

# Zihe Zhou

PHD APPLICANT — COMPUTER SCIENCE (FALL 2027) · SCALABLE MACHINE LEARNING ON LARGE NETWORKS

zhchen.zhou@mail.utoronto.ca · (+1) 365-992-9518 · Toronto, ON, Canada

Website [www.zhouchen.ai](http://www.zhouchen.ai) · ORCID [0009-0009-9000-7436](https://orcid.org/0009-0009-9000-7436) · GitHub [github.com/AmbitiousK](https://github.com/AmbitiousK) · arXiv [2607.14531](https://arxiv.org/abs/2607.14531)

## RESEARCH INTERESTS

---

Machine learning on large, complex networks that is **both scalable and structurally faithful** — overlapping community detection, large-scale graph algorithms, and reproducible benchmarking. Emerging interest in **interpretable graph-structured representations**.

## EDUCATION

---

### University of Toronto

Sept 2025 – Aug 2026 (expected)

M.Eng., Mechanical & Industrial Engineering (concentration: Data Analytics & Machine Learning) — GPA 4.0/4.0

- Five completed graduate courses, **all graded A+**: Foundations of Data Analytics and Machine Learning; Data Science Methods and Statistical Learning; Introduction to Machine Learning; Cloud-Based Data Analytics; Positive Psychology for Engineers.

### University of Waterloo

Sept 2020 – June 2025

Bachelor of Computer Science, Honours, *with Distinction*

- Computer Science Major · **Artificial Intelligence Specialization** · **Computational Mathematics Minor**.

## PUBLICATIONS & PREPRINTS

---

Zihe Zhou, S. Aref. “Overlapping Network Community Detection Using Sparse Backbones” (the *Highway* algorithm).

Accepted, IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) 2026 (Springer conference proceedings; conference Aug 24–27 2026). Preprint: arXiv [2607.14531](https://arxiv.org/abs/2607.14531) (Jul 15 2026).

- Journal extension **in preparation** for *Social Network Analysis and Mining* (SNAM, Springer).

## RESEARCH EXPERIENCE

---

### Graduate Researcher, University of Toronto — Dept. of Mechanical & Industrial Engineering

Sept 2025 – present

Advisor: Samin Aref

Defined and executed a complete, largely independent research cycle in overlapping community detection:

- **Formulated the problem.** Built *TRIAD*, a quadratically constrained program with node-, edge-, and community-level terms that explicitly diagnoses and penalizes structural degeneracy. It produced accurate, structurally reliable assignments but did not scale beyond a few hundred nodes — which turned scalability into the central research question.
- **Designed a scalable algorithm.** Built *Highway*: rather than inference over the full graph, it constructs a sparse *backbone* (hybrid modularity-inspired + local Jaccard edge scoring), selects distributed degree-greedy anchors, propagates memberships under a **neighbor-only** rule (a node cannot reinforce its own membership), and calibrates communities from pattern- and neighbor-level consensus.
- **Designed the evaluation.** Benchmarked against **10 existing methods + an ablation** on ~3,000 synthetic graphs (LFR; ABCD+o<sup>2</sup>) and 3 real SNAP networks up to **1.13M nodes / 2.99M edges**, across 5 metrics (Q<sub>ov</sub>, FRI, Dice, F\*, ONMI).

- **Stress-tested under critique.** Ran additional overlap-sweep experiments to directly characterize how detection quality holds up as overlap increases.

## OPEN-SOURCE SOFTWARE & SYSTEMS

---

- **Highway** — C++ backend + Python interface, on-demand compilation. Only method to finish all three real SNAP networks within 300s (Amazon 3.7s / DBLP 3.8s / YouTube 23.3s); **7.34×** faster than the strongest competitor on the largest instance (**~60×** vs COPRA); **Q<sub>ov</sub> + 18.9% (rank 1)** on ABCD+o<sup>2</sup>. **Integrated into** the open-source [cdlib](#) community-detection library.
- **Open reproducible OCD benchmark** — lets third parties plug in their own algorithm for apples-to-apples comparison across baselines and metrics. [github.com/AmbitiousK/highway-ocd-benchmark](https://github.com/AmbitiousK/highway-ocd-benchmark)
- **HELM** — a multi-client financial data-and-analytics platform: a governed data spine, reproducible analytics pipelines, audit trails, and Web / iPad / iPhone interfaces (grown from a graduate course project).

## TALKS & PRESENTATIONS

---

- **Conference talk**, WAW 2026 (Workshop on Algorithms and Models for the Web Graph) — *Highway* and its precursor *TRIAD*, in person, June 18 2026. [Recording](#).
- **Scheduled talk**, Toronto Metropolitan University — *Highway*, Aug 12 2026.
- **Poster (abstract accepted)**, MIE Graduate Research Symposium 2026 — *Highway*, University of Toronto (Hart House Great Hall), May 4 2026. Poster subsequently selected for the MIE hallway display.

## HONORS & AWARDS

---

- Bachelor of Computer Science awarded **with Distinction** (University of Waterloo).
- University of Toronto M.Eng. — all five completed graduate courses graded A+.

## TECHNICAL SKILLS

---

**Languages:** C++, Python, Java, JavaScript, SQL

**Areas:** graph algorithms · overlapping community detection · large-scale network analysis · reproducible benchmarking · machine learning · full-stack systems