

Mining Graph Data

COURSE INTRODUCTION

Teacher: Le Ngoc Thanh

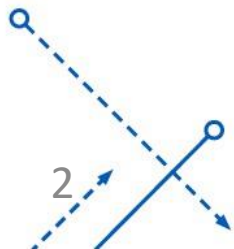
Email: lnthanh@fit.hcmus.edu.vn



fit@hcmus

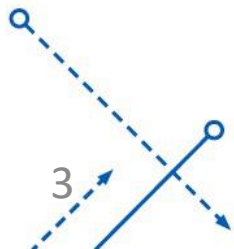
Contents

- Introduction of subjects and topics
- Rule
- Networks and graphs
- Graph mining
- Maths and applications



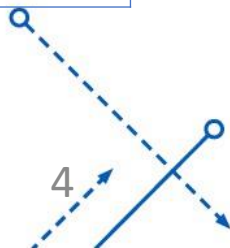
About the subject

- Subject name : Mining Graph Data
- Theory/Practice: 45/30.
- Rate of listening to lectures and self-studying: 40/60.
- References :
 - Lecture slides
 - Aggarwal, Charu C., and Haixun Wang, eds. Managing and mining graph data. Vol. 40. New York: Springer, 2010.
 - Easley, David, and Jon Kleinberg. "Networks, crowds, and markets: Reasoning about a highly connected world." Significance 9 (2012): 43-44.
 - Ketmaneechairat, Hathairat. "Graph Mining Laws, Tools and Case Studies." Journal of Digital Information Management 12, no. 6 (2014): 446
 - ...



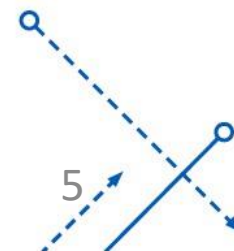
Subjects

Week	Subject
1	Introduction to Graph Data Mining, Algorithms and Applications
2	Review the knowledge related to graphs and data mining
3	Patterns in static and dynamic graphs; generate graph.
4	Indexing Graph and Ranking
5	Mining Graph Pattern
6	Graph Classification



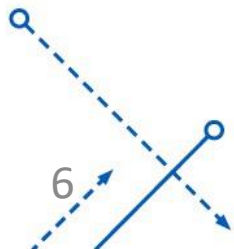
Subjects

Week	Subject
7	Clustering and community detection
8	Link prediction
9	Embedding graphs
10-15	Seminar topics includes: deep learning for graphs, graph summarization, recommendation systems, anomaly detection, large size graphs.



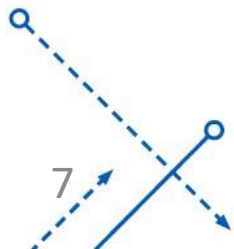
Contents

- Introduction of subjects and topics
- Rule
- Networks and graphs
- Graph mining
- Maths and applications



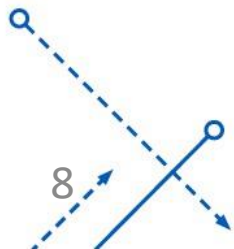
Regulations and academic assessment

- Theory: 40%
 - Final Exam: 40%
- Lab: 30%
- Seminar Project: 30%
- Regulations:
 - Copy, cheats → not be allowed to take the final exam
 - Actions that disrupt the classroom → not be allowed to take the final exam.



Contents

- Introduction of subjects and topics
- Rule
- **Networks and graphs**
- Graph mining
- Maths and applications

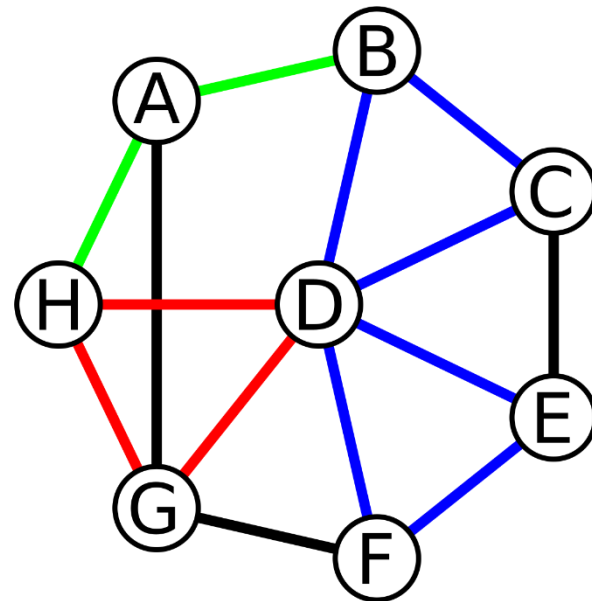


Networks and graphs

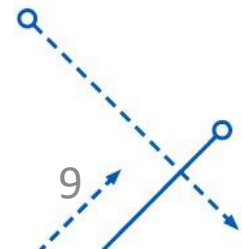
- **Network** often used to represent the natural relationship of objects in the real world.
- Meanwhile, **graphs** show the relationship generated through the automatic process.



Social Network



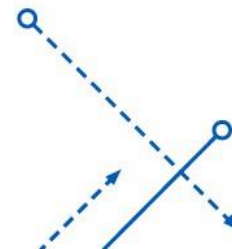
Graphs



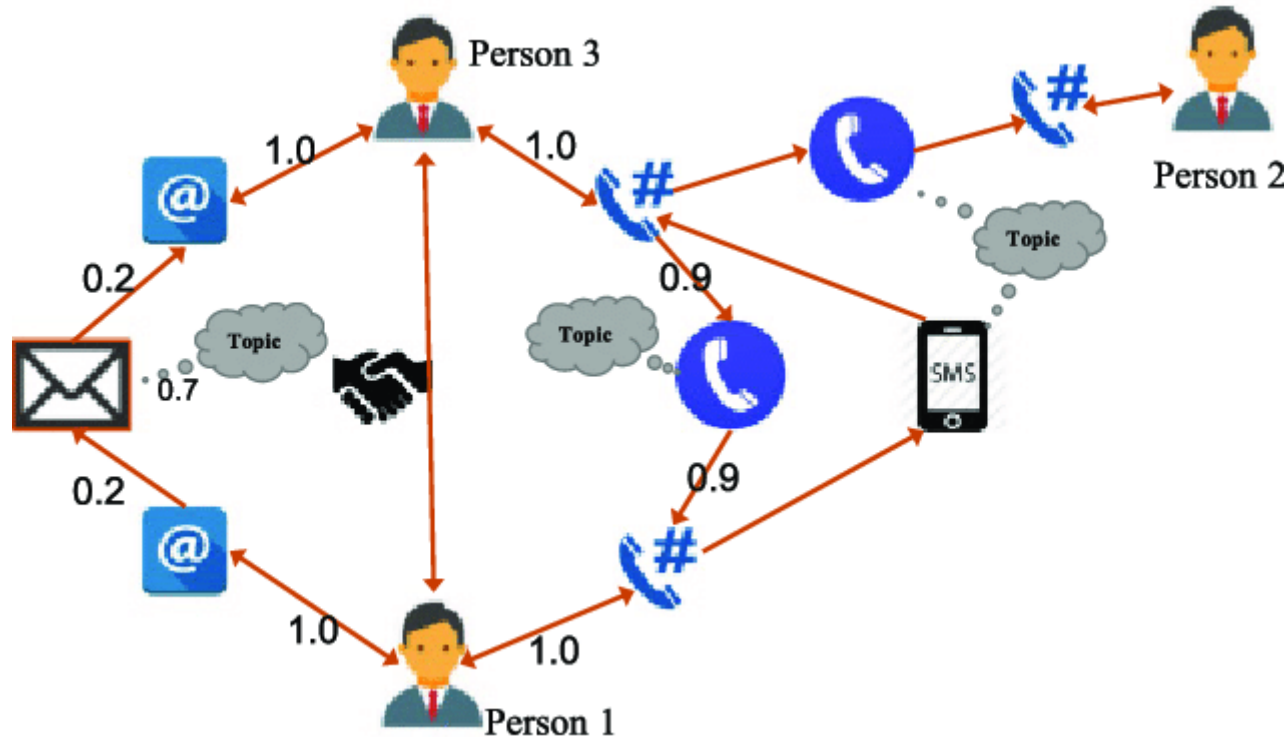
Social Network



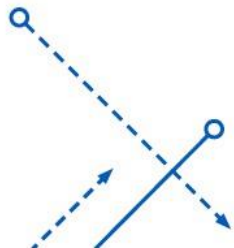
- **Vertices:** People
- **Edges:** Friendships



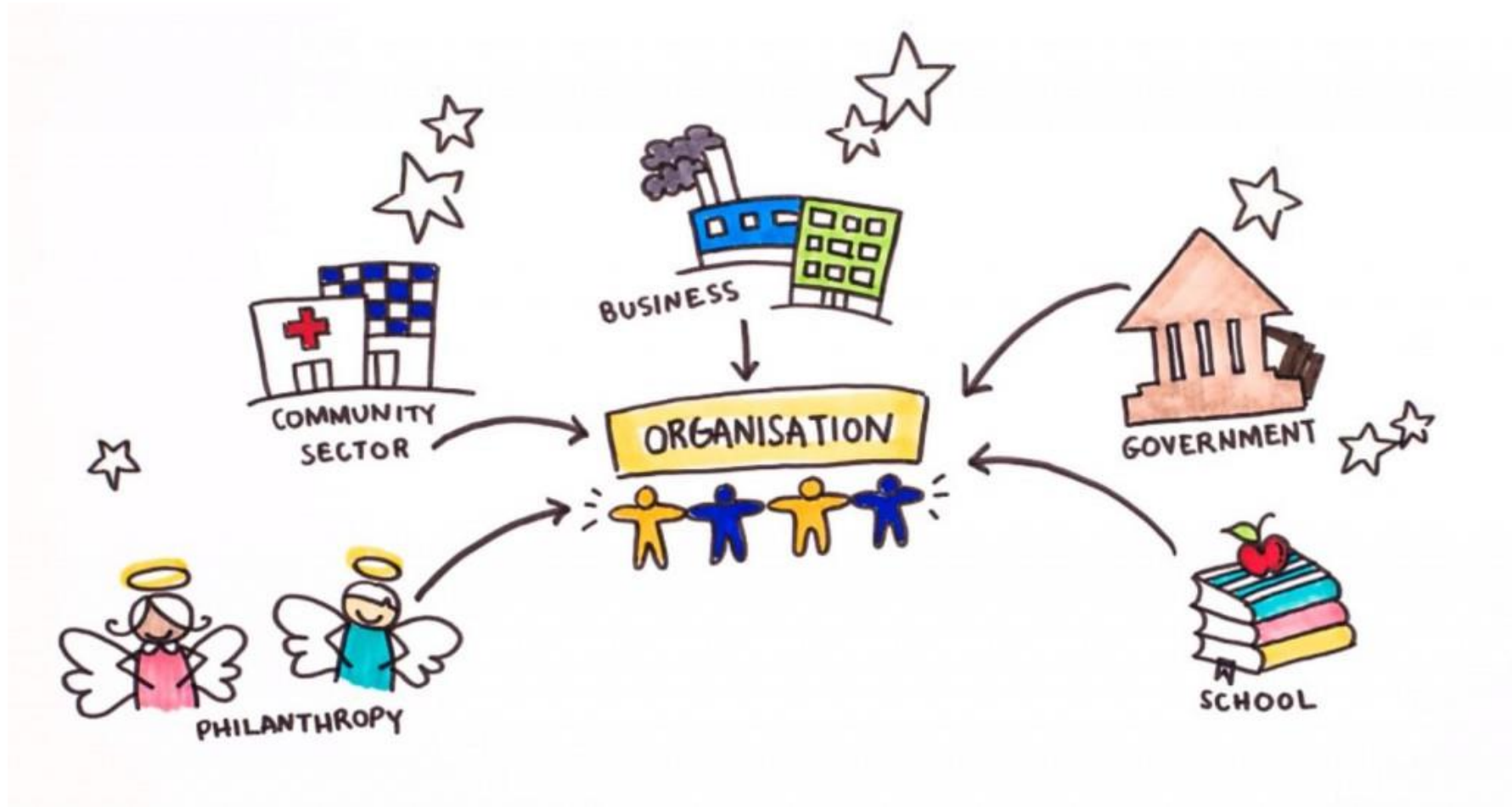
Communication network



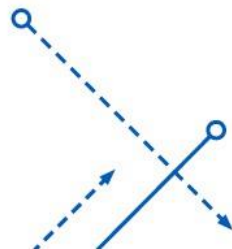
- **Vertices:** People
- **Edges:** email exchange



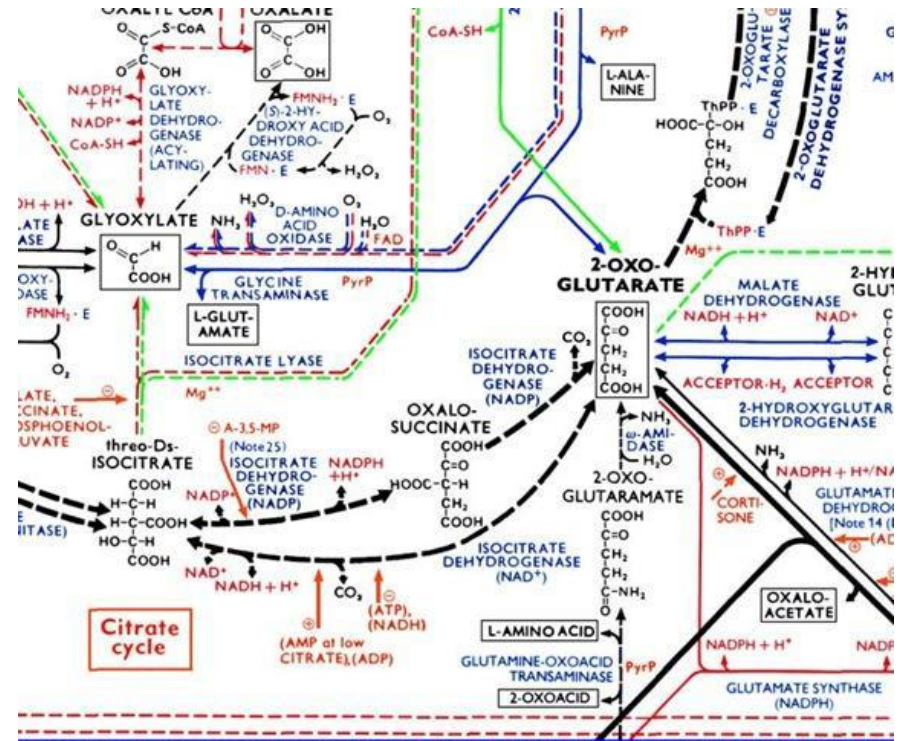
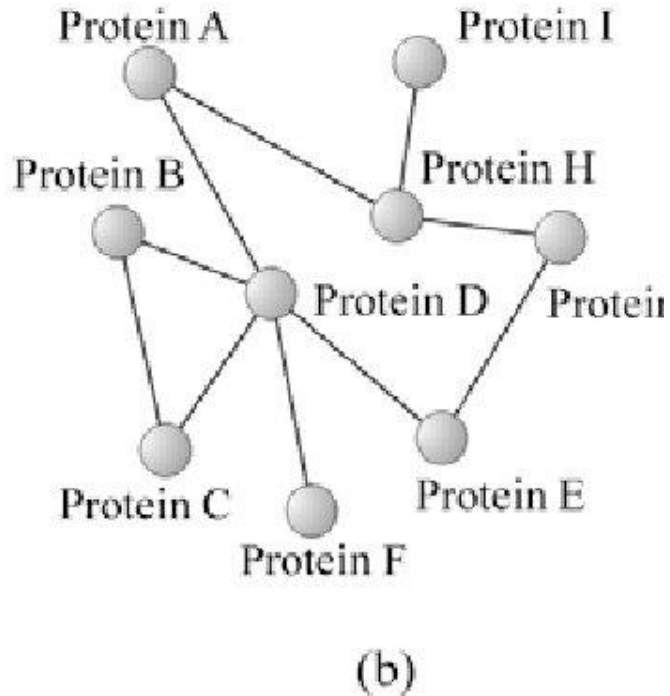
Business network



- **Vertices:** Companies
- **Edges:** relationships (financial, collaboration)



Biological network



- **Vertices:** Proteins
- **Edges:** interactions

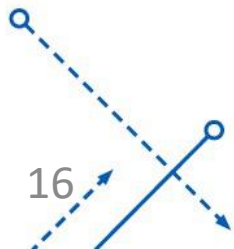
- **Vertices:** metabolites, enzymes
- **Edges:** chemical reactions



- **Vertices:** Web Pages
- **Edges:** Links

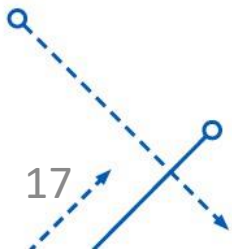
Contents

- Introduction of subjects and topics
- Rule
- Networks and graphs
- **Graph mining**
- Maths and applications



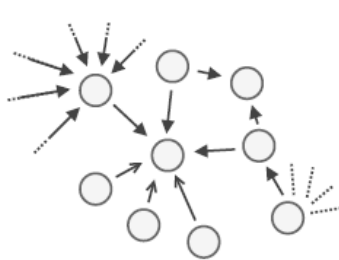
Why is network analysis important?

- The system is connected by many components, if we only focus on understanding a single individual, we cannot grasp the whole system.
- There are 2 big questions :
 - What are the structural properties of the network?
 - What interactive process is happening in the network?

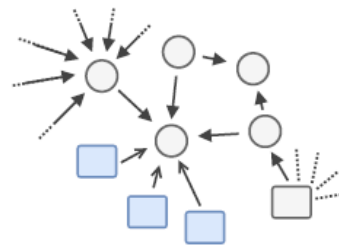


Research in the net

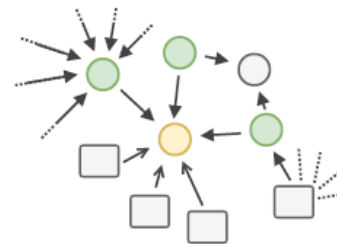
- In the field of network analysis, people focus on studying network behaviors such as human-to-human behavior in social networks.
 - Predict behavior based on its measurable properties.



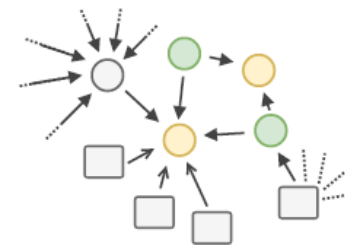
RAW Graph



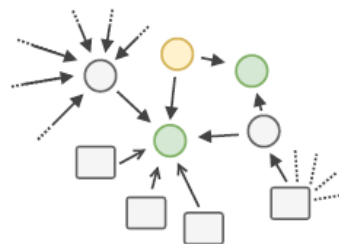
Co-Authorship
(Collaboration)



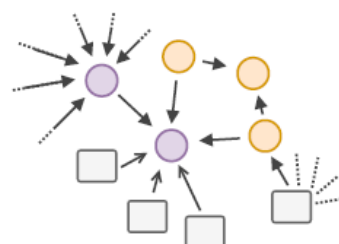
Citations



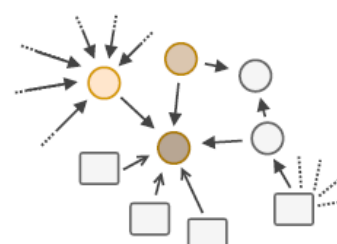
Co-Citations



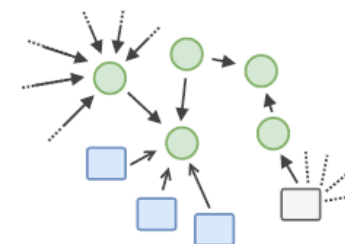
Bibliographical
Coupling



Topics



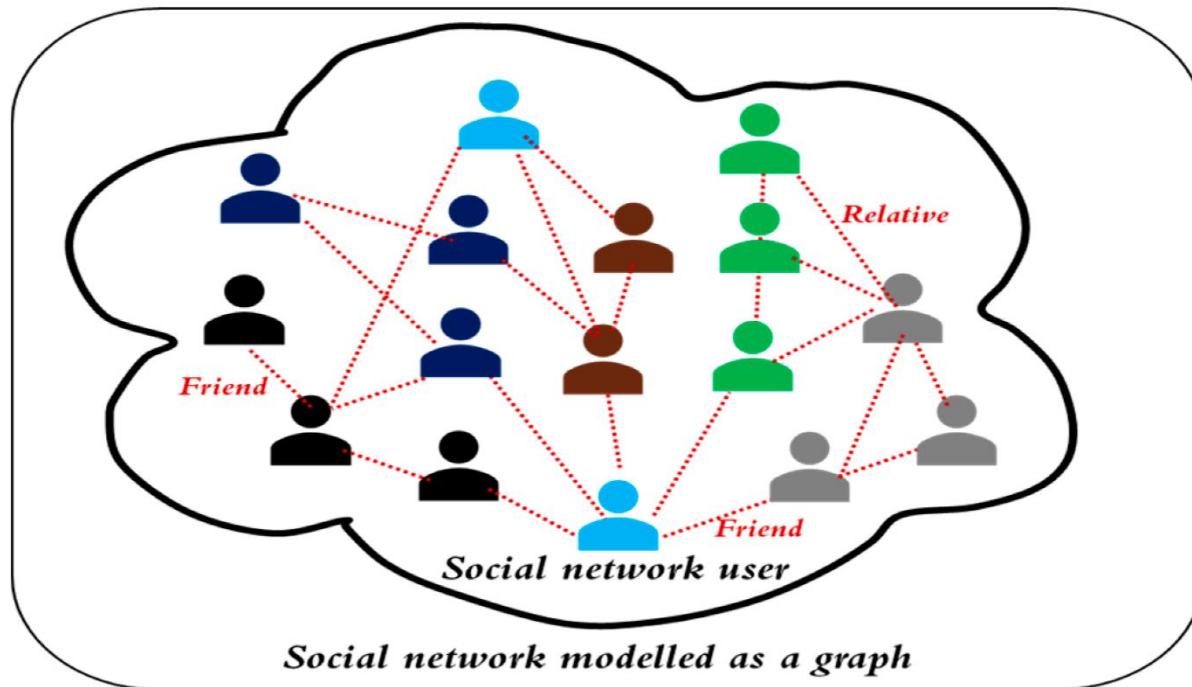
Co-Words



Heterogeneous
Networks

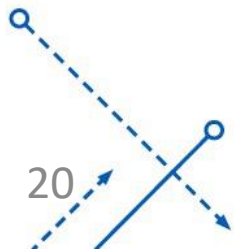
Modeling network with graphs

- Networks are not separate from graphs, they can be **re-modeled as graphs** and take advantage of its theoretical foundations.



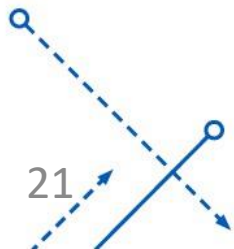
Mining Network's Challenges

- Normal graph:
 - Large size, very very large size (massive)
 - Too sparsity/ too density
 - small diameter
 - dynamic
- Requires efficient algorithms for storage and computation.



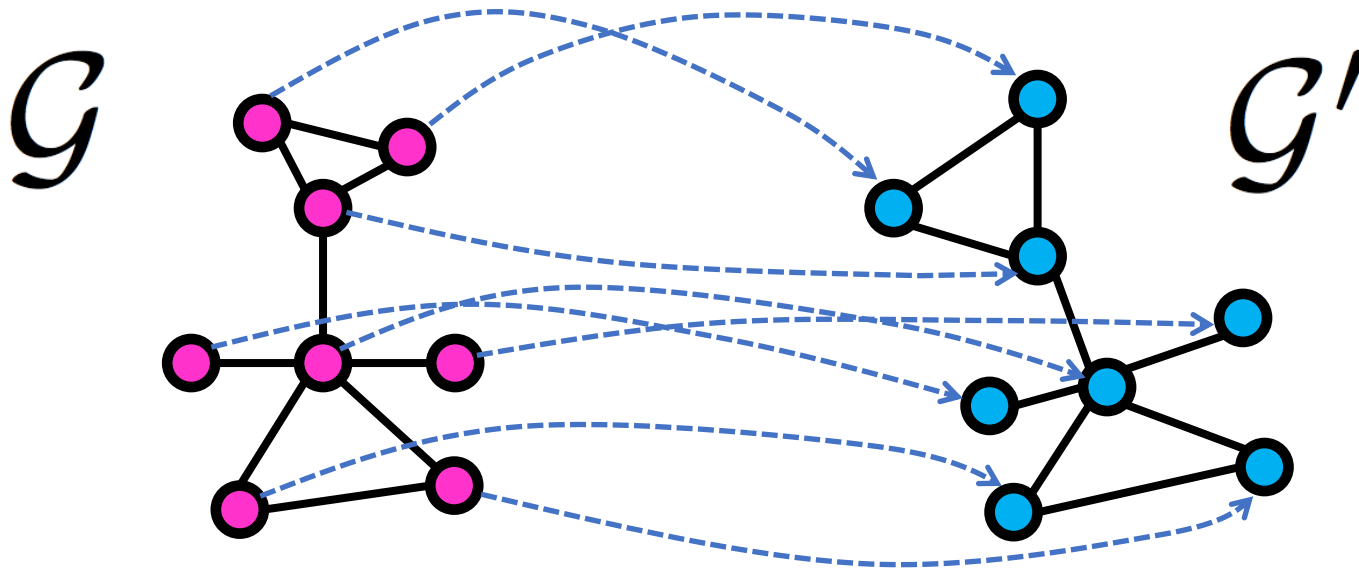
Contents

- Introduction of subjects and topics
- Rule
- Networks and graphs
- Graph mining
- Maths and applications



Maths and applications

- Graph matching

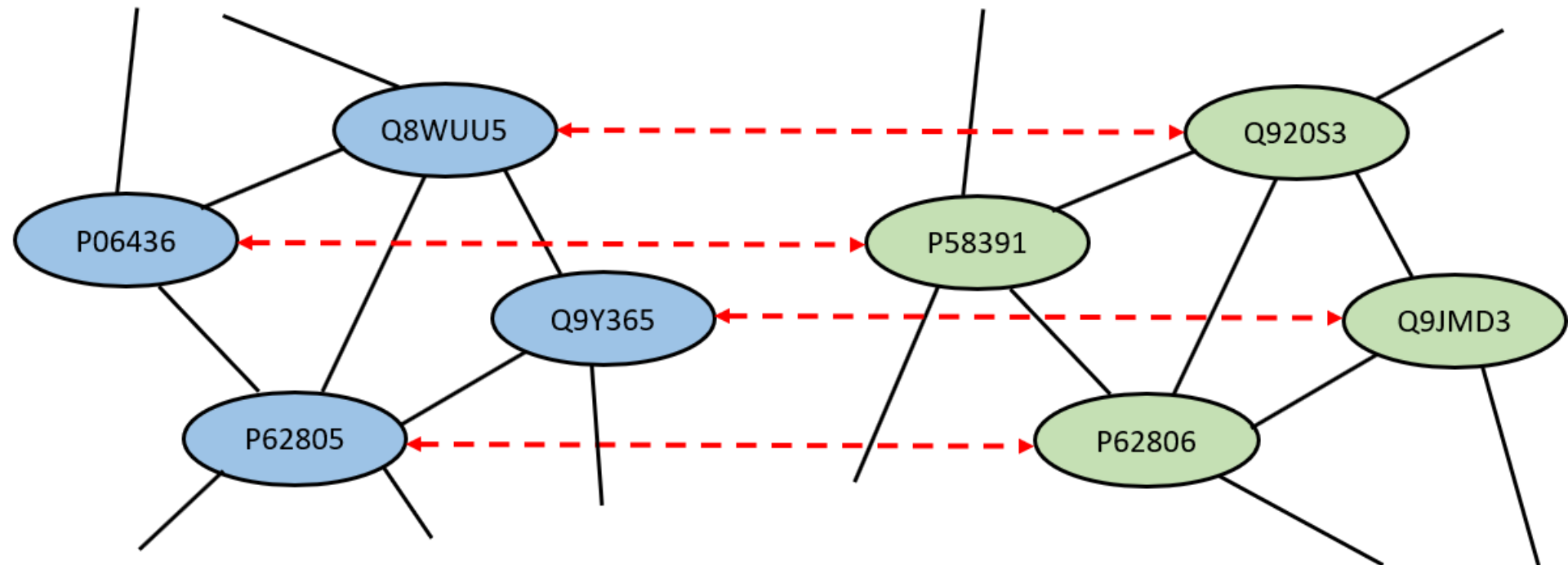


Maths and applications

- **Graph matching**

Human Network

Mouse Network

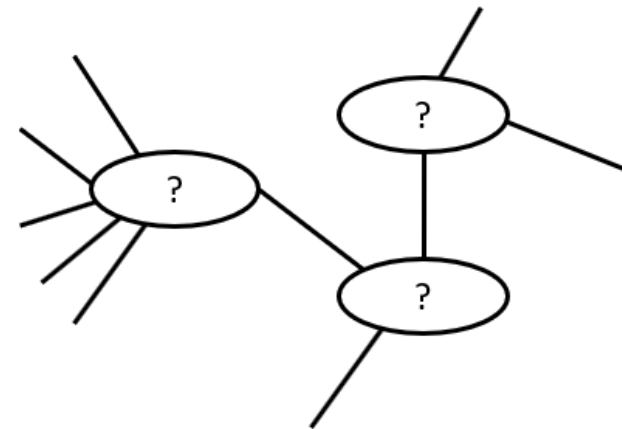
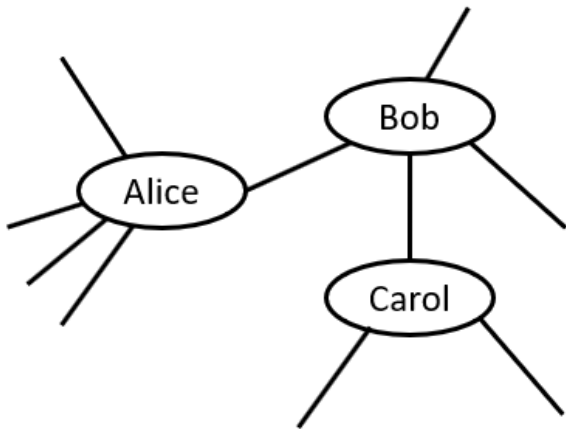


Finding proteins with the same function across different species based on their interaction networks

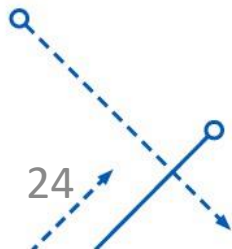
Maths and applications

- **Graph matching**

LinkedIn

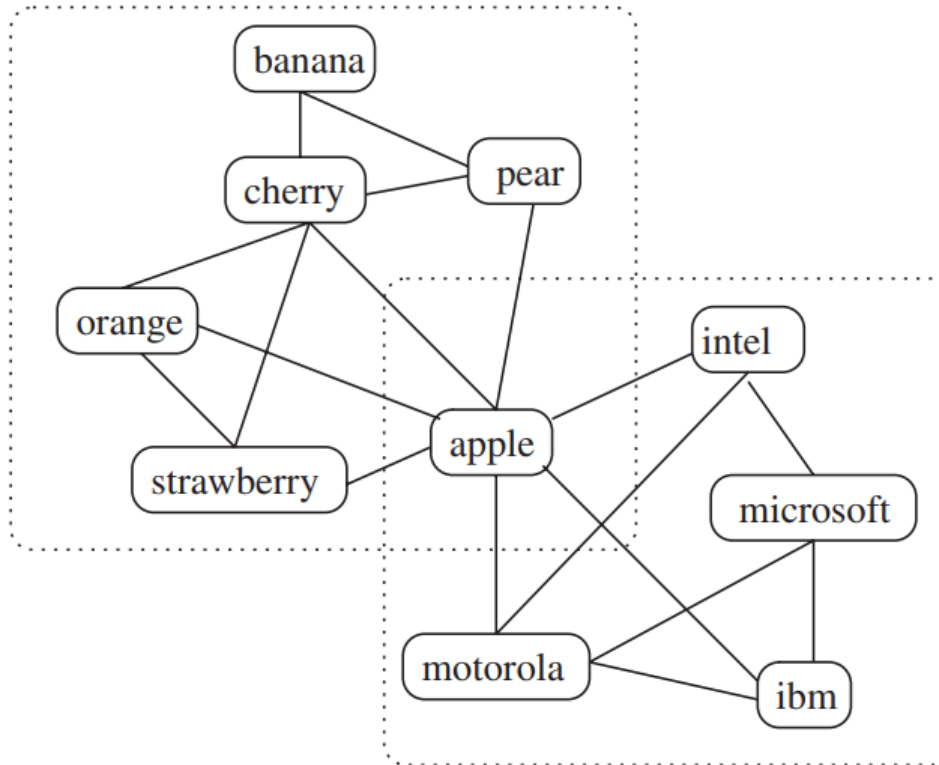


Using graphs in a network to identify hidden identities in a social network



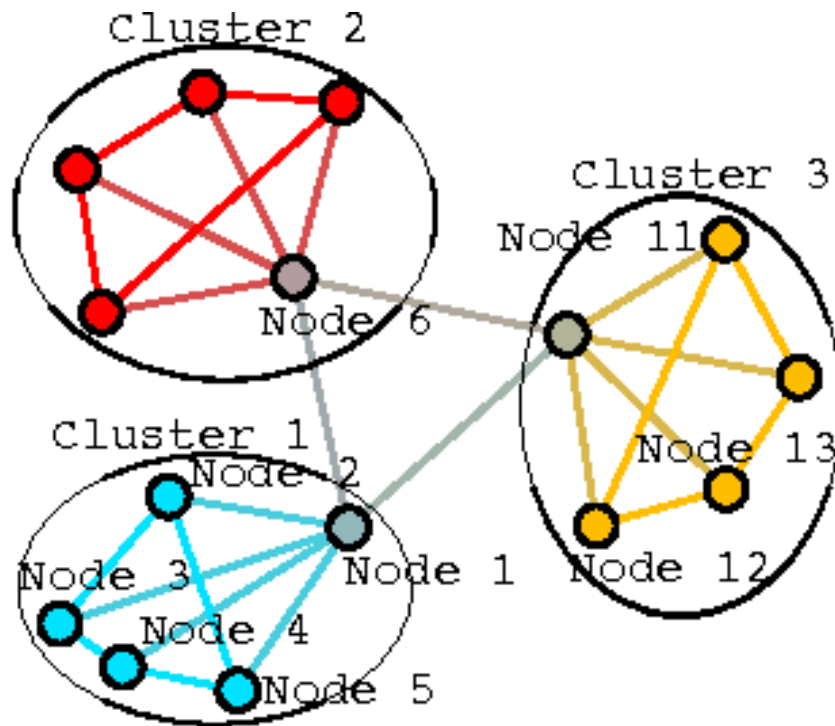
Maths and applications

- Semantic class



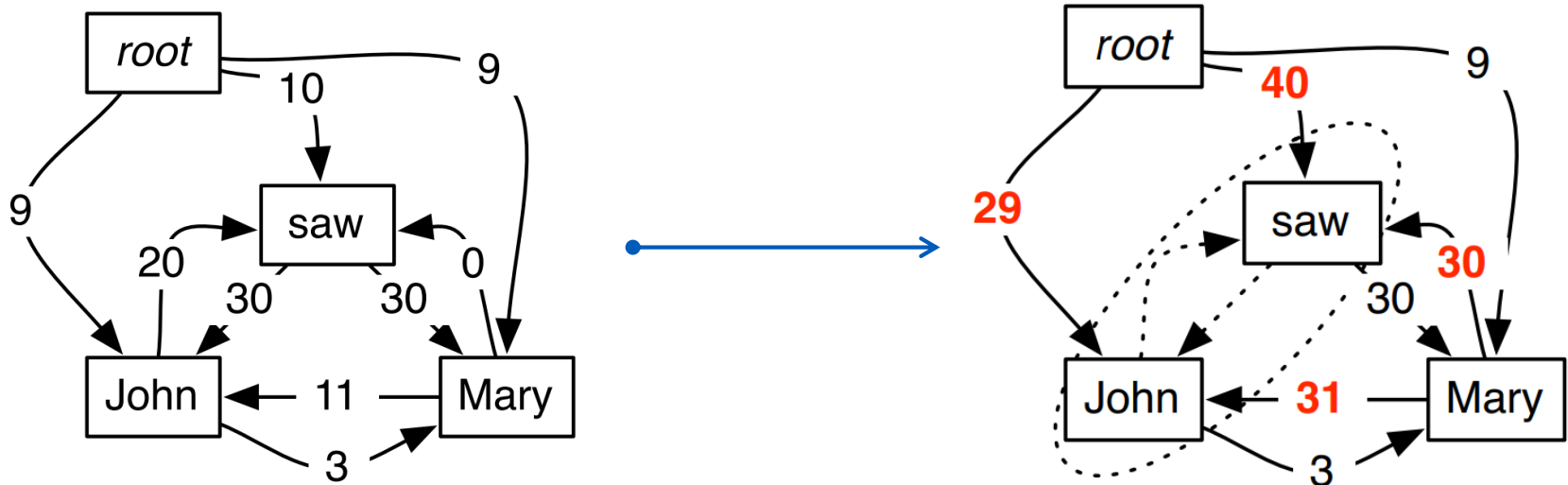
Maths and applications

- **Graph clustering**

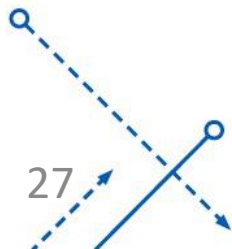


Maths and applications

- **Graph clustering**

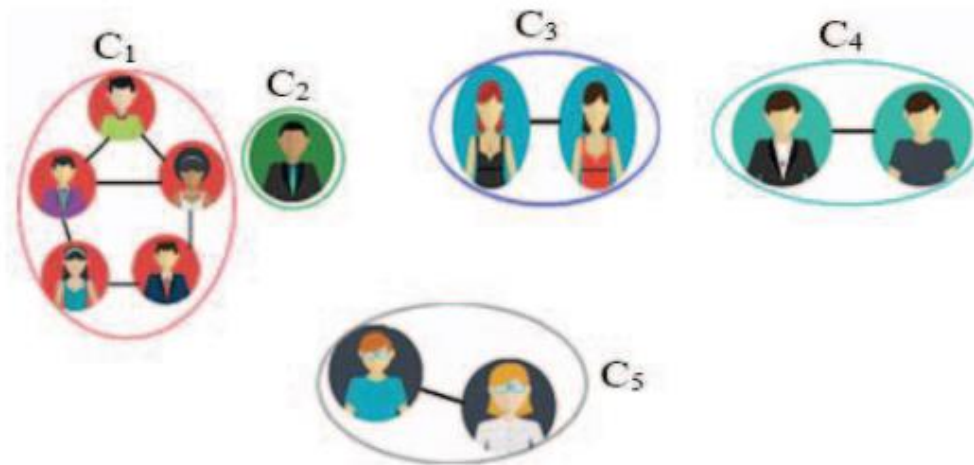


Grouping to reduce graph complexity



Maths and applications

- **Graph clustering**

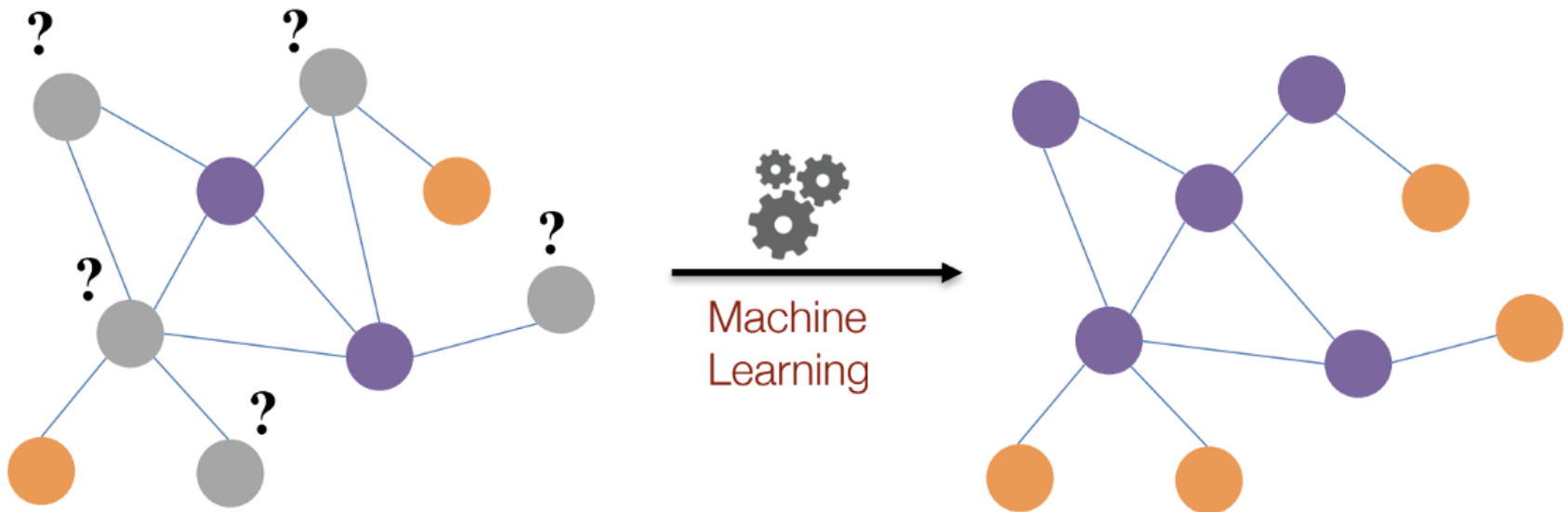


Community detection

Maths and applications

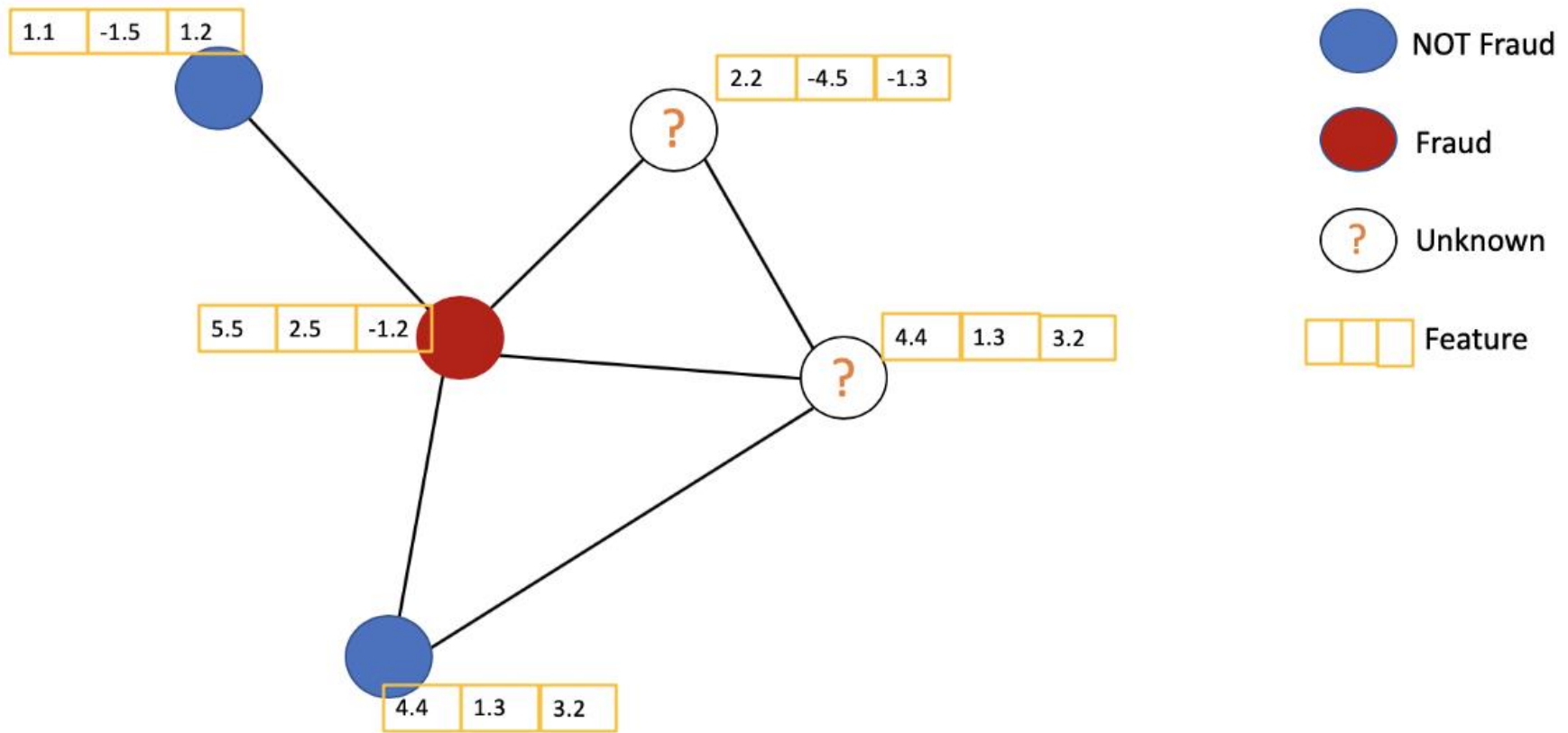
- **Graph classification**

- Labeling the top
- Labeling the link
- Labeling the graph/subgraph



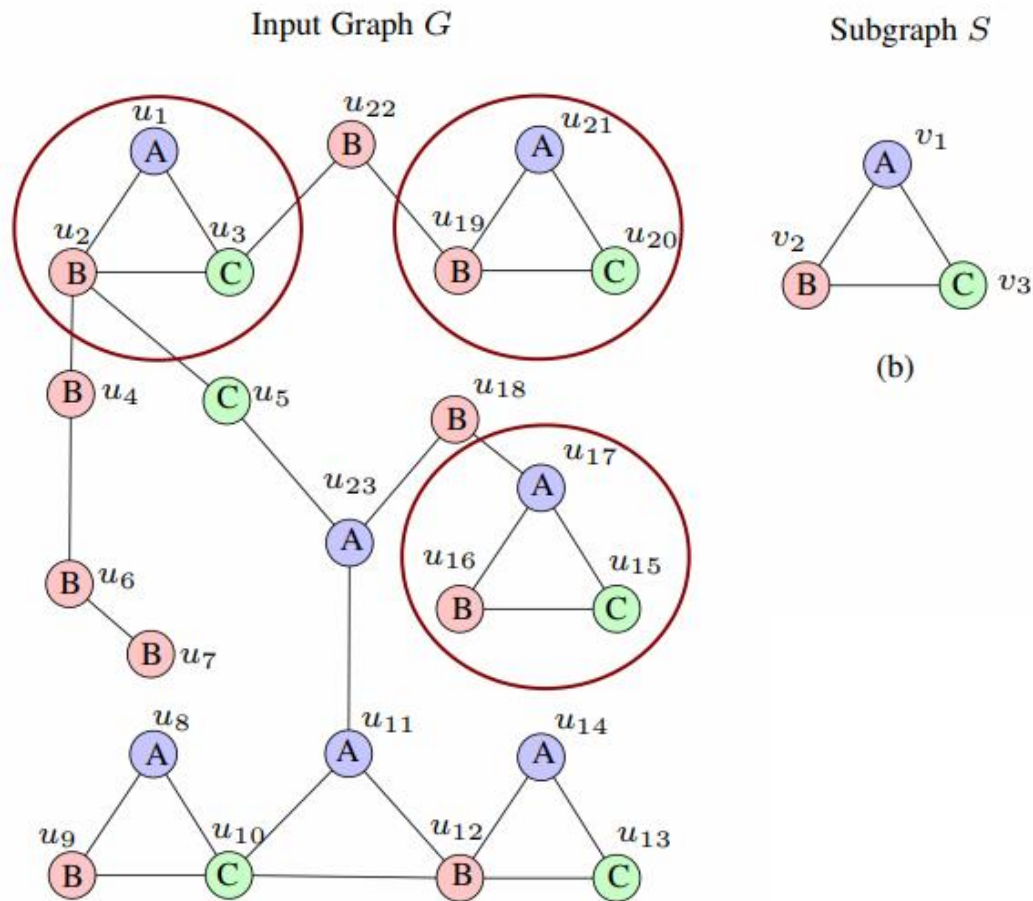
Maths and applications

- **Graph classification**



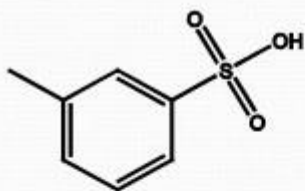
Maths and applications

- Frequent pattern mining in graph)

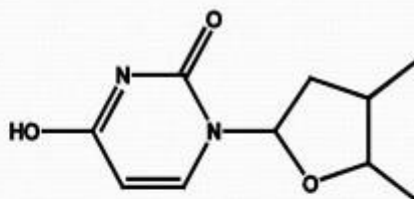


Maths and applications

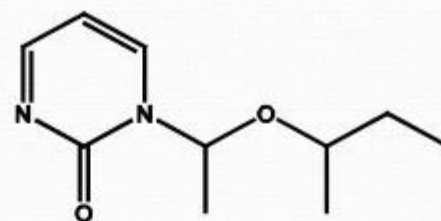
- Frequent pattern mining in graph



(A)



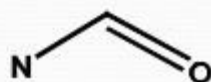
(B)



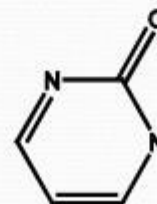
(C)

FREQUENT PATTERNS
(MIN SUPPORT IS 2)

(1)



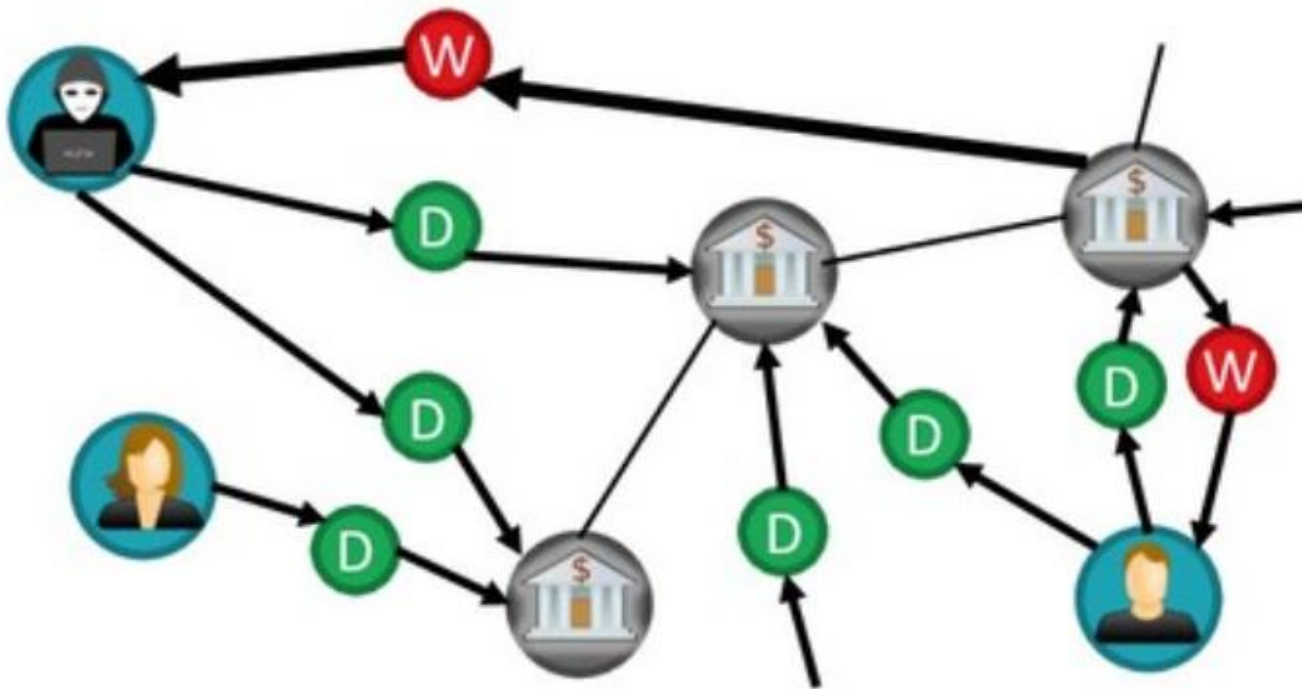
(2)



Common chemical bond string pattern

Maths and applications

- Frequent pattern mining in graph

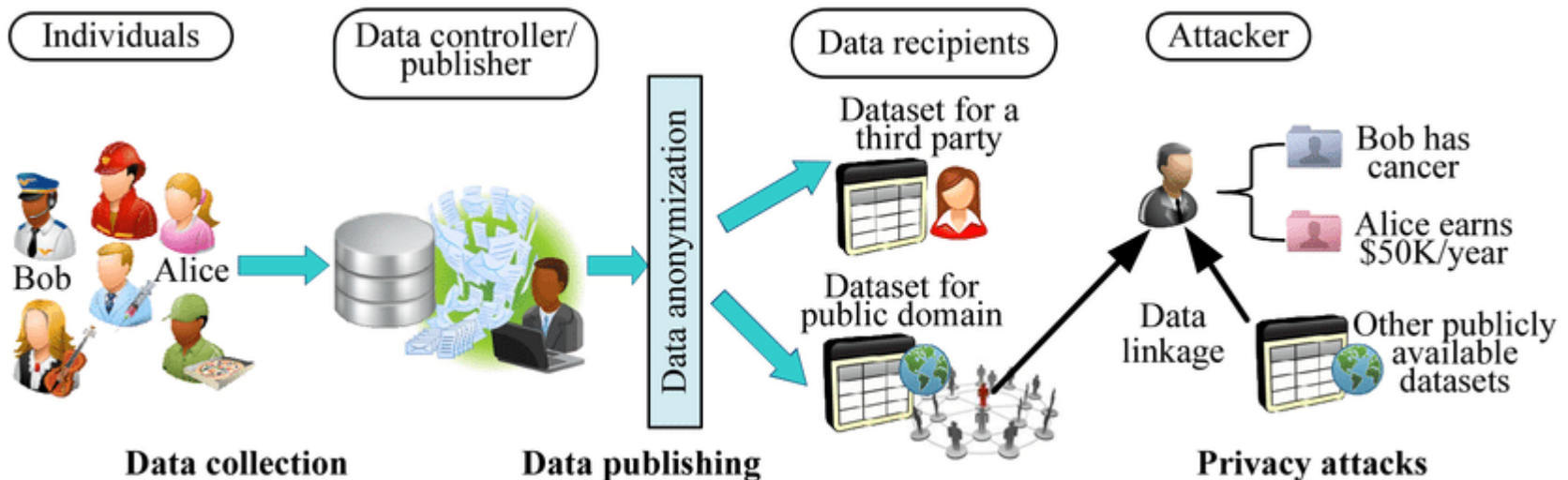


Withdrawal and deposit form

Maths and applications

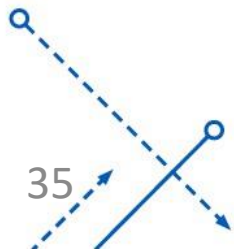
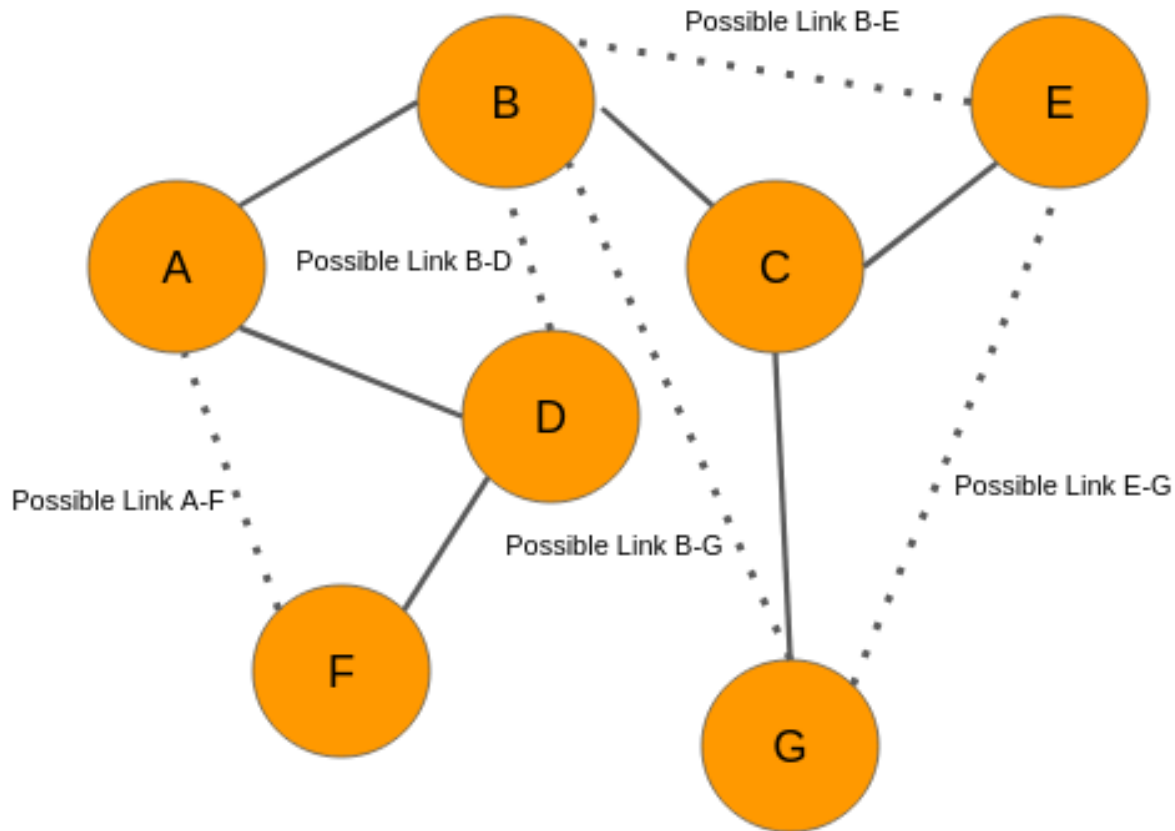
- **privacy-preserving in graph**

- It may not be enough to remove the identifying information, since the information can be interpolated from known vertices.
- How to mask identifier information without breaking the overall structure of the graph?



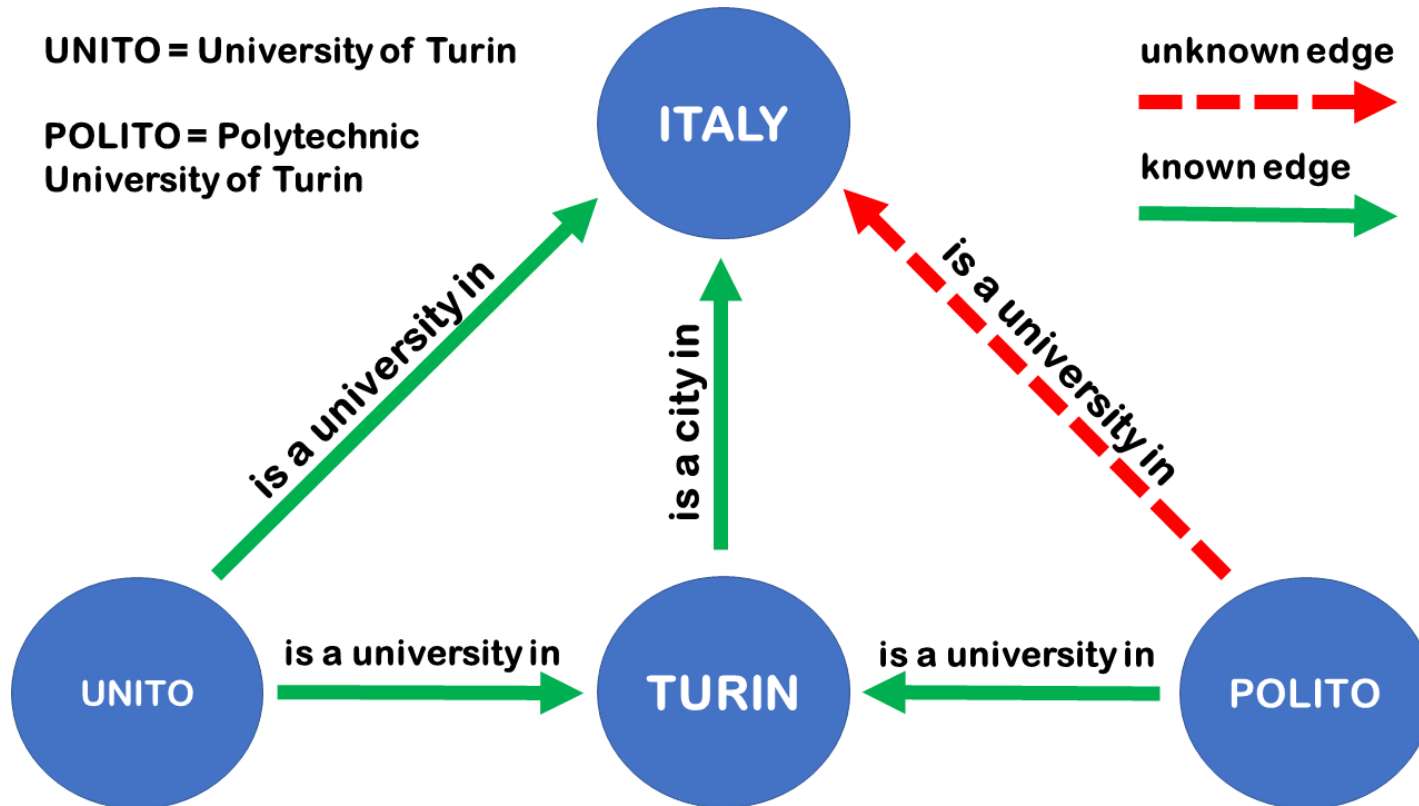
Maths and applications

- **Link prediction**



Maths and applications

- Link prediction



Maths and applications

Graph analysis task	Application	Field
Graph clustering	Data storage Data compression	Database systems
	Popularity prediction Tag recommendation	Social network analysis
	Substructure identification Network usage optimization	Computer networks
Graph matching	2D,3D Image analysis Face recognition Face verification Object registration/retrieval	Computer vision
	Document analysis	Language engineering
	Molecular structure study	Computational chemistry
Random walks	Enumeration	Multiple
	Volume computation	Computational geometry
	Mobile agent modelling	Distributed systems
	Web crawling	Internet computing
Anomaly detection	System intrusion detection Network attack detection	Computer security
	Financial fraud detection	Law enforcement
	Influential individual detection	Social network analysis

References

- Aggarwal, Charu C., and Haixun Wang, eds.
Managing and mining graph data

