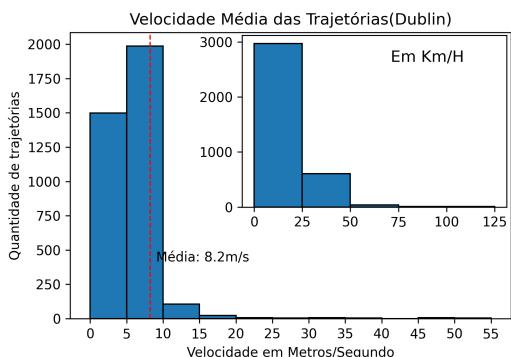
#### 4.1.4. Velocidade média por trajetória

Os histogramas de velocidade média das trajetórias expõem quantas trajetórias possuem uma velocidade média dentro de um intervalo específico e são representadas nas Figuras 10, no caso de Roma, e 11, no caso de Dublin. Da mesma forma que a métrica avaliada anteriormente, os resultados destoaram, portanto a escala dos gráficos internos foi diferenciada, onde Roma vai de 0 a 250 Km/H e Dublin, 0 a 125 Km/H.

**Figura 10. Velocidade média por trajetória (Roma)**



**Figura 11. Velocidade média por trajetória (Dublin)**



A diferença entre as médias não é muito grande entre as duas bases, sendo a de Roma pouco maior. Como ônibus são veículos de rotação interna na cidade, é normal que a velocidade média seja mais baixa por ser controlada por empresas e órgãos de trânsito, o que pode ser confirmado pelo gráfico com a maior parte das trajetórias de velocidade

média até 20Km/H. Os táxis por terem mais liberdade de movimentação e velocidade, tendem a se mover mais rápido, o que também é demonstrado nos gráficos com uma quantidade um pouco maior de trajetórias cuja média de velocidade é acima dos 20Km/H.

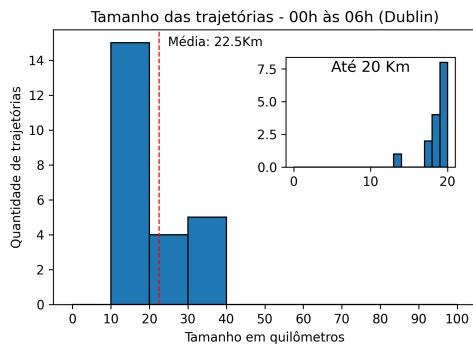
## 4.2. Análise Espaço-Temporal

Nesta Seção, é realizada uma análise espaço-temporal das duas bases. Para isso, foi necessário dividir os dados em 4 (quatro) intervalos de tempo de 6 (seis) horas de duração ao longo de um dia. Nesses intervalos foram plotados gráficos consistindo do comprimento das viagens, além dos mapas de calor de cada cidade nos mesmos intervalos para melhor avaliação da movimentação de veículos.

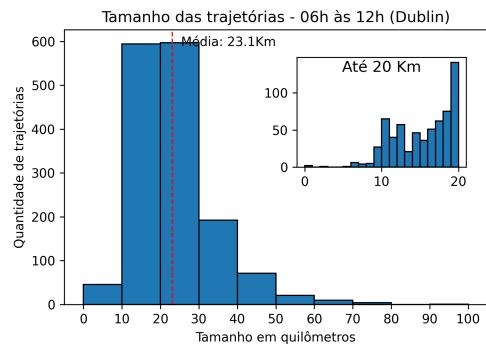
### 4.2.1. Comprimento das viagens por hora do dia - Dublin

Para um melhor entendimento da movimentação veicular na análise espaço-temporal, foram construídos gráficos que descrevem o comprimento das viagens de Dublin ao longo de intervalos de horário. Para esta análise os 4 (quatro) intervalos foram definidos, sendo: 00h às 06h, 06h às 12h, 12h às 18h e 18h às 00h.

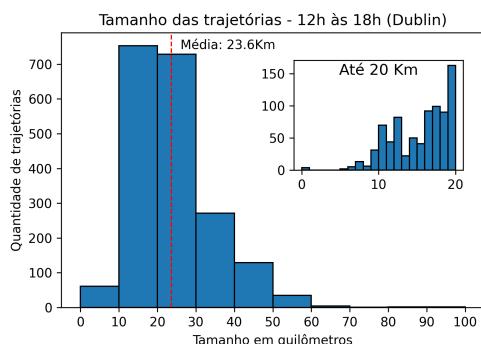
**Figura 12. Histograma de tamanho das trajetórias 00h-06h (Dublin)**



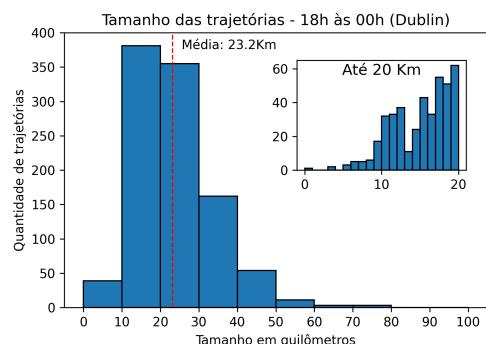
**Figura 13. Histograma de tamanho das trajetórias 06h-12h (Dublin)**



**Figura 14. Histograma de tamanho das trajetórias 12h-18h (Dublin)**



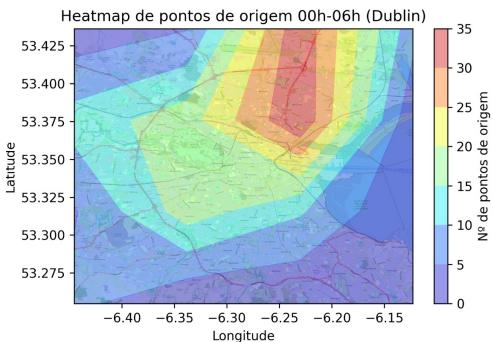
**Figura 15. Histograma de tamanho das trajetórias 18h-00h (Dublin)**



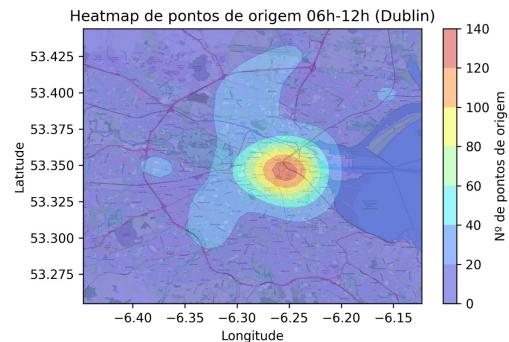
Nas Figuras 12, 13, 14 e 15 pode-se observar uma média de distância de viagens estável, entorno de 23Km. O que é esperado para transportes públicos que tendem a manter suas trajetórias ao longo do dia, não havendo muitas alterações que não sejam quantidade de ônibus à disposição. É visível um concentração na quantidade de viagens na parte da manhã e tarde em relação à noite e madrugada. Isso indica uma maior circulação na linha de ônibus durante o dia com trajetórias de tamanho fixo, o que é esperado no comportamento de transportes coletivos.

#### 4.2.2. Dublin - Origem

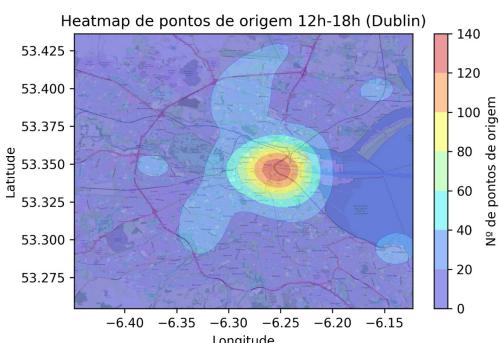
**Figura 16. Mapa de Calor dos pontos de origem 00h-06h (Dublin)**



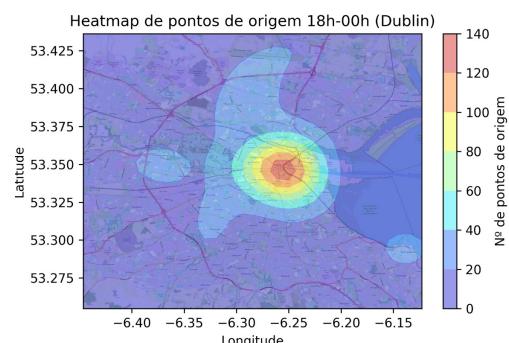
**Figura 17. Mapa de Calor dos pontos de origem 06h-12h (Dublin)**



**Figura 18. Mapa de Calor dos pontos de origem 12h-18h (Dublin)**



**Figura 19. Mapa de Calor dos pontos de origem 18h-00h (Dublin)**



Na Figura 16 pode-se visualizar uma distribuição de pontos de origem mais homogênea na região da cidade de Dublin. Isso indica que durante esse intervalo de horários, a tendência é de uma movimentação menos concentrada em pontos específicos, isto é, mais dispersa. A causa disso é a menor quantidade de deslocamentos, comum ao horário, levando a uma menor frota de ônibus durante o período. O foco da origem de viagens durante esse período é a região nordeste da cidade, provavelmente devido ao aeroporto na região ou possivelmente sendo a localização das garagens dos ônibus.

Durante o período de 06h00 às 12h00, demonstrado na Figura 17, inicia-se uma maior concentração de pontos de origem de deslocamento em algumas áreas específicas da cidade. A maior parte dos pontos é contida no centro da cidade, mas há formações em alguns pontos específicos. Na Figura 17 especificamente, é visível uma bolha a oeste da região central, onde se encontra uma área industrial da cidade. A nordeste da região central, também há uma bolha localizada onde é localizado o hotel do aeroporto da cidade. Da região central surgem duas formações estendendo na direção sudoeste e norte. Ambas são regiões residenciais, mas no caso da região norte, sua formação se completa no aeroporto de Dublin.

Na Figura 18, que descreve o intervalo de 12h-18h, ocorre uma nova formação que está presente na região sudeste da cidade. Essa nova região formada está localizada precisamente no segundo porto da cidade. As demais regiões mantém proporções semelhantes.

Na Figura 19 há uma maior diferença com relação às anteriores. Há um aumento de concentração na formação da região industrial, indicando maior saída de pessoas da região neste horário. Também é visível uma diminuição na concentração em áreas residenciais.

#### 4.2.3. Dublin - Destino

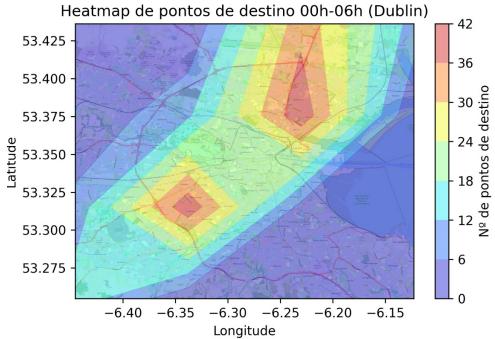
Com relação aos pontos de destino nos mesmos períodos, ocorrem algumas mudanças em comparação aos pontos de origem. Primeiramente, no horário de 00h-06h (Figura 20) ocorre um novo foco a oeste da região central, próximo à área industrial. O que demonstra uma propensão maior de movimentação de veículos em direção ao aeroporto e à região industrial, partindo das áreas residenciais e do próprio aeroporto.

Na Figura 21, há um foco maior na região portuária (formação a sudeste da região central) se comparado à Figura 17. Isso indica uma maior entrada que saída de veículos no porto da cidade durante esse período.

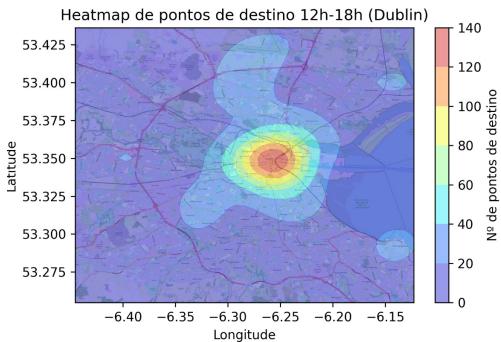
No horário de 12h-18h, representado na Figura 22, comparando-a à Figura 18, que descreve a origem dos veículos no mesmo período de tempo, percebe-se uma grande diminuição da concentração na região industrial, o que indica que nesse intervalo de horário ocorre saída de veículos da região industrial, oeste à região central.

Por fim, na Figura 23, observa-se o desaparecimento de todas regiões formadas fora da região central, se comparada à Figura 19. O que demonstra uma tendência de saída dos veículos das regiões industriais e portuárias para a região central e regiões residenciais da cidade.

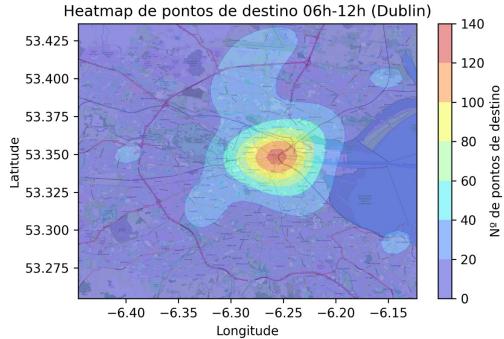
**Figura 20. Mapa de Calor dos pontos de destino 00h-06h (Dublin)**



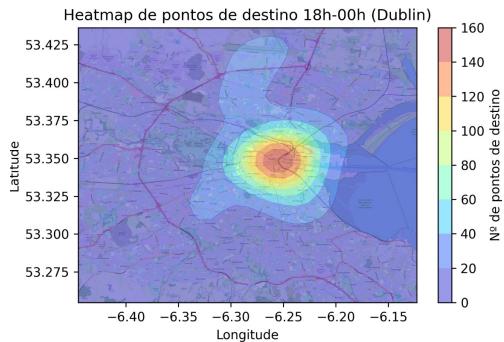
**Figura 22. Mapa de Calor dos pontos de destino 12h-18h (Dublin)**



**Figura 21. Mapa de Calor dos pontos de destino 06h-12h (Dublin)**



**Figura 23. Mapa de Calor dos pontos de destino 18h-00h (Dublin)**

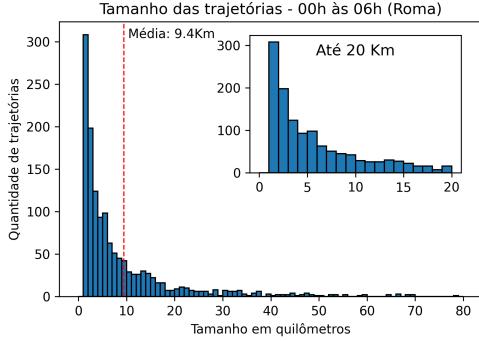


#### 4.2.4. Comprimento das viagens por hora do dia - Roma

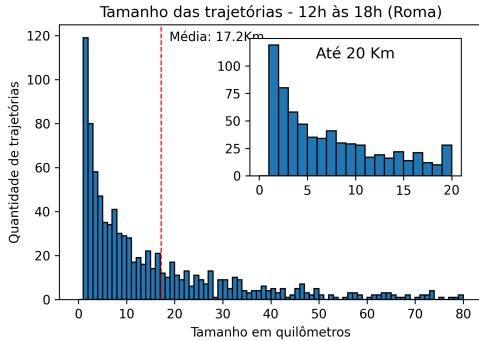
Para Roma também foram plotados gráficos descrevendo tamanho das trajetórias de acordo com os quatro intervalos de horas investigados. Nas Figuras a seguir, vê-se um comportamento típico para veículos de transporte não coletivo, no caso táxis, onde têm-se uma grande mudança de comportamento ao longo de um dia.

Nas Figuras 24, 25, 26 e 27 é possível notar a grande mudança de comportamento das viagens de acordo com o horário. Há uma tendência de aumento da distância percorrida por um veículo numa trajetória conforme o horário se aproxima da madrugada, enquanto a quantidade de trajetórias diminui. Na parte da manhã, quando as trajetórias ainda são de curta distância, há uma maior quantidade de viagens. Na parte da tarde e noite (horários de pico) há uma menor quantidade de trajetórias, porém com uma média de distância bem maior. Na madrugada ocorrem poucas viagens.

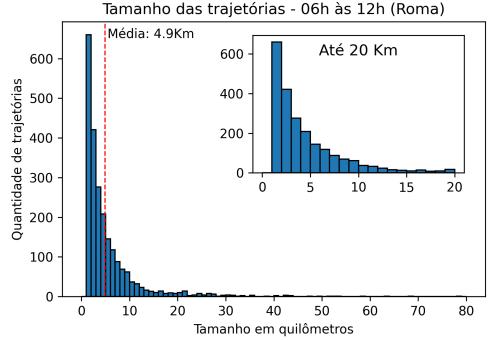
**Figura 24. Histograma de tamanho das trajetórias 00h-06h (Roma)**



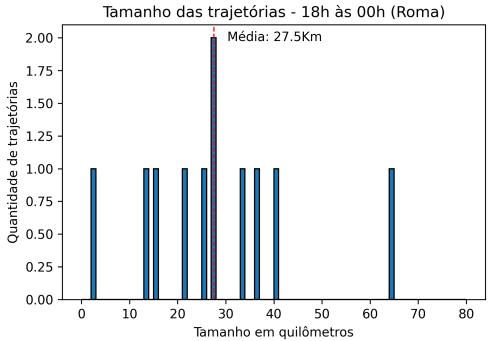
**Figura 26. Histograma de tamanho das trajetórias 12h-18h (Roma)**



**Figura 25. Histograma de tamanho das trajetórias 06h-12h (Roma)**



**Figura 27. Histograma de tamanho das trajetórias 18h-00h (Roma)**

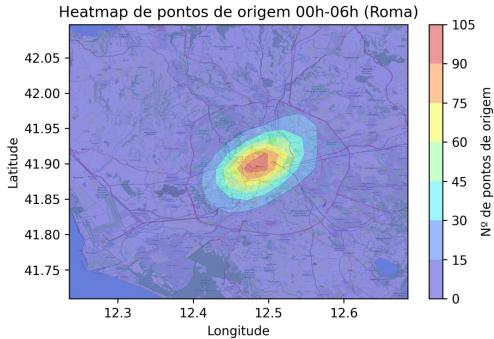


#### 4.2.5. Roma - Origem

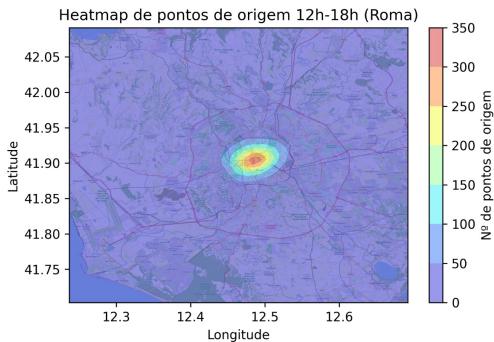
No caso de Roma os dados de visualização da movimentação veicular são bem diferentes dos dados obtidos pela base de Dublin. Nas Figuras 28, 29, 30 e 31, ocorrem poucas mudanças entre os intervalos de horário escolhidos. É perceptível um eixo sudoeste-nordeste no horário de 00h-01h, o que pode indicar uma tendência de movimentação da cidade nessa direção. Nos horários de 06h-12h e 12h-18h há uma diminuição nas regiões periféricas e grande concentração na área mais central da cidade (Figuras 29 e 30).

Na Figura 31, representando o horário de 18h-00h, a região de maior concentração se mantém, porém há uma expansão na área de movimentação. Essa formação possui uma pequena protuberância saindo em direção ao sul.

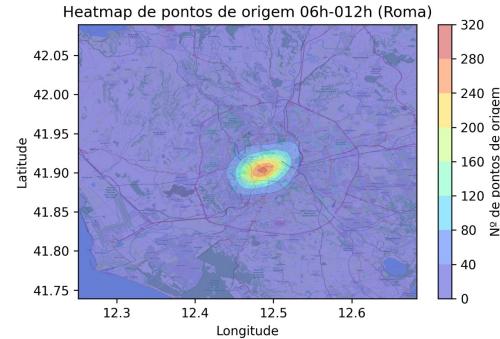
**Figura 28. Mapa de Calor dos pontos de origem 00h-06h (Roma)**



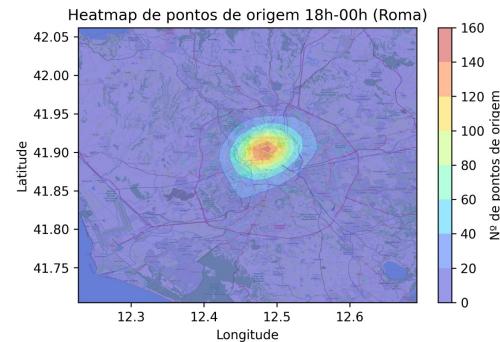
**Figura 30. Mapa de Calor dos pontos de origem 12h-18h (Roma)**



**Figura 29. Mapa de Calor dos pontos de origem 06h-12h (Roma)**



**Figura 31. Mapa de Calor dos pontos de origem 18h-00h (Roma)**

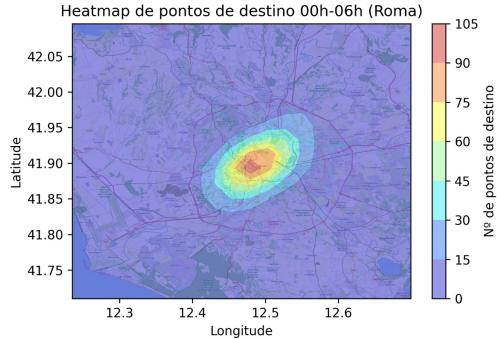


#### 4.2.6. Roma - Destino

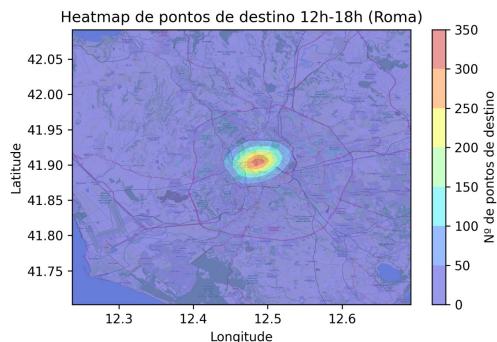
Avaliando as regiões das Figuras que representam as áreas de destino em Roma, em comparação individual às Figuras homólogas das áreas de concentração de pontos de origem, percebem-se pouca ou nenhuma diferença. Para comparação, são apresentadas a seguir as Figuras 32, 33, 34 e 35, representando respectivamente os intervalos de horários de 00h-06h, 06h-12h, 12h-18h e 18h-00h.

Nelas, são vistos os mesmos padrões. Primeiro há um grande volume de veículos chegando na região interna da cidade em um eixo na direção sudoeste-nordeste (Figura 32). Em seguida ocorre uma grande concentração de chegada de veículos na região mais central da cidade entre 06h-12h (Figura 34). No próximo intervalo, ainda há grande concentração de veículos, porém ocorre um pequeno deslocamento do ponto central para o noroeste da região central (Figura 34). Por fim, há um aumento dessa área de concentração durante o horário de 18h-00h (Figura 35), demonstrando uma maior dispersão nos destinos dos veículos no horário em que normalmente as pessoas tendem a retornar às suas casas.

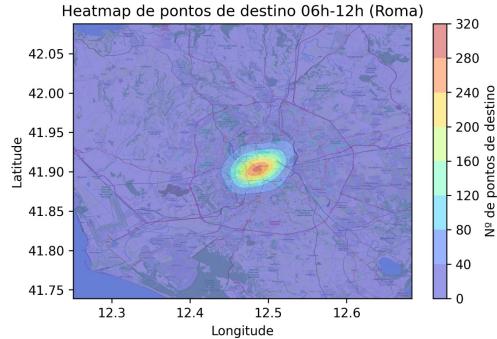
**Figura 32. Mapa de Calor dos pontos de destino 00h-06h (Roma)**



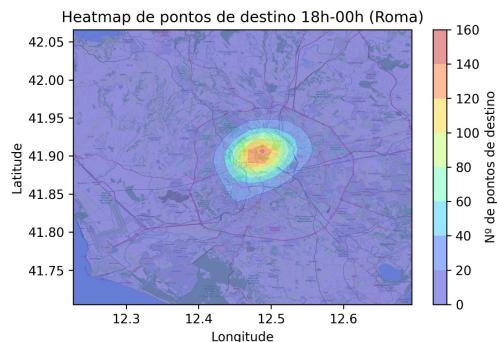
**Figura 34. Mapa de Calor dos pontos de destino 12h-18h (Roma)**



**Figura 33. Mapa de Calor dos pontos de destino 06h-12h (Roma)**



**Figura 35. Mapa de Calor dos pontos de destino 18h-00h (Roma)**



## 5. Conclusão e Trabalhos Futuros

Neste trabalho, foram aplicadas técnicas de estudo de bases de dados veiculares para a realização de uma comparação estatística do comportamento de uma base de táxis de Roma e uma base de ônibus de Dublin. Foram feitas análises de métricas de cada base, bem como uma análise espaço-temporal do comportamento dos veículos ao longo do dia em suas respectivas cidades.

Dos dados apresentados, conclui-se que há uma diferença de comportamento entre as duas bases. Essa diferença pode se dar pelo comportamento natural específico de cada tipo de veículo (táxis e ônibus), que possuem diferentes objetivos e funções no transporte municipal. Outro fator seria o contexto histórico das cidades. Dublin, por exemplo, é uma cidade portuária desde que foi fundada, o que influencia na movimentação de seus veículos. Já Roma é uma cidade histórica, o que atrai movimentação principalmente ao seu centro histórico. Esses pontos demonstram que estudos de movimentação veicular são fundamentais para futuras implementações de redes veiculares inteligentes, devido às particularidades na movimentação de cada cidade.

Para futuros trabalhos que estendam o escopo deste artigo, pode-se investigar uma heurística para secção de *traces* mais eficaz do que a que foi aplicada à base de Roma.

Além disso, pode-se utilizar a Matriz OD a fim de extrair informações acerca das áreas mais movimentadas, definir rótulos para as mesmas e avaliar sua importância na movimentação veicular em uma cidade. Por fim, também pode ser feito um estudo para aplicação de técnicas de Inteligência Artificial na Matriz OD com o objetivo de prever a movimentação veicular.

## Referências

- Nardine Basta, Amal ElNahas, Hans Peter Grossmann, and Slim Abdennadher. Generic geo-social mobility model for vanet. In *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*, pages 1–5. IEEE, 2016.
- Clayson Celes, Azzedine Boukerche, and Antonio AF Loureiro. Towards understanding of bus mobility for intelligent vehicular networks using real-world data. In *2019 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6. IEEE, 2019.
- Marcos Cintra. A crise do trânsito em são paulo e seus custos. *GV EXECUTIVO*, 12(2): 58–61, 2013.
- Ana Gainaru, Ciprian Dobre, and Valentin Cristea. A realistic mobility model based on social networks for the simulation of vanets. In *VTC Spring 2009-IEEE 69th Vehicular Technology Conference*, pages 1–5. IEEE, 2009.
- Lei Gong, Hitomi Sato, Toshiyuki Yamamoto, Tomio Miwa, and Takayuki Morikawa. Identification of activity stop locations in gps trajectories by density-based clustering method combined with support vector machines. *Journal of Modern Transportation*, 23(3):202–213, 2015.
- Xiangjie Kong, Feng Xia, Zhaolong Ning, Azizur Rahim, Yinqiong Cai, Zhiqiang Gao, and Jianhua Ma. Mobility dataset generation for vehicular social networks based on floating car data. *IEEE Transactions on Vehicular Technology*, 67(5):3874–3886, 2018.
- Quannan Li, Yu Zheng, Xing Xie, Yukun Chen, Wenyu Liu, and Wei-Ying Ma. Mining user similarity based on location history. In *Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems*, pages 1–10, 2008.
- Wenshuang Liang, Zhuorong Li, Hongyang Zhang, Shenling Wang, and Rongfang Bie. Vehicular ad hoc networks: architectures, research issues, methodologies, challenges, and trends. *International Journal of Distributed Sensor Networks*, 11(8):745303, 2015.
- Zhaolong Ning, Feng Xia, Noor Ullah, Xiangjie Kong, and Xiping Hu. Vehicular social networks: Enabling smart mobility. *IEEE Communications Magazine*, 55(5):16–55, 2017.