

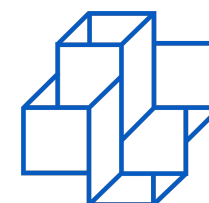


Detecção de Anomalias no Transporte Rodoviário Urbano

**Ana Beatriz Cruz¹, João Ferreira¹, Bernardo Monteiro¹,
Rafaelli Coutinho¹, Fabio Porto², Eduardo Ogasawara¹**

¹ CEFET/RJ

² LNCC



Laboratório
Nacional de
Computação
Científica

Introdução

- Aumento da População Urbana
- Problemas relacionados a mobilidade
 - Acidentes
 - Poluição
 - Trânsito



Estudos de comportamento de trânsito (perspectiva de trajetória)

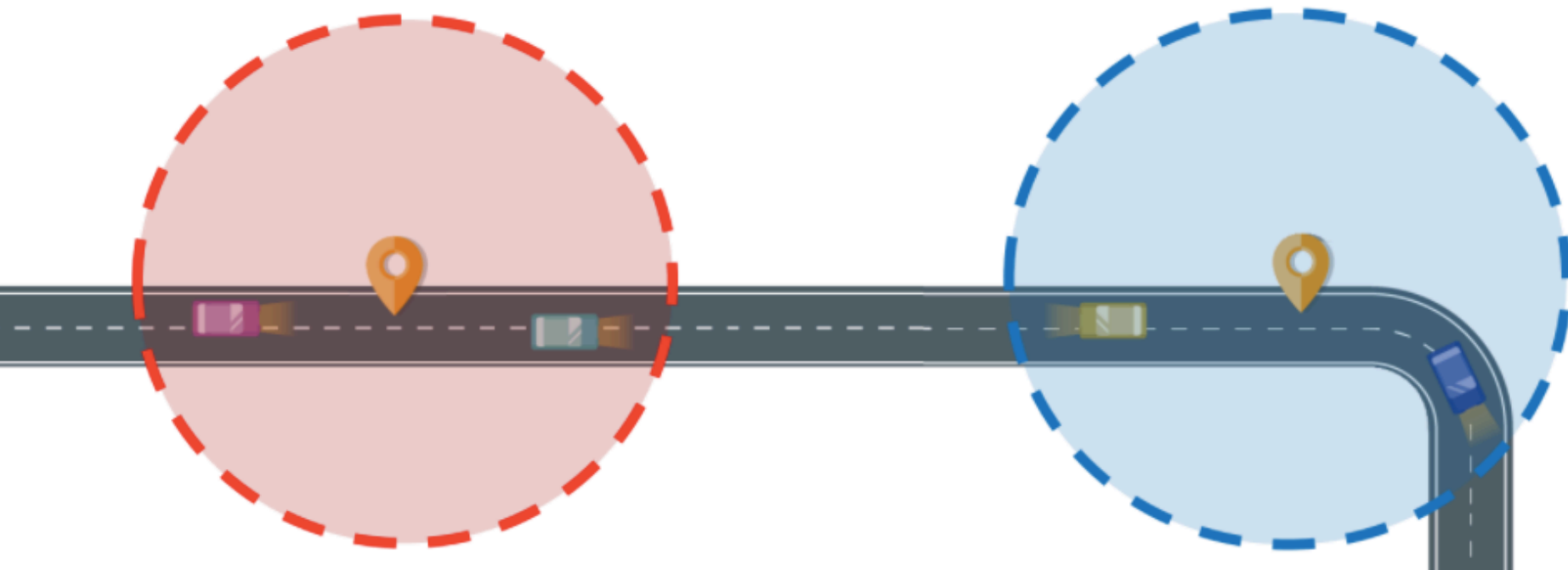
- Compreensão do comportamento do trânsito visando minimizar impactos de engarrafamentos
- Ferreira et al. [2013] usaram dados de GPS de taxi NY para ter análise visual das trajetórias
- Verhein and Chawla [2008] usaram dados de celulares, RFID e satélites para identificar trajetórias de trânsito intenso
- Pan et al. [2013] analisaram as variações nas movimentações das trajetórias para observar comportamento anômalo em rotas

Estudos de comportamento de trânsito (perspectiva de agregações espaço-temporais)

- Andrienko e Andrienko [2008,2011] propuseram métodos de agregação espaço-temporal para trajetórias visando formar mapas de fluxo do espaço urbano
 - Favorece visão sumarizada dos dados
 - Associação de observação de objetos móveis a objetos permanentes

Agregação espaço-temporal

- Carência de estudos que analisam a mobilidade por uma perspectiva de agregações de séries espaço-temporais
 - Propicia uma melhor compreensão?
 - Propicia uma compreensão complementar?



Trajetória

- Dados gerados por sensores como GPS e RFID
- Observação:
 - $\hat{tr}_o = (p, v)$
 - $\hat{tr}_o.v \in \mathbb{R}$
 - $\hat{tr}_o.p \in \mathbb{R}^2$
- Trajetória:
 - $tr_k = \langle \hat{tr}_1, \dots, \hat{tr}_{co} \rangle$
- Visão Individualizada do objeto que se move no espaço
- *Dataset* de trajetórias:
 - $TR = \bigcup_{k=1}^{ck} tr_k$



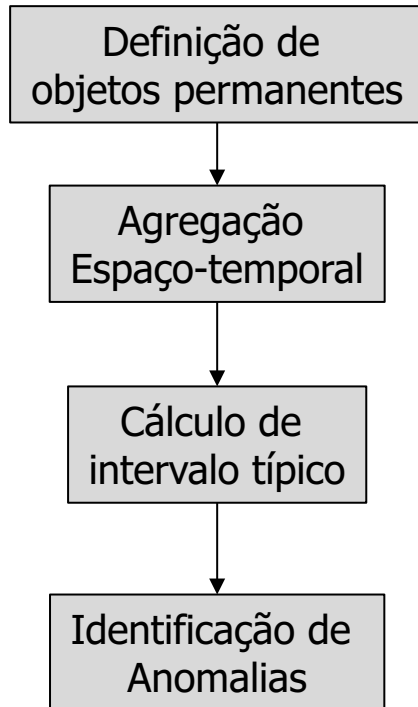
Agregação de séries espaço-temporais de objetos permanentes a partir das séries de trajetórias

- Partição do espaço S em ci regiões $S = \bigcup_{i=1}^{ci} \hat{S}_i$
 \hat{S}_i é uma região do espaço: $\hat{S}_i = [(xl_i, yl_i), (xu_i, xu_i)]$
- Partição do tempo T em cj intervalos $T = \bigcup_{j=1}^{cj} \hat{t}_j$
 \hat{t}_j é um intervalo de tempo: $\hat{t}_j = [(j - 1) \cdot ma + 1, j \cdot ma]$
- Dataset ST de séries espaço-temporais: $ST = \bigcup_{i=1}^{ci} st_i$
 $st_i = \langle \hat{u}_{i,1}, \dots, \hat{u}_{i,cj} \rangle$
- Agregação:
$$\forall \hat{S}_i \in S, \hat{t}_j \in T \left(\hat{u}_{i,j} = \Gamma_{\theta(tr_k.\hat{t}r_o.v)} \left(\sigma_{o \in \hat{t}_j, tr_k.\hat{t}r_o.p \in \hat{S}_i} tr_k.\hat{t}r_o \right) \right)$$

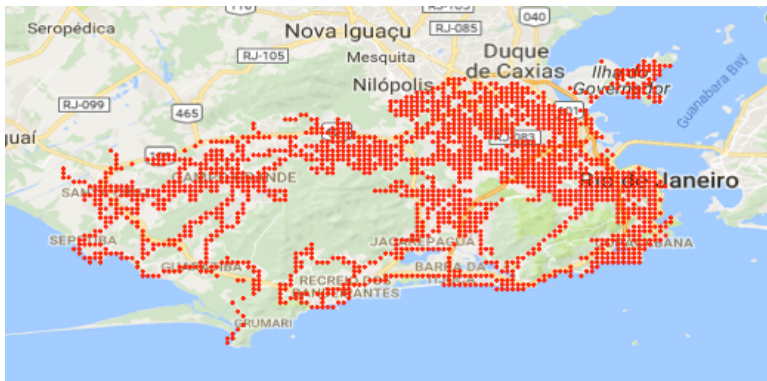
Valor esperado, intervalo típico e anomalias

- Considere ma o intervalo de particionamento do tempo em minutos
 - Um dia tem obs observações: $obs = \frac{1440}{ma}$, ma divisor de 1440
- Para $w \in [1, obs]$ as observações $\hat{u}_{i,w+d \cdot ma}$ corresponde a observações no mesmo horário w nos diferentes dias d
- O valor esperado na posição i e no horário w para um intervalo de dias D é $st_{i,w,D}$
- O intervalo típico $IT_{i,w,D}$:
 - $IT_{i,w,D} = [q_1 - 3 \times IQR, q_3 + 3 \times IQR]$
- Anomalia são observações em ST no intervalo de dias D que não sejam típicas

Metodologia



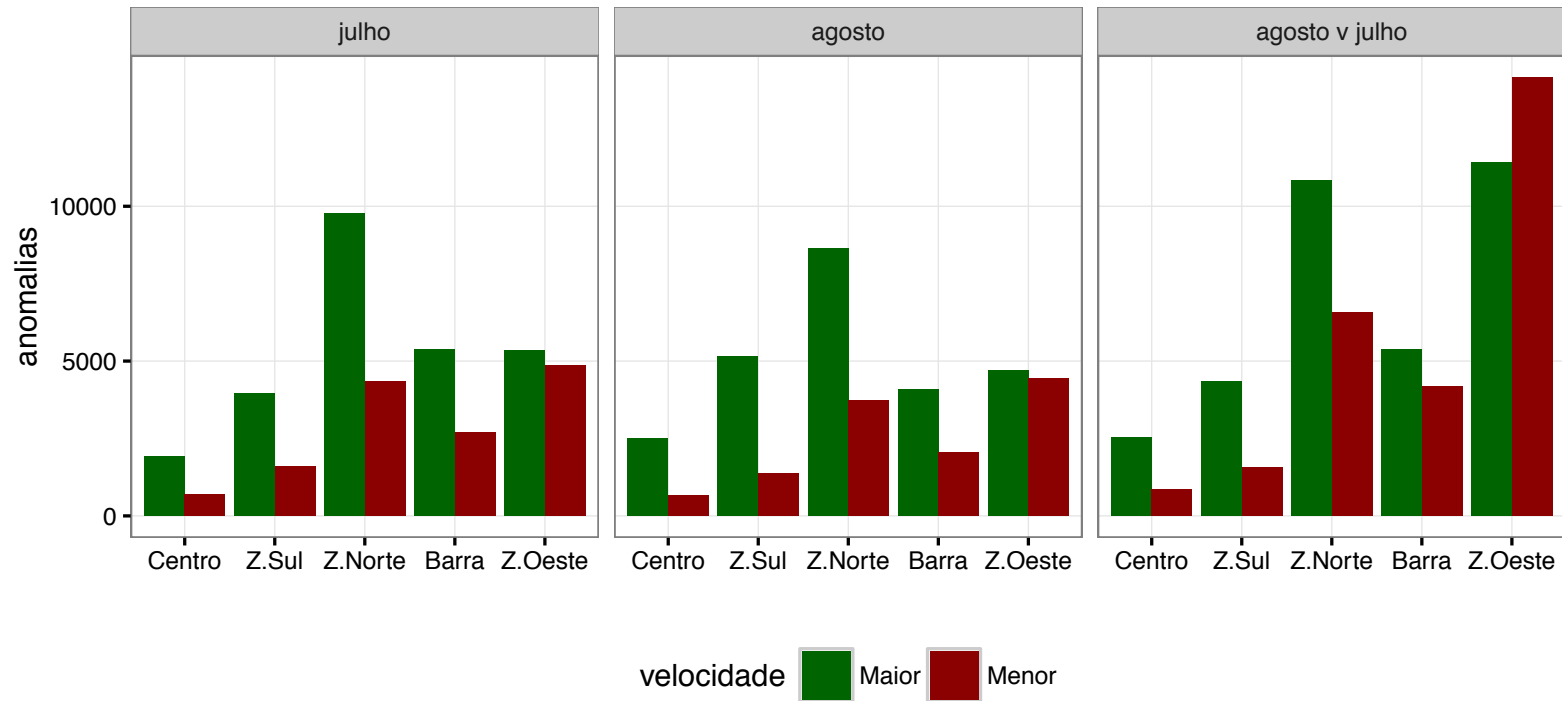
- Particionamento do espaço em malha de 500m x 500m
- Seleção dos quadrantes relevantes
- Intervalo de tempo (ma) de 10 minutos
- Agregação espaço-temporal
- Cálculo de intervalo típico para cada mês



Avaliação preliminar

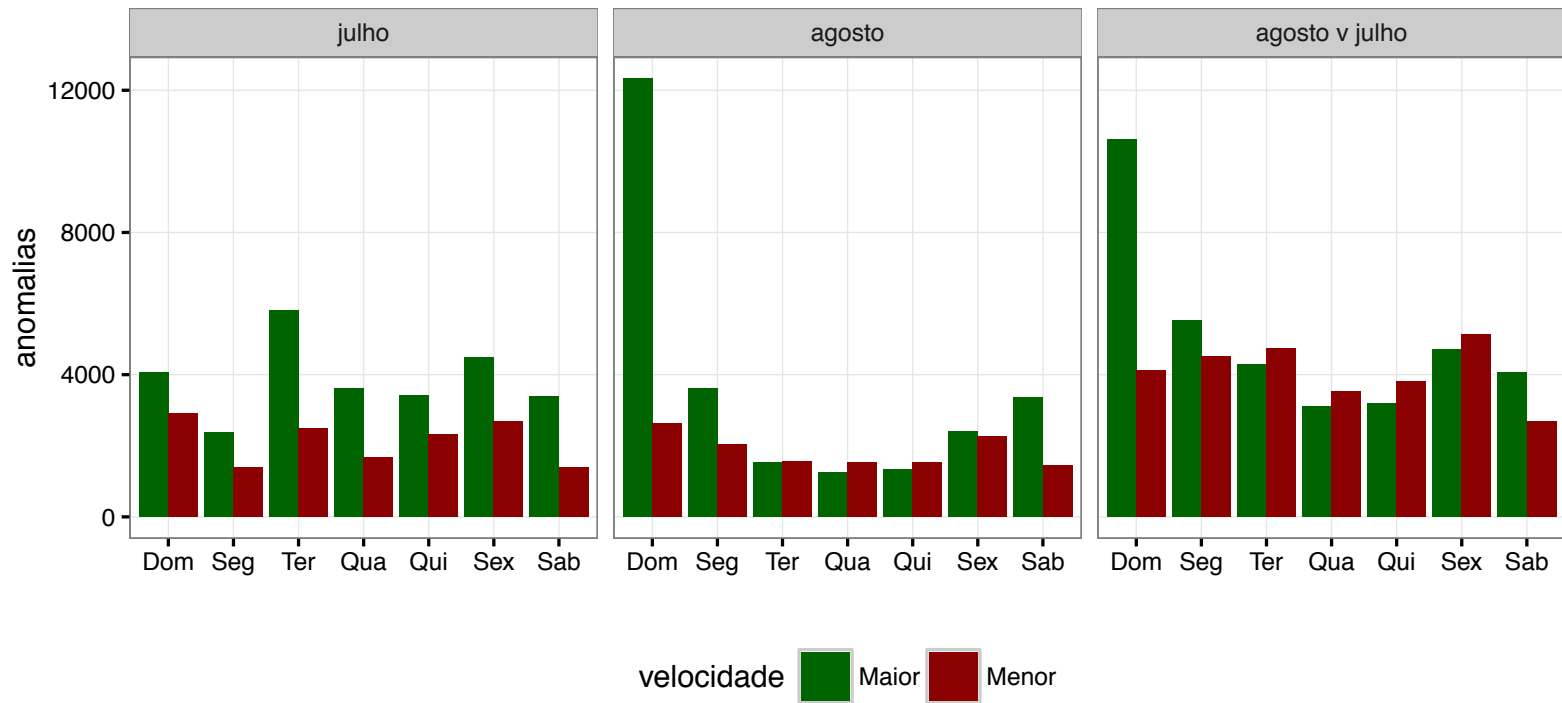
- Metodologia aplicada em Julho 2014 e Agosto de 2014
- Julho/2014:
 - 2.961.726 observações por dia
 - 42.050 anomalias identificadas (1.03%)
- Agosto/2014:
 - 3.403.180 observações por dia
 - 38.843 anomalias identificadas (1.07%)
- Comparação relativa Agosto versus Julho
 - 64.041 anomalias identificadas (1.76%)

Anomalias por Região Administrativa



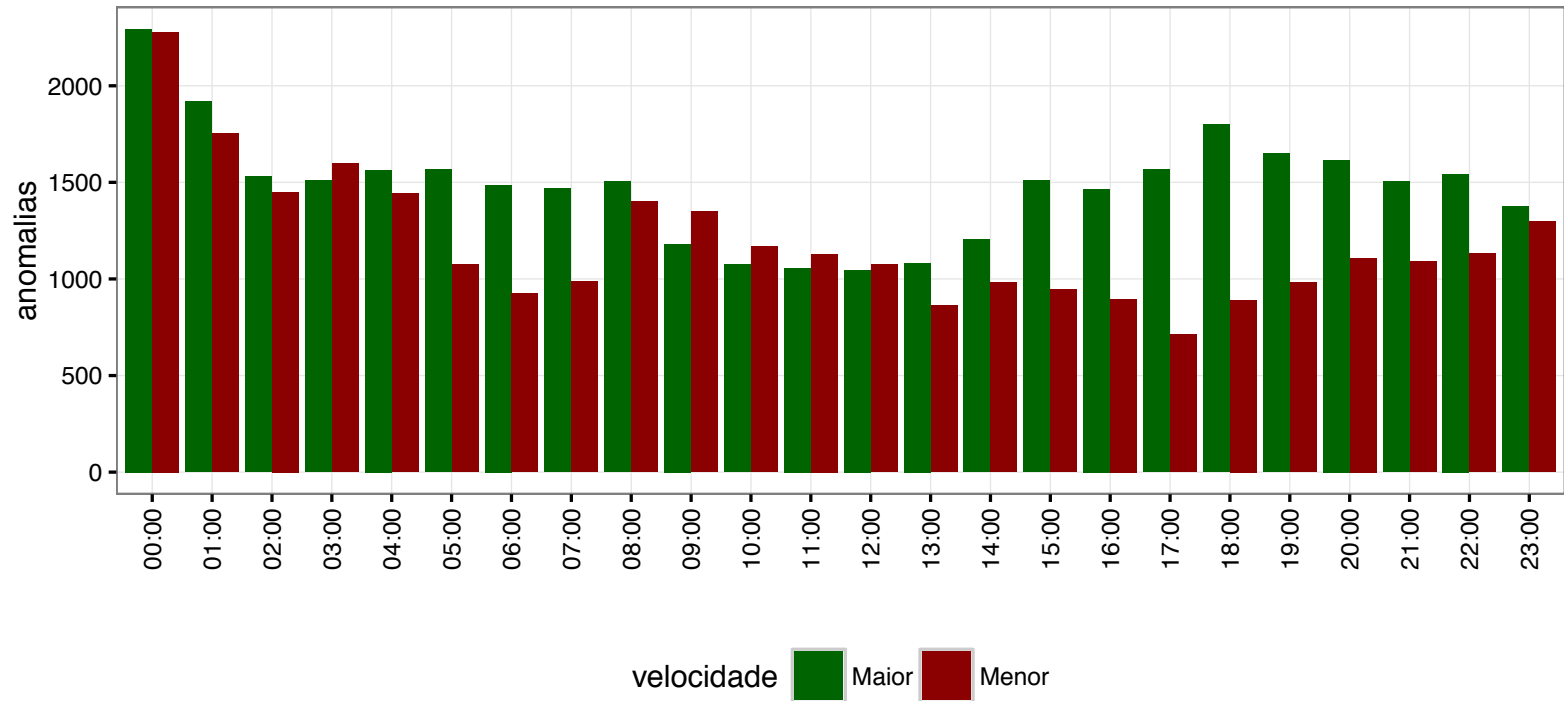
Julho e Agosto: Maior quantidade de anomalias (+) na zona norte e (-) na zona oeste
Na comparação relativa, há uma quantidade grande de anomalias (-) na zona oeste

Anomalias por Dia da Semana



Anomalias (+): domingos (esperado) e nas terças de julho (Copa do Mundo)
Anomalias (-): sextas (esperado) e domingo (manifestações, alt. transito)

Anomalias por horário (agosto versus julho)



Anomalias (+): valores altos 17h às 20h (*)
Anomalias (-): valores altos das 9h às 12h

Considerações Finais

- Análise Preliminar de anomalias em séries espaço-temporais do objetos permanentes
- 1% de anomalias por mês estudado
- Anomalias pontuais versus anomalias recorrentes
- Aplicação de padrões frequentes
 - Aumento do período estudado

	pre	cons	support	confidence	lift
2	{dt=2014-05-08}	{tipo=Menor}	0.01101123	0.9992925	2.598525

Greve de motoristas de ônibus no Rio deixa trens do Metrô superlotados 

Gustavo Maia
Do UOL, no Rio 08/05/2014 | 09h48 > Atualizada 09/05/2014 | 10h21

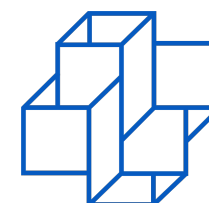
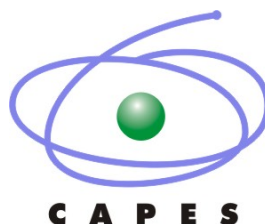


Detecção de Anomalias no Transporte Rodoviário Urbano

Ana Beatriz Cruz¹, João Ferreira¹, Bernardo Monteiro¹,
Rafaelli Coutinho¹, Fabio Porto², Eduardo Ogasawara¹



¹ CEFET/RJ
² LNCC



Laboratório
Nacional de
Computação
Científica