Yiyue Qian

https://yiyueqian.github.io/

yqian5@nd.edu
+1-216-370-8373

EDUCATION

University of Notre Dame — South Bend, USA
PhD in Computer Science — *Jan. 2022 - present*
Research Interest: Graph Representation Learning, Multi-modal Learning, Large Language Models (LLMs), Hypergraph Representation Learning, Self-supervised Learning, Contrastive Learning, Graph Imbalance Learning, Few-shot Learning, Adversarial Learning, Deep Learning, Machine Learning, Data Mining, Social Network Mining, Fraud Detection.

Case Western Reserve University — Cleveland, USA
PhD in Computer Science — *Sept. 2019 - Dec. 2021*
Research Interest: Graph Representation Learning, Self-supervised Learning, Machine Learning, Deep Learning, Data Mining.

University College London — London, UK
Master in Statistics — *Sept. 2016 - Sept. 2017*
Courses: Machine Learning, Statistical Models and Data Analysis, Applied Bayesian Methods, Statistics Inference, Decision and Risk.
Thesis: Classifying Three Propensities in Orange Telecom Company

EXPERIENCE

Amazon — Seattle, USA
Applied Scientist II — *Dec. 2023 - present*

- **Generative Learning:** Design and develop advanced generative learning methods, LLMs, and multi-modal learning research to address real-world problems for AWS customers.

Amazon — Seattle, USA
Applied Scientist (Intern) — *May. 2023 - Sept. 2023*

- **Self-Supervised Graph Pretraining for Collusion Prevention:** Designed foundational graph self-supervised contrastive learning models called CollusionEmbed over *unlabeled data* on Amazon to learn the pre-trained collusion embeddings (i.e., seller embedding, buyer embedding, and asin embedding) that can be easily fine-tuned to detect any downstream Collusion risks (The AUC improvement is more than 20% over downstream abuse detection tasks, e.g., rank abuse detection).

- **Two-Tower Multi-Modal Fusion Model for Collusion Unearthing:** Designed and developed a multi-modal two-tower framework called 2TowerFusion to fuse activities among buyer groups, sellers, and products for unveiling collusion among buyer groups and sellers on Amazon. (The AUC improvement is almost 40% compared with the production models.)

- **Fine-tune LLMs for Fraud Detection:** Prompted engineering and fine-tuned large language models (e.g., LLaMa 2 and Hugchat), on fraudulent datasets within Amazon to identify redirecting traffic websites and suspicious ASIN products.

Amazon — Seattle, USA
Data Scientist (Intern) — *May. 2022 - Aug. 2022*

- **Ring of Abuser Detection:** Developed and **deployed** a universal multi-modal GNN framework called **U-ROAD** to detect the ring of abusers (e.g., seller-buyer collusion) on *Amazon* over heterogeneous graphs over seller, buyer, and product with multi-modal features (i.e., text and image). (The annual bad debt reduction on Amazon is estimated at $600k.)
- **Publication: Qian, Y.**, Chen, P., Cui, S., Chen, D.. U-ROAD: Universal Ring-of-Abuser Detection via Multi-Modal Heterogeneous Graph Learning, **KDD** RelKD workshop, 2023.

Ping An Technology Co., Ltd. — Shenzhen, China
Algorithm Engineer in NLP (Full-time) — *Jan. 2018 - Aug. 2019*

- **Model Building:** Built NLP-related models to capture main topics, recognize named entities, analyze relationships among entities, and judge sentiment from unstructured text for Ping An FinTech AI analysis platform.

- **Risk Event Monitoring:** Developed models on Linux to analyze the topic of global daily financial news such as Reuters News and monitor risk events at different levels (bankruptcy, violation of laws, affairs of executives, etc.). This launched function on the AI analysis platform is estimated to save more than $1 million per year for Ping An Group.

China Vanke Co., Ltd. — Guangzhou, China
Data Analyst (Full-time) — *Oct. 2017 - Dec. 2017*

- **Model Building:** Built machine learning models to analyze impact factors of commercial buildings and predict trends in real estate prices in big cities.

- **Marketing Analysis:** Visualized results for marketing insights and analyzed the trend on the model results.

Research Experience

- **University of Notre Dame** — South Bend, USA
  Research Assistant (Full-time) — *Jan. 2022 - present*
  - **Drug Trafficking Detection:** To detect drug traffickers on *social media*, developed a novel framework MetaHG by integrating graph representation learning and meta-learning over the built heterogeneous graph which incorporates multi-modal data and relation information among nodes in graph. (93% of drug traffickers in our dataset can be detected.)

    To study drug trafficking among different *darknet markets*, developed an adversarial framework dStyle-GAN that combines writing and photography styles of drug posts, graph embedding learning, and adversarial learning to identify the same drugs posted among different darknet markets. (89.3% of posted drug pairs among multiple darknet markets in our dataset can be identified correctly.)
  - **Malicious Repository Detection:** Built a set of novel models (i.e., Meta-AHIN, Rep2Vec, and CLA-HG) integrating graph learning models, self-supervised learning methods (i.e., attribute-masking and adversarial graph contrastive learning), and meta-learning models (i.e., MAML) to analyze the repositories on *GitHub* (i.e., malicious repository detection and repository recommendation). (87.3% of malicious repositories in our dataset can be successfully detected.)

- **Case Western Reserve University** — Cleveland, USA
  Research Assistant (Full-time) — *Sept. 2019 - Dec. 2021*
  - **α-Satellite Website for Assessing Risk during COVID-19:** Developed a website called α-Satellite with a dynamic map and an analysis board to provide a risk index at multiple levels (e.g., county-level) and information on public perception about the pandemic; Built a model to learn the area embedding based on the heterogeneous graph augmented by cGAN for calculating the risk index. (https://covid-19.yes-lab.org, more than 50k visits in half years.)

Selected Publication

- **Qian, Y.**, Ma, T., Zhang, C., and Ye, Y.. Hypergraph Contrastive Learning for Drug Trafficking Community Detection, **ICDM**, 2023. (co-first author)
- **Qian, Y.**, Chen, P., Cui, S., Chen, D.. U-ROAD: Universal Ring-of-Abuser Detection via Multi-Modal Heterogeneous Graph Learning, **KDD** RelKD workshop, 2023.
- **Qian, Y.**, Zhang, C., Wen, Q., Zhang, Y., Ye, Y. and Zhang, C.. Co-Modality Graph Contrastive Learning for Imbalanced Node Classification, **NeurIPS**, 2022.
- **Qian, Y.**, Zhang, Wen, Q., Y., Ye, Y. and Zhang, C.. Rep2Vec: Repository Embedding via Heterogeneous Graph Adversarial Contrastive Learning, **KDD**, 2022.
- **Qian, Y.**, Zhang, Y., Ye, Y. and Zhang, C.. Malicious Repositories Detection with Adversarial Heterogeneous Graph Contrastive Learning, **CIKM**, 2022.
- **Qian, Y.**, Zhang, Y., Ye, Y. and Zhang, C.. Distilling Meta Knowledge on Heterogeneous Graph for Drug Trafficker Detection on Social Media, **NeurIPS**, 2021.
- **Qian, Y.**, Zhang, Y., Ye, Y. and Zhang, C.. Adapting Meta Knowledge with Heterogeneous Information Network for COVID-19 Themed Malicious Repository Detection, **IJCAI**, 2021.
- Zhang, Y., **Qian, Y.**, Ye, Y. and Zhang, C.. Adapting Distilled Knowledge for Few-shot Relation Reasoning over Knowledge Graphs, **SDM**, 2022.
- Zhang, Y., **Qian, Y.**, Fan, Y., Ye, Y., Li, X., Xiong, Q. and Shao, F.. dStyle-GAN: Generative Adversarial Network based on Writing and Photography Styles for Drug Identification in Darknet Markets, **ACSAC**, 2020.
- Wen, Q., Ouyang, Z., Zhang, J., **Qian, Y.**, Ye, Y., and Zhang, C.. Disentangled Dynamic Heterogeneous Graph Learning for Opioid Overdose Prediction. **KDD**, 2022.
- Ye, Y., Fan, Y., Hou, S., Zhang, Y., **Qian, Y.**, Sun, S., Peng, Q., Ju, M., Song, W. and Loparo, K.. Community Mitigation: A Data-driven System for COVID-19 Risk Assessment in a Hierarchical Manner. **CIKM**, 2020.

Skills Summary

- **Languages and Platforms**   Python (PyTorch, TensorFlow, Keras), R, C++, JavaScript, SQL, Linux, AWS, Colab
- **Soft Skills**   Leadership, Team-Spirited, Problem Solving, Motivated, Critical Thinking, Creative Thinking
- **Languages**   Mandarin, English

Awards

- NeurIPS Scholar Award, 2022
- CIKM Student Travel Award, 2022

Services

- **Program Committee**   IJCAI, ICML, KDD, NeurIPS, CIKM, SDM, ICDM, ICLR, AAAI, WWW.
- **Reviewer**   Journal of Neurocomputing, Transactions on Big Data, Transactions on Dependable and Secure Computing, Transactions on Dependable and Secure Computing, ACM Transactions on Privacy and Security, Transactions on Neural Networks and Learning System.
- **Outreach Program Judger**   Learning Science Sphere Event at Northpoint Elementary School in 2023.
- **Teaching Assistant**   Computer Security (CSE-40567, CSDS-344).