Tiago da Silva

tdsh97@gmail.com | github.com/tiagodsilva | tiagodsilva.github.io | linkedin.com/in/tiagodasilvah

Education

PhD in Applied Mathematics 2024/03 - 2024/12

School of Applied Mathematics, Brazil. Advisor: Prof. Diego Mesquita.

- Thesis title: Streaming, Distributed, and Asynchronous Amortized Inference.
- Eligible to fast-track graduation due to recognized academic excellence.

MSc in Applied Mathematics 2024/03 – 2024/12

School of Applied Mathematics, Brazil. Advisor: Prof. Diego Mesquita.

- Thesis title: Human-aided Causal Discovery of Ancestral Graphs.

BSc in Data Science 2020/03 – 2023/12

School of Applied Mathematics, Brazil

- GPA: 9.9/10.0. 1st in class and 1st in the entrance exam. Received a fully-funded scholarship.

Employment

Rei do Pitaco 2025/01 – ongoing

- Data Scientist.
  - Created a scalable FastAPI-based solution for serving the company's internal bet building product.
  - The service generates over 30% of the company total's monthly revenue and roughly 50% of revenue from core products.
  - Complete product ownership: from designing the frontend (TypeScript) and backend (Python), to provisioning infrastructure (Terraform) and setting up observability (Datadog, Airflow).
  - Developed the first company-wide MCP to streamline operational management of our bet building product.

Proffer (Price Monitoring & Optimization startup) 2024/04 – ongoing

- Software Engineer.
  - Extended the company's API to support larger workloads via asynchronous processing and caching.
  - Developed new features that facilitated the integration of the company's services with the customers' ERPs.
  - Contributed in the migration of the company's infrastructure between cloud providers.
  - Helped refactor the company's ETL pipeline to ensure the codebase scaled sublinearly with cutomer growth.

Rei do Pitaco (largest fantasy sports company in Brazil) 2023/01 – 2023/07

- Data Science intern.
  - Designed predictive models to define the opening lines of bets on the outcomes of sport events (bookmaking).
  - Deployed and upheld the created models within applications serving thousands of concurrent users.

Honors & Awards

Award for Academic Excellence, Brazilian Society of Applied and Computational Mathematics. 2023

First place, School of Applied Mathematics entrance exam. 2020

I was awarded 19 prizes in scientific competitions during high school, including:

William Glenn Whitley Prize for achieving the highest score on the State Mathematical Olympiad. 2019

Top score in the country, Brazilian Mathematical Olympiad of Public Schools. 2019

Top score in the country, Brazilian Mathematical Olympiad of Public Schools. 2018

Gold medal, Brazilian Chemistry Olympiad. 2018

Gold medal, Brazilian Mathematical Olympiad of Public Schools. 2017

Gold medals, State Chemistry Olympiad. Highest score in 2019. 2016-2019

Selected Publications

1. **When do GFlowNets learn the right distribution?** ICLR 2025 (spotlight)
   - da Silva, T., Silva, E., Alves, R., Souza, A., Garg, V., Kaski, S., Mesquita, D.
   - TL;DR: We show that a GFlowNet's learning objective may have an unattainable global minimum, which is undetectable by standard diagnostics. To address this, we propose the first tractable metric for assessing GFlowNets.

2. **Generalization and Distributed Learning of GFlowNets?** ICLR 2025
   - da Silva, T., Souza, A., Rivasplata, O., Garg, V., Kaski, S., Mesquita D.
   - TL;DR: We devise the first PAC-Bayesian generalization bounds for GFlowNets. Inspired by them, we also introduce the first general-purpose distributed learning algorithm for GFlowNets, drastically accelerating training convergence.

3. **Streaming Bayes GFlowNets** NeurIPS 2024
   - da Silva, T., Souza, D., and Mesquita, D.
   - TL;DR: We design a method to update GFlowNets trained on a streaming Bayesian posterior. Experiments show a drastic reduction in training time when compared against learning from scratch a model based on the entire dataset.

4. **On Divergence Measures for Training GFlowNets** NeurIPS 2024
   - da Silva, T., Silva, E., and Mesquita, D.
   - TL;DR: We empirically show that the inefficacy of divergence-based objectives for GFlowNets is due to their large gradient variance. We then develop variance reduction techniques that significantly accelerate training convergence.

5. **Embarrassingly Parallel GFlowNets** ICML 2024
   - da Silva, T., Souza, A., Carvalho, L., Kaski, S., and Mesquita, D.
   - TL;DR: We propose a divide-and-conquer approach to train a log-pool of GFlowNets in an embarrassingly parallel fashion. Results show a significant speed up in learning when the unnormalized target is expensive to evaluate.

6. **Exploring scientific literature by textual and image content using DRIFT** Computer & Graphics 2022
   - Pocco, X., da Silva, T., Poco, J., Nonato, L. G., Gomez-Nieto, E.
   - TL;DR: We developed a text- and image-driven visualization-based search engine for scientific literature.

Preprints & Workshops

1. **Human-aided Causal Discovery of Ancestral Graphs** LatinX @ NeurIPS 2024
   - da Silva, T., Silva, E., Góis, A., Heider, D., Kaski, S., Mesquita, D., Ribeiro, A.
   - TL;DR: We devise a Bayesian human-in-the-loop algorithm for causal discovery under latent confounding.

Research Experience

Green AI Lab, Brazil 2022/08 – ongoing

Working with probabilistic ML. My recent efforts were mostly directed towards leveraging GFlowNets for asynchronous and approximate Bayesian inference. I have also worked on geometric deep learning, learning theory, variational autoencoders, diffusion probabilistic models, and PINNs. Our current research led to publications at ICML and NeurIPS.

Aalto University, Finland 2024/07 – 2024/10

I was a visiting scholar on the Probabilistic Machine Learning group under the supervision of Prof. Vikas Garg and Prof. Sami Kaski. I worked on developing (non-vacuous) statistical guarantees for GFlowNets and on geometric deep learning.

Visual Data Science Lab, Brazil 2020/08 – 2023/01

Research assistant supervised by Prof. Jorge Poco. I assisted the development of a framework for reverse engineering of visualizations (see the open-source library REV) and of a platform for image-based literature search (see our C&G paper).

Languages

Portuguese (Native), English

Skills

Computer languages: Proficient with Python and SQL. Competent with R and Stan. Familiar with C++ and JavaScript.

Scientific computing frameworks: PyTorch, PyTorch Geometric, GPyTorch, NumPy, SciPy.

Technologies: Git, Linux.

Data visualization: Matplotlib, Altair, Vega-lite, D3.

Computer vision libraries: OpenCV, YOLOv5, SAM.

References

Diego Mesquita diego.mesquita@fgv.br | Amauri Souza amauri.souza@aalto.fi