k-DynMix: Um Mecanismo de Proteção Dinâmica de Privacidade em Mix-Zones

Ekler P. de Mattos
Augusto Domingues
Fabrício Silva
Heitor Ramos
Antonio A. F. Loureiro

UFMG
UFV
UFMS

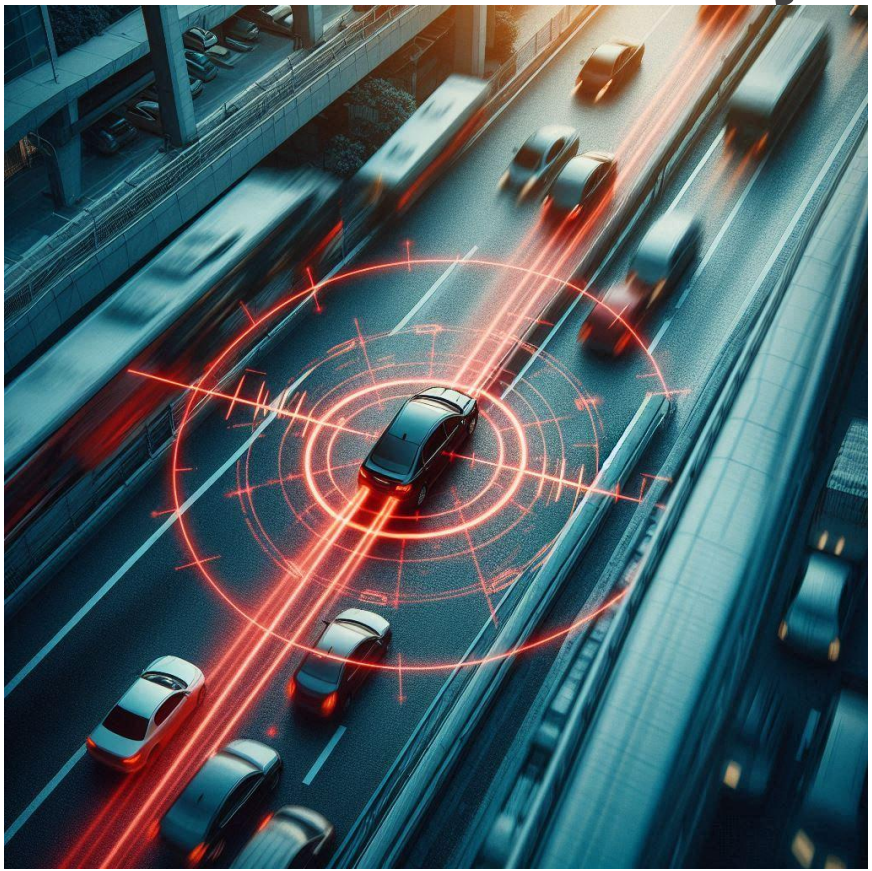
Universidade Federal de Minas Gerais (DCC-UFMG)
Universidade Federal de Mato Grosso do Sul (UFMS)
Universidade Federal de Viçosa (UFV)

Smart Cities: dados de localização e mobilidade

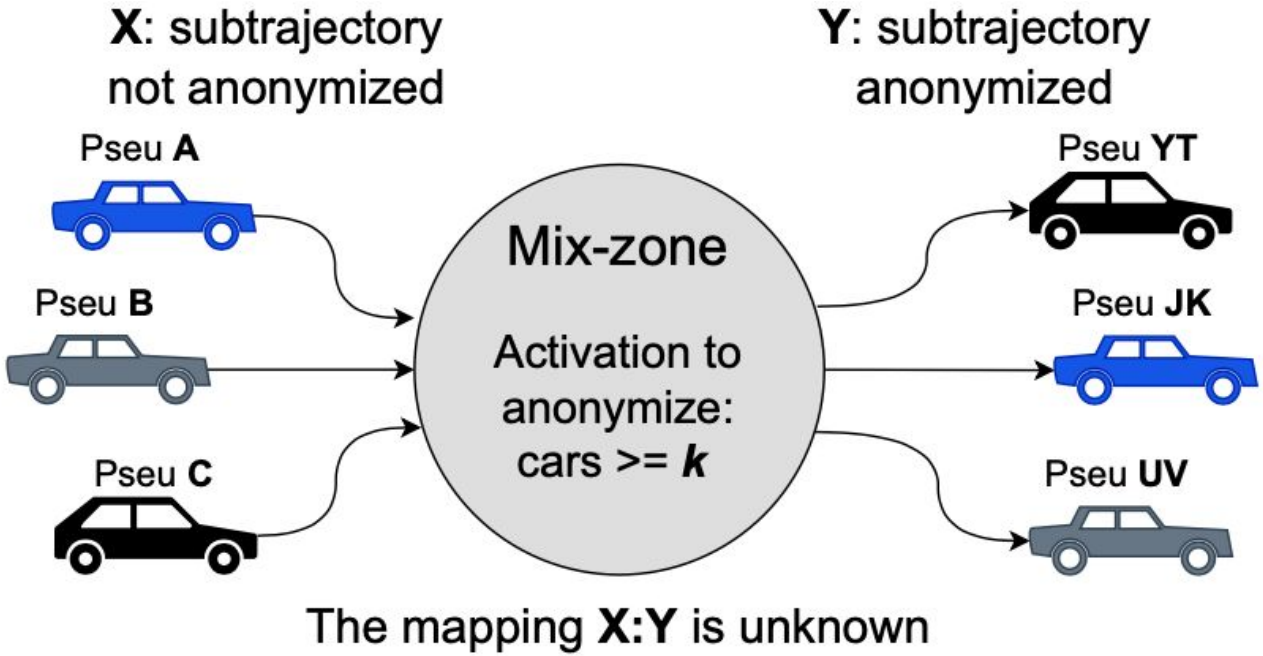
- Smart Cities
- IoT e IoV
- Conectar pessoas e veículos
- Dados de localização
- $$$
- Privacidade

Ameaças e LPPMs - Anonimização

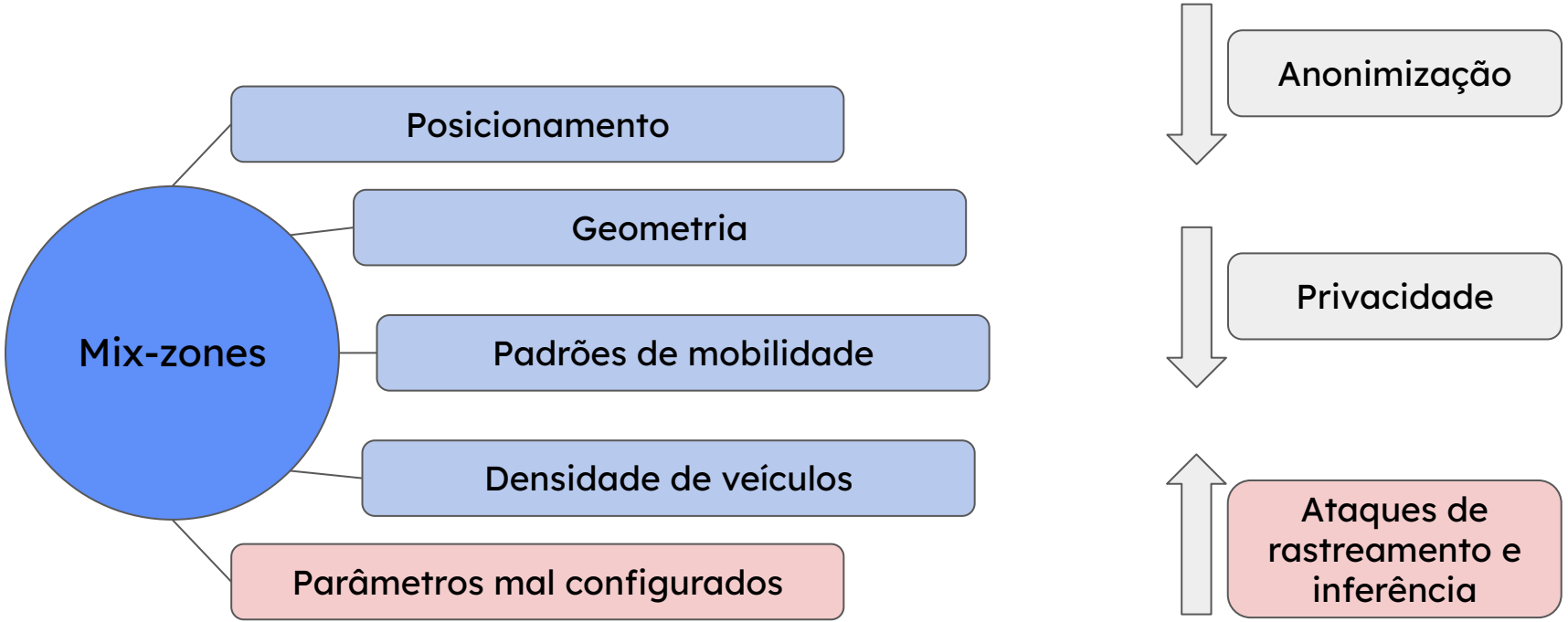
- Informações latentes
- Pontos de interesse
- Comportamento individual
- **Identidade**



Mix-zones


X: subtrajectory not anonymized
Y: subtrajectory anonymized
Pseu A
Pseu B
Pseu C
Mix-zone
Activation to anonymize: cars >= k
Pseu YT
Pseu JK
Pseu UV
The mapping X:Y is unknown

Problema - Mix-zones - Fatores



Problema - Mix-zones - nível de privacidade (*k*)

- Um *k* mal calibrado em uma mix-zone (M) pode não acompanhar as flutuações de tráfego ao longo do tempo.

- Consequência: não atingir o mínimo de número de veículos em M para ativá-la e gerar a anonimização.

Desafio - Mix-zones - nível de privacidade (*k*)

Como melhorar o desempenho das mix-zones em termos de *anonimização*, *eficácia*, *privacidade* e *qualidade de anonimização* diante das flutuações de tráfego de veículos ao longo do tempo?

k-Dynamic Mix-zone (*k*-DynMix)

- Um mecanismo de mix-zone dinâmica que ajusta o nível de privacidade *k* ao longo do tempo, em modo online e complexidade linear, com eventos como as flutuações no tráfego de veículos.

  - **Objetivo**: alcançar a noção de **maior anonimização (*Higher Anonymization* (HA))**.

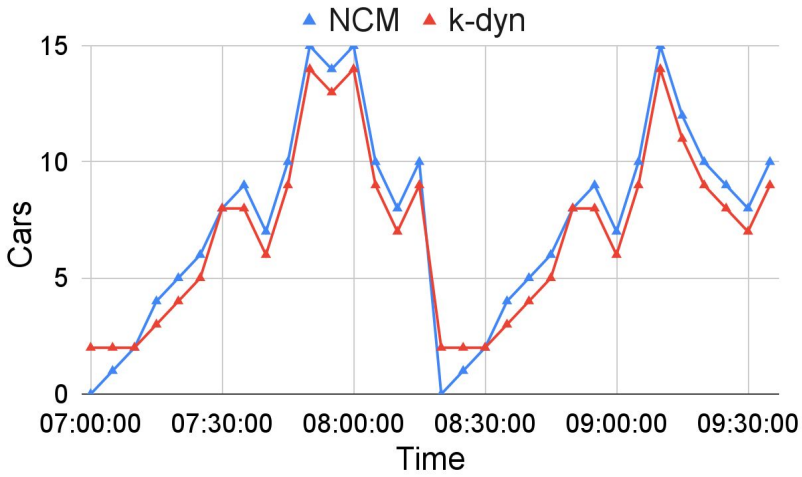
k-DynMix: Conceitos

- *Number of Cars on Mix-zone* (NCM)
- *Mix-zone Activation* (MA):

$$MA \leftarrow NCM \geq k$$

- *Higher Anonymization* (HA):

$$(NCM_t \geq k_t \wedge k_t \approx NCM_t) \implies MA_t^H$$



k-DynMix: Ideia Geral

- Controlar o k_t baseado em eventos que ocorrem nas mix-zones.
- Ajustar *k* como um limite inferior, mas próximo de NCM o mais rápido possível para atender a HA.
- Eventos:
  - *Mix-zone Activation (MA)*
  - *Mix-zone Deactivation (MD)*
  - *Timeout of Arriving Cars in Mix-zones (TAC)*

k-DynMix: Predição de *k*

- Evento MA:
  - *k* cresce exponencialmente até atingir ρ.
  - *k* cresce linearmente até atingir o NCM: evitar perda de privacidade com MD.
- Eventos MD e TAC:
  - $\rho = k/2$,
  - *k* diminui para a privacidade inicial k_b de modo a atingir MA.
-

Experimentos

I - Análise da Acurácia das Técnicas de Predição de k

k-DynMix vs. SMA e WEMA

II - Análise de Cobertura das Mix-zones

k-DynMix vs. Mix-zones estáticas

III - Análise Qualidade da Anonimização (AQ)

- Métricas: funcionamento interno da mix-zone
- Reflete: eficácia, privacidade, anonimização dos dados no momento em que a mix-zone é ativada.

k-DynMix vs. Mix-zones estáticas

Datasets

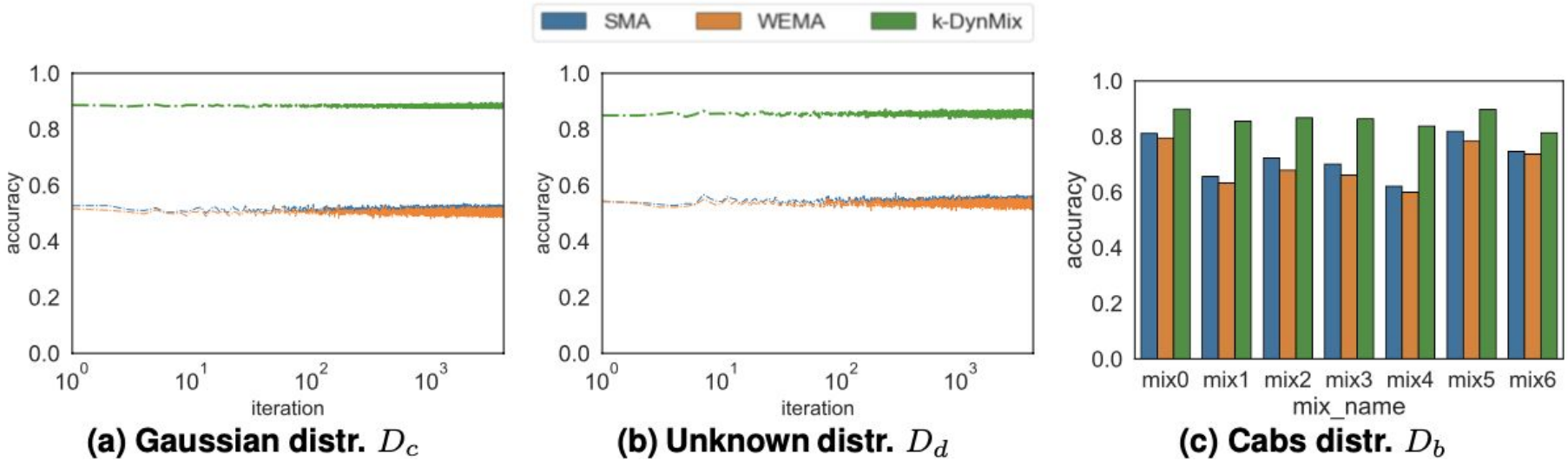
- Real: Cabspotting, SF-EUA, 25 dias, 500, taxis, 440.000 viagens.

Amostra	Dia	Usuários	Registros	Viagens
D_a	18/05/2008	442	366.951	1.770
D_b	19/05/2008	454	417.781	2.036

- Sintético: séries temporais que simulam o fluxo de tráfego NCM em uma mix-zone.

Amostra	Tipo	Registros	NCM
D_c	Gaussiana	2000	μ=10, σ=2,5
D_d	Aleatória	2000	0-21

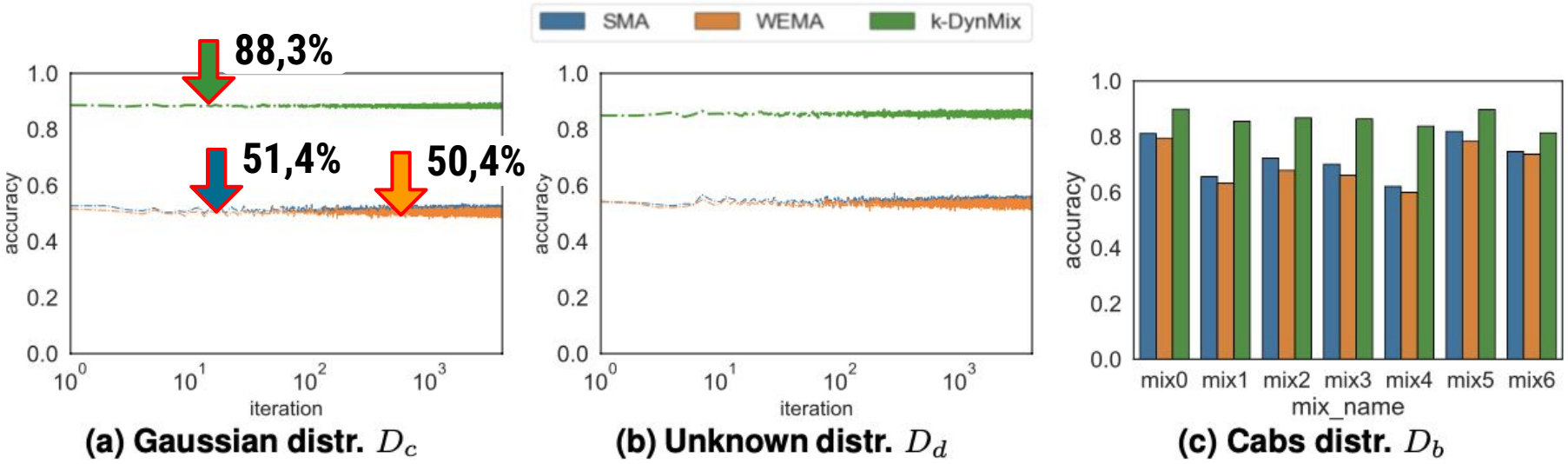
I- Acurácia em prever k



(a) Gaussian distr. D_c (b) Unknown distr. D_d (c) Cabs distr. D_b

- $ACC_{pred} = |MA|/(|MA| + |MD|)$
- Figs (a) e (b) - *Bootstrap*, 4000 reamostragens da distr. NCM.

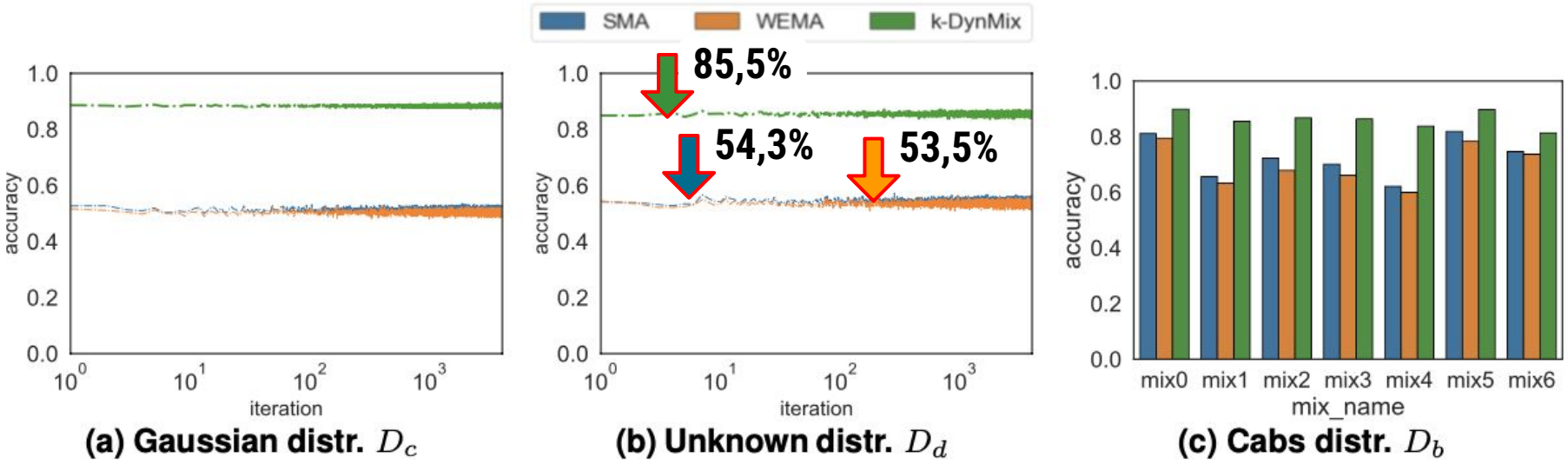
I- Acurácia em prever k



(a) Gaussian distr. D_c (b) Unknown distr. D_d (c) Cabs distr. D_b

- Figs (a) e (b) - *Bootstrap*, 4000 reamostragens da distr. NCM.

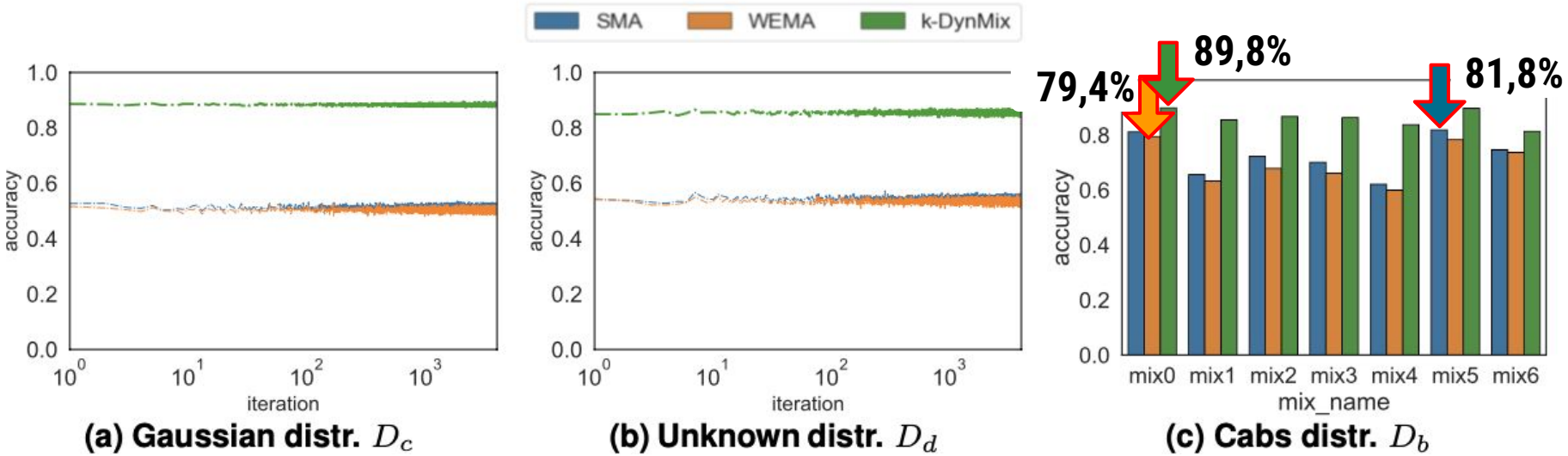
I- Acurácia em prever k



(a) Gaussian distr. D_c (b) Unknown distr. D_d (c) Cabs distr. D_b

- Figs (a) e (b) - *Bootstrap*, 4000 reamostragens da distr. NCM.

I- Acurácia em prever k



(a) Gaussian distr. D_c (b) Unknown distr. D_d (c) Cabs distr. D_b

- Fig (c) - NCM de 7 mix-zones posicionadas.
- Conjunto de dados Cabspotting - D_b

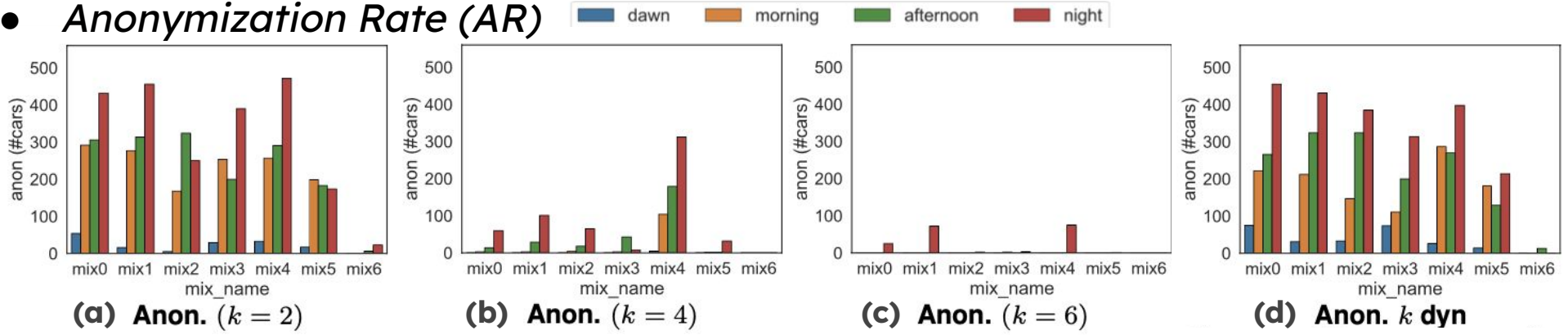
II- Cobertura das Mix-zones

- D_a anonimizado com 7 mix-zones posicionamento
    - k = 2,4,6 vs. k-DynMix.

- Métricas:
    - *Non-Anonymization Rate (NAR)*
    - *Anonymization Rate (AR)*
    - *Mix-zone Efficacy (ME)*
- Janela de Tempo: *dawn, morning, afternoon, and night*

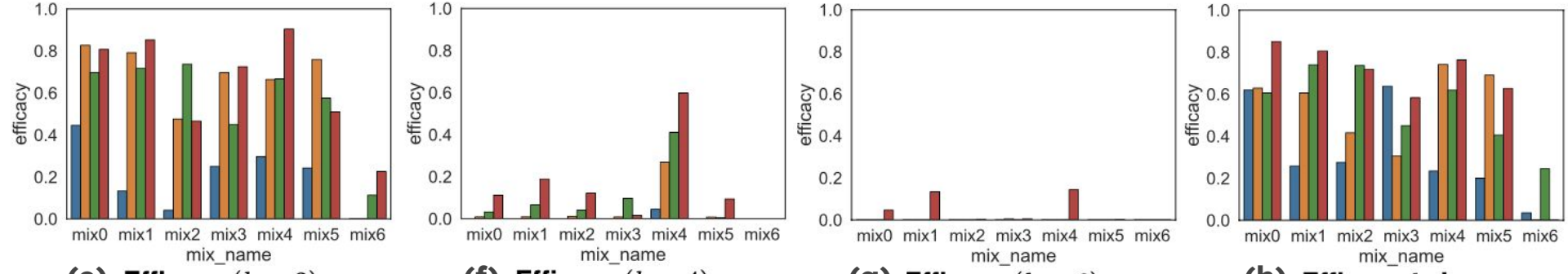
II- Cobertura das Mix-zones

- *Anonymization Rate (AR)*



(a) **Anon.** $(k = 2)$ (b) **Anon.** $(k = 4)$ (c) **Anon.** $(k = 6)$ (d) **Anon.** k **dyn**

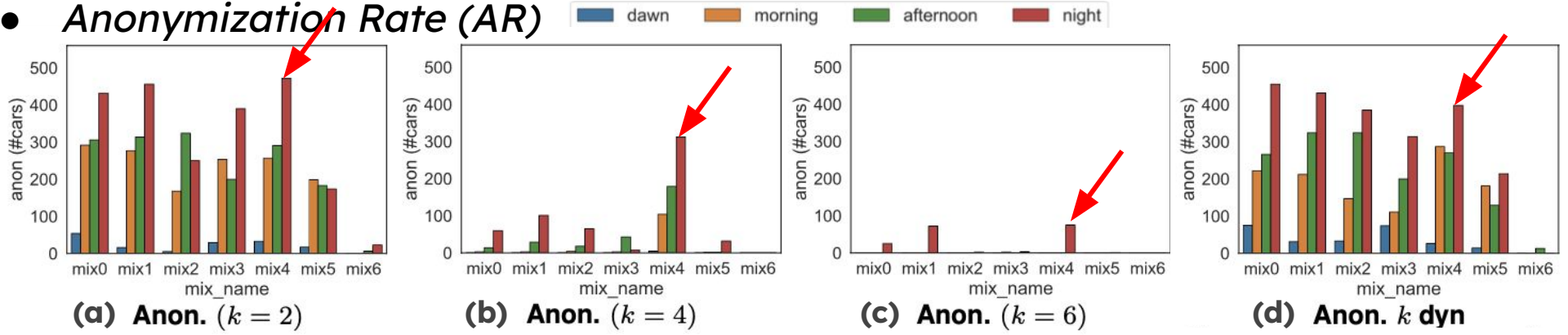
- *Mix-zone Efficacy (ME)*



(e) **Efficacy** $(k = 2)$ (f) **Efficacy** $(k = 4)$ (g) **Efficacy** $(k = 6)$ (h) **Efficacy** k **dyn**

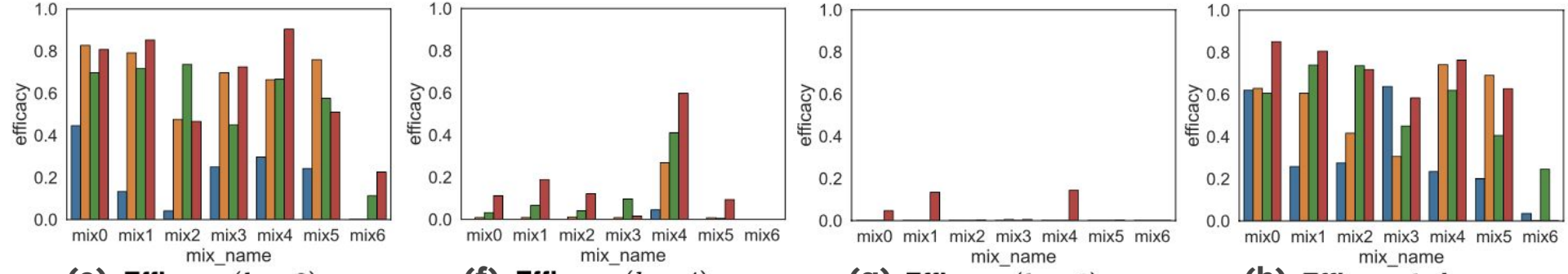
II- Cobertura das Mix-zones

- *Anonymization Rate (AR)*



(a) **Anon.** $(k = 2)$ (b) **Anon.** $(k = 4)$ (c) **Anon.** $(k = 6)$ (d) **Anon.** k **dyn**

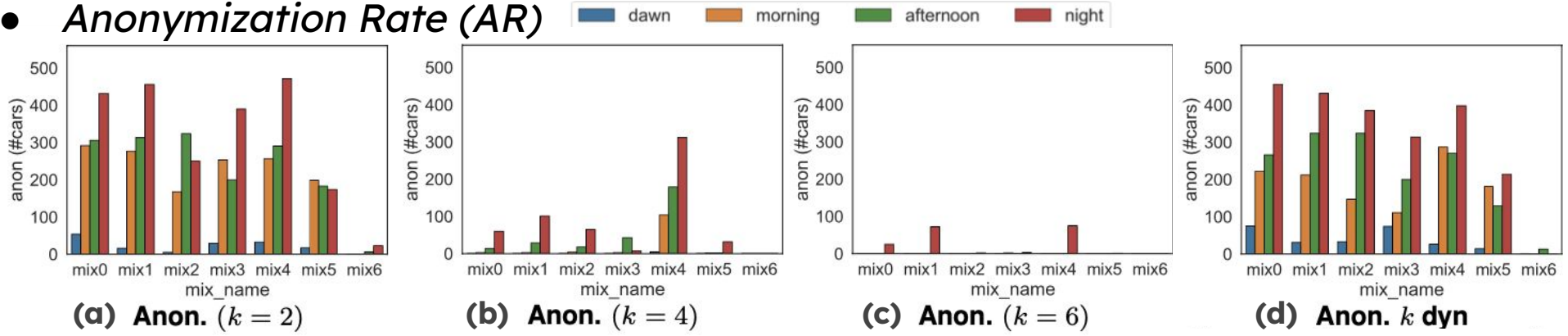
- *Mix-zone Efficacy (ME)*



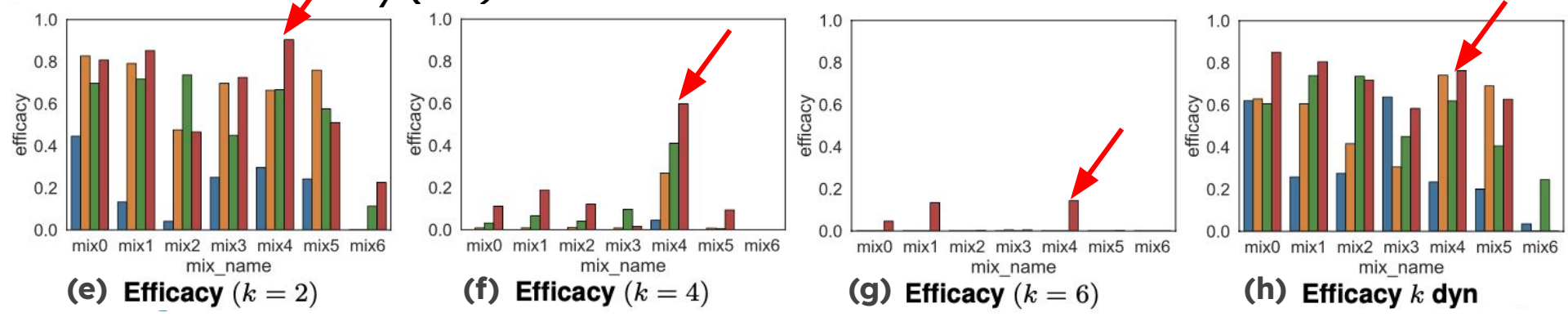
(e) **Efficacy** $(k = 2)$ (f) **Efficacy** $(k = 4)$ (g) **Efficacy** $(k = 6)$ (h) **Efficacy** k **dyn**

II- Cobertura das Mix-zones

- *Anonymization Rate (AR)*

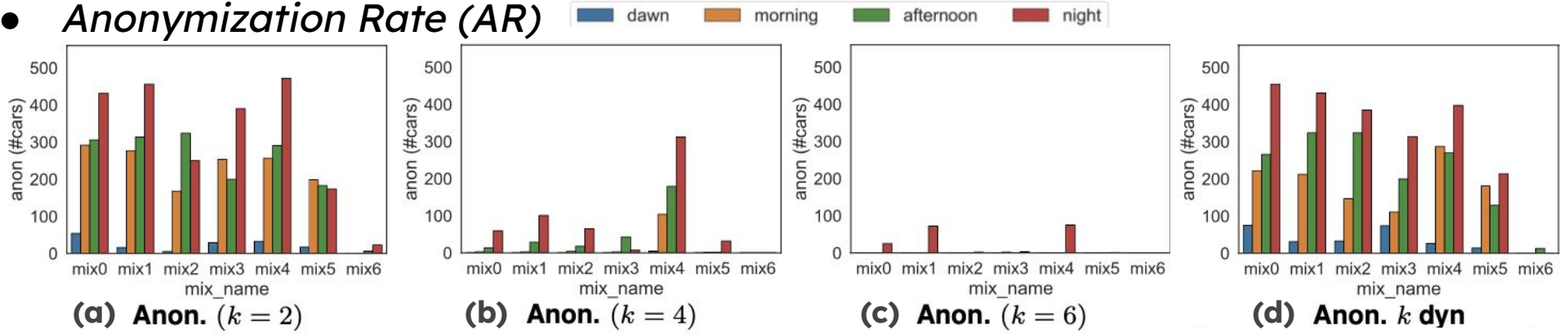


- *Mix-zone Efficacy (ME)*



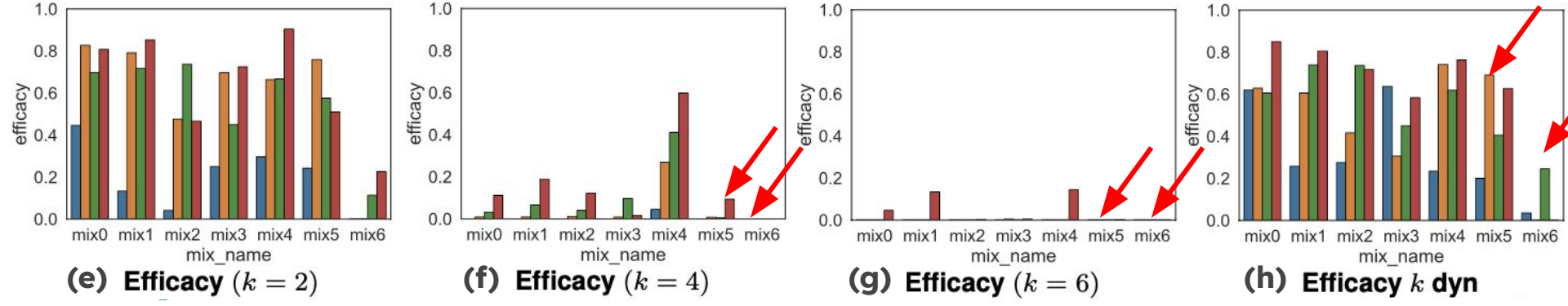
II- Cobertura das Mix-zones

- *Anonymization Rate (AR)*



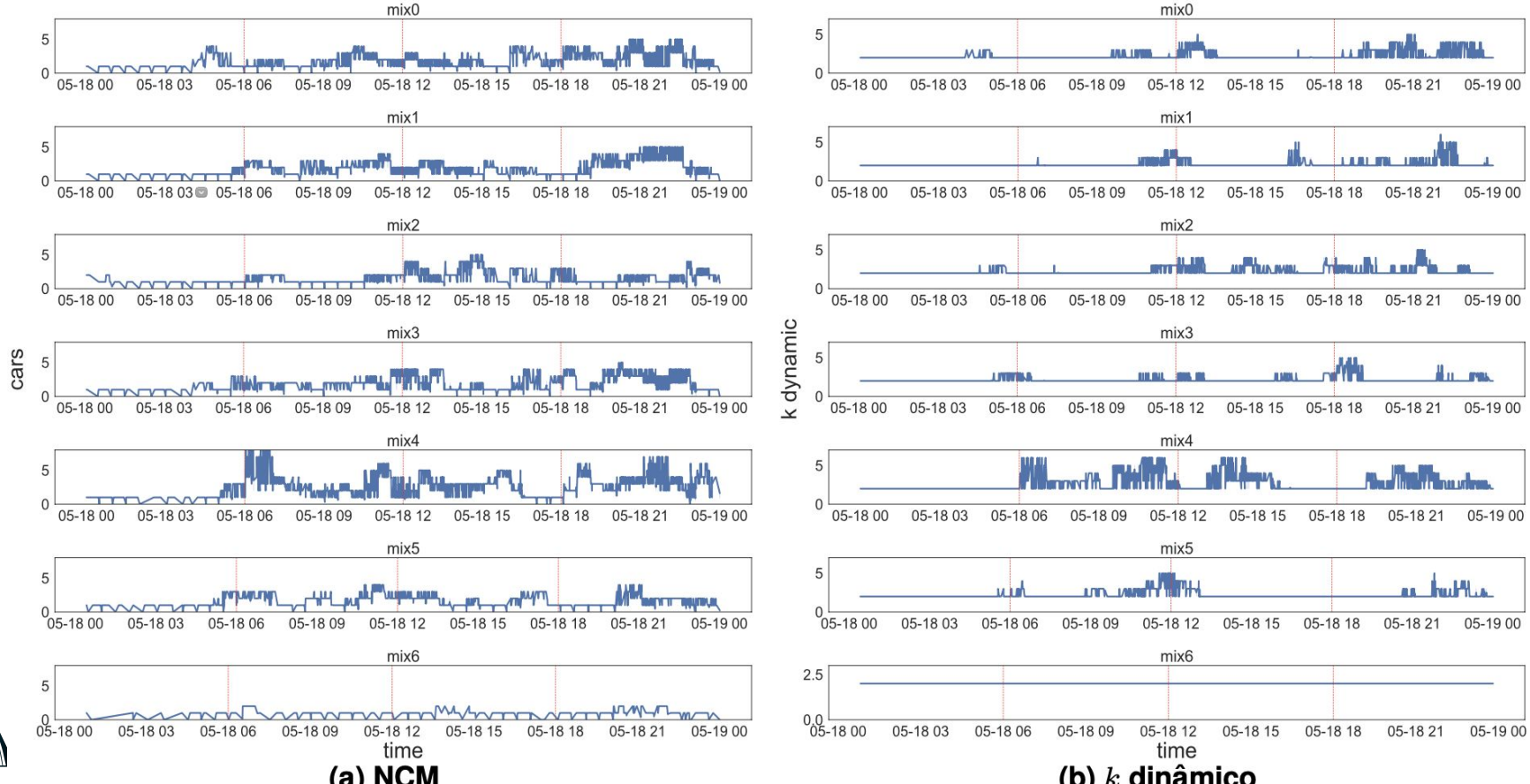
(a) **Anon.** $(k = 2)$ (b) **Anon.** $(k = 4)$ (c) **Anon.** $(k = 6)$ (d) **Anon.** k **dyn**

- *Mix-zone Efficacy (ME)*



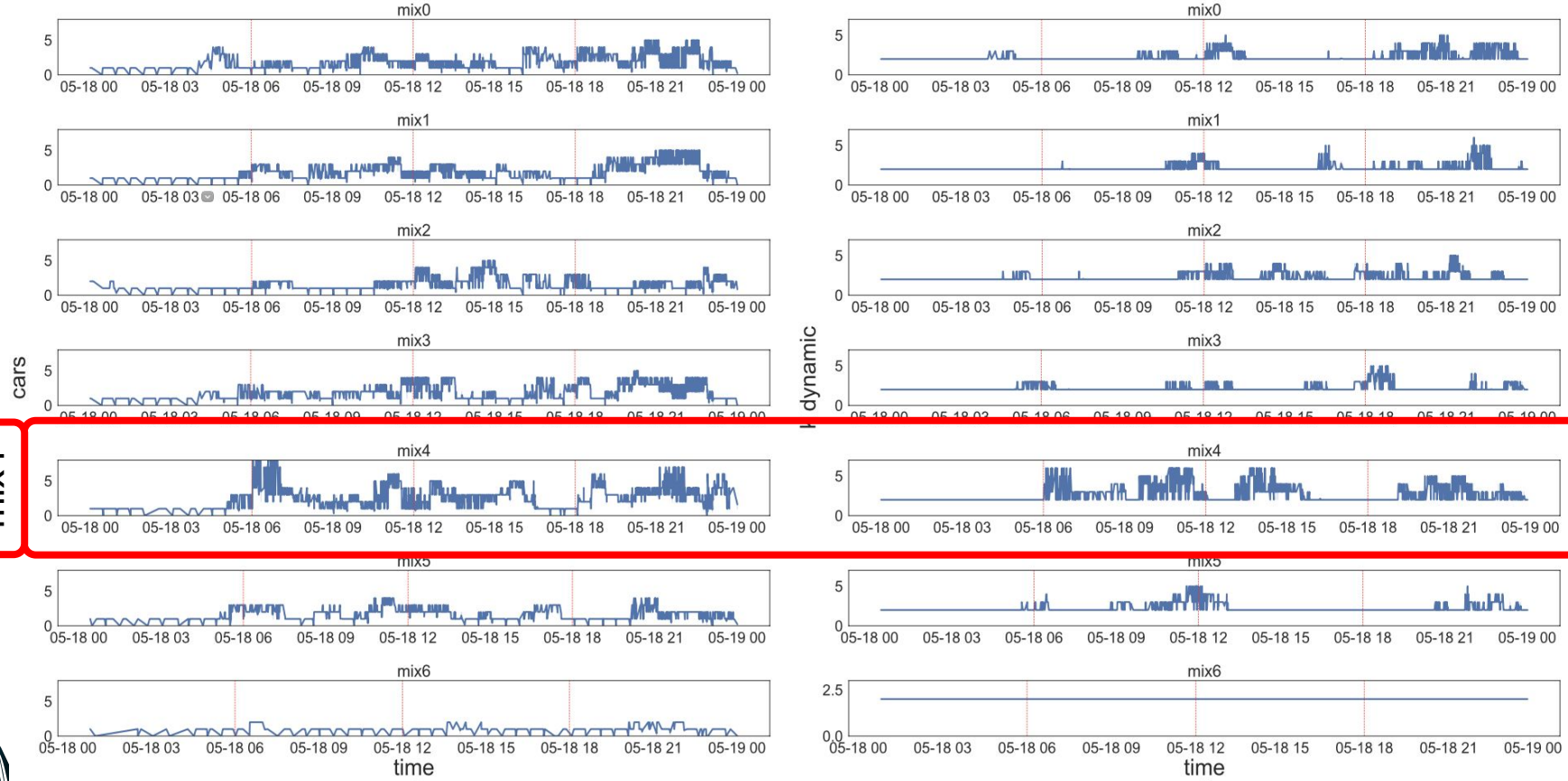
(e) **Efficacy** $(k = 2)$ (f) **Efficacy** $(k = 4)$ (g) **Efficacy** $(k = 6)$ (h) **Efficacy** k **dyn**

III- AQ: Number of Cars on Mix-zone (NCM)

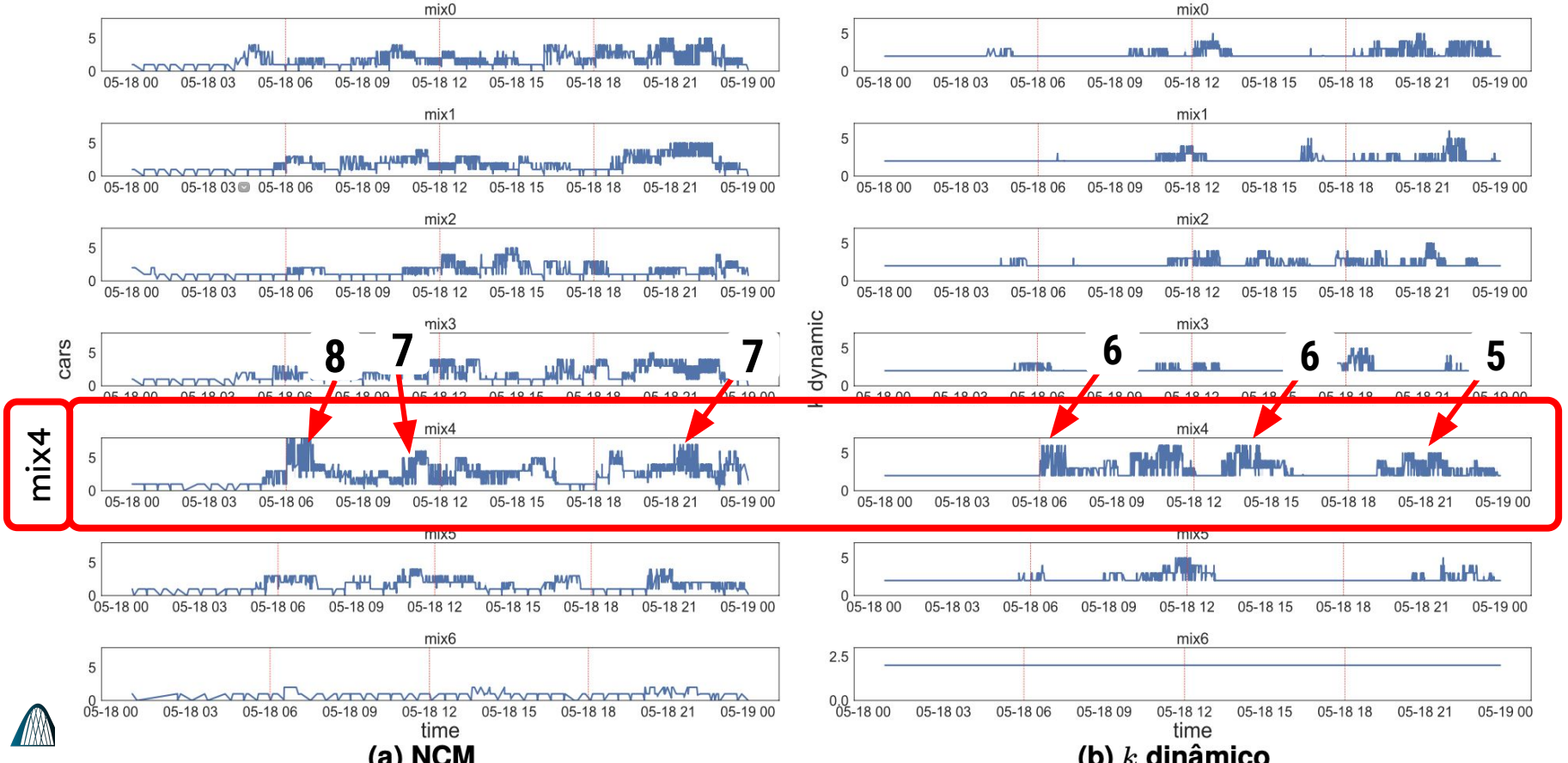


(a) NCM (b) k dinâmico

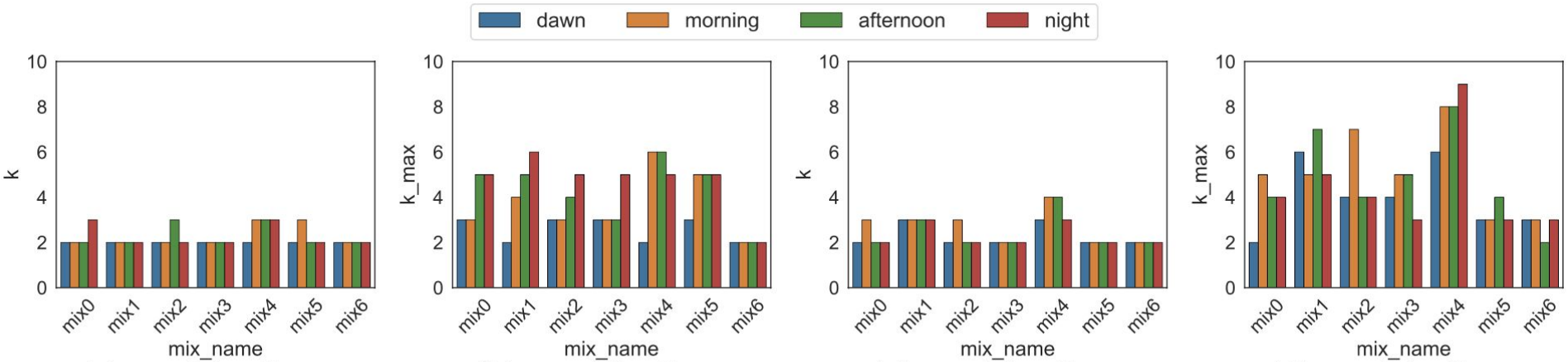
III- AQ: Number of Cars on Mix-zone (NCM)



III- AQ: Number of Cars on Mix-zone (NCM)



III- k médio e máximo



(a) avg(k), dia

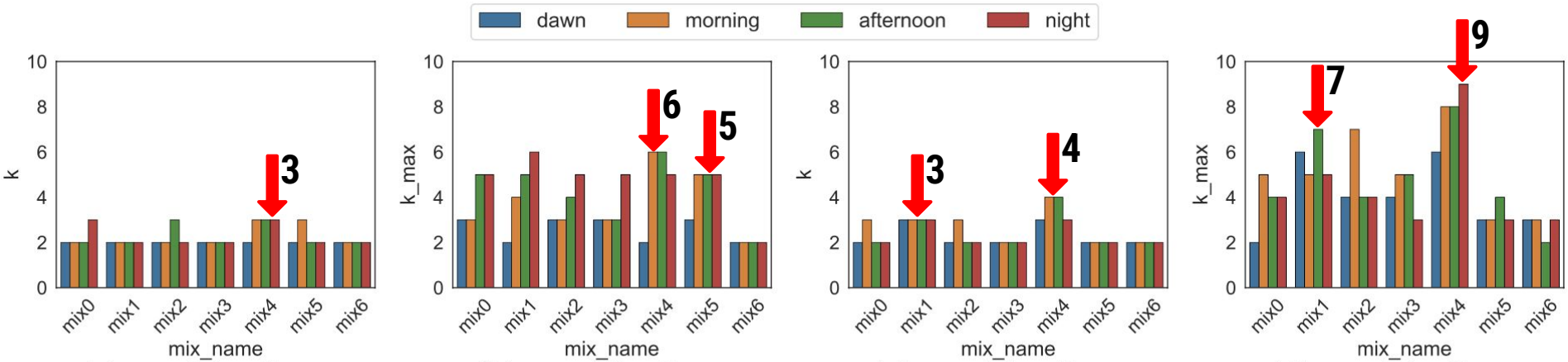
(b) max(k), dia

(c) avg(k), dia

(d) max(k), dia

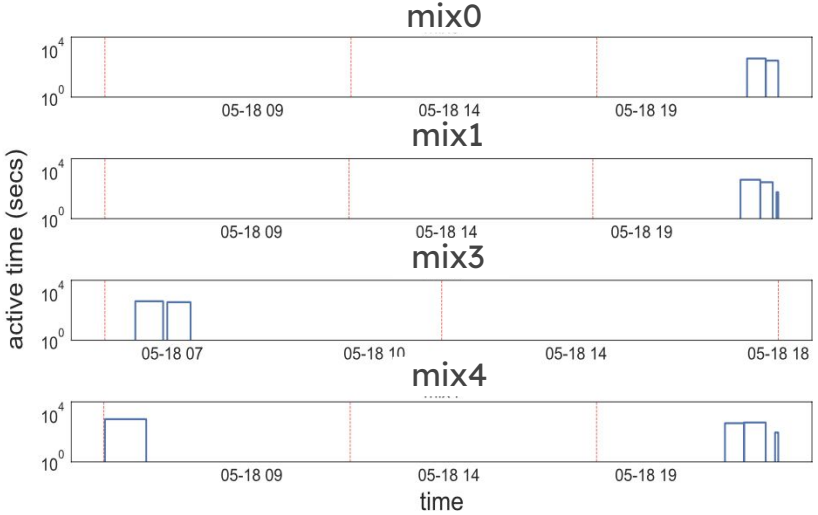
	18^o	18^o	19^o	19^o
Datasets	D_a	D_a	D_b	D_b

III- k médio e máximo

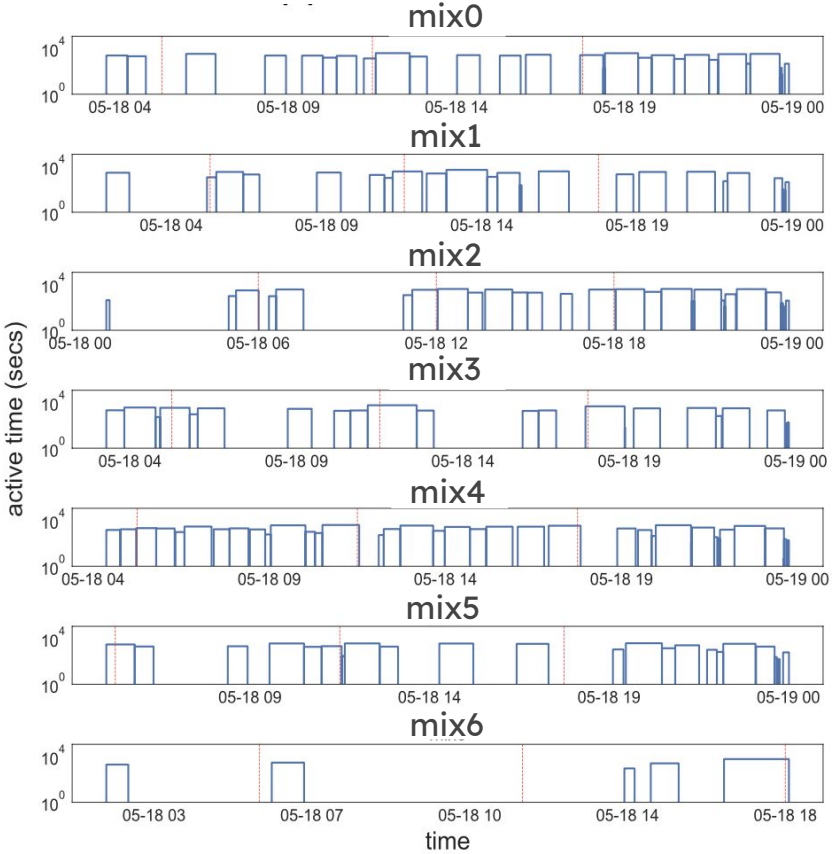


	(a) avg(k), dia	(b) max(k), dia	(c) avg(k), dia	(d) max(k), dia
	18^o	18^o	19^o	19^o
Datasets	D_a	D_a	D_b	D_b

III- AQ: Activation Time of the Mix-zone (ATM)

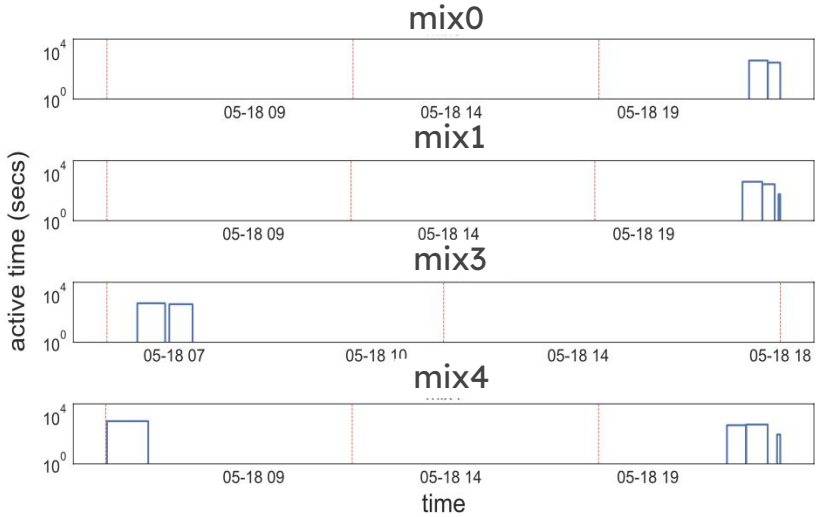


(e) ATM para $k = 6$

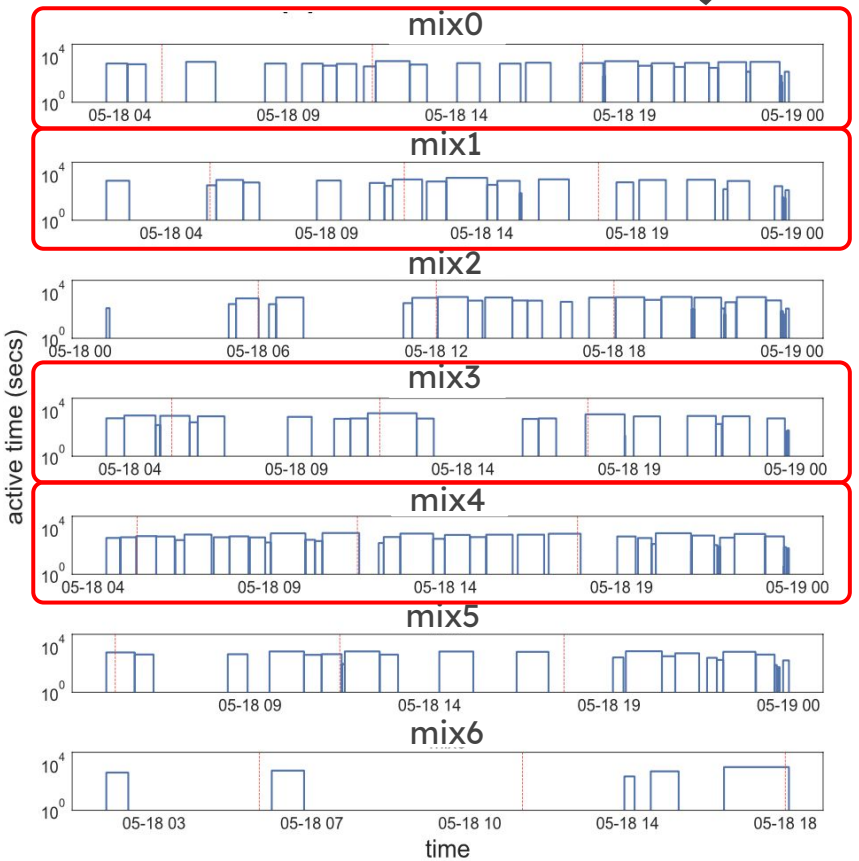


(f) ATM para k dinâmico

III- AQ: Activation Time of the Mix-zone (ATM)



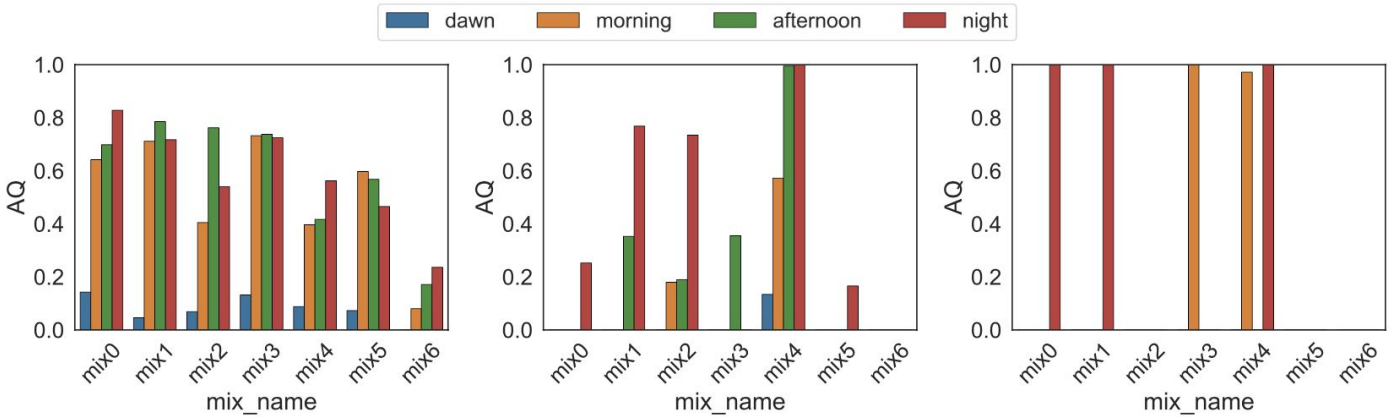
(e) ATM para $k = 6$



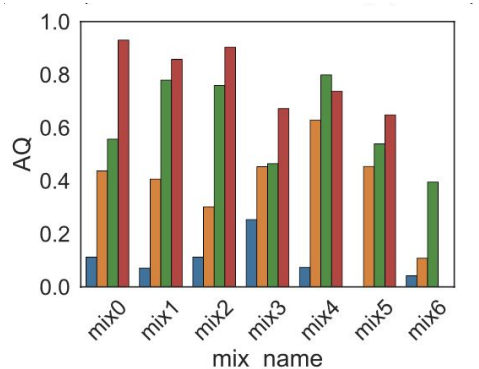
(f) ATM para k dinâmico

- Não houveram ATMs:
- mix2, mix5, mix6

III- Qualidade da Anonimização (AQ)

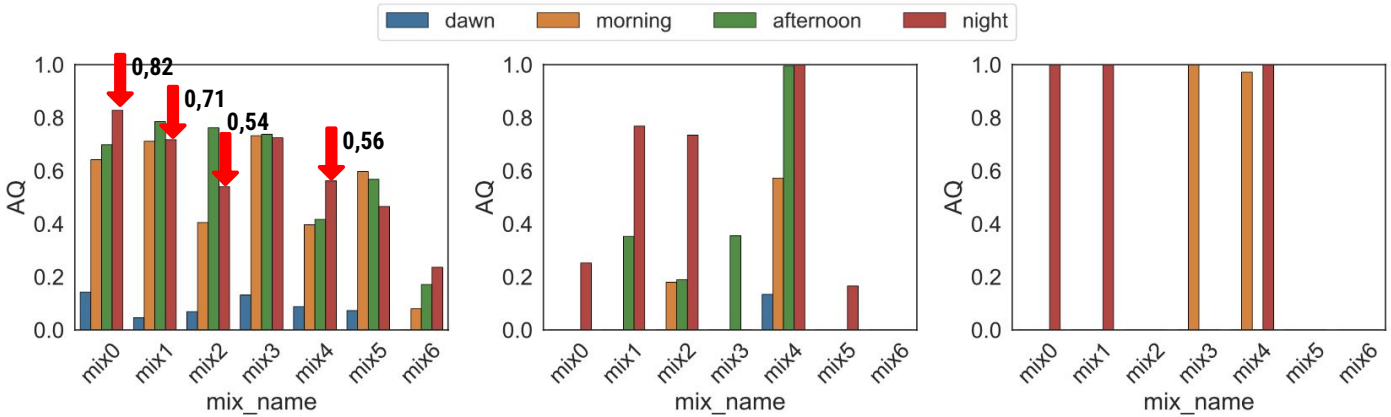


(a) AQ $(k = 2)$**.** **(b) AQ** $(k = 4)$**.** **(c) AQ** $(k = 6)$**.**

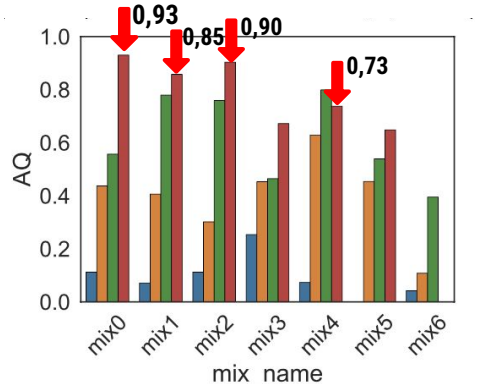


(d) AQ k **dyn.**

III- Qualidade da Anonimização (AQ)



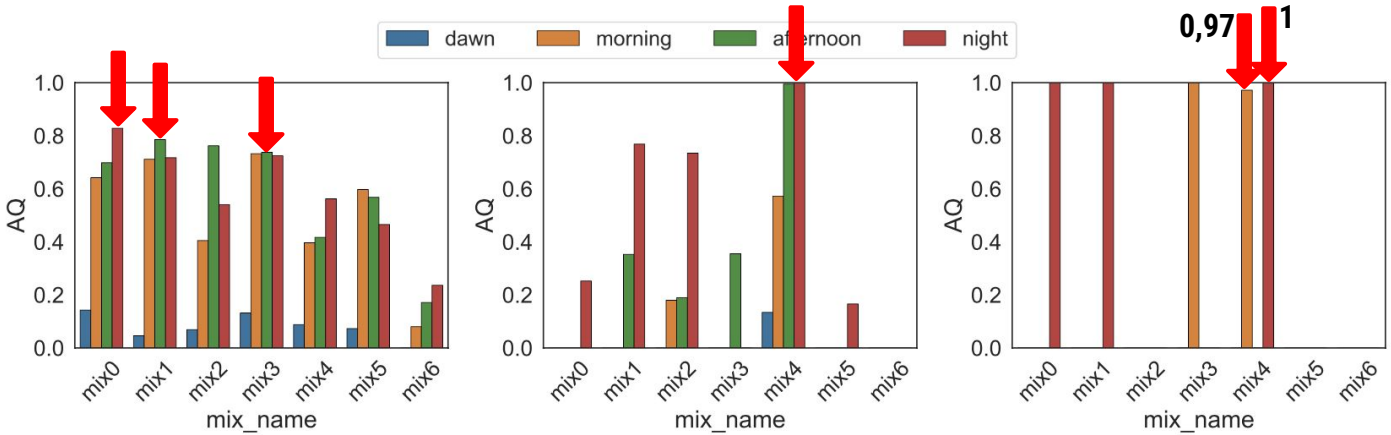
(a) AQ $(k = 2)$. (b) AQ $(k = 4)$. (c) AQ $(k = 6)$.



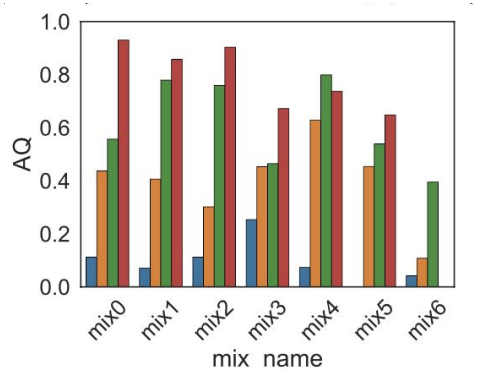
(d) AQ k dyn.

0,73

III- Qualidade da Anonimização (AQ)

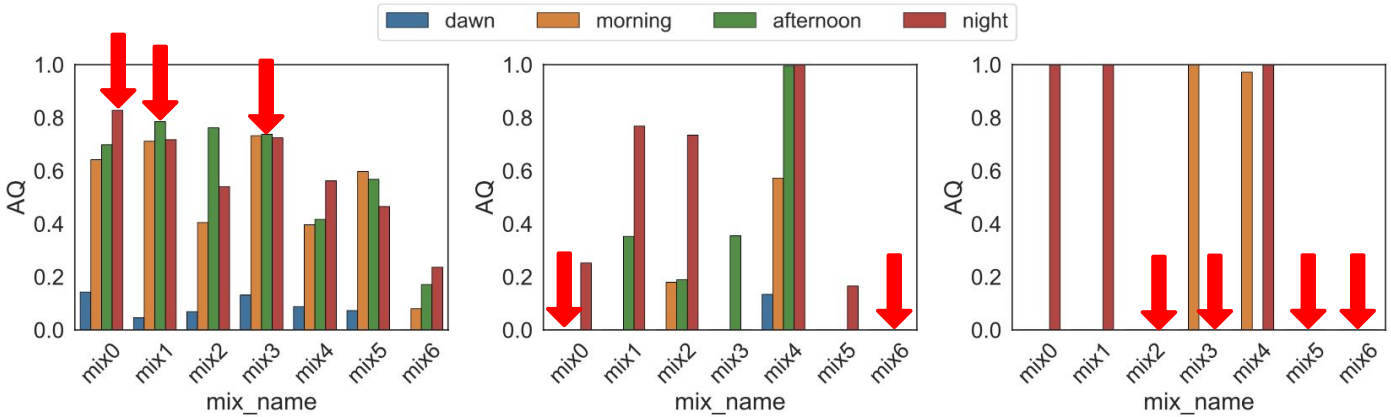


(a) AQ $(k = 2)$. (b) AQ $(k = 4)$. (c) AQ $(k = 6)$.

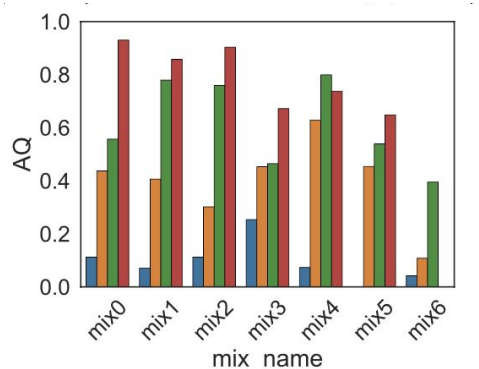


(d) AQ k dyn.

III- Qualidade da Anonimização (AQ)

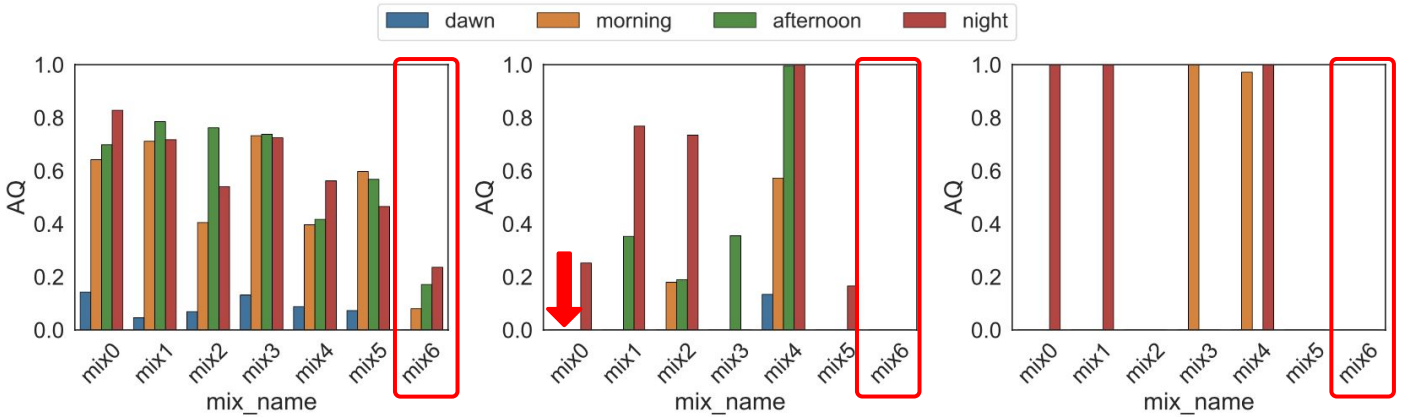


(a) AQ $(k = 2)$. (b) AQ $(k = 4)$. (c) AQ $(k = 6)$.

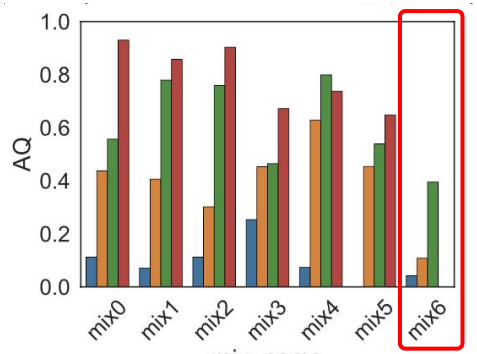


(d) AQ k dyn.

III- Qualidade da Anonimização (AQ)



(a) AQ $(k = 2)$. (b) AQ $(k = 4)$. (c) AQ $(k = 6)$.



(d) AQ k dyn.

Considerações finais

- k-DynMix superou os modelos de predição em estimar a privacidade.
- Resultados semelhantes ao melhor resultado das mix-zones clássicas
  a. Métricas de cobertura e AQ
- Superou as mix-zones clássicas
  a. Maximizou a privacidade ao melhor possível.
  b. Mix-zones de baixo tráfego.

Trabalhos Futuros

- Comparar k-DynMix com técnicas de predição para séries temporais.
- Aplicar ataques de rastreamento para testar a robustez da nossa solução.
- Testar a solução para datasets de diferentes modais.

Obrigado!

- Ekler Paulino de Mattos

ekler.mattos@dcc.ufmg.br