YIFAN HOU

Email: yifan.hou@inf.ethz.ch Homepage: https://yifan-h.github.io/

Education

ETH Zurich 2020 - Present
Department of Computer Science and Engineering
Ph.D. in Computer Science and Engineering
Advisor: Prof. Mrinmaya Sachan Anticipated Graduation: June, 2024

The Chinese University of Hong Kong (CUHK) 2018 - 2020
Department of Computer Science and Engineering GPA: 3.87/4.0
M.Phil. in Computer Science and Engineering
Advisors: Prof. James Cheng & Prof. Ming-Chang Yang

Huazhong University of Science and Technology (HUST) 2014 - 2018
School of Electronic Information and Communications GPA: 3.59/4.0
B.Eng. in Information Science for Advanced Class in Mathematics and Physics

Experience

Toyota Technological Institute at Chicago *(Summer Intern)* June 2020 - September 2020
Extracting and Analyzing the Structured Multimodal Information on Webpages
Advisor: Prof. Mrinmaya Sachan

National University of Singapore *(Summer Intern)* June 2019 - September 2019
Analyzing Graph Neural Networks based on Information Theory
Advisor: Prof. Richard T. B. Ma

University of Illinois at Urbana-Champaign *(Summer Exchange)* July 2017 - August 2017
Information Science and Engineering Summer School Program (Network Analysis)

Current Research Focus

model interpretation, knowledge graphs, and graph-based machine learning,

Selected Publications (Google Scholar)

Yifan Hou, Guoji Fu, Mrinmaya Sachan. Understanding Knowledge Integration in Language Models with Graph Convolutions. Submitted to *ICML*, 2022.

Yifan Hou, Mrinmaya Sachan. Bird's Eye: Probing for Linguistic Graph Structures with a Simple Information-Theoretic Approach. In *ACL-IJCNLP*, 2021.

Yifan Hou, Jian Zhang, James Cheng, Kaili Ma, Richard T. B. Ma, Hongzhi Chen, Ming-Chang Yang. Measuring and improving the use of graph information in graph neural networks. In *ICLR*, 2020.

Yifan Hou, Hongzhi Chen, Changji Li, James Cheng, and Ming-Chang Yang. A representation learning framework for property graphs. In *SIGKDD*, 2019.

Hongzhi Chen, Xiaoxi Wang, Chenghuan Huang, Juncheng Fang, **Yifan Hou**, Changji Li, James Cheng. Large scale graph mining with G-Miner. In *SIGMOD*, 2019.

Hongzhi Chen, Changji Li, Juncheng Fang, Chenghuan Huang, James Cheng, Jian Zhang, **Yifan Hou**, Xiao Yan. Grasper: A high performance distributed system for OLAP on property graphs. In *SoCC*, 2019.

Selected Research Projects

Multilingual Knowledge Graph Integration Still working on
Department of Computer Science (ETH Zurich)

- Proposing a method to enhance MLLMs (e.g., mBERT, XLM) with multilingual knowledge graphs (from Wikidata).
- Testing the enhanced MLLMs from two sides: 1. KG side (e.g., multilingual knowledge graph completion, entity alignment); 2. general MLLM downstream tasks (cross-lingual tasks).

Knowledge Integration Understanding Submitted to ICML 2022
Department of Computer Science (ETH Zurich)

- Proposing a novel analysis model, which uses graph convolution to simulate the knowledge integration process.
- Interpreting knowledge-enhanced models (K-Adapter, ERNIE), finding out how much and which types of knowledge in the knowledge graph is integrated successfully.

Graph Structure Probing in Language Models Published in ACL 2021
Department of Computer Science (ETH Zurich)

- Proposing an information-theoretic probe for general graph structure (entire structure and local structure).
- Detecting and understanding the linguistic structures in BERT with proposed probing method.
- Evaluating syntactic and semantic knowledge in large pre-trained language models (BERT, ELMo).

Graph Neural Network Analysis Published in ICLR 2020
Department of Computer Science and Engineering (CUHK)

- Proposing a general GNN framework and defining the information gain from neighborhood in aggregation.
- Proposing two graph smoothness metrics to measure the quantity and quality of information gain from graph data.
- Proposing an improved GNN model using two smoothness metrics to maximize information gain and reduce noises.
- Evaluating existing graph algorithms: struc2vec, GraphWave, Label Propagation, GCN, GraphSAGE and GAT.

Graph Representation Learning for Property Graphs Published in SIGKDD 2019
Department of Computer Science and Engineering (CUHK)

- Proposing a three-step graph embedding framework to capture the differences among neighbors.
- Extending the framework to support edge properties and direction in neighborhood aggregation.
- Evaluating graph representation learning algorithms: DeepWalk, node2vec, GCN and GraphSAGE.

Graph Query Processing Systems Published in SIGMOD 2019 and SoCC 2019
Department of Computer Science and Engineering (CUHK)

- Cleaning large-scale connected data (up to 2TB) from twitter, amazon, and converting them to the property graph format using distributed system Pregel+.
- Evaluating graph databases (e.g., OrientDB) on query latency and throughput under the distributed setting.

Selected Awards

Faculty Outstanding MPhil Thesis Award (CUHK): The **best thesis award** of Faculty of Engineering 2020

Student Travel Award (KDD) 2019

Outstanding Graduates Awards (HUST): The highest student achievement at HUST 2018

The National Scholarship: The highest award for students in China 2016

Skills

Languages: Mandarin (native), English (fluent)

Programming Languages: Python (Expert), C++ (Intermediate), MATLAB (Intermediate)