**Xinyue Feng**
xinyue.feng@rutgers.edu | +1(*732)522-6621*
Personal Website LinkedIn

## EDUCATION

**Rutgers University** — **New Jersey, United States**
*Ph.D. in Computer Science, GPA: 3.96/4.0* — **2022 – 2026/2027 (Expected)**
**Nanjing University** — **Nanjing, China**
*M.S. in Statistics,* **GPA:** *3.86/4.0 (Ranking: **1st/37**)* — **2019 – 2022**
**Sun Yat-sen University** — **Guangzhou, China**
*B.S. in Mathematics,* **GPA:** *3.75/4.0 (Ranking: 3rd/75),* ***Excellent Graduation Thesis*** — **2015 – 2019**

## RESEARCH INTERESTS

Large Language Model, Graph Neural Networks, Data Mining

## WORK&RESEARCH EXPERIENCE

**JD Logistics**
*Research Intern* — **Apr.2025– Present**
**Project: A multi-agent LLM system for Phone-Based AI Sales**

- Sales aim to contact logistics customers to gather their information, discover their logistics demands, and promote our logistics services. However, human sales are always High cost, low efficiency, making it necessary to propose a LLM-based Sales System.

**JD Logistics**
*Research Intern* — **Apr.2022– Feb.2025**
**Project: Heterogeneous Graph-based Customer Expansion**

Over the past three years, I served as one of the ***leading contributors*** to the Customer Expansion project at JD Logistics. We successfully unlocked the value of over ***10 billion logistics data*** by constructing a ***billion-scale heterogeneous graph***, ultimately improving the customer acquisition success rate by nearly ***192%*** and bringing in over ***ten million new orders*** in the past year.

**Background: Limitations of JD's Original System**

Customer expansion aims to seek high-value customers willing to sign long-term contracts with our logistics company to ensure stable revenue streams. However, the existing system in JD faced several limitations:

- Relied heavily on manually crafted rules to identify target customers.
- Focused only on customer profiles, ignoring the rich interaction signals between customers and other entities (e.g., companies, logistics stations, other customers).

We proposed constructing a **heterogeneous graph** to capture relationships between these entities and leverage Heterogeneous Graph Neural Networks (HGNNs) for target customer identification.

**Challenges & Solutions**

- **Large-Scale, Noisy, and Fragmented Data**: Over 10 billion records across hundreds of Spark tables.
  **Solution**: Build a scalable pyspark pipeline for efficient data cleaning and alighment, enabling construction of a billion-scale heterogeneous graph.
- **Extremely Sparse Positive Labels**: Only <0.07% of customers are signed, making it difficult for HGNNs to learn meaningful patterns. Even multi-task learning still struggle to improve performance.
  **Solution**: *(KDD'25) Hierarchical Structure Sharing Empowers Multi-task HGNNs for Customer Expansion*
- **Low efficiency and Information-Loss in Billion-Scale Graphs Training**: Training with all neighbor interactions in HGNNs was computationally infeasible. Existing sub-sampling methods reduce computational load by neighbor sub-samping but suffer from substantial information loss.
  **Solution**: *(CIKM'25) NeighSqueeze: Compact Neighborhood Grouping for Efficient Billion-Scale Heterogeneous Graph Learning*

- **Missing Important Relations**: The presence of some links (e.g., customer-company) significantly boosts prediction performance, but such links are missing for most customers.
  **Solution**: *(KDD'24) Paths2pair: Meta-path based link prediction in billion-scale heterogeneous graphs*
- **Complex Relationships Modeling** : Existing meta-paths could not express non-linear relations.
  **Solution**: *(VLDB'24) Complex-Path: Effective and Efficient Node Ranking with Paths in Billion-Scale Heterogeneous Graphs*

**Real-world Impact**

We developed a billion-scale heterogeneous customer prediction framework and it has been deployed for over six months, boosting ***acquisition success by 192%*** and generating ***over ten million new orders.***

## PUBLICATIONS

- **Xinyue Feng,** Shuxin Zhong, Jinquan Hang, et al. ***NeighSqueeze: Compact Neighborhood Grouping for Efficient Billion-Scale Heterogeneous Graph Learning***. arXiv 2025.5.
- **Xinyue Feng**, Shuxin Zhong, Jinquan Hang, et al. ***Hierarchical Structure Sharing Empowers Multi-task Heterogeneous GNNs for Customer Expansion***. KDD 2025.
- Jinquan Hang, Zhiqing Hong, **Xinyue Feng**, et al. ***Paths2pair: Meta-path based link prediction in billion-scale commercial heterogeneous graphs***. KDD 2024
- Jinquan Hang, Zhiqing Hong, **Xinyue Feng**, et al. ***Complex-Path: Effective and Efficient Node Ranking with Paths in Billion-Scale Heterogeneous Graphs.*** VLDB 2024.

## AWARDS (SELECTED)

| | | |
|---|---|---|
| · | 26/1681 (Top 2%). CVPR2021 Security AI Challenger | **2021** |
| · | First Prize (Top 5%), Excellent Student Scholarship . NJU | **2019-2021** |
| · | University-level Excellent Graduation Thesis. SYSU | **2019** |
| · | First Prize (Top 1%), China Undergraduate Mathematical Contest in Modeling | **2017** |
| · | First Prize (Top 5%), Excellent Student Scholarship | **2016 – 2021** |

## SKILLS

- **Programming Language:** *Python, C/C++, MATLAB, R, SQL*
- ***Tools&Fameworks:*** *PyTorch, TensorFlow, PySpark*