

# Simpósio de Integração Acadêmica

“Bicentenário da Independência: 200 anos de ciência, tecnologia e inovação no Brasil e 96 anos de contribuição da UFV”

SIA UFV 2022



## SENDAS: SCALABLE ENRICHMENT FOR MOBILITY DATA SETS

Henrique S. Santana, Fabrício A. Silva - [henrique.s.santana,fabricio.asilva@ufv.br](mailto:henrique.s.santana,fabricio.asilva@ufv.br)  
Universidade Federal de Viçosa, Instituto de Ciências Exatas e Tecnológicas – Campus Florestal

Palavras-Chave: mobilidade humana, Big Data, análise de mobilidade

Área Temática: Ciência da Computação - Grande Área: Ciências Exatas e Tecnológicas  
Categoria: Pesquisa - Financiamento: CNPq

### Introdução

O avanço e popularização da tecnologia nos anos recentes fez crescer a disponibilidade de dados de mobilidade, coletados de diferentes fontes. Entretanto, dados de trajetórias têm seguido a tendência de Big Data, e as ferramentas atualmente existentes não cumprem com o requisito de escalabilidade. Para cobrir essa falta, nós propomos o Sendas (*Scalable ENrichment for mobility DATA Sets*), uma biblioteca em Scala construída com o *framework* Apache Spark[1], com o objetivo de conceder execução paralela e distribuída para técnicas consolidadas de análise de mobilidade.

### Objetivos

- Elaborar uma biblioteca escalável e extensível para tratar grandes volumes de dados de mobilidade;
- Prover implementações de algoritmos do estado-da-arte;
- Apresentar melhorias em técnicas de cálculo de fluxo e identificação de padrões (*motifs*) de mobilidade;
- Avaliar a solução proposta em conjuntos de dados reais e comparar com soluções já existentes.

### Material e Métodos

Formalizamos definições sobre a terminologia de dados de trajetória, o cálculo de fluxo de mobilidade, e a identificação de *motifs* de mobilidade[2]. A principal delas é a de que trajetórias podem ser representadas de diferentes formas[3], através de uma representação espacial e uma visão de trajetória – um formato que organiza no tempo os registros espaciais. No cálculo de fluxo, introduzimos a noção de divisão e categorização de tempo, para a inspeção de aspectos temporais do fluxo (ex.: fluxo médio por dia). Na identificação de *motifs*, propusemos uma nova técnica de rotulação dos grafos que representam cada padrão, levando em conta suas características específicas, levando a um algoritmo de complexidade de tempo linear.

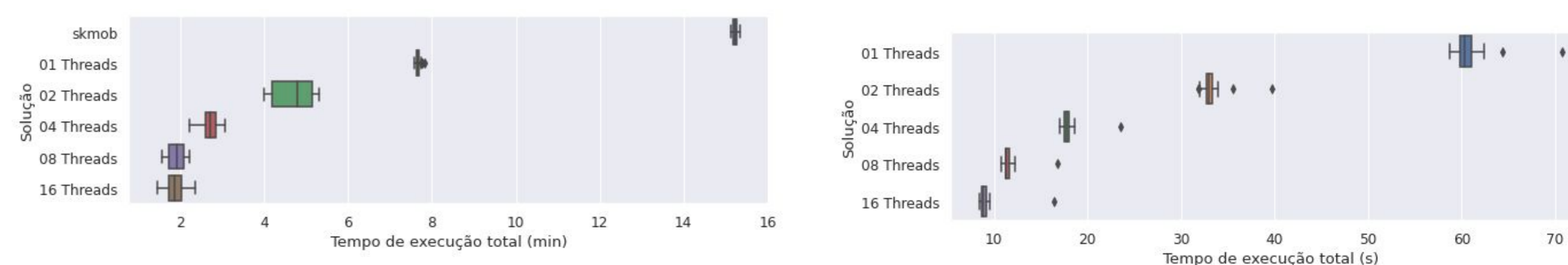
### Apoio Financeiro

CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico.



### Resultados e Discussão

Para avaliar a proposta, aplicamos as funcionalidades apresentadas num dataset de 49.790 usuários únicos e cerca de 8,5 milhões de registros. Foi comparado o tempo de execução de cada funcionalidade considerando a execução sequencial e a paralela, com diferentes números de threads, sendo o cálculo de fluxo também comparado com o desempenho da biblioteca Scikit-Mobility[4]. Os resultados obtidos mostraram ganhos de 4 a 6 vezes no tempo de execução paralela da biblioteca em relação à execução não-paralela. O primeiro gráfico se refere ao fluxo, e o segundo se refere aos *motifs*.



### Conclusões

Assim, as contribuições deste trabalho são múltiplas: refinamos conceitos da análise de mobilidade, propusemos técnicas de melhor desempenho e elaboramos uma nova ferramenta de utilidade para a comunidade científica.

### Bibliografia

- [1] ZAHARIA, M. et al. Resilient distributed datasets: A fault-tolerant abstraction for in-memory cluster computing. In: **9th USENIX Symposium on Networked Systems Design and Implementation (NSDI 12)**. San Jose, CA: USENIX Association, 2012. p. 15–28. ISBN 978-931971-92-8.
- [2] SCHNEIDER, C. M. et al. Unravelling daily human mobility motifs. **Journal of The Royal Society Interface**, The Royal Society, v. 10, n. 84, p. 20130246, 2013.
- [3] GRASER, A. MovingPandas: Efficient structures for movement data in Python. **GI\_Forum – Journal of Geographic Information Science** 2019, v. 7, p. 54–68, 2019.
- [4] PAPPALARDO, L. et al. **scikit-mobility: a Python library for the analysis, generation and risk assessment of mobility data**. 2021.

### Agradecimentos

Pesquisa desenvolvida com o auxílio do Cluster da UFV (Universidade Federal de Viçosa).

