Cheng Xin

email: cx122@rutgers.edu || phone: +1 (610) 597-1901 || website: jackal092927.github.io

Bio: Cheng Xin is a Post-Doctoral Associate of the Center for Discrete Mathematics and Theoretical Computer Science (DIMACS) at Rutgers University. His research focuses on the intersection of computational topology, geometry, machine learning, and artificial intelligence. Dr. Xin earned his Ph.D. in Computer Science from Purdue University, where his doctoral research centered on topological data analysis and graph representations. His work explores innovative approaches to data representation and analysis, seeking to uncover hidden patterns and structures in complex datasets. Through his research, Dr. Xin aims to develop novel algorithms and techniques that leverage the power of topology and geometry to enhance machine learning and AI models. By bridging the gap between these disciplines, he strives to create more efficient, interpretable, and robust models for a wide range of applications.

EDUCATION	**Purdue University**, West Lafayette, Indiana, USA

- Ph.D. in Computer Science — Aug 2020 – Aug 2023
  - Adviser: Prof. Tamal K. Dey
  - Focus: Topological data analysis, Graph representation learning
  - Thesis: Decomposition and Stability of Multiparameter Persistence Modules

The Ohio State University, Columbus, Ohio, USA

- Ph.D. in Computer Science — Aug 2016 – Aug 2020
  - Adviser: Prof. Tamal K. Dey
  - Focus: Computational topology, Topological data analysis

Lehigh University, Bethlehem, Pennsylvania, USA

- M.S. in Computer Science — Jan 2014 – May 2016
  - Advisor: Prof. Xiaolei Huang
  - Thesis: Machine Learning Techniques for Cervigram Image Analysis

Tongji University, Shanghai, China

- B.Eng. in Software Engineering — Sep 2009 – Jul 2013

PROFESSIONAL EXPERIENCE

Center for Discrete Mathematics and Theoretical Computer Science (DIMACS)
Rutgers University

- Postdoctoral Researcher, Advisor: Prof. Jie Gao — Oct 2024 – Present
  - Topological data analysis, Graph generative model, Explainable graph neural networks.
  - Non-Euclidean representation [8], Cooperative learning in social networks [5].

Department of Computer Science, Purdue University

- Research Assistant, Advisor: Prof. Tamal K. Dey — Aug 2020 – Aug 2023
  - Deep generative model for graphs and point cloud data [6].
  - Rectangular Approximation and Stability of 2-parameter Persistence Modules [3].

Department of Computer Science and Engineering, The Ohio State University

- Research Assistant, Advisor: Prof. Tamal K. Dey — Jan 2017 – May 2020
  - Generalized persistence algorithm for multi-parameter persistence modules. [4]
  - Computing bottleneck distance on multi-parameter persistence modules. [2]

Department of Computer Science and Engineering, Lehigh University
IDEA lab, Advisor: Prof. Xiaolei Huang

- Project: Behavior Analysis of Students Enrolled in MOOC — May 2015 – Aug 2015
  - Study topological features of distributions of different student behavior.
  - Predicted dropout rate, 90% accuracy achieved (top 10%).

- Project: Machine Learning Methods for Computer Aided Diagnosis [11] — Jan 2015 – May 2015
  - Contruct texture and color features to represent medical images.
  - A comprehensive comparison among various machine learning models on medical images.

- Project: Deep Learning Methods on Medical Image [10] — Aug 2014 – Dec 2014
  - Developed paralleled MLP networks with early machine learning frameworks (Theano/Pylearn.)
  - Combined paralleled MLP with SVM and Random Forest to do the classification.

PUBLICATIONS   # References

[1]   Tamal K. Dey and Cheng Xin. "Computing Bottleneck Distance for 2-D Interval Decomposable Modules". In: *Symposium on Computational Geometry*. 2018.

[2]   Tamal K. Dey and Cheng Xin. *Computing Bottleneck Distance for Multi-parameter Interval Decomposable Persistence Modules*. 2019. arXiv: 1803.02869 [cs.CG].

[3]   Tamal K. Dey and Cheng Xin. *Rectangular Approximation and Stability of 2-parameter Persistence Modules*. 2021. arXiv: 2108.07429 [cs.CG].

[4]   Tamal K. Dey and Cheng Xin. "Generalized persistence algorithm for decomposing multiparameter persistence modules". In: *Journal of Applied and Computational Topology* 6.3 (Sept. 2022), pp. 271–322. ISSN: 2367-1734. DOI: 10.1007/s41468-022-00087-5. URL: https://doi.org/10.1007/s41468-022-00087-5.

[5]   Shahrzad Haddadan, Cheng Xin, and Jie Gao. "Optimally Improving Cooperative Learning in a Social Setting". In: *Proceedings of the 41st International Conference on Machine Learning*. Ed. by Ruslan Salakhutdinov et al. Vol. 235. Proceedings of Machine Learning Research. PMLR, 21–27 Jul 2024, pp. 17148–17188. URL: https://proceedings.mlr.press/v235/haddadan24a.html.

[6]   Lu Ling et al. "Dl3dv-10k: A large-scale scene dataset for deep learning-based 3d vision". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024, pp. 22160–22169.

[7]   Soham Mukherjee et al. *D-GRIL: End-to-End Topological Learning with 2-parameter Persistence*. 2024. arXiv: 2406.07100 [cs.LG]. URL: https://arxiv.org/abs/2406.07100.

[8]   "Neuc-MDS: Non-Euclidean Multidimensional Scaling Through Bilinear Forms". In: *NeurIPS 2024 (to appear)*.

[9]   Cheng Xin et al. "GRIL: A 2-parameter Persistence Based Vectorization for Machine Learning". In: *Proceedings of 2nd Annual Workshop on Topology, Algebra, and Geometry in Machine Learning (TAG-ML)*. Ed. by Timothy Doster et al. Vol. 221. Proceedings of Machine Learning Research. PMLR, 28 Jul 2023, pp. 313–333. URL: https://proceedings.mlr.press/v221/xin23a.html.

[10] Tao Xu et al. "A New Image Data Set and Benchmark for Cervical Dysplasia Classification Evaluation". In: *International Workshop on Machine Learning in Medical Imaging*. Springer. 2015, pp. 26–35.

[11] Tao Xu et al. "Multi-feature based benchmark for cervical dysplasia classification evaluation". In: *Pattern recognition* 63 (2017), pp. 468–475.

INDUSTRIAL EXPERIENCE

Machine Learning Scientist Intern, Big data group

- Electronic Arts, Redwood City, CA, USA   May 2018 – Aug 2018
  - Project: Feature evaluation and selection.
    In order to reduce the cost of query and storage on a huge database with possibly redundant data, we apply feature selection techniques in data mining to evaluate values of each features in the database. The motivation is that, for some attributes or features, we just care about their coarse-grained statistical features instead of exact values for each data point. Then we can use their statistical properties and their relations with other attributes or features instead of the fine-grained raw data. In practice, we found that a lot of attributes or features stored in the database are redundant to some extent which might be ignored in the development stage. We can use the above idea to evaluate each attributes or features to check how redundant they are and furthermore provide some suggestions for developer to check if they are redundant as we expected. This tool is built on Spark.

Software Development Engineer Intern, AWS infrastructure group

- Amazon, Seattle, Washington, USA   May 2015 – Aug 2015
  - Built BGP monitoring protocol messages management tools for receiving, parsing, storing, retrieving BMP messages.

Develop Engineer, Platform group

- Koal, Shanghai, China   Jul 2013 – Dec 2013
  - Big Data Process: data collection, data preparation and preprocessing, Chinese phrase identification and analysis.
  - Designed and managed Database; designed and implemented back-end features and interfaces, front-end webpage UI.

Developer Support Intern, Database business intelligence group

- Microsoft, Shanghai, China   Jul 2012 – Nov 2012
  - Studied basic theories and designs about the SQL Server, including relational engine and storage engine.
  - Used trouble shooting tools for dealing with problem cases in practice.

COURSES	■ **Algorithm & Theory** Computability and Complexity, Algorithms, Advanced Algorithms, Pseudorandomness ■ **AI & Machine Learning** Machine learning on Graphs, Statistical Learning Theory, Pattern Recognition, Data Mining, ■ **Maths & Physics** Abstract Algebra, General Topology, Algebraic Topology, Algebraic Geometry, Quantum Information, Combinatorics and Graph Theory
LANGUAGES	■ Chinese (Native), English (Fluent)
SKILLS	■ Python, Pytorch, Keras, Spark, Java, C, C++, MATLAB, R
HONERS AND REWARDS	■ The first place winner of 2017 Microsoft's college code competition at the Ohio State University ■ The first place winner of 2015 Microsoft's college code competition at Lehigh University ■ Third Prize of 2013 Tongji coding competition ■ Third Prize Scholarship of Tongji University
INTERESTS	■ Climbing, Hiking, Basketball, GO, Quantum Computation, Animation
REFERENCES	■ **Dr. Jie Gao** Professor of Computer Science Department, Rutgers University email: jg1555@cs.rutgers.edu ■ **Dr. Tamal K. Dey** Professor of Computer Science Department, Purdue University email: tamaldey@purdue.edu ■ **Dr. Pan Li** Assistant Professor of School of Electrical and Computer Engineering, Georgia Institute of Technology email: panli@gatech.edu ■ **Dr. Tao Xu** Research scientists of Meta, email: xutao@meta.com ■ **Dr. Feng Luo** Professor of Department of Mathematics, Rutgers University email: fluo@math.rutgers.edu