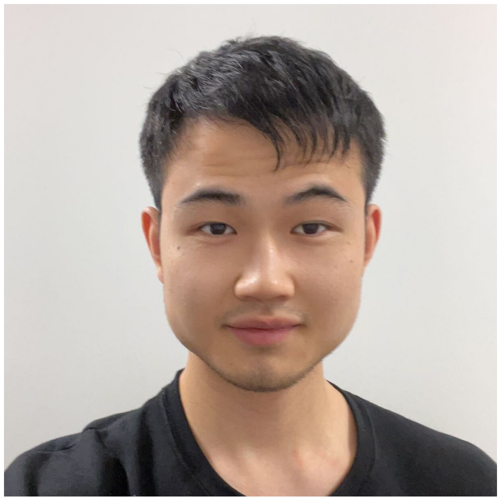
Mining & Learning on Graphs

Course Overview and Logistics

Yu Wang, Ph.D.
Assistant Professor
Computer and Information Science
University of Oregon
CS 410/510 - Fall 2024

Welcome

Welcome to the CS410/510 - Mining & Learning on Graphs!



Yu (Jack) Wang

You

Contact:

yuwang@uoregon.edu

Vanderbilt University, Nashville
The Home Depot Intern, Atlanta
Adobe Intern, San Jose
University of Oregon, Eugene

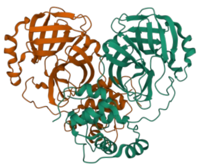
Welcome

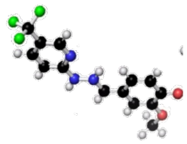
Our lab is currently recruiting!

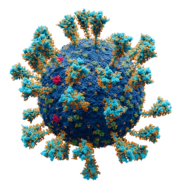
- Data Mining and Machine Learning
- Graph and Geometric Machine Learning
- Data-centric + GenAI
- Trustworthy + Textual-attributed Graph
- AI/ML Application: Information Retrieval/Science/Cyber-security

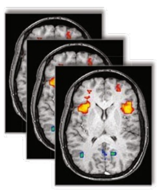
Graph-structured Data is Everywhere!

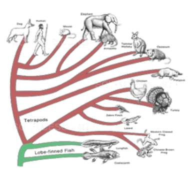
Scientific Graph

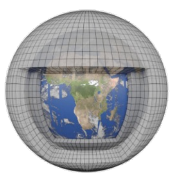

Protein


Small Molecule


Virus

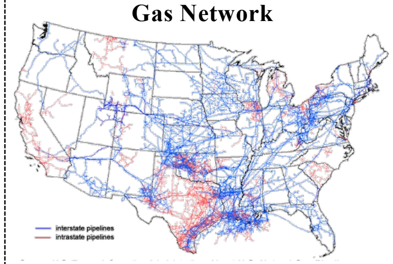

Brain Neural


Phylogenetic Tree

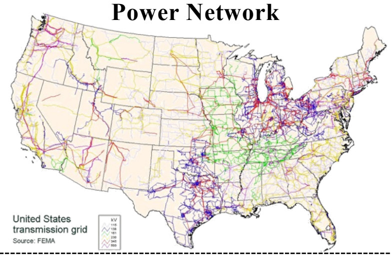

3D Grid

Infrastructure Graph

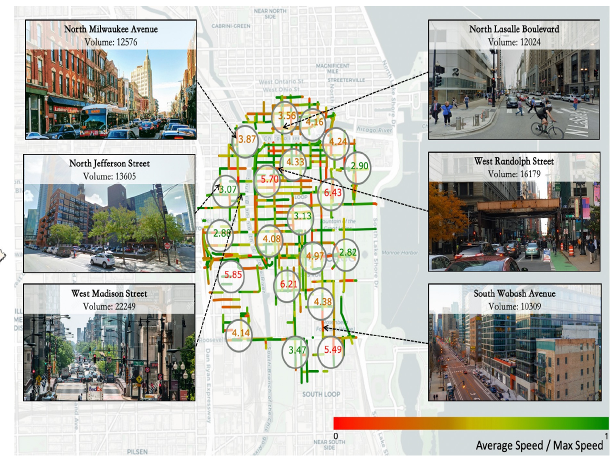
Gas Network



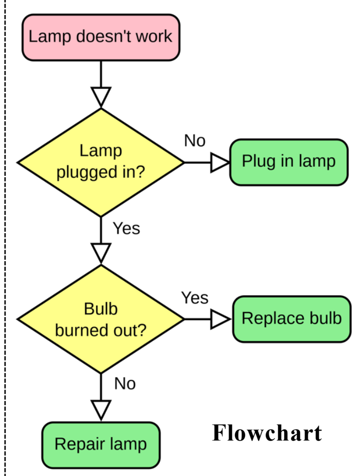
Power Network



Transportation Network

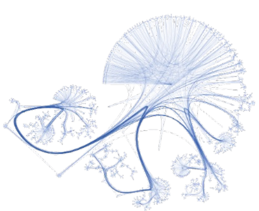


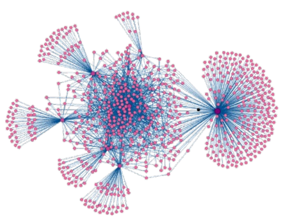
Decision Graph

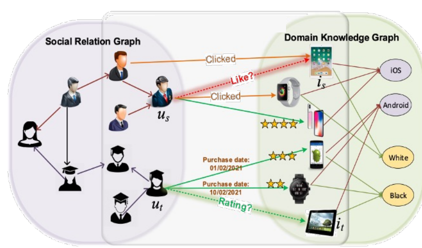


Flowchart

Social Interaction Graph


Citation Network

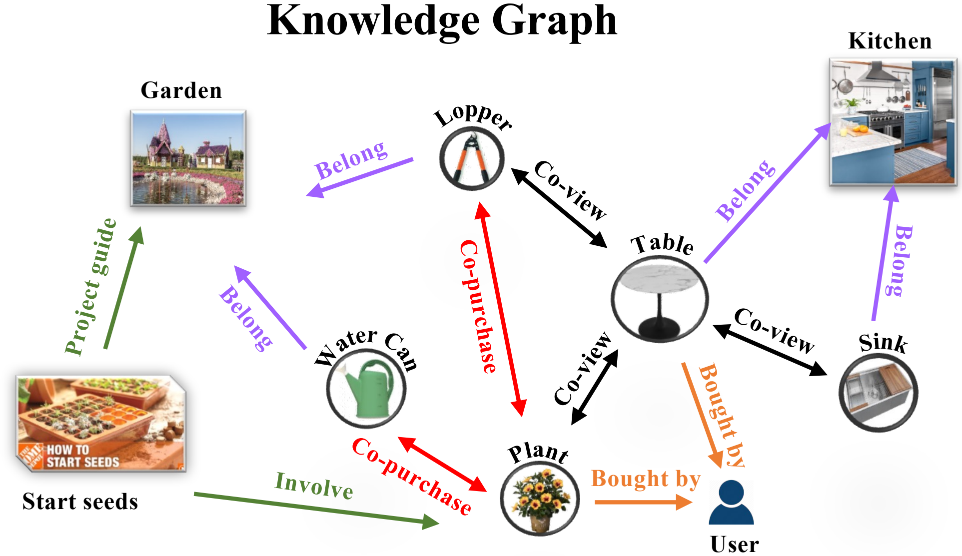

Transaction Network



User-Entity Interaction Graph

Virtual Village with AI Agents

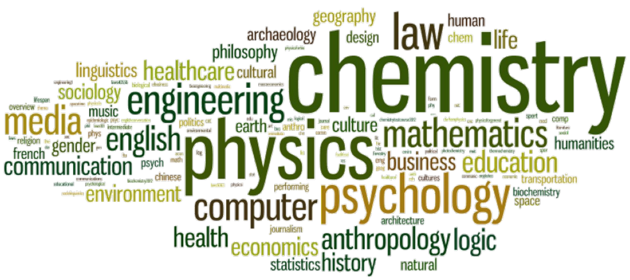
Knowledge Graph



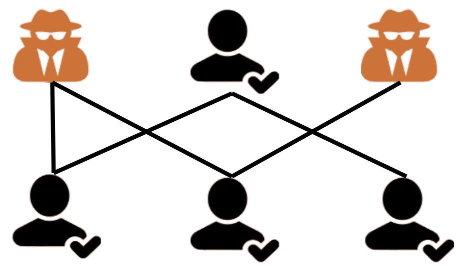
The Home Depot Product Knowledge Graph

Graph-based Task is Everywhere!

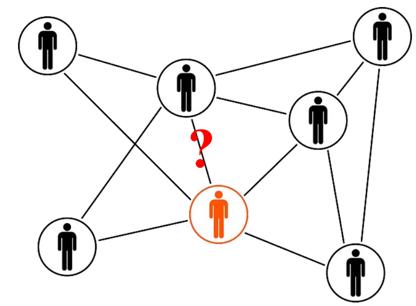
Paper Management

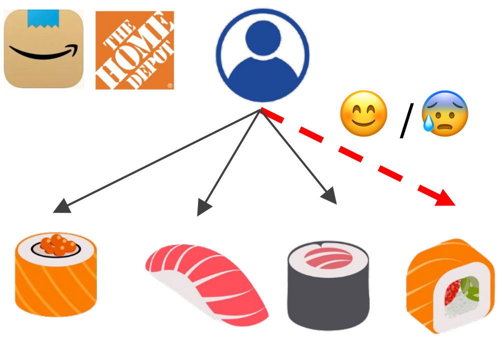


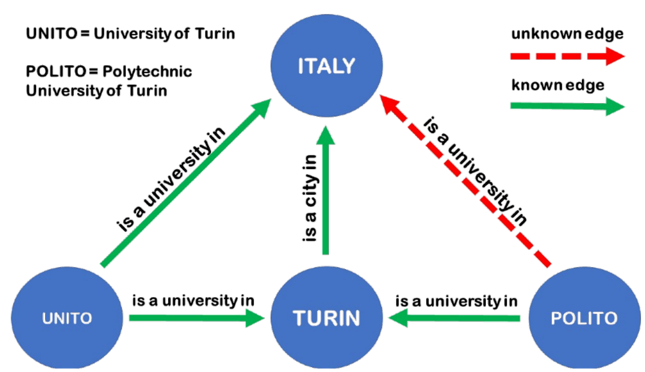
Anomaly Detection



Social Prediction





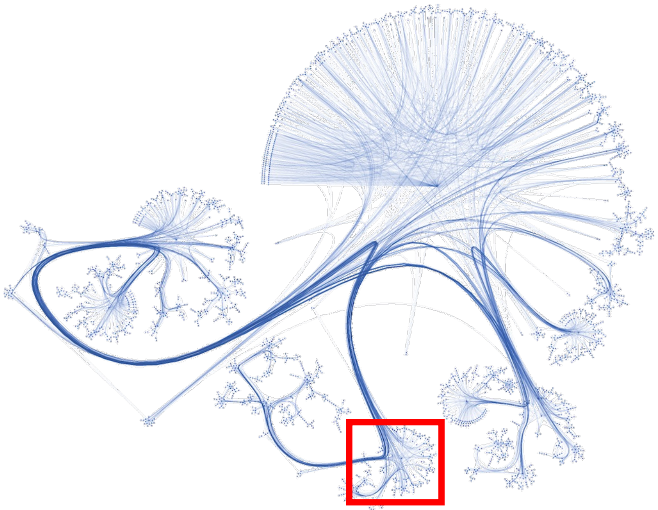


Recommender system

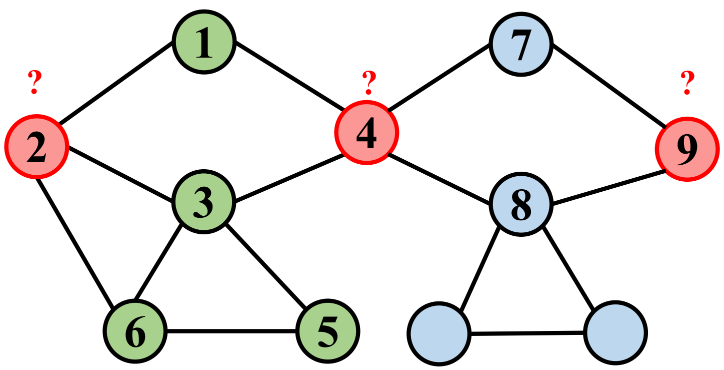
Question Answering

Example 1: Academic Paper Management

Cora - Paper Citation Networks



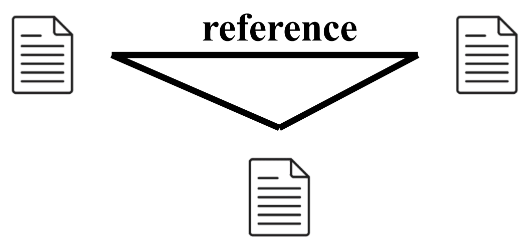
Zoom in



1 INTRODUCTION

We consider the problem of classifying nodes (such as documents) in a graph (such as a citation network), where labels are only available for a small subset of nodes. This problem can be framed as graph-based semi-supervised learning, where label information is smoothed over the graph via some form of explicit graph-based regularization (Zhu et al., 2003; Zhou et al., 2004; Belkin et al., 2006; Weston et al., 2012), e.g. by using a graph Laplacian regularization term in the loss function:

$$\mathcal{L} = \mathcal{L}_0 + \lambda \mathcal{L}_{\text{reg}}\,, \quad \text{with} \quad \mathcal{L}_{\text{reg}} = \sum_{i,j} A_{ij} \| f(X_i) - f(X_j) \|^2 = f(X)^\top \Delta f(X)\,. \qquad (1)$$



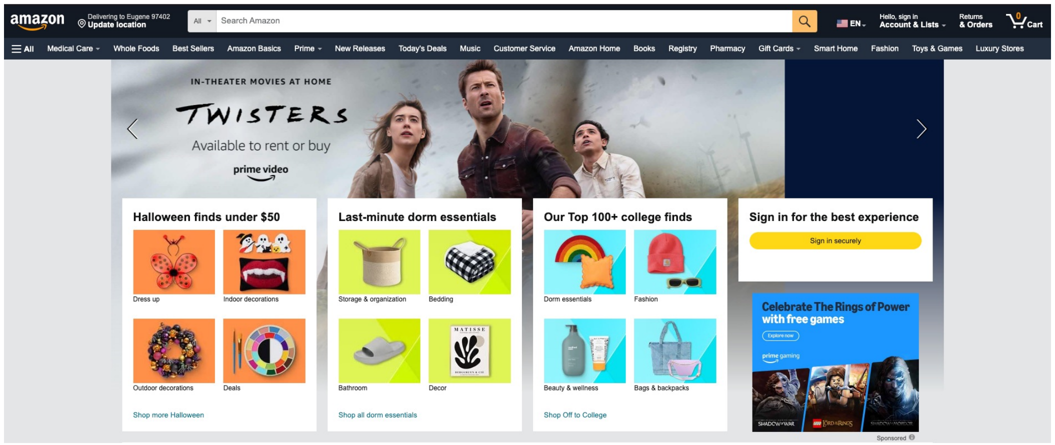
Machine Learning

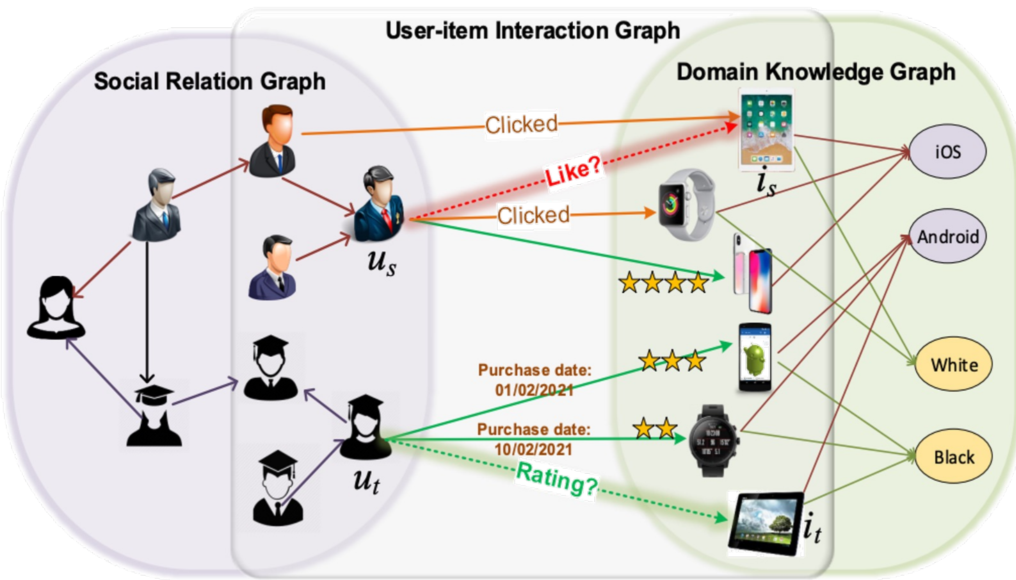
Computer System

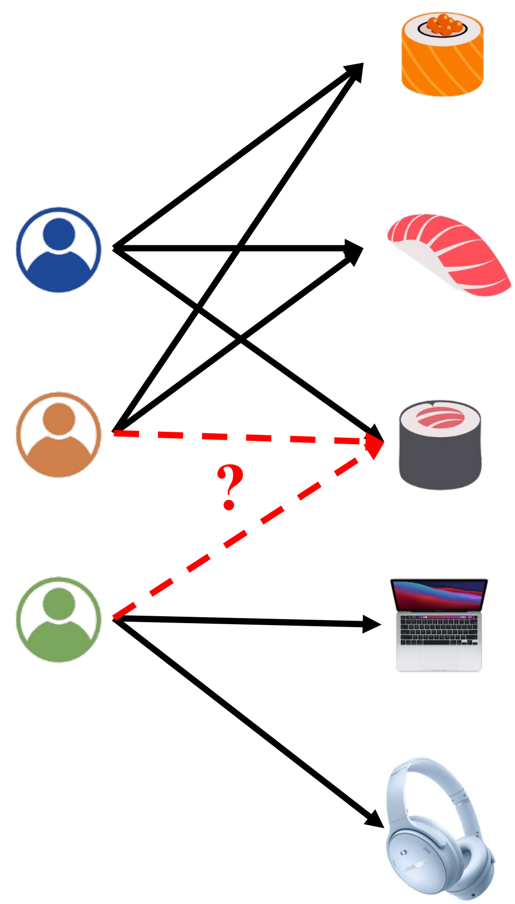
What do you think?

Example 2: Personalization

Amazon – Customer-Product Network

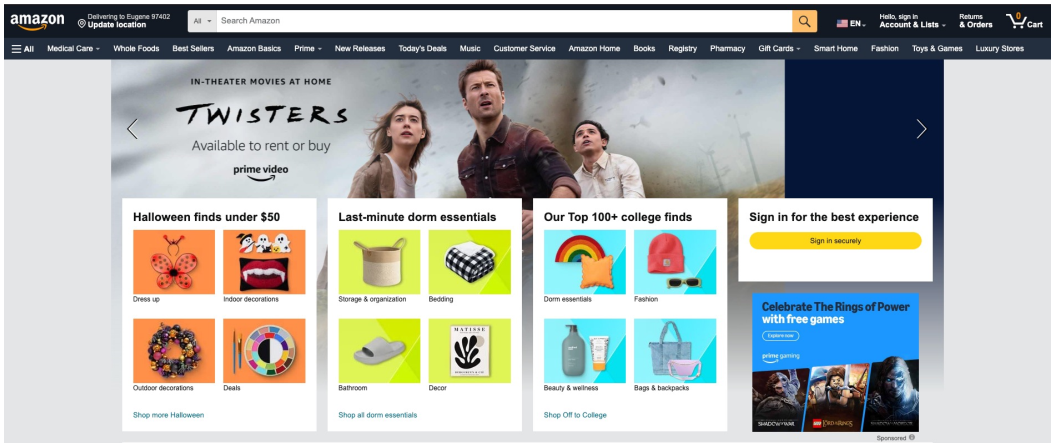


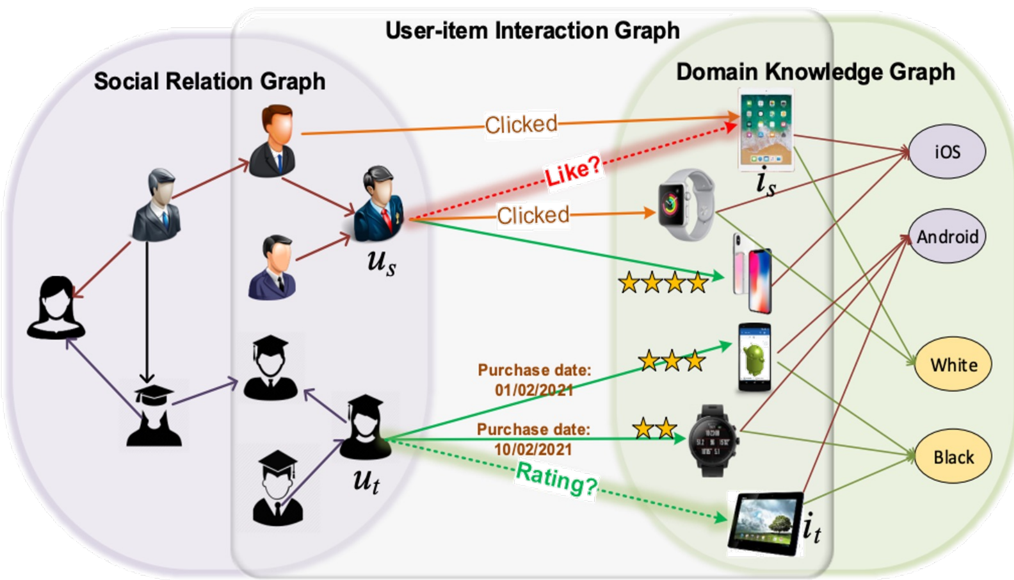


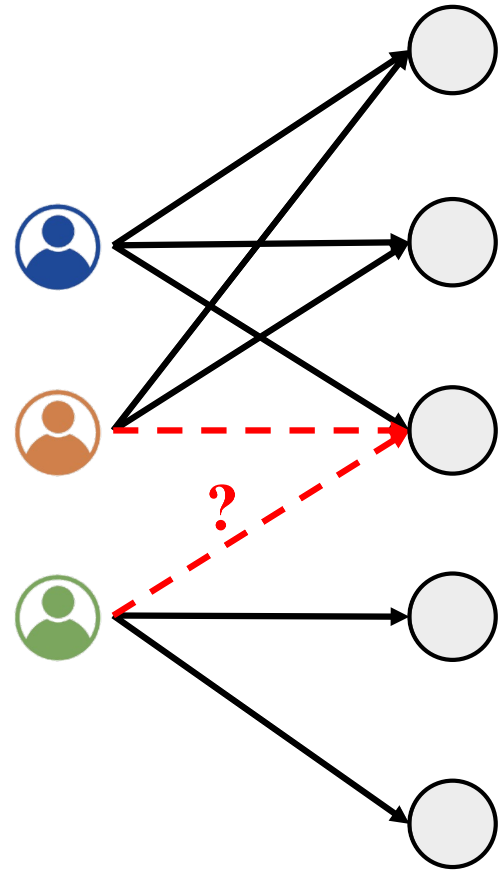


Example 2: Personalization

Amazon – Customer-Product Network



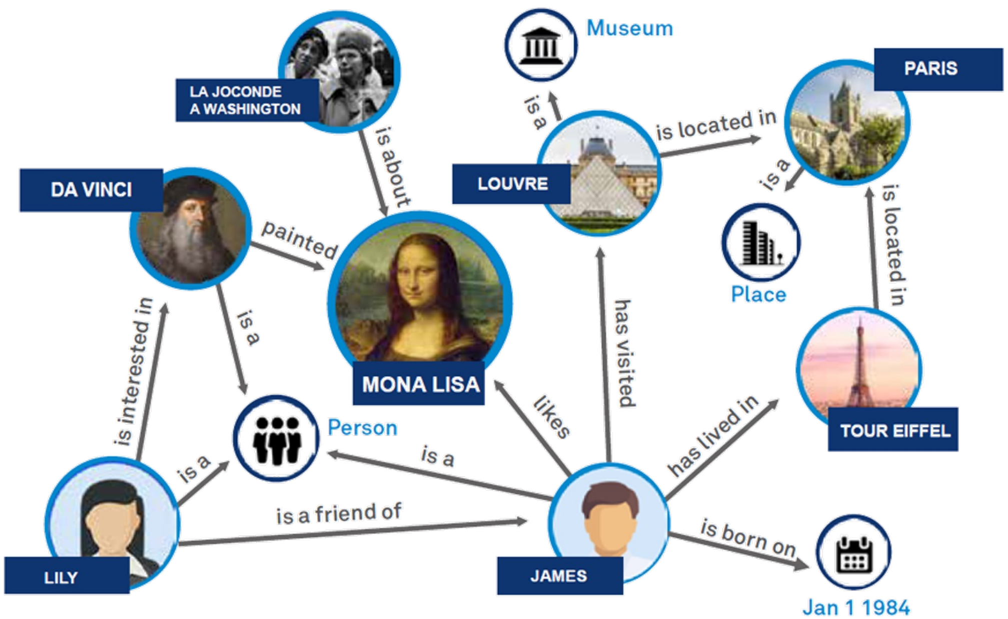




Recommendation based on customer-product interaction

Example 3: Question-answering

Knowledge Graph



Store a large amount of factual knowledge in a symbolic format

Who painted Mona Lisa?

(1) Locate Mona Lisa

(2) Find her 1-hop neighbor

(3) Da Vinci Painted Mona Lisa

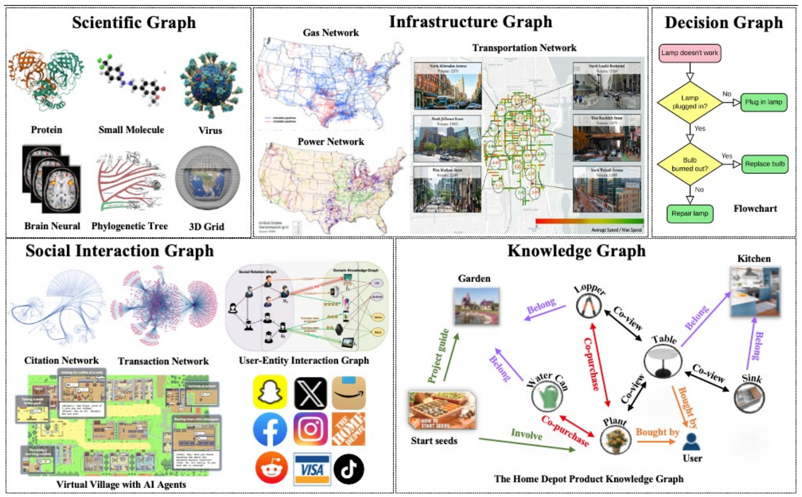
How about LLMs?

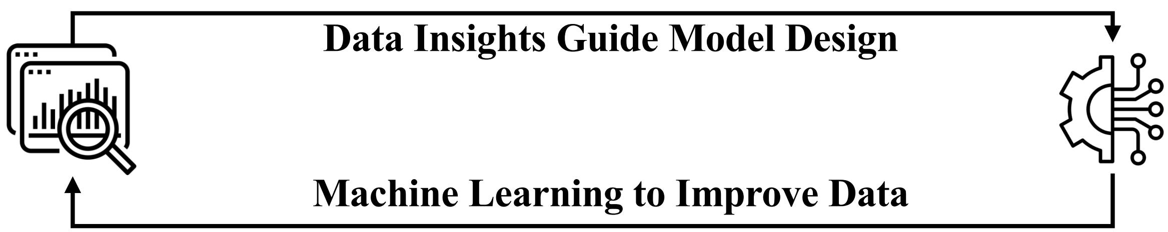
Who painted Mona Lisa? Answer in 5 words

Leonardo da Vinci painted Mona Lisa.

LLMs store knowledge in their parameters

Data Mining and Machine Learning on Graphs



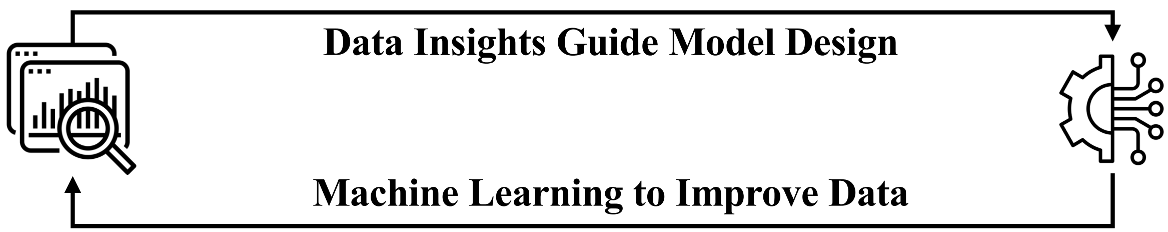


Data mining
Analyze data
Derive patterns and relationships
Solve real-world problems

Machine Learning
Design Model
Allow Computer to Learn and Improve
Without being explicit programmed

Data Mining and Machine Learning on Graphs

Data mining **Machine Learning**



Real-world Applications

Data Structure

- Image
- Language
- Time-series Data
- Graph
  - Document Graph
  - Infrastructure Graph
  - Scientific Graph
  - Social Graph
  - ……

Purpose

- Utility - Performance
- Fairness
- Privacy
- Diversity
- Robustness
- Adversarial attack
- Eco-friendly
- Efficiency

Data Mining and Machine Learning on Graphs

Data mining | **Machine Learning**

Data Insights Guide Model Design

Machine Learning to Improve Data

Real-world Applications

Data Structure

- Image
- Language
- Time-series Data
- Graph
  - Document Graph
  - Infrastructure Graph
  - Scientific Graph
  - Social Graph
  - ……

Purpose

- Utility - Performance
- Fairness
- Privacy
- Diversity
- Robustness
- Adversarial attack
- Eco-friendly
- Efficiency

Data Mining & Machine Learning on Graphs

Network Analysis

- **Background**
  - Graph Theory
  - Linear Algebra
- **Statistical Graph Model**
  - Erdos-Renyi, Barabasi-Albert
  - Small-World, Chung-Lu
- **Network Analysis**
  - Degree, Closeness
  - Betweeness, Katz
  - Eigenvector, PageRank

Computation Methods

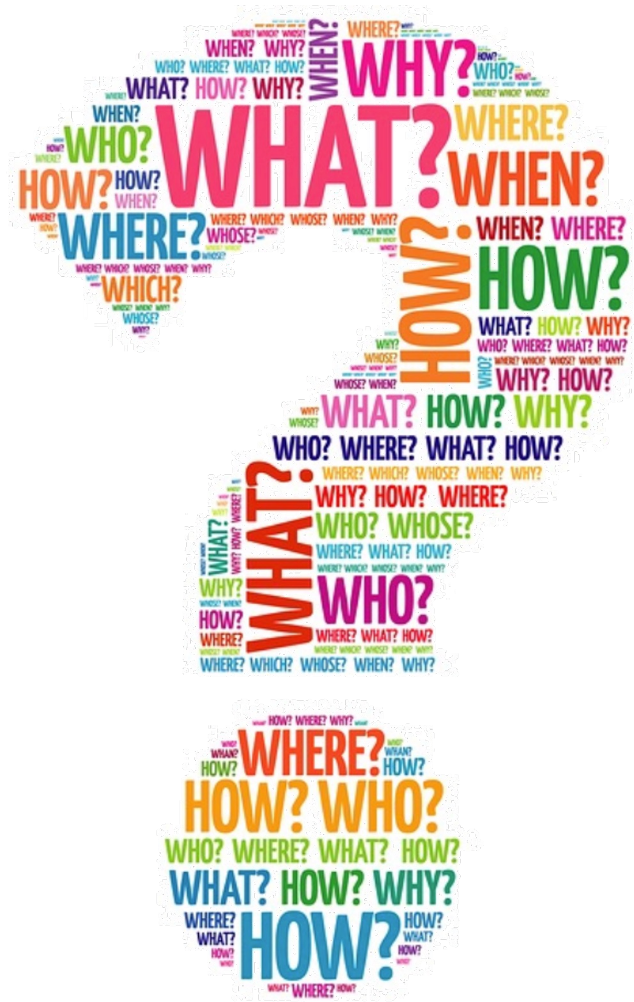
- **Link Prediction**
- **Node Classification**
- **Graph Classification**
- **Network Diffusion**
- **Graph Clustering**

Machine Learning on Graph

- **Network Embedding**
- **Graph Neural Networks**
- **Self-supervised Learning**
- **Trustworthy Issue**
- **Data-quality Issue**

10/01 **10/22** **11/07** **12/03**

Any Question?



Course Logistics - Overview

Website

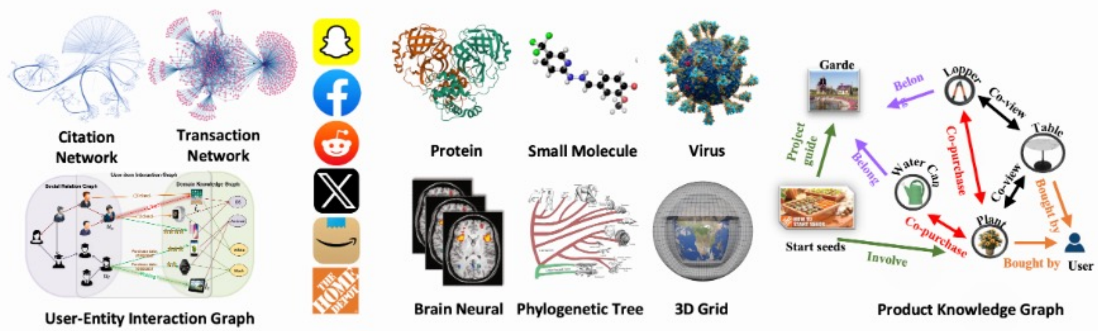
https://ml-graph.github.io/fall-2024/

Department of Computer Science, University of Oregon

Data Mining & Machine Learning on Graphs

Fall-2024

SYLLABUS | SCHEDULE | ASSIGNMENTS | PROJECT | MATERIALS | GRADE



Course Description

Graph-structured data is ubiquitous worldwide, e.g., social networks like Facebook, e-commerce platforms like Amazon, infrastructures like transportation networks, and chemical graphs like molecules. This course explores basic analytical techniques, computational methods, and graph machine learning models for graph-related applications.

Topics include:

- **Graph Foundations**: Basic Graph Theory, Statistical Graph Models, Network Properties.
- **Graph Computational Methods**: Link Prediction, Node/Graph Classification, Diffusion, and Clustering.
- **Deep Graph Models**: GNNs, Self-supervised Learning, Data Quality and Trustworthy Issues.
- **Real-world Applications**: Academic Paper Management, Recommender System, Drug Discovery.

Students will complete assigned homework, a midterm exam, and a team-based course project.

All details are in this website!

Course Logistics – Goals and Requirements

Goals:

- Broad overview of basic knowledge and algorithm foundations of ML/DM on Graphs
- Hands-on experience with solving GML/DM problems
- Master real-world GML/DM applications

Requirements:

- Little to no background in ML
- Basic linear algebra, probability and statistics, and calculus
- Programming – Python
- Jupyter Notebooks for homework assignments

Course Logistics – Basic Contents

Times:

- **Classes:** Tuesday/Thursday 4:00-5:20 pm, 132 GSH
- **Office hours:** Friday 3:30-5:00 pm PST, other time by appointment
- **Zoom:** https://uoregon.zoom.us/j/4052006678

Components:

- 3 homework assignments (35%)
- Midterm (30% - CS410, 25% - CS510)
- Final Project (30%)
- Participation (5%)
- Paper Presentation (5% - CS510)
- Homework submitted with Overleaf (5% Bonus)

Course Logistics – Homework Assignments

Assignments

- Writing Assignment 1
- Programming Assignment 2
- Programming Assignment 3

No collaborations are allowed for assignments unless otherwise specified. Late Assignments will receive:

- 20% reduction, if submitting within (0, 24) hours late
- 40% reduction, if submitting within [24, 48) hours late
- 100% reduction, if submitting within [48, ∞) hours late, unless having documental special circumstances

You can download the assignments here. Also check out each assignment page for any additional info.

Assignment 1

Assignment 2

Assignment 3

Will Release Soon!

Course Logistics – Paper Presentation

https://ml-graph.github.io/fall-2024/paper/

1. **Introduction and Background** – What is the general impact and background of the topic?

2. **Motivation and Problem** – What is the core research problem and why do we study it?

3. **Related Work and Challenges** – How did previous works on this problem and what are some challenges?

4. **Proposed Solutions/Methods and Rationale** – What are the proposed methods/techniques and why propose them? What specific reasons that solving this problem would require these porposed methods/technqiues

5. **Experimental Setting, Results and Analysis** – What experiments are designed to verify the proposed method? How are results being discussed and analyzed? Are there any interesting findings?

5. **Conclusion and Future Work**

Course Logistics – Paper Presentation

Paper

Graph Retrieval Augmented Generation

[ArXiv 2024] **Graph Retrieval-Augmented Generation: A Survey** [Paper]
[ArXiv 2024] **From Local to Global: A Graph RAG Approach to Query-Focused Summarization** [Paper]

Social Network Analysis

[ArXiv 2024] **Exploring Collaboration Mechanisms for LLM Agents: A Social Psychology View** [Paper]
[ArXiv 2024] **Scaling Large-Language-Model-based Multi-Agent Collaboration** [Paper]
[ArXiv 2024] **Network Formation and Dynamics Among Multi-LLMs** [Paper]
[ArXiv 2024] **Large Language Models Empowered Agent-based Modeling and Simulation: A Survey and Perspectives** [Paper]
[ArXiv 2024] **LLMs generate structurally realistic social networks but overestimate political homophily** [Paper]

GraphAI for Science

[NeurIPS 2024] **Learning to Group Auxiliary Datasets for Molecule** [Paper]
[ICLR 2023] **DiffDock: Diffusion Steps, Twists, and Turns for Molecular Docking** [Paper]
[ICML 2022] **Equivariant Diffusion for Molecule Generation in 3D** [Paper]
[ICLR 2023] **DiGress: Discrete Denoising diffusion for graph generation** [Paper]

GraphAI for Cybersecurity and System

[JSAC 2020] **RouteNet: Leveraging Graph Neural Networks for Network Modeling and Optimization in SDN** [Paper]
[ASCE 2022] **Graph Neural Networks for State Estimation in Water Distribution Systems: Application of Supervised and Semisupervised Learning** [Paper]
[ASCE 2022] **Optimal Power Flow using Graph Neural Networks** [Paper]
[ArXiv 2024] **PowerGraph: A power grid benchmark dataset for graph neural networks** [Paper]
[IEEE SmartGridComm] **On Graph Theory vs. Time-Domain Discrete-Event Simulation for Topology-Informed Assessment of Power Grid Cyber Risk** [Paper]

Course Logistics – Grades

Course Assessment and Grading Scale

Category	CS-410 (%)	CS-510 (%)
Assignment	35%	35%
Midterm Exam	30%	25%
Final Project	30%	35%
Participation	5%	5%
Paper Presentation	0%	5%

Grade	Range
A	A+: 98-100, A: 93-97, A-: 90-92
B	B+: 87-89, B: 83-86, B-: 80-82
C	C+: 77-79, C: 73-76, C-: 60-72
F	F: <60

Any Question?