

Kaan Sancak

SENIOR RESEARCH SCIENTIST · META, RECOMMENDATION SYSTEMS

1 Hacker Way, Sunnyvale, 94089, CA

☎ (+1) 404 429 01 44 | ✉ kaansnck@gmail.com | 🏠 www.kaansancak.com | 🌐 kaansancak | 🌐 kaansancak

Areas of Interest

- Recommendation Systems, Generative Recommendation, Large-Scale Machine Learning
- Machine Learning Systems & AI Infrastructure
- High-Performance Computing, Parallel & Distributed Algorithms, GPU Systems

Education

Georgia Institute of Technology

PH.D. IN COMPUTER SCIENCE, GPA: 4.00/4.00

Atlanta, GA, USA

Aug. 2019 - Aug. 2024

- Advisor: **Umit V. Catalyurek**. Focus: ML Systems, High-Performance Computing, Graph Learning.

Bilkent University

B.SC. IN COMPUTER ENGINEERING, GPA: 3.82/4.00

Ankara, Turkiye

Aug. 2015 - May 2019

- *Summa Cum Laude* – Ranked 4th among 231 engineering students.

Industry Experience

Meta

SENIOR RESEARCH SCIENTIST, META RECOMMENDATION SYSTEMS

Sunnyvale, CA, USA

Feb. 2026 - Present

- Building next-generation recommendation models and ML infrastructure powering Meta's core ranking and personalization systems at scale.

Meta

RESEARCH SCIENTIST, AI/ML, ADS RANKING & FOUNDATIONAL AI

Sunnyvale, CA, USA

Aug. 2024 - Feb. 2026

- Bridged the **model-infrastructure** gap via model/infra co-design: built scalable relational models on Meta's ML stack, partnering with infra teams to unlock new capabilities and improve efficiency.
- Architected model infra integrating Meta's real-time graph engine with a billion-scale Ads recommendation model; improved data freshness from **days**→**minutes**, boosted training QPS by **20%**, and cut feature storage cost by **3x**.
- Reduced graph model dataloader latency by **2x** by decoupling graph sampling from the feature service.
- Lead contributor to Ads **Graph Foundational Model (GFM)**; designed and implemented the end-to-end framework for graph-based ad recommendation.

Meta

RESEARCH SCIENTIST INTERN, RANKING & FOUNDATIONAL AI

Sunnyvale, CA, USA

May 2023 - Aug. 2023

- Developed scalable graph-based models for efficient ad ranking without compromising quality; published in **AAAI'25**. Contributed to mini-batch Graph Transformer with node tokenization, published in **ICLR 2024**.

Meta

PART-TIME STUDENT RESEARCHER, AI SYSTEMS HW/SW CO-DESIGN

Menlo Park, CA, USA

Aug. 2022 - Dec. 2022

- Extended caching mechanisms for Meta's ML training platform, reducing data serving computations by up to **99%** for key models.

Meta

RESEARCH SCIENTIST INTERN, AI SYSTEMS HW/SW CO-DESIGN

Menlo Park, CA, USA

May 2022 - Aug. 2022

- Built caching infrastructure for Meta's data ingestion pipelines, eliminating redundant computation across large-scale model training runs.

Pacific Northwest National Laboratory

RESEARCH INTERN, HPC & SYSTEMS

Richland, WA, USA

May 2021 - Aug. 2021

- Developed distributed graph algorithms and high-performance data structures on the SHAD framework.

Facebook

Menlo Park, CA, USA

RESEARCH INTERN, AI SYSTEMS HW/SW CO-DESIGN

May 2020 - Aug. 2020

- Improved Facebook's graph engine performance via novel partitioning schemes: **10%** query throughput gain, up to **5×** end-to-end speedup on random-walk queries, and **10×** reduction in server-side query generation.
- Integrated the graph engine with Instagram Ads; infrastructure continues to serve Instagram, Threads, and Facebook.

National Resource for Network Biology (NRNB)

Remote

GOOGLE SUMMER OF CODE DEVELOPER

May 2018 - Aug. 2018

- Developed collaborative pathway editing tools for *cBioPortal for Cancer Genomics* (Memorial Sloan Kettering Cancer Center).

IBM

Ankara, Turkey

SOFTWARE ENGINEERING INTERN

Jun. 2017 - Aug. 2017

- Built cloud-based file transfer and object recognition applications using IBM Bluemix, Python, and Kafka.

Academic Research Experience

Georgia Institute of Technology, TDALab

Atlanta, GA, USA

GRADUATE RESEARCH ASSISTANT, ADVISOR: [UMIT V. CATALYUREK](#)

Aug. 2019 - Aug. 2024

- Developed scalable graph-based models as efficient alternatives to Graph Transformers, capable of processing million-node graphs; published in [AAAI'25](#).
- Collaborated with Meta AI on a mini-batch Graph Transformer with node tokenization; published at [ICLR 2024](#).
- Developed a family of models enhancing skip connections for GNNs as zero-overhead plug-ins; published in [LoG 2024](#).
- Created MG-GCN, a multi-GPU GCN training framework that significantly reduces training time and memory; published at [ICPP 2022](#).
- Co-designed Elga, a system for elastic distributed dynamic graph analysis on streaming graphs; published at [SC 2021](#).
- Designed fast heuristic algorithms for the symmetric rectilinear partitioning problem; published in [JEA 2022](#).

Bilkent University, i-Vis Research Lab

Ankara, Turkey

UNDERGRADUATE RESEARCHER, ADVISOR: [UGUR DOGRUSOZ](#)

Sept. 2017 - May 2019

- Contributed to [Newt](#) (biological pathway editor) and [cBioPortal](#) for cancer genomics (Memorial Sloan Kettering Cancer Center).

Teaching Experience

Georgia Institute of Technology

Atlanta, GA, USA

GRADUATE TEACHING ASSISTANT FOR CSE 6230: HIGH-PERFORMANCE PARALLEL COMPUTING

Jan. 2024 - May 2024

- Held office hours, designed and graded assignments and exams, and delivered lectures.

Georgia Institute of Technology

Atlanta, GA, USA

GRADUATE TEACHING ASSISTANT FOR CSE 6220: INTRO TO HIGH-PERFORMANCE COMPUTING

Jan. 2021 - May 2021

- Held office hours, designed and graded assignments and exams.

Bilkent University

Ankara, Turkey

TEACHING ASSISTANT FOR CS 102: ALGORITHMS AND PROGRAMMING II

Sept. 2017 - Jan. 2018

- Graded assignments, programming projects, and exams.

Bilkent University

Ankara, Turkey

TEACHING ASSISTANT FOR CS 114: INTRODUCTION TO PROGRAMMING

Jan. 2017 - May 2017

- Graded assignments, programming projects, and exams.

Publications

Under Review

- Kaan Sancak**, M. Fatih Balin, Umit V. Catalyurek, "Exploring Graph Sampling to Scale Graph Transformers"

Published

- Mufei Li, Dongqi Fu, Limei Wang, Si Zhang, Hanqing Zeng, Kaan Sancak, Ruizhong Qiu, Haoyu Peter Wang, Xiaoxin He, Xavier Bresson, Yinglong Xia, Chonglin Sun, Pan Li”Haystack Engineering: Context Engineering for Heterogeneous and Agentic Long-Context Evaluation”, arXiv’2025.
- Mufei Li, Dongqi Fu, Limei Wang, Si Zhang, Hanqing Zeng, Kaan Sancak, Ruizhong Qiu, Haoyu Peter Wang, Xiaoxin He, Xavier Bresson, Yinglong Xia, Chonglin Sun, Pan Li”Haystack Engineering: Context Engineering Meets the Long-Context Challenge in Large Language Models”, NeurIPS 2025 Workshop on Evaluating the Evolving LLM Lifecycle: Benchmarks, Emergent Abilities, and Scaling.
- **Kaan Sancak**, Zhigang Hua, Jin Fang, Yan Xie, Bo Long, Andrey Malevich, M. F. Balin, U. V. Catalyurek, ”A Fast and Effective Alternative to Graph Transformers”. *The AAAI Conference on Artificial Intelligence (AAAI)*, Feb 2025.
- **Kaan Sancak**, M. Fatih Balin, Umit V. Catalyurek, ”Do We Really Need Complicated Graph Learning Models? – A Simple but Effective Baseline” *The Learning on Graphs Conference (LoG 2024)*.
- Dongqi Fu, Zhigang Hua, Yan Xie, Jin Fang, Si Zhang, **Kaan Sancak**, Hao Wu, Andrey Malevich, Jingrui He, Bo Long, ”VCR-Graphormer: A Mini-batch Graph Transformer via Virtual Connections”, *International Conference on Learning Representations (ICLR)*, May 2024.
- **Kaan Sancak***, M. Fatih Balin*, Umit V. Catalyurek, ”MG-GCN: Scalable Multi-GPU GCN Training Framework”, *International Conference on Parallel Processing (ICPP)*, Aug 2022.
- A. Yasar, M. F. Balin, X. An, **Kaan Sancak**, Umit V. Catalyurek, ”On Symmetric Rectilinear Matrix Partitioning”, *Journal of Experimental Algorithmics (JEA)*, Sep 2022.
- Kasimir Gabert, **Kaan Sancak**, M Yusuf Özkaya, Ali Pinar, Ümit V Çatalyürek, ”Elga: Elastic and Scalable Dynamic Graph Analysis”, *International Conference for High Performance Computing, Networking, Storage and Analysis*, Nov 2021.

Talks

51st International Conference on Parallel Processing (ICPP)

PRESENTER FOR [MG-GCN: SCALABLE MULTI-GPU GCN TRAINING FRAMEWORK](#)

Virtual

Aug. 2022

- Presented our paper on scalable GNN training in the [International Conference on Parallel Processing \(ICPP\)](#) main conference.

Service

Reviewer

- [International Conference on Learning Representations \(ICLR\)](#): 2025
- [Conference on Neural Information Processing Systems \(NeurIPS\)](#): 2024
- [Conference on Neural Information Processing Systems \(NeurIPS\) Benchmarks and Datasets](#): 2024
- [ACM SIGKDD Conference on Knowledge Discovery and Data Mining \(KDD\)](#): 2024
- [Learning on Graphs Conference \(LoG\)](#): 2024
- [International Conference on Web Search and Data Mining \(WSDM\)](#): 2024
- [IEEE Transactions on Big Data \(TBD\)](#): 2021

Skills

- **Programming:** Python, C++, PyTorch, CUDA, MPI, Spark, Java, JavaScript, LaTeX, Git
- **Frameworks:** PyTorch Geometric (PyG), DGL, NCCL, Distributed Training
- **Languages:** English, Turkish

Honors & Awards

- 2019 **Summa Cum Laude**, Department of Computer Science, Bilkent University
- 2016-2019 **High Honor Student**, Department of Computer Science, Bilkent University

Ankara, Turkey

Ankara, Turkey