

# Joël Mathys

jmathys@ethz.ch | joelmathys.com | linkedin.com/in/joelmathys



## Education

---

### ETH Zurich

Graduation within next 6 to 12 months

*PhD Student, Distributed Computing Group (DISCO)*

- I focus on building better graph learning architectures and teaching neural networks to reason properly. More broadly, I am interested in how we can achieve generalization, particularly in extrapolation settings using reasoning and computation depth.
- Advisor: Prof. Dr. Roger Wattenhofer

### ETH Zurich

September 2021

*MSc in Computer Science (Grade: 5.75/6, graduated with distinction)*

- Thesis: Labeling Schemes for Reachability in Directed Graphs

*BSc in Computer Science (Grade: 5.46/6)*

September 2020

- Thesis: A Speed Accurate Continuous Attractor Neural Network Model of Head Direction Cells, supervised by Prof. Dr. Angelika Steger

### Hong Kong University of Science and Technology

Fall 2018

*Exchange Programme in Computer Science (GPA: 3.94/4, Dean's List)*

## Experience

---

### NEC Laboratories Europe

Heidelberg, Germany

*Research Internship, Intelligent Software Systems Group*

November 2024 - April 2025

- Hosted by Dr. Federico Errica. Research on long-range interactions in graph neural networks.

### ETH Zurich

Zurich, Switzerland

*Teaching*

Since March 2022

- Responsible for planning, preparation, and execution of lectures and exams, e.g. Hands-on Deep Learning and Principles of Distributed Computing.
- Supervised 30+ student projects
- Reviewed papers for NeurIPS, ICLR, ICML, WSDM, ECML, FC, and PODC, amongst others.

## Selected Works

---

- *N-vium: Mixture-of-Exits Transformer for Accelerated Exact Generation.*  
We extend standard transformers to directly learn a mixture of distributions, where sampling remains exact so tokens learn to use depth adaptively. Because tokens can commit early, we can parallelize the remaining computation with later tokens, achieving up to 57.9% speedup at 1.5B scale with no perplexity cost.
- *From Message-Passing to Linearized Graph Sequence Models.*  
We show how to decouple information propagation and processing depth in graph neural networks (GNNs). This allows us to recast key architectural graph learning aspects into the design of the input sequence.
- *LRIM: A Physics-Based Benchmark for Provably Evaluating Long-Range Capabilities in Graph Learning. ICLR 2026.*  
We introduce a provable long-range graph benchmark based on physics simulations, which we theoretically and empirically analyse across common message-passing and graph transformer architectures.
- *Abstract Visual Reasoning Enabled by Language. NVVLR @ CVPR 2023.*  
First to showcase that LLMs can be applied on the ARC-AGI benchmark to solve abstract visual reasoning problems by using language structure.

## Research Interests

---

- **What are the boundaries of machine learning architectures?** I investigate the limits of graph learning architectures, what makes certain tasks fundamentally hard, and how architectural choices shape what and how well models can learn.
- **How can we teach models to generalize?** Classical algorithms generalize well but are hard to design or even learn, while machine learning models learn easily but often fail to generalize. I aim to explore the intersection through neural reasoning: how we can incentivize it, and leverage more computation to generalize to much larger and more complex instances beyond the training distribution.
- **How can we capture and quantify long-range interactions?** I am interested in methods that go beyond local-only information to mimic reasoning or real-world systems that require long-range interactions. I try to quantify progress towards this by designing appropriate benchmarks and proposing architectures.

## Skills

---

- **Experienced in Deep Learning:** Transformer Architectures, Geometric Deep Learning, Message-Passing, modern State-Space Models, Graph Neural Networks
- Projects involving Large Language Models, Discrete Diffusion, Structural Biological Data, Simulation and Synthetic Data and Analysis Evaluation with Large Datasets
- **Programming:** Python (PyTorch), C++, Bash,  $\LaTeX$ , Git, HPC (Slurm)
- **Languages:** German (native), English (fluent), French (basic)

## Extracurricular

---

### Swiss Olympiad in Informatics (SOI)

2014 – 2022

*Participant, Organizer, Deputy Leader*

- Participated in SOI from 2014 to 2016. Represented Switzerland at CEOI 2015 (Brno, CZ) and IOI 2016 (Kazan, RU).
- Active member of the organizing team, preparing competitions and camps and teaching high school students. Deputy Leader of the Swiss Delegation for IOI 2018 (Tsukuba, JPN).

## Personal Interests

---

Badminton · Reading · Cooking · Competitive Programming · Ultimate Frisbee

## Publications

---

**Joël Mathys**, Henrik Christiansen, Federico Errica, Takashi Maruyama, and Francesco Alesiani. Lrim: a physics-based benchmark for provably evaluating long-range capabilities in graph learning. In *The Fourteenth International Conference on Learning Representations*, April 2026a. URL <https://openreview.net/forum?id=IAZXEX1dVV>.

**Joël Mathys**, Basil Rohner, Saku Peltonen, and Roger Wattenhofer. From message-passing to linearized graph sequence models. May 2026b. URL <https://arxiv.org/abs/2605.12358>. Preprint, arXiv:2605.12358.

Aleksander Lorenc, Frédéric Berdoz, **Joël Mathys**, and Roger Wattenhofer. N-vium: Mixture-of-exits transformer for accelerated exact generation. May 2026. URL <https://arxiv.org/abs/2605.13190>. Preprint, arXiv:2605.13190.

**Joël Mathys** and Federico Errica. Learn to jump: Adaptive random walks for long-range propagation through graph hierarchies. In *Combining Bayesian and Neural approaches for Structured data (ComBayNS@IJCNN25)*, June 2025.

- Florian Grötschla, **Joël Mathys**, Loïc Holbein, and Roger Wattenhofer. Reinforcement learning for locally checkable labeling problems. In *The Seventeenth Workshop on Adaptive and Learning Agents (ALA@AAMAS25)*, May 2025.
- Joël Mathys**, Andreas Plesner, Jorel Elmiger, and Roger Wattenhofer. Synthetic Data for Blood Vessel Network Extraction. In *Will Synthetic Data Finally Solve the Data Access Problem? (SynthData@ICLR2025)*, Singapore, April 2025.
- Niccolò Grillo, Andrea Toccaceli, Benjamin Estermann, **Joël Mathys**, Stefania Fresca, and Roger Wattenhofer. Beyond Interpolation: Extrapolative Reasoning with Reinforcement Learning and Graph Neural Networks. In *1st Workshop on Neural Reasoning and Mathematical Discovery (NEURMAD@AAAI25)*, Philadelphia, USA, March 2025.
- Joël Mathys**, Florian Grötschla, Kalyan Varma Nadimpalli, and Roger Wattenhofer. Rethinking message passing for algorithmic alignment. In *NeurIPS 2024 Workshop on Symmetry and Geometry in Neural Representations*, December 2024.
- Florian Grötschla, **Joël Mathys**, Christoffer Raun, and Roger Wattenhofer. GraphFSA: A Finite State Automaton Framework for Algorithmic Learning on Graphs. In *27th European Conference on Artificial Intelligence (ECAI)*, Santiago de Compostela, Spain, October 2024a.
- Florian Grötschla, **Joël Mathys**, Robert Veres, and Roger Wattenhofer. Core-GD: A hierarchical framework for scalable graph visualization with GNNs. In *The Twelfth International Conference on Learning Representations*, May 2024b. URL <https://openreview.net/forum?id=vtyasLn4RM>.
- Stefan Künzli, Florian Grötschla, **Mathys, Joël**, and Roger Wattenhofer. Surf: A generalization benchmark for gnn predicting fluid dynamics. In Soledad Villar and Benjamin Chamberlain, editors, *Proceedings of the Second Learning on Graphs Conference*, volume 231 of *Proceedings of Machine Learning Research*, pages 13:1–13:23. PMLR, 27–30 Nov 2024. URL <https://proceedings.mlr.press/v231/kunzli24a.html>.
- Julian Minder, Florian Grötschla, **Joël Mathys**, and Roger Wattenhofer. SALSA-CLRS: A sparse and scalable benchmark for algorithmic reasoning. In *The Second Learning on Graphs Conference*, 2023. URL <https://openreview.net/forum?id=PRapGjDGFQ>.
- Giacomo Camposampiero, Loïc Houmard, Benjamin Estermann, **Mathys, Joël**, and Roger Wattenhofer. Abstract visual reasoning enabled by language. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 2643–2647, June 2023.
- Florian Grötschla, **Joël Mathys**, and Roger Wattenhofer. Learning Graph Algorithms With Recurrent Graph Neural Networks. In *Workshop on Graphs and more Complex structures for Learning and Reasoning (GCLR@AAAI)*, Washington D.C., USA, February 2023.
- Moritz Neun, Christian Eichenberger, Henry Martin, Markus Spanring, Rahul Siripurapu, Daniel Springer, Leyan Deng, Chenwang Wu, Defu Lian, Min Zhou, Martin Lumiste, Andrei Ilie, Xinhua Wu, Cheng Lyu, Qing-Long Lu, Vishal Mahajan, Yichao Lu, Jiezhong Li, Junjun Li, Yue-Jiao Gong, Florian Grötschla, **Mathys, Joël**, Ye Wei, He Haitao, Hui Fang, Kevin Malm, Fei Tang, Michael Kopp, David Kreil, and Sepp Hochreiter. Traffic4cast at neurips 2022 – predict dynamics along graph edges from sparse node data: Whole city traffic and eta from stationary vehicle detectors. In Marco Ciccone, Gustavo Stolovitzky, and Jacob Albrecht, editors, *Proceedings of the NeurIPS 2022 Competitions Track*, volume 220 of *Proceedings of Machine Learning Research*, pages 251–278. PMLR, 28 Nov–09 Dec 2022. URL <https://proceedings.mlr.press/v220/neun23a.html>.
- Florian Grötschla and **Joël Mathys**. Hierarchical Graph Structures for Congestion and ETA Prediction. In *Traffic4cast@NeurIPS22*, December 2022.
- Mathys, Joël**, Robin Fritsch, and Roger Wattenhofer. Decentralized graph processing for reachability queries. In Weitong Chen, Lina Yao, Taotao Cai, Shirui Pan, Tao Shen, and Xue Li, editors, *Advanced Data Mining and Applications*, pages 505–519, Cham, 2022. Springer Nature Switzerland. ISBN 978-3-031-22064-7.