COSI241
Advanced Topics in Graph Mining

Chuxu Zhang
Computer Science Department
Brandeis University

8/27/2020

Course Information

❑ COSI241A - Advanced Topics in Graph Mining

Covers recent core techniques and advances in graph mining research.
Main content: network embedding, graph neural network, recommendation with graphs, knowledge graph, and others.
Students read and discuss literature, make presentation, and work on related research projects.

❑ Instructor: Dr. Chuxu Zhang

Assistant Professor
General research: data mining, applied machine learning, deep learning
Specific research: graph and network mining, recommender systems, time series and spatial-temporal data analysis, interdisciplines (e.g., computational chemistry)
Contact: chuxuzhang@brandeis.edu
Webpage: https://chuxuzhang.github.io

Course Information

- Time: Tuesday/Thursday 4:00pm – 5:30pm
- Location: Goldsmith 300 (In-Person)
- Office Hour: By Appointment
- Teaching Assistant: Peizhao Li (peizhaoli@brandeis.edu)
- LATTE Forum (Announcement, Discussion, Submission, etc.)
- Course Website: https://chuxuzhang.github.io/course/COSI241.html

Course Logistics

❑ Prerequisites

Applied Linear Algebra (MATH 15A) is preferred, Background in Data Mining and Machine Learning (e.g., COSI 126A, COSI 123A, or COSI 101A) is required, Programming experience in {**Python**, C/C++, or Matlab} is required.

❑ Course Objectives

Be familiar with the recent advances in graph mining research.
The skills, techniques, and implementations of algorithms and models of related study.
The potentiality of developing new algorithms and models to improve previous study or creating new inspiring research problems and corresponding solutions.
Students are expected to start graph mining or other related research during or after this course.

Course Logistics

❑ Course Materials

There is no required textbook as it is a seminar course. A list of suggested papers and tutorials will be provided. Class lecture slides will be provided by the instructor and students. Students should make slides for their paper presentations.

❑ Special Statement for Covid-19

Must wear masks, keep social distance during lectures

Remote office hours via Zoom (By appointment)

Current in-person class, may change to remote in case of bad situation

Student's preference: in person/remote

Course Logistics

❑ Grading

Class attendance and participation (10%)
Paper reading, presentation, and discussion (40%)
A course project with presentation (50%: midterm report 10%, final presentation 10%, final paper 30%)

Grades (4 points in Brandeis)
A (A+) = 4.00; A- = 3.67; B+ = 3.33; B = 3.00; B- = 2.67; C+ = 2.33; C = 2.00;
C- = 1.67; D+ = 1.33; D = 1.00; D- = 0.67 and E = 0.00

Grades (100 points)
A (A+) = [90, 100]; A- = [87, 90); B+ = [84, 87); B = [81, 84); B- = [78, 81); C+ = [75, 78);
C = [72, 75); C- = [69, 72); D+ = [66, 69); D = [63, 66); D- = [60, 63); and E = [0, 60)

More About Paper Presentation

❑ Paper presentation (10%*4)

Use suggest paper list (topics of each month are summarized in course syllabus).

Presentation (**scheduled by TA**), read papers, make slides, present and discuss the paper.

Contact TA for your presentation schedule in each month, no repeat presentation for the same paper, students who present earlier have more choices, each student is expected to present 1 paper every month.

Grading is based on your understanding of paper, slides, presentation, and discussion.

More About Paper Presentation

❑ Presentation content (30 mins)

Includes following content:

Introduction/Background
Problem
Model/Method/Algorithm
Experiments
Conclusion

Provide sufficient details for audiences

❑ Discussion (15 mins)

At the end of your presentation, should raise your thought/question/discussion about the presentation work.

More About Course Project

- **Course Project (midterm pre. 10%, final pre. 10%, final paper 30%)**

  Research project chosen by students and should be related to this course.

  Could be a follow up work of current papers, or a new problem and its solution. (**implementation of current work is not a research project**)

  Group project: up to 2 students. (Form group by yourself)

  Midterm proposal/presentation: mid-October, Final report/presentation: last week.

  The final report is expected to be submitted to a DM/NLP/AI/ML conference.

  Grading based on quality of your project, completion, report, and presentation.

Course Logistics

❑ Schedule

Aug. 27	Introduction	Course Introduction
Sep. 1/3	Graph Representation Learning Tutorial	
Sep. 8	Network Embedding	
Sep. 15/17	Network Embedding	
Sep. 22/24	Graph Neural Network	
Sep. 29	Graph Neural Network	
Oct. 1	Project Proposal Presentation	Project Proposal Due
Oct. 6/8	Recommendation with Graphs Tutorial	
Oct. 13/15	Recommendation with Graphs	
Oct. 20/22	Recommendation with Graphs	
Oct. 27/29	Recommendation with Graphs	
Nov. 3/5	Knowledge Graphs Tutorial	
Nov. 10/12	Knowledge Graphs	
Nov. 17/19	Knowledge Graphs	
Nov. 24/26	No Class	Thanksgiving Holiday
Dec. 1/3	Project Presentation	Project Final Report Due

Course Logistics

Q & A

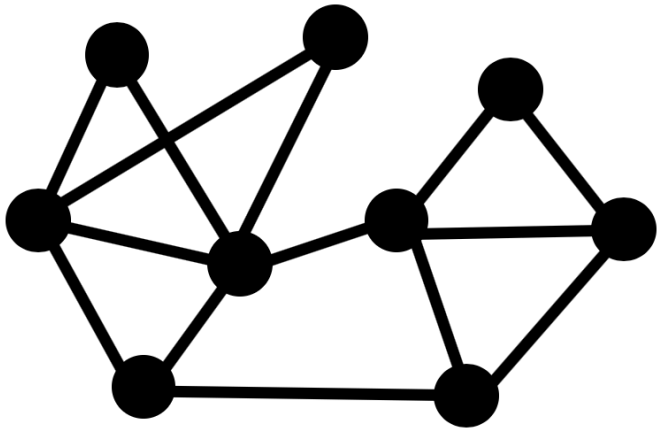
Introduction

❑ What is graph (network)

A data type of data mining/machine learning.
matrix data, set data, time series data, spatial data, **graph/network data**, …

A general language for describing real-world complex systems of interacting entities.

Basic components: node/edge.

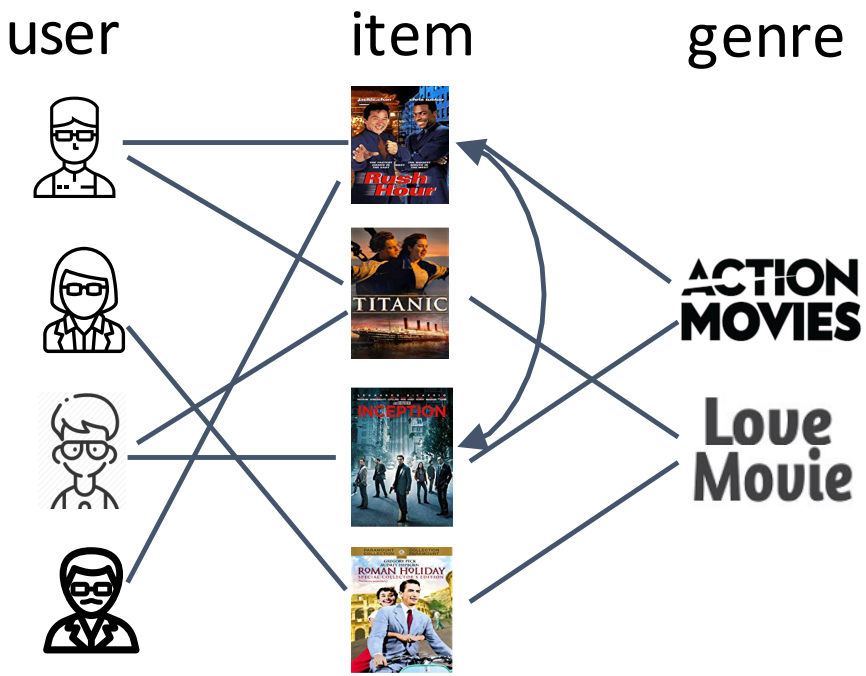


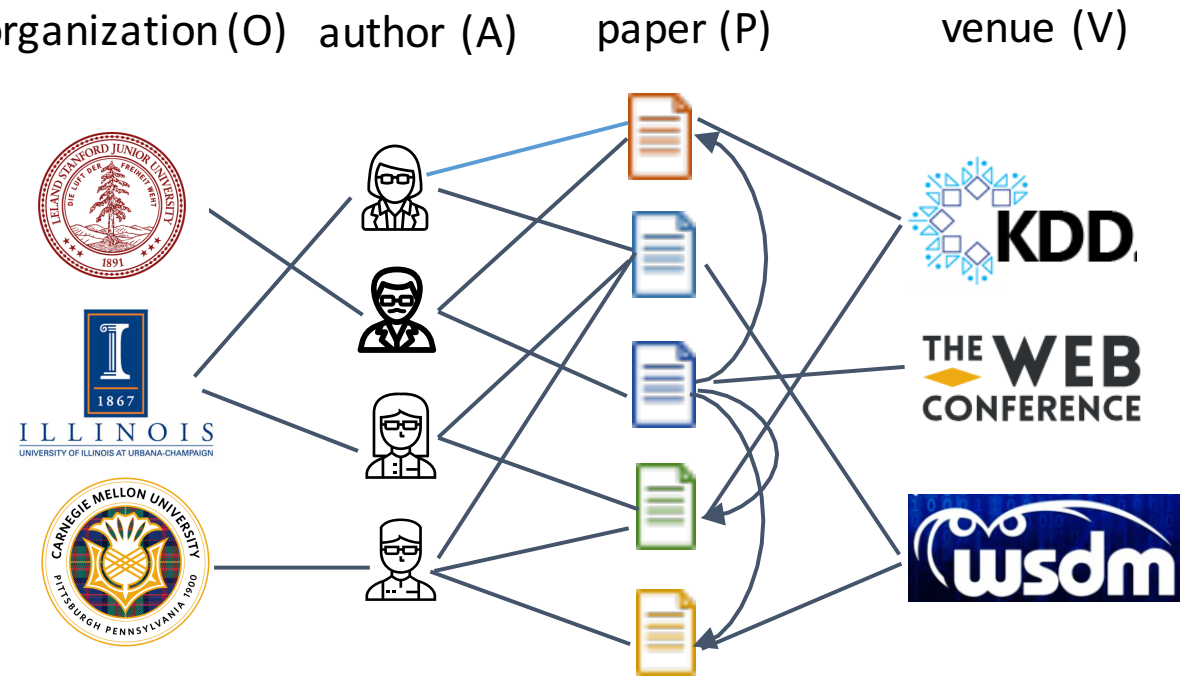
Introduction

- What is graph (network)

Introduction

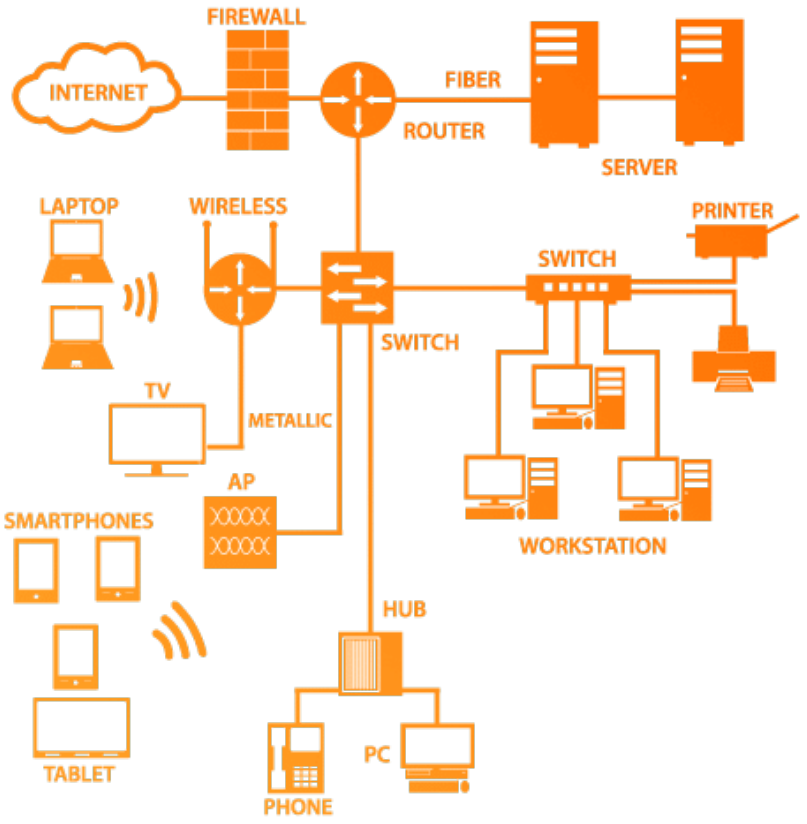
- What is graph (network)





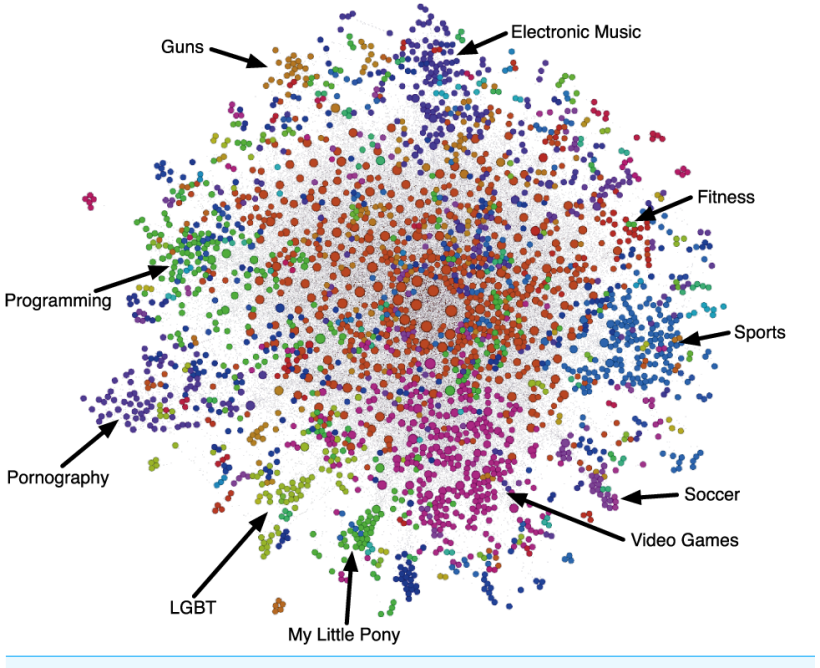
Introduction

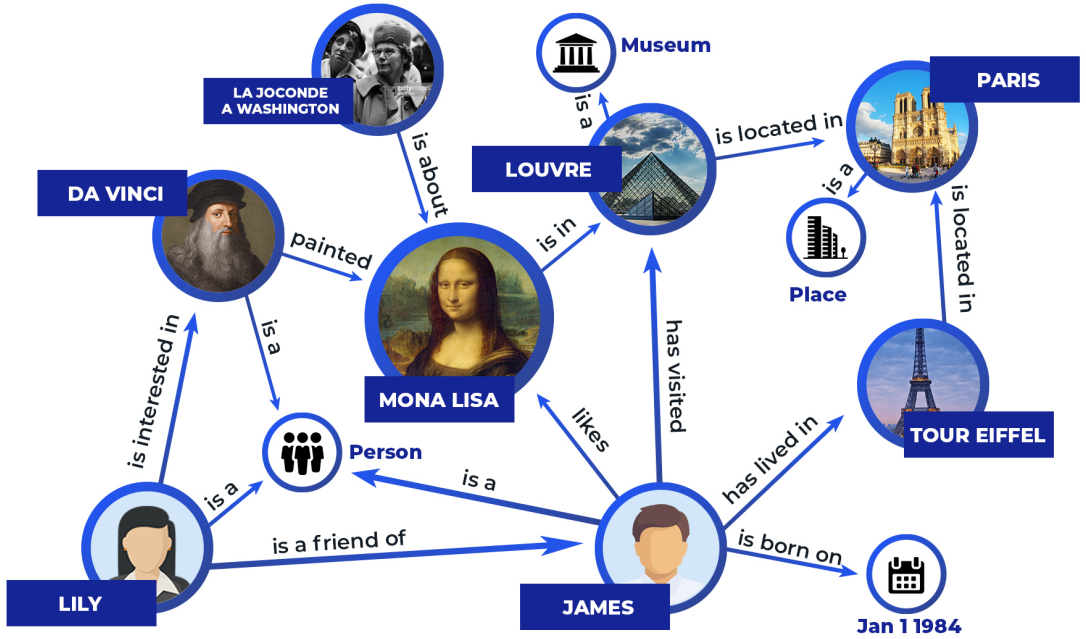
- What is graph (network)



Introduction

- What is graph (network)





Introduction

- What is graph (network)

Behind many real-world complex systems there is an intricate diagram, **a graph (network)**, that describing the interactions between the entities.

Introduction

❑ Some concepts

Graph/Network: mathematical form/real system

(Un)weighted Graph: edge weight

(In)directed Graph: direction

Bipartite/Tripartite Graph: two/three kinds of nodes

Homogeneous/Heterogeneous Graph: one/some kinds of nodes/edges

Introduction

❑ Why use graph (network)

Universal language for describing complex data
Many real-world complex systems are represented as graphs (networks)

Shared vocabulary between fields
Same for computer science, social science, chemistry, biology, …

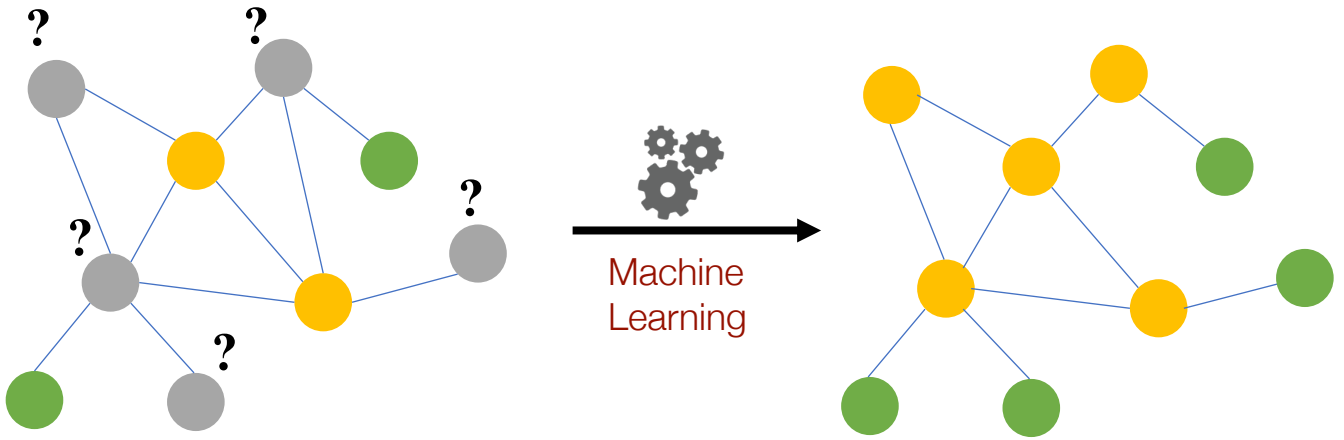
Data availability
Web, chemistry/biology, mobile, health, …

Model graph/network to predict complex systems for different applications

Introduction

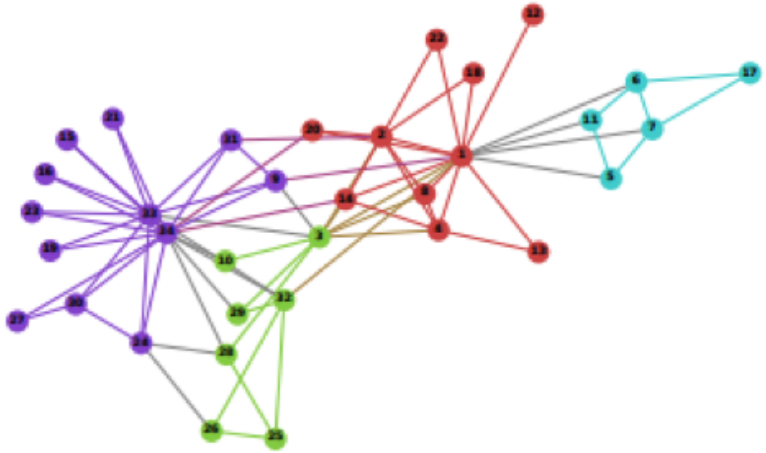
❑ Graph (network) application

Predict the type/color of
a given node
(**node classification**)

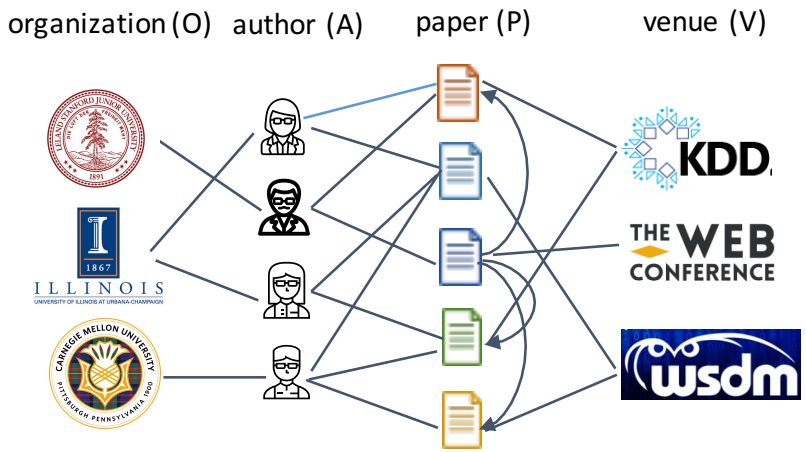


Introduction

❑ Node classification



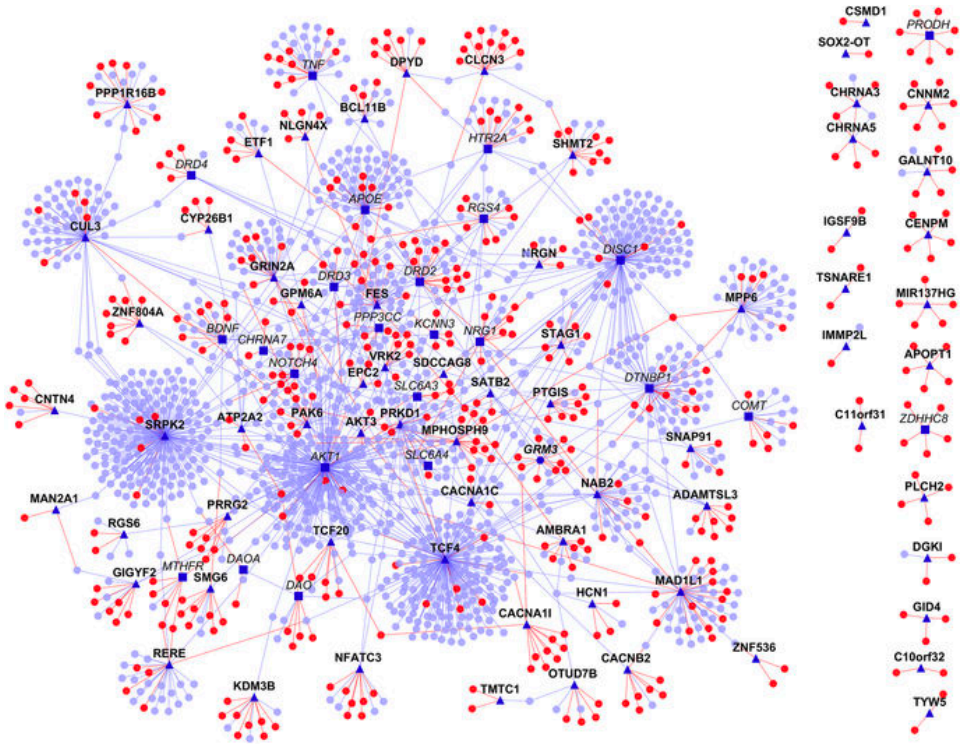
group label of each user
in social network



author/paper category
in academic

Introduction

- Node classification

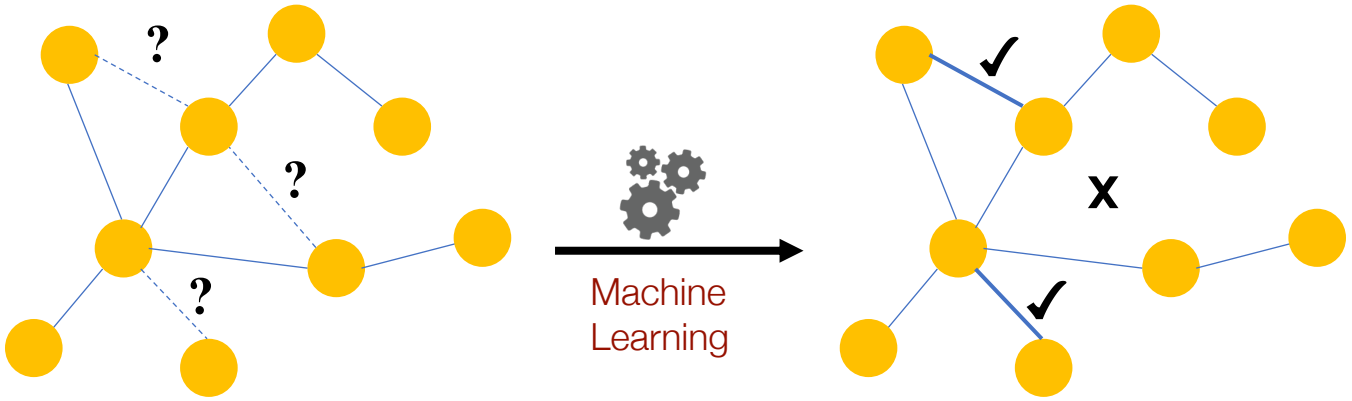


function of proteins
in the interactome

Introduction

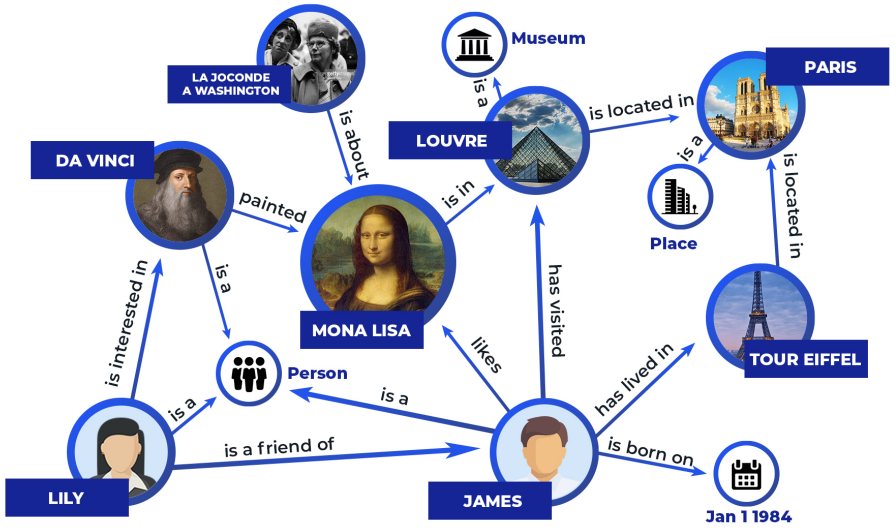
❑ Graph (network) application

Predict whether two nodes are linked (**link prediction**)

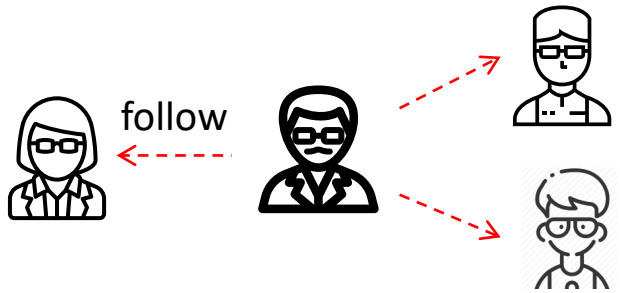


Introduction

- Link prediction



relation reasoning
in knowledge base



friendship prediction
in social media

Introduction

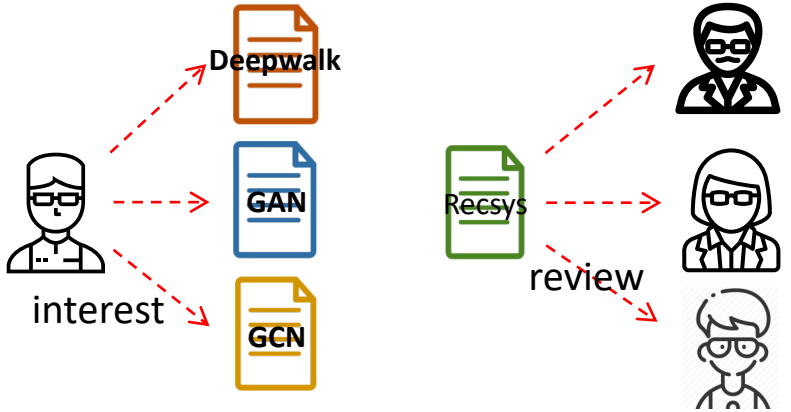
❑ Graph (network) application

Predict whether two nodes are linked
(**node recommendation**)

E-commerce

item recommendation

Web Academic System

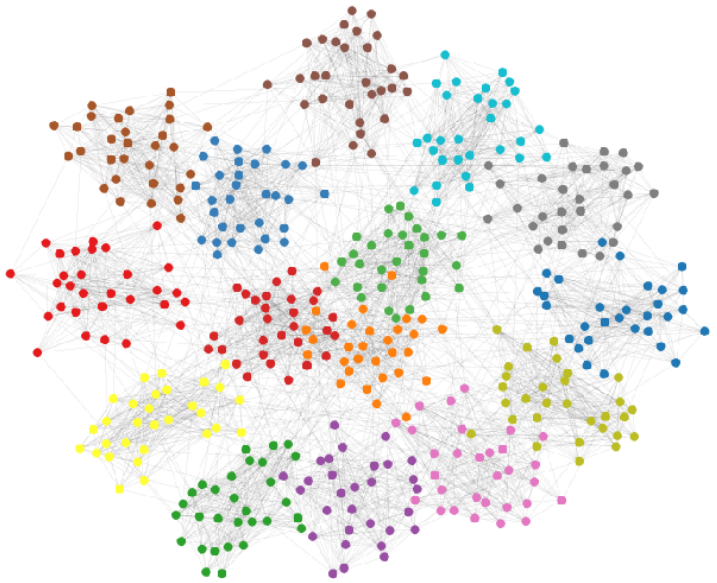


paper recommendation
reviewer recommendation

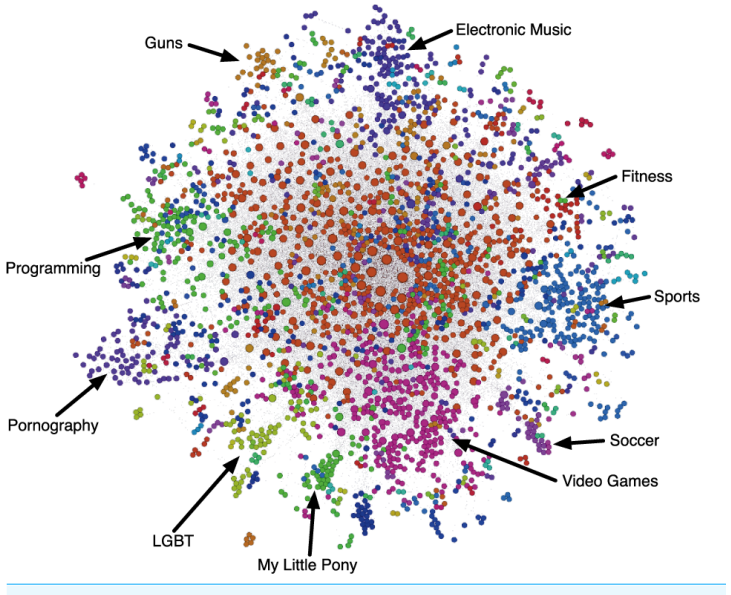
Introduction

❑ Graph (network) application

Identify densely linked clusters of nodes (**community detection**)



user group in social media



topic group in online forum

Introduction

❑ Graph (network) application

Measure similarity of two nodes (**similarity search**)

Rank	KDD	SIGMOD	SIGIR	WWW	WSDM
0	KDD	SIGMOD	SIGIR	WWW	WSDM
1	SDM	PVLDB	TREC	CIKM	WWW
2	ICDM	ICDE	CIKM	SIGIR	SIGIR
3	DMKD	TODS	IPM	KDD	KDD
4	KDD E	VLDBJ	IRJ	ICDE	AIRWeb
5	PKDD	PODS	ECIR	TKDE	CIKM
6	PAKDD	EDBT	TOIS	VLDB	WebDB
7	TKDE	CIDR	WWW	TOTT	ICDM
8	CIKM	TKDE	JASIST	SIGMOD	VLDB
9	ICDE	ICDT	JASIS	WebDB	VLDBJ
10	TKDD	DE Bull	SIGIR F	WISE	SDM

similar venue in academic

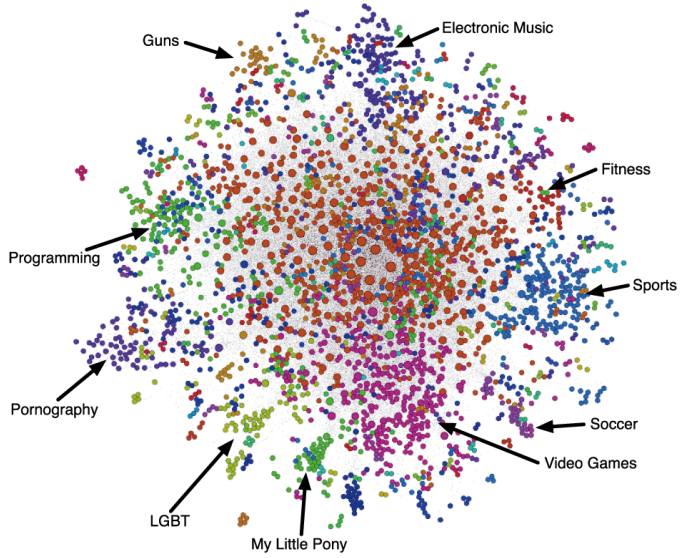
similar item search

Introduction

❑ Graph (network) application

Predict the type/attribute
of a given graph
(**graph classification**)

reactivity identification
in drug discovery

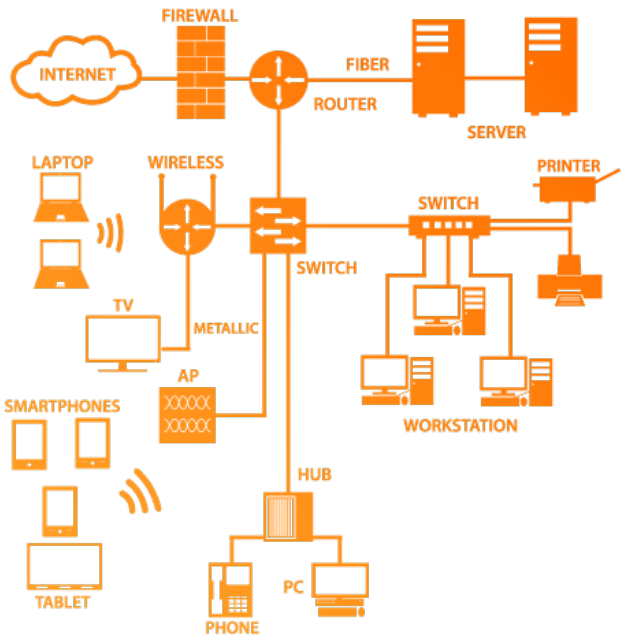


topic group
in online forum

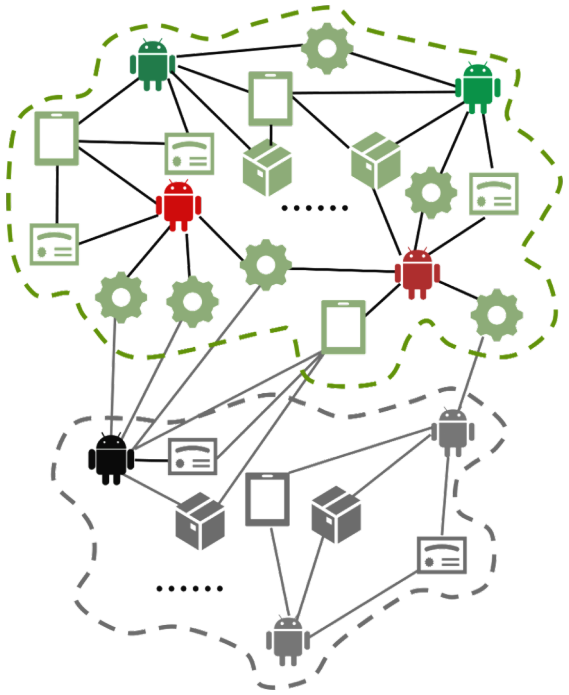
Introduction

❑ Graph (network) application

Predict outlier node of a given graph (**anomaly detection**)



device error in IT system



malware detection

Course Logistics

❑ Schedule

Aug. 27	Introduction	Course Introduction
Sep. 1/3	Graph Representation Learning Tutorial	
Sep. 8	Network Embedding	
Sep. 15/17	Network Embedding	
Sep. 22/24	Graph Neural Network	
Sep. 29	Graph Neural Network	
Oct. 1	Project Proposal Presentation	Project Proposal Due
Oct. 6/8	Recommendation with Graphs Tutorial	
Oct. 13/15	Recommendation with Graphs	
Oct. 20/22	Recommendation with Graphs	
Oct. 27/29	Recommendation with Graphs	
Nov. 3/5	Knowledge Graphs Tutorial	
Nov. 10/12	Knowledge Graphs	
Nov. 17/19	Knowledge Graphs	
Nov. 24/26	No Class	Thanksgiving Holiday
Dec. 1/3	Project Presentation	Project Final Report Due

Introduction

❑ What you should do now (contact TA)

Schedule your paper presentation for September

Topics in September: Network embedding, Graph neural network

Form course project group and design the suitable project

Determine your choice between in-person class or remote class

Introduction

Q & A