João Pedro Rodrigues Mattos

linkedin.com/in/joaopedromattos

Email : jrm28@rice.edu
Mobile : +19735838844

Education

- **Rice University** Aug 2022 – May 2027 (expected)
  PhD in Computer Science focused on Machine Learning and Graphs; GPA: 4.00

- **University of São Paulo (USP), ICMC** Feb 2017 – Dec 2021
  Bachelor's degree in Computer Science; GPA: 3.36 (8.4/10 - Top 16%)

Papers

- *Attribute-Enhanced Similarity Ranking for Sparse Link Prediction* - **João Pedro Rodrigues Mattos**, Zexi Huang, Mert Kosan, Ambuj Singh, Arlei Lopes da Silva - KDD 2025
- *Semi-Supervised Graph Attention Networks for Event Representation Learning* - **João Pedro Rodrigues Mattos** and Ricardo Marcacini - ICDM 2021
- *Smart toys and children's privacy: usable privacy policy insights from a card sorting experiment* - André de Lima Salgado, Felipe Silva Dias, **João Pedro Rodrigues Mattos**, Renata Pontin de Mattos Fortes, Patrick CK Hung - SIGDOC 2019

Work Experience

- **eBay Inc.** May 2024 - Aug 2024
  Applied Research Intern - PhD
  - **Graphs and Large Language Models - TraceXplainer**: Research project reduces incident analysis from hours to minutes with 75% expert approval by combining LLMs with graph techniques for explainability and root cause analysis of microservices. (Currently in production)
  - **Data collections and explanation quality**: Managed a team of 20 expert annotators to collect over 100 prompt reviews and annotations for few-shot learning. Our method generates expert-level textual explanations of traces and is inspired by Graph Neural Networks.

- **Experian** Jun 2021 - Jun 2022
  Data Scientist I
  - **Analytics Profile Segmentation Tools**: Designed an internal library of pipelines used by the entire Marketing Analytics to enable scalable data preprocessing and model training experiments using Apache Spark.
  - **ReLaTeX**: Reduced report creation time from 3 days to minutes by designing and incorporating a LaTeX parser at the end of the model evaluation pipeline previously developed.

- **Experian** Oct 2020 - Jun 2021
  Data Science Intern
  - **Machine Learning**: Segmented consumer profile data of 210 million Brazilian customers through developing and deploying more than 100 ML models (Logistic Regression, LGBM, XGBoost).
  - **ETL**: Automated the three main Marketing Analytics pipelines, enabling data processing speed-ups by migrating from SAS to Python and using Apache Spark / Hadoop.

Research Projects and Extension

- **Imaflora** Nov 2020 - Apr 2022
  Research Internship
  - **Graph embedding and Markov Chain**: Ranked the reliability of wood transportation chains by modeling Amazon Rainforest timber flux data (2014–2020) using Graph Embedding techniques and Markov Chain models, enhancing insights for government monitoring.
  - **Data Visualization and ETL**: Developed a real-time data visualization tool that processed over 1 million wood merchant transactions by building Spark ETL pipelines and Julia scripts for parallel computing, enabling detailed exploration of model-classified chains and networked data.

- **GNNs and Event Embedding - Universidade de São Paulo (USP)** Oct 2020 - Oct 2021
  Research Assistant
  - **Deep Learning and Graph Attention Networks**: Developed GNEE, a novel Event Graph Embedding technique supporting heterogeneous networks by leveraging Graph Attention Networks (GAT) in a semi-supervised training regime, under the guidance of Professor Ricardo Marcacini.
  - **Paper and Undergraduate Thesis extension**: GNEE was accepted in ICDM 2021 (International Conference on Data Mining) as a short paper. These results were extended in my Undergraduate Thesis, which compared the performance of our proposed method in this task with other techniques such as GAT and GCN in streams of data, finding that GNEE is, in many cases, superior in Accuracy and F1 while maintaining a linear computational cost.
- **ModCovid - 19 (CeMEAI)** Apr 2020 - Jul 2020
  Research Internship - supported by the Serrapilheira Institute
  - **Machine Learning with Time Series**: Assisted in designing efficient Covid-19 mitigation strategies for cities by developing an automatic control system using time series models (ARIMA, Prophet), with contributions acknowledged in the "Robot Dance" paper.
  - **Data Visualization and ETL**: Built an automated pipeline to scrape, preprocess, and visualize medical form data using Selenium, Pandas, and D3.js, supporting various experiments by feeding data into Machine Learning models and dashboards.
- **Privacy Policies and Joker - Universidade de São Paulo (USP)** Jul 2018 - Jul 2020
  Research Assistant - supported by Fapesp (Sao Paulo State Research Support Foundation)
  - **Human Computer Interaction (HCI) and Data Mining**: Developed Joker, an open-source card sorting tool for clustering usability datasets, enabling researchers to conduct remote experiments with large participant pools via crowdsourcing.
  - **Results and collaborations**: Published a research paper in collaboration with Ontario Tech University in SIGDOC 2019, showcasing findings from this project.
- **Data ICMC - Extracurricular Data Science Group** Jan 2019 - Dec 2021
  Coordinator
  - **Introduction to Machine Learning**: Taught the Introduction to Machine Learning course to over 120 students at USP through Data, our extracurricular Data Science and Machine Learning group. The online lectures, with over 20,000 views on YouTube, became one of the largest Machine Learning courses in Portuguese. The curriculum was inspired by top courses from Stanford, MIT, and Cornell.
  - **Project supervision**: Conducted the development of more than 10 Machine Learning and Deep Learning projects about Computer Vision, Audio, and Natural Language Processing.
  - **Events**: Organized lecture series and panels covering areas in the frontiers of Deep Learning, such as "Understanding Deep Learning" and "Open-Endedness Panel". Both had lecturers from DeepMind, Google Brain, UCL, Oxford and many other affiliations.

Fellowships and Awards

- **Best Research Presentation - Computer Science Department - 1st place - Spring 2024**: The award is given to the best research talk of the semester at the Computer Science Department seminar course. Obtained 1st place out of 15 students competing, according to the majority of the students in the department.
- **D2K Research Mentoring Fellowship - Spring 2024 - 1st Place Winner**: The award is given to 10 PhD Students from 2nd year and above to mentor a data science capstone team, working on a project from academia, industry, or non-profit. All projects compete for 1st place in the end-of-semester poster showcase.
- **D2K Research Mentoring Fellowship - Fall 2023**: The award is given to 10 PhD Students from 2nd year and above to mentor a data science capstone team, working on a project from academia, industry, or non-profit.
- **Experian Recognition Award**: The award is given to the most efficient Experian team in Brazil by the end of the fiscal year.
- **Finalist of Interhack**: 3rd out of 180 participants in Brazil's biggest hackathon for students.

Technical Skills

Languages: C/C++, Python, JavaScript, SQL
DS/ML: Numpy, SciPy, Scikit-Learn, PySpark, PyTorch (PyTorch Geometric, DGL), networkx, graph-tool
Others: BeautifulSoup, Scrapy, Django, Git, Docker, NodeJS, Linux, Hadoop

Teaching

- **Reinforcement Learning** Spring 2024
  - **Teaching Assistant**: Academic service as a Teaching Assistant at Rice University. The course provides an introduction to reinforcement learning (RL), a machine learning paradigm for solving sequential decision-making problems and designing autonomous agents, covering both classical and recent algorithms for reinforcement learning and imitation learning.
- **Machine Learning with Graphs** Spring 2024
  - **Teaching Assistant**: Academic service as a Teaching Assistant at Rice University. The course overviews traditional and more recent graph-based machine learning algorithms, from spectral methods to state-of-the-art Graph Neural Networks (GNNs).
- **Teaching Assistant Scholarship** Jan 2018 - Jul 2018
  - **Teaching**: Assisted in practical classes in the Algorithms and Data Structures course, given to 60 students. I was also involved in elaborating and grading the assignments.

Volunteering

- **Elder digital inclusion course (6 months)**: Taught the basics of how to use and interact with modern technologies, such as smartphones, cloud storage, and productivity software (Microsoft Word, Excel, etc.).