

Zexi Huang

SENIOR MACHINE LEARNING SCIENTIST AT TIKTOK RECOMMENDATION | Ex-AMAZON | AI PHD AT UCSB | 8+ YOE IN ML RESEARCH

[✉ zexihuang.phd@gmail.com](mailto:zexihuang.phd@gmail.com) | [🏡 zexihuang.com](http://zexihuang.com) | [🔗 zexihuang](https://zexihuang.com) | [🔗 zexihuang](https://zexihuang.com) | [🔗 Zexi Huang](https://zexihuang.com)

Education

College of Engineering, University of California, Santa Barbara (UCSB)

Santa Barbara, USA

DOCTOR OF PHILOSOPHY AND MASTER OF SCIENCE IN COMPUTER SCIENCE

Sep. 2018 - Jun. 2023

- Course GPA: 4.0/4.0, UCSB Computer Science *Outstanding Scholar* Fellow (top 4 out 63 PhD students).
- Advisor: Prof. Ambuj Singh. Dissertation: Learning Representations for Information-rich Graphs.

Yingcai Honors College, University of Electronic Science and Technology of China (UESTC)

Chengdu, China

BACHELOR OF ENGINEERING IN COMPUTER SCIENCE AND TECHNOLOGY (HONORS)

Sep. 2014 - Jun. 2018

- GPA: 3.96/4.0, Avg. Score: 92.79/100 (1/90 in freshman year, 1/87 in sophomore year, 1/93 in junior year).
- Recipient of the *Most Outstanding Students Awards* of UESTC (top 10 out of 5,000 seniors).
- Advisors: Prof. Junming Shao and Prof. Sinno Jialin Pan. Thesis: Transfer Learning for Community Detection in Multiplex Networks.

Industry Experience

Senior Machine Learning Scientist, Livestreaming Recommendation, TikTok

San Jose, USA

SCALING BILLION-SCALE RECOMMENDATION MODELS WITH MULTI-DOMAIN MULTI-MODAL SIGNALS FOR TIKTOK LIVE

Jan. 2025 - Present

- Designed and deployed a novel feature interaction architecture that scales up the ranking baseline by 50x, achieving +1%/-1% in CTR AUC/Logloss and +10%/+8% in Quality Watch Session/Duration / User, with superior scaling laws compared to state-of-the-art methods.
- Incorporated user behavioral information from other domains (short videos, e-commerce, and search) with graph representation learning.
- Encoded visual, ASR, and metadata information from the content with vision-language models to improve model generalizability.

Machine Learning Scientist, Livestreaming Recommendation, TikTok

San Jose, USA

RANKING AND HOST EXPERIENCE OPTIMIZATION FOR TIKTOK LIVE RECOMMENDATION

May. 2023 - Dec. 2024

- Owned the iterations of the core ranking models of TikTok Live Recommendation, leveraging representation learning, multi-task learning, contrastive learning, regression modeling, and sequence modeling, with +7% Watch Duration / User gains from online A/B experiments.
- Developed the host go-live model that captures the relationship between watch-live and go-live with causal inference and uplift modeling to motivate authorized hosts to go-live, achieving +2% Host Session / User in online A/B experiments.

Applied Scientist Intern, Books Tech, Amazon

San Diego, USA

STOCHASTIC INVENTORY MANAGEMENT FOR PRINT-ON-DEMAND AND GRAPH-BASED TEXT CLASSIFICATION FOR CONTENT INTELLIGENCE

Jun. 2022 - Sep. 2022

- Developed a stochastic inventory management model based on dynamic programming for Amazon's Print-On-Demand business, optimizing the ordering strategy for 42.2M units of books and realizing an annual saving of \$10.4M.
- Designed and implemented a graph-based NLP model for long text embedding and classification using graph neural networks, leading to superior performance to state-of-the-art transformer-based models for Kindle book contents.

Applied Scientist Intern, Books Tech, Amazon

San Diego, USA

GRAPH-BASED FRAUD DETECTION IN KINDLE DIRECT PUBLISHING

Jun. 2021 - Sep. 2021

- Designed graph-based machine learning models for fraud detection based on multi-modal signals in Kindle Direct Publishing.
- Implemented the models in an end-to-end fashion and deployed them to graphs with millions of nodes and billions of edges in production.
- Validation results show that the models can surface fraud rings undetected by existing processes with an estimated annual value of \$2.4M.

Applied Scientist Intern, Books Tech, Amazon

San Diego, USA

GRAPH-BASED FRAUD DETECTION IN KINDLE DIRECT PUBLISHING

Jun. 2020 - Sep. 2020

- Proposed to augment existing fraud detection methods with graph-based machine learning models for Kindle Direct Publishing.
- Designed and implemented various heuristics and an embedding framework for attributed heterogeneous multiplex networks.
- The models are deployed into production and results show that they surface up to 15 times more fraud compared to the existing processes.

Research Intern, Computational Intelligence Lab, Nanyang Technological University

Singapore

TRANSFER LEARNING FOR COMMUNITY DETECTION IN MULTIPLEX NETWORKS

Sep. 2017 - Feb. 2018

- Proposed to refine community detection results in some layers with transferred knowledge from other layers in multiplex networks.
- Designed a representation-based community detection framework and implemented it with an extended symmetric NMF approach.
- Our algorithm outperforms other representation-based community detection algorithms, especially when the target layer is noisy.

Publications

[KDD-2025] João Mattos, **Zexi Huang**, Mert Kosan, Ambuj Singh, Arlei Silva. **Attribute-Enhanced Similarity Ranking for Sparse Link Prediction**. ACM SIGKDD Conference on Knowledge Discovery & Data Mining, 2025.

[TIST-2025] Mert Kosan*, **Zexi Huang***, Sourav Medya, Sayan Ranu, Ambuj Singh. **GCFExplainer: Global Counterfactual Explainer for Graph Neural Networks**. ACM Transactions on Intelligent Systems and Technology, 2024.

[AAAI-2024] Aritra Bhowmick, Mert Kosan, **Zexi Huang**, Ambuj Singh, Sourav Medya. **DGCLUSTER: A Neural Framework for Attributed Graph Clustering via Modularity Maximization**. AAAI Conference on Artificial Intelligence, 2024.

[WSDM-2023] **Zexi Huang***, Mert Kosan*, Sourav Medya, Sayan Ranu, Ambuj Singh. **Global Counterfactual Explainer for Graph Neural Networks**. ACM International Conference on Web Search and Data Mining, 2023.

[WSDM-2022] **Zexi Huang**, Arlei Silva, Ambuj Singh. **POLE: Polarized Embedding for Signed Networks**. ACM International Conference on Web Search and Data Mining, 2022.

[KDD-2021] **Zexi Huang**, Arlei Silva, Ambuj Singh. **A Broader Picture of Random-walk Based Graph Embedding**. ACM SIGKDD Conference on Knowledge Discovery & Data Mining, 2021.

Research Projects

Attribute-Enhanced Similarity Ranking for Sparse Link Prediction

ADVISOR: PROF. AMBUJ SINGH, COLLABORATORS: JOÃO MATTOS, MERT KOSAN, ARLEI SILVA

May. 2021 - Aug. 2024

- Scrutinized the training and evaluation of supervised link prediction methods and identify their limitations in handling class imbalance.
- Proposed a novel topology-centric framework that combines graph learning, topological heuristics, and an N-pair loss for link prediction.
- Results showed that the proposed method is 145% more accurate and trains/infers 11/6,000 times faster than the state-of-the-art methods.

Global Counterfactual Explanation for Graph Neural Networks

ADVISOR: PROF. AMBUJ SINGH, COLLABORATORS: MERT KOSAN, SOURAV MEDYA, SOURAV MEDYA, SAYAN RANU

Oct. 2021 - Jun. 2024

- Formulated the novel problem of global counterfactual reasoning/explanation of graph neural networks for graph classification.
- Proposed GCFExplainer, the first global explainer powered by vertex-reinforced random walks on an edit map with a greedy summary.
- Results showed that GCFExplainer not only provides crucial high-level insights but also outperforms existing methods in recourse quality.

Signed Embedding for Polarized Graphs

ADVISOR: PROF. AMBUJ SINGH, COLLABORATOR: ARLEI SILVA

Oct. 2020 - Aug. 2021

- Designed a novel polarization measure for signed graphs and showed that existing methods fail in polarized signed link prediction.
- Proposed a polarized embedding algorithm that captures both topological and signed similarity jointly via signed autocovariance.
- Extensive experiments showed that the proposed model outperforms state-of-the-art methods by up to one order of magnitude.

Multiscale Graph Convolution via Neural Diffusions

ADVISOR: PROF. AMBUJ SINGH, COLLABORATOR: WEI YE, YUNQI HONG

Apr. 2020 - Oct. 2021

- Interpreted the layer-wise propagation rule of GCN from the perspective of power iteration and analyzed its converging process.
- Designed a novel GCN architecture that learns to aggregate multiscale information based on graph diffusions with a neural network.
- Illustrated the effectiveness and efficiency of the proposed model by extensive comparative studies with state-of-the-art methods.

Graph Representation Learning Based on Random-walks

ADVISOR: PROF. AMBUJ SINGH, COLLABORATOR: ARLEI SILVA

Sep. 2018 - Feb. 2021

- Presented a unified view of embedding, covering different random-walk processes, similarity metrics, and embedding algorithms.
- Showed both theoretical and empirical evidence of the superiority of the novel autocovariance embedding in link prediction.
- Illustrated ways to exploit the multiscale nature of random-walk similarity to further optimize embedding performance.

Teaching Experience

Lead Teaching Assistant, Department of Computer Science, UCSB

Sep. 2020 - Sep. 2021

Teaching Assistant, CS 130A, CS 8, CS 174A/174N, Department of Computer Science, UCSB

Sep. 2018 - Dec. 2019

Instructor, Machine Learning Workshops for LMU/UCSB Nanotech PhD Symposium

May. 2019 - Jun. 2019

Academic Services

Registration Chair KDD'23

Program Committee AAAI'23-25, KDD'22, SDM'22

Reviewer TNNLS'23, Neural Networks'23, TIST'22-25, TKDD'21-24, KDD'20-21, WebConf'21