Chengyuan Deng

Address: Hill Center
Department of Computer Science
Rutgers University, NJ, USA

Email: malus.dns@gmail.com
Phone: (+1) 848-391-2312

Education

Rutgeres University, New Jersey, USA — 2021 Fall - present
Ph.D in Computer Science, Theory of Computing Group

- Advisor: Prof. Jie Gao

Rutgers University, New Jersey, USA — Sep 2018 - Dec 2020
Master of Science in Computer Science

- **Graduate with Honor:** Outstanding Publication Award

Tongji University, Shanghai, China — Sep 2014 - Sep 2018
Bachelor of Engineering in Electronics and Information Engineering
Minor in Applied Mathematics

Research Interest

My primary research interest revolves around **Algorithms** and **Machine Learning**. For instance: Graph algorithms and combinatorics; Computational geometry; Sublinear algorithms; Differential privacy; Machine Learning with provable guarantees on privacy, fairness, explainability, etc.
I am tentatively exploring: LLMs and Mathematical aspects of music such as combinatorics and geometry.

Theory Papers

(In submission) "Property testing for Structural Balance and Correlation Clustering".

(In submission) "Johnon-Lindenstrauss Lemma beyond Euclidean Geometry".

"On the Price of Differential Privacy for Hierarchical Clustering". With Jie Gao, Jalaj Upadhyay, Chen Wang, Samson Zhou, **ICLR 2025**

"Low-sensitivity Hopsets". With Vikrant Ashvinkumar, Aaron Bernstein, Jie Gao, Nicole Wein, **ITCS 2025**

"Neuc-MDS: Non-Euclidean Multi-dimensional Scaling Through Bilinear Forms". With Jie Gao, Kevin Lu, Feng Luo, Hongbin Sun, Cheng Xin, **NeurIPS 2024**

"The Discrepancy of Shortest Paths". With Greg Bodwin, Jie Gao, Gary Hoppenworth, Jalaj Upadhyay and Chen Wang, **ICALP 2024**

"Evaluating Stability in Massive Social Networks: Efficient Streaming Algorithms for Structural Balance". With Vikrant Ashvinkumar, Sepehr Assadi, Jie Gao and Chen Wang, **RANDOM 2023**

"Differentially Private Range Query on Shortest Paths". With Jie Gao, Jalaj Upadhyay and Chen Wang, *Symposium on Algorithms and Data Structures,* **WADS 2023**

"Impossibility of Depth Reduction in Explainable Clustering". With Surya Teja Gavva, Karthik C. S., Parth Patel and Adarsh Srinivasan, **Information and Computation**

ML Papers

(In submission)"LEMMA-RCA: A Large Multi-modal Multi-domain Dataset for Root Cause Analysis".

"Correlation-aware Online Change Point Detection", **CIKM 2025**

"Deconstructing The Ethics of Large Language Models from Long-standing Issues to New-emerging Dilemmas". **AI and Ethics**

"Domain Specialization as the Key to Make Large Language Models Disruptive: A Comprehensive Survey". **Journal of Computing Surveys**

"$\mathbb{E}^{FWI}$: Multiparameter Benchmark Datasets for Elastic Full Waveform Inversion of s Geophysical Properties", *Website: efwi-lanl.github.io*, **NeurIPS 2023**

"OpenFWI: Large-Scale Multi-Structural Benchmark Datasets for Seismic Full Waveform Inversion", *Website: openfwi-lanl.github.io*, **NeurIPS 2022 (Spotlight)**

"On the Global Self-attention Mechanism for Graph Convolutional Networks", **ICPR 2020 (Oral)**

"SAG-VAE: End-to-end Joint Inference of Data Representations and Feature Relations", **IJCNN 2020 (Oral)**

"Imbalance-XGBoost: Leveraging Weighted and Focal Loss for Imbalanced Binary Classification with XGBoost", **Pattern Recognition Letter**

Research Experiences

Ph.D student, Theory of Computing Group, Rutgers University
Advised by Prof.Jie Gao Sep 2021 - present

- Topic: Graph Algorithms, Differential Privacy, Learning Theory

Applied scientist intern, Amazon
Hosted by Dr.Chengwei Su, Kechen Qin, Emre Barut May 2023 - present

- Topic: Large language models

Research intern, NEC Labs
Hosted by Dr.Zhengzhang Chen Jan 2023 - present

- Topic: Time series analysis, AIOps

Research Associate & Student Guest, Los Alamos National Lab
Hosted by Dr.Youzuo Lin April 2021 - present

- Topic: AI for science, Deep learning Three projects completed.

Previous Machine Learning Research Experiences
Rutgers University 2018-2021

- Self-motivated research during Master at Rutgers, four papers published, worked on Graph Neural Networks, Generative models, etc.
- Research Assistant at WINLAB Rutgers, worked on VR and system optimization
- Research intern on Natural Langeuage Processing at Recurrent.ai, worked on Text-to-speech

Selected Projects

Distribution Testing in Multi-pass Streaming Model
Advised by Prof. Sepehr Assadi, course project for CS514: Sublinear Algorithms

- Literature Review of Distribution Testing in Single-pass Streaming Model and Distributed Communication Model.
- Proved a lower bound in Multi-pass Streaming Model for Uniformity Testing

Imbalance-XGBoost
Self-motivated

- **Open-source python library**, availabe on Github and PyPi, **star 240+**.
- The library leverages weighted and focal loss for imbalanced binary classification with XGBoost. State-of-the-art performances were achieved on a recently collected Parkinson disease dataset by Focal-XGBoost. Paper published.

(Kaggle) Intersection Congestions Prediction
Advised by Prof. Saed Sayad

- Implemented multiple regression models, neural networks, CatBoost, LightGBM, XGBoost to predict waiting time and distance at intersections in four cities: Atlanta, Boston, Chicago and Philadelphia.
- XGBoost outperformed other approaches, **leaderboard 25/432**.

Industrial Experiences

Machine Learning Intern, Newark — Jan 2020 - Mar 2020
Haystack.ai

- Engaged in developing deep learning models from cutting-edge academic papers for real-world applications
- Example Project: Selfie-to-anime. Collected anime images for training, implemented a cutting-edge paper published in ICLR 2020 "Unsupervised Generative Attentional Networks" and built the API with Flask.

Data Analysis Intern, Shanghai — Jun 2016 - Sep 2016
Haitong Securities, International

- Analyzed the daily stock quotation and cyclical data by setting up models then predicted trends
- Proposed financial models for cutting-edge companies and wrote reports, with **200+** pageviews daily

Talks & Posters

◇ "The Discrepancy of Shortest Paths", *DIMACS Theory of Computation Seminar*
◇ "Differentially Private Range Query on Shortest Paths", *DIMACS Graph Algorithm Workshop*
◇ "Physics-augmented Representation Learning for Full-Waveform Inversion" *AGU*

Honors and Awards

◇ Rizvi Research Excellence Award 2025
◇ School of Arts and Science Research Travel Award
◇ Nomination of IBM Fellowship
◇ Travel Award, *WADS/CCCG 2023*
◇ Scholar Award, *NeurIPS 2022*
◇ Yamaha Asian Music Scholarship of Honorable Mention, **Piano Performance**, *2017 Shanghai.*
◇ ACM Programming Contest, Shanghai Regional, **First Prize**, *2018 Shanghai.*
◇ Mathematical Modeling Invitation of U.S.A. **Second Prize**, *2018 Shanghai.*
◇ National Undergraduate Contest in Mathematical Modeling, **First Prize**, *2017 Shanghai.*

Professional Services

◇ Reviewer, NeurIPS 2023, 2024; ICLR 2024, 2025; SDM 2024, WWW 2024, WSDM 2024.
◇ Reviewer, IEEE Transaction on Neural Networks and Learning System (TNNLS)