

Data Warehousing

Lecture 7 – Intro to Graph and Graph Databases

CITS3401
CITS5504

Yuanyi Luo

Computer Science and
Software Engineering

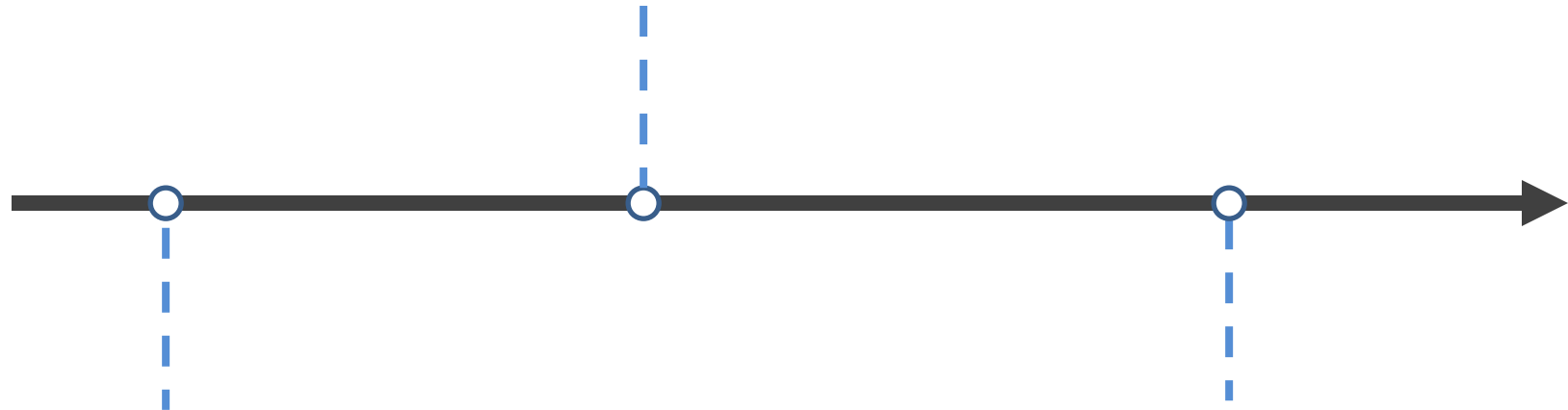
School of Maths, Physics
and Computing

Acknowledgement: The lecture slides are adopted from online sources.

- **Yuanyi Luo**
- **Fourth-year PhD candidate**
- **Interest in multimodal machine learning and pattern recognition**
- **Researching on multimodal graph representation**

Three Questions of Today's Lecture

How to manage these
graphs composed of
large datasets?

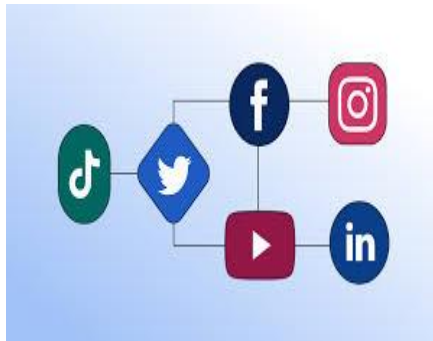


What should we do if we
want to summarize and
clarify many complex
relationships in our lives?

How to introduce graph
into multimodal machine
learning?

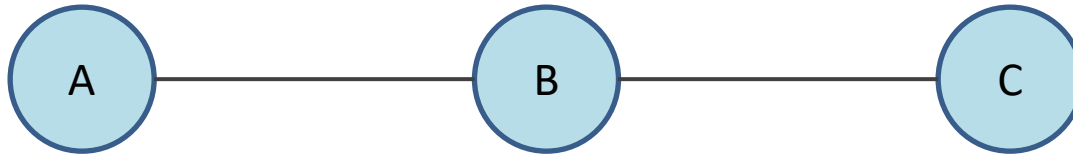
Unstructured
data

Telling stories



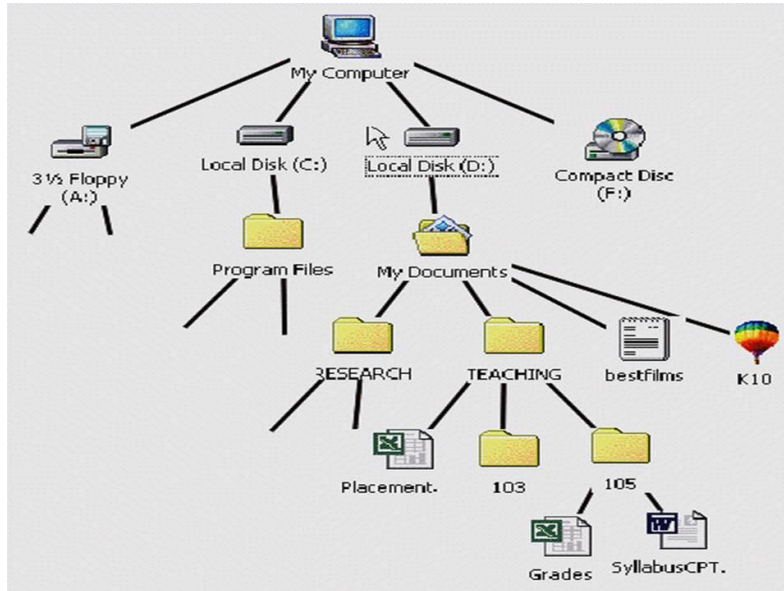
FOOD MENU	
Paucek and Lage Restaurant	
MAIN COURSE	
Cheesburger	\$34
Cheese sandwich	\$22
Chicken burgers	\$23
Spicy chicken	\$33
Hot dog	\$24
APPETIZERS	
Fruit Salad	\$10
Cocktails	\$12
Nuggets	\$14
Sandwich	\$10
French Fries	\$5
BEVERAGES	
Milk Shake	\$3
Iced Tea	\$2
Orange Juice	\$4
Lemon Tea	\$3
Coffee	\$5
 08 9444 1000  122 Anson Road, City	

Linear structure

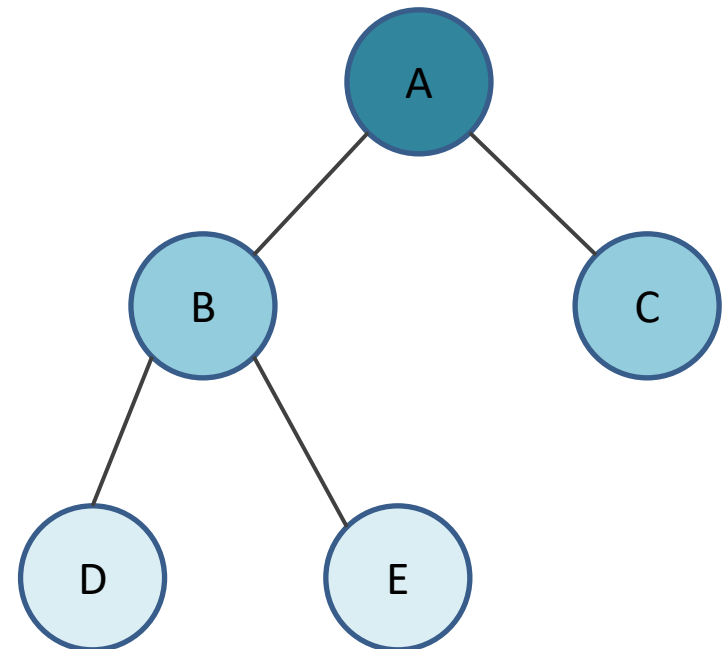


- Linear structure

Tree structure



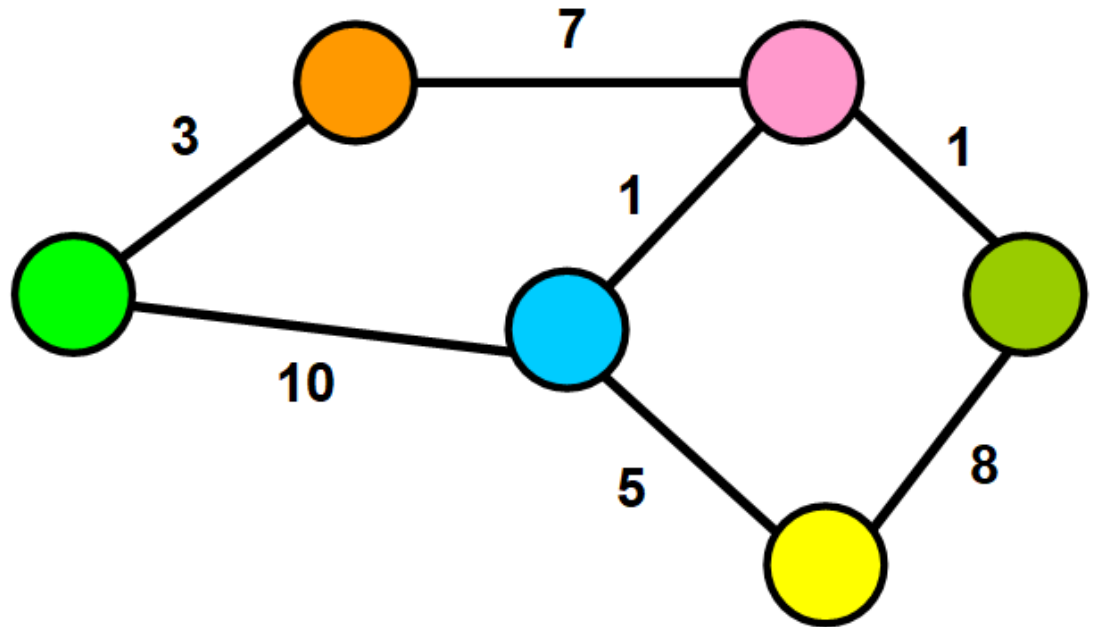
Hierarchical
data



- Tree structure

Graph structure

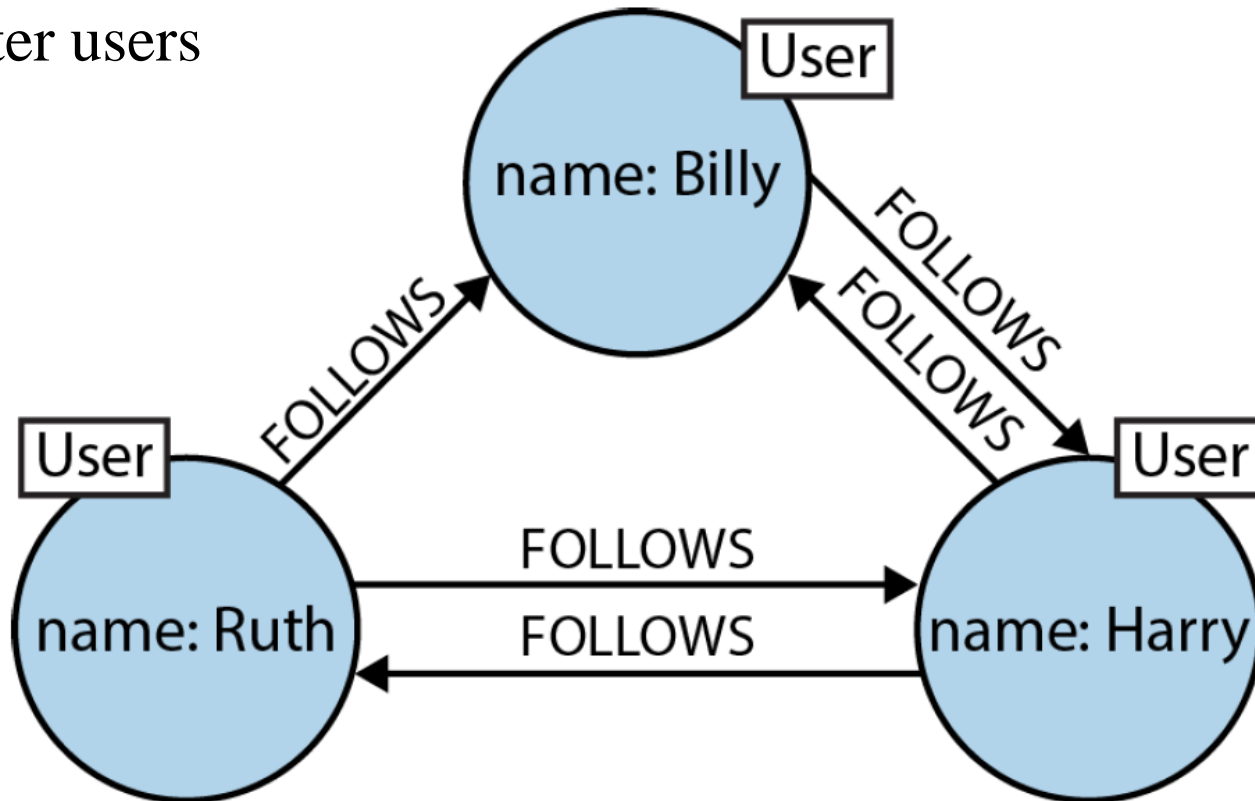
- Complex networks with multiple interconnections
- Graph provides us with a convenient way to make choices
- Graphs aren't bound by rigid structural principles
- Graphs serve as the ideal tool for visualizing these connections in a clear and comprehensive manner



- Graph structure

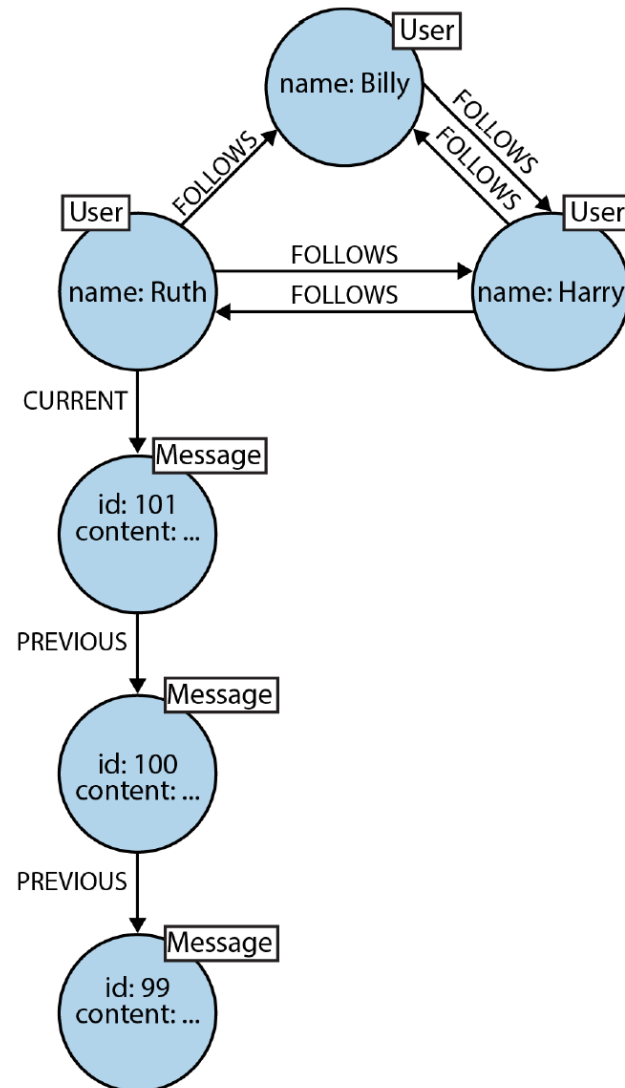
Graph example - Twitter

- Small network of Twitter users



Graph example - Twitter

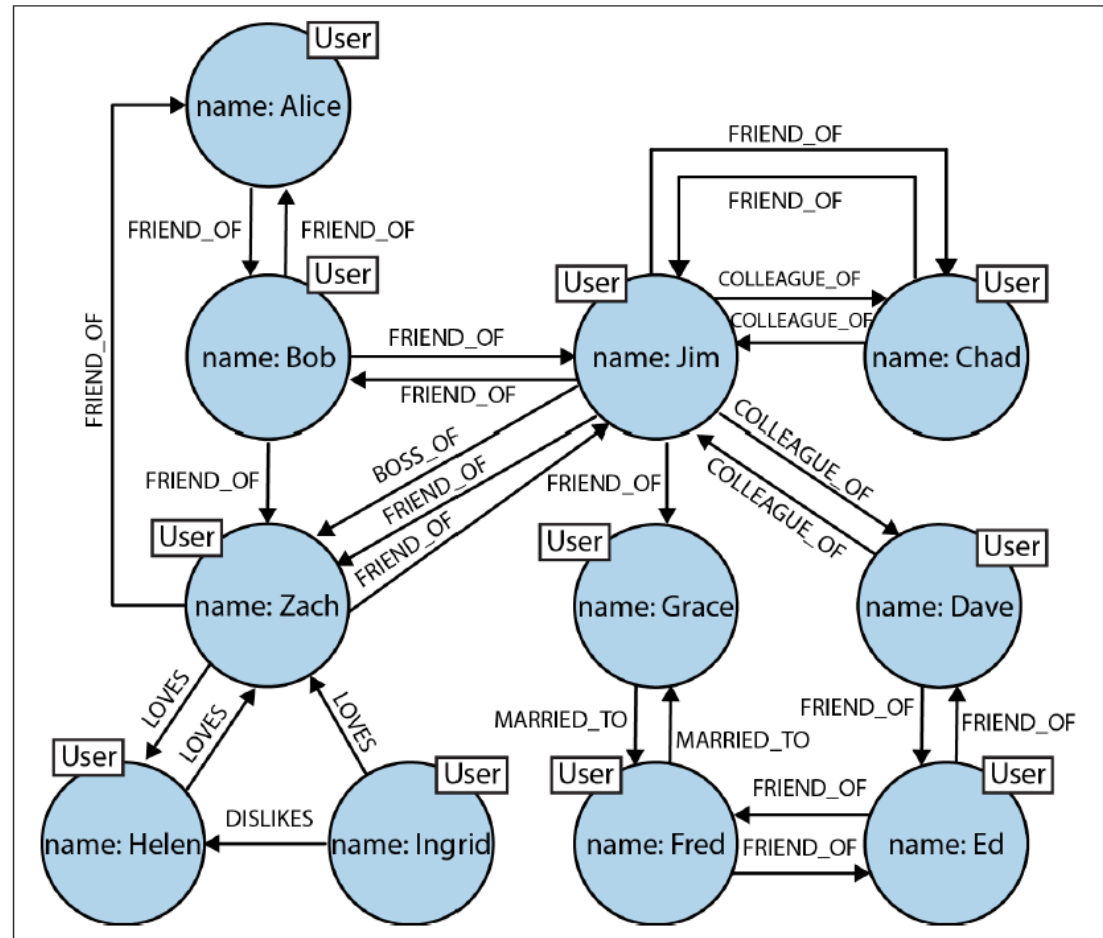
- Add new nodes without any rigid structured principles
- Expand the graph to include the messages



- The most common form of graph model is the **property graph model**, whereby:
 - The graph contains **nodes** and **relationships**.
 - A node may have zero or more **properties** (key-value pairs).
 - Nodes can be labelled with one or more **labels**.
 - Relationships can be **named** and **directed**, and always have a start and end node.
 - Relationships can also contain **properties**.

Labels and Relationships

- The graph on the right is a more complex example of a social network.
- Note that the relationships between entities **do not exhibit uniformity**.
- New nodes and relationships were added **without compromising the existing network** or migrating data (*flexibility*).



Graph for searching

Connected Papers: a graph-based tool for finding scientific papers



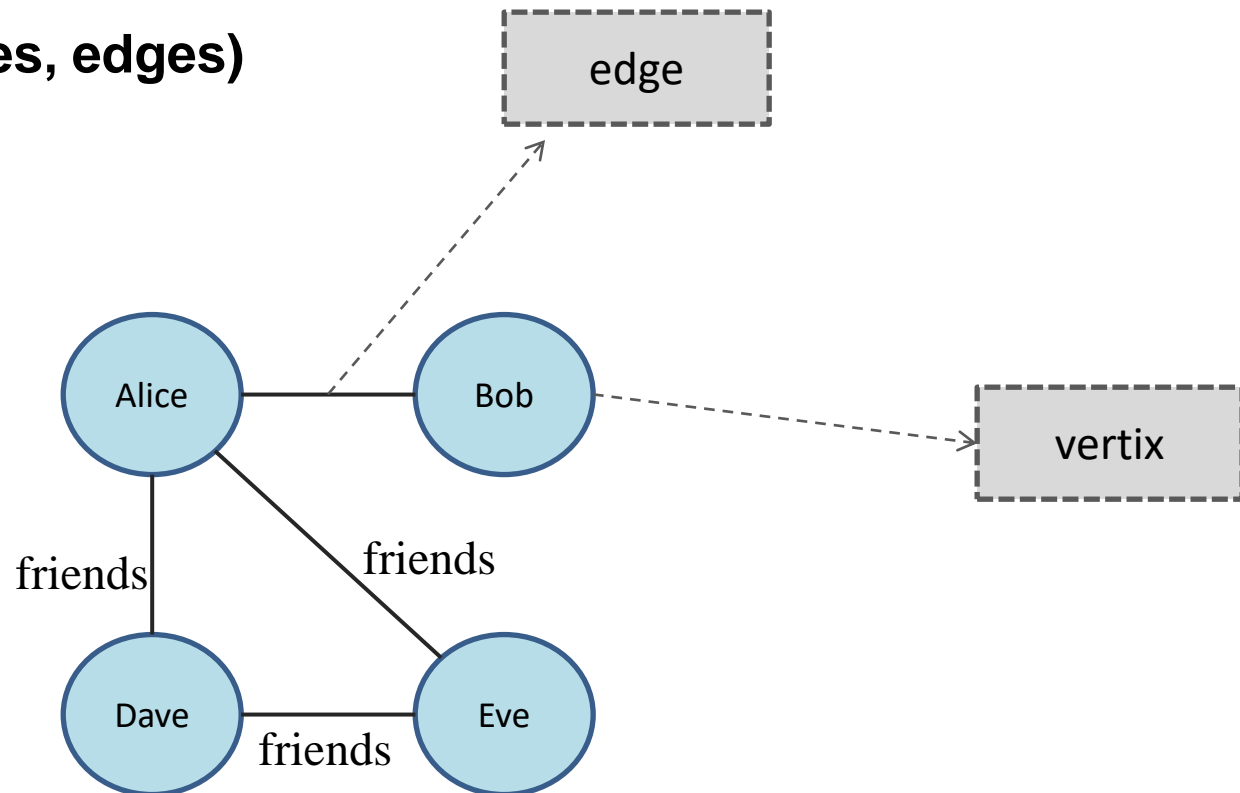
Answer the first question

What should we do if we want to summarize and clarify many complex relationships in our lives?

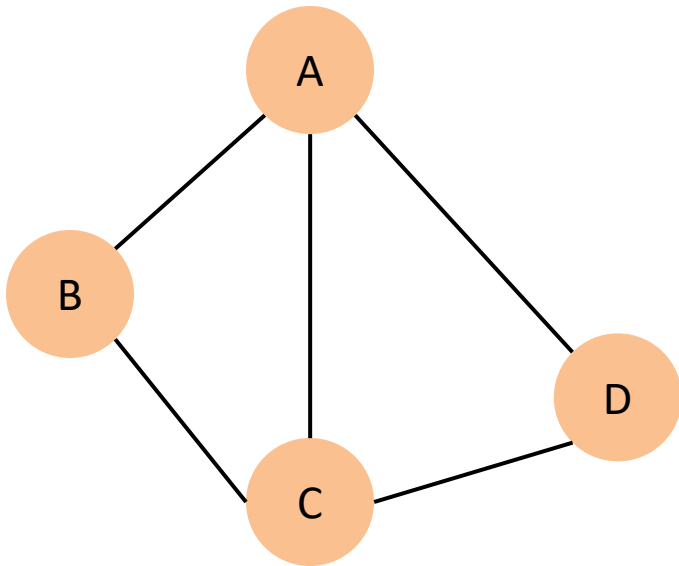
The definition of a graph

- A graph is a collection of **vertices** and **edges**, also known as **nodes** and **relationships**.

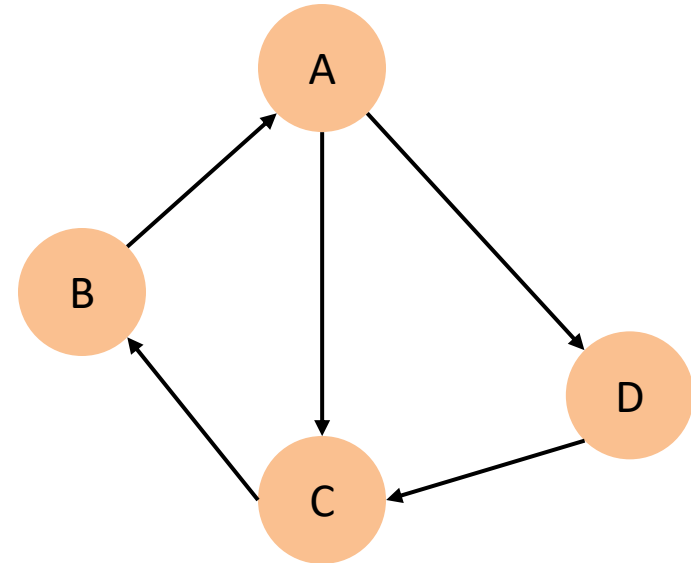
Graph = (vertices, edges)



The definition of a graph

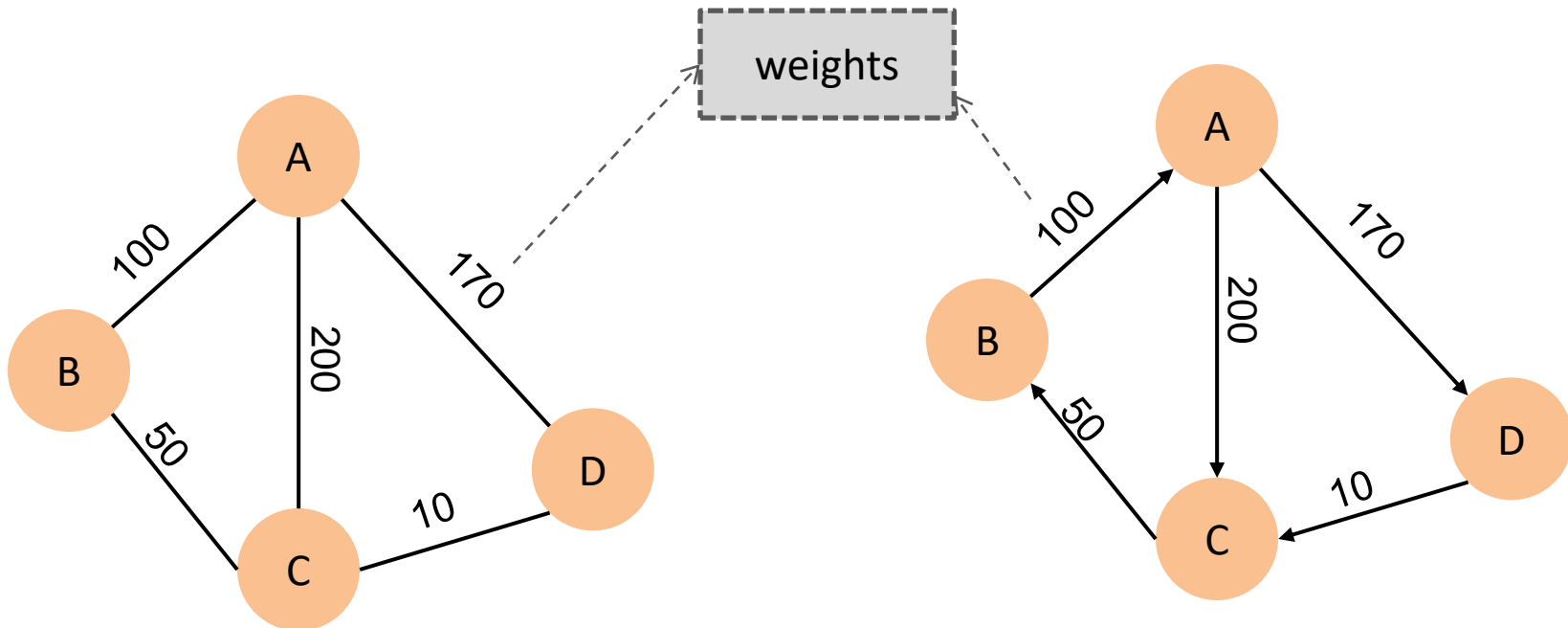


- Undirected graph



- directed graph

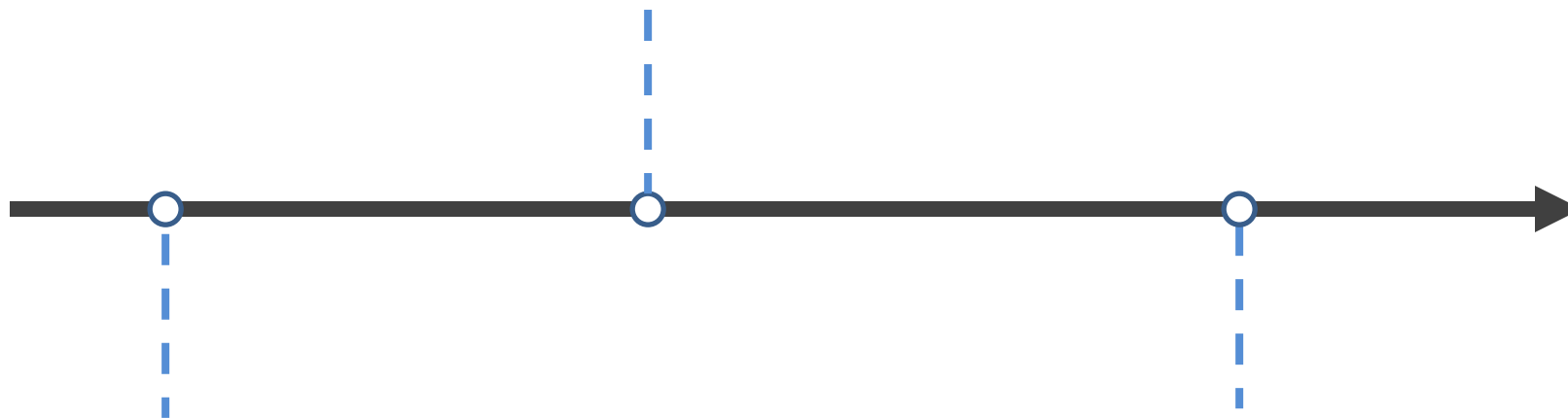
The definition of a graph



- different weights of graphs have different means

Second Questions of Today's Lecture

How to manage these
graphs composed of
large datasets?

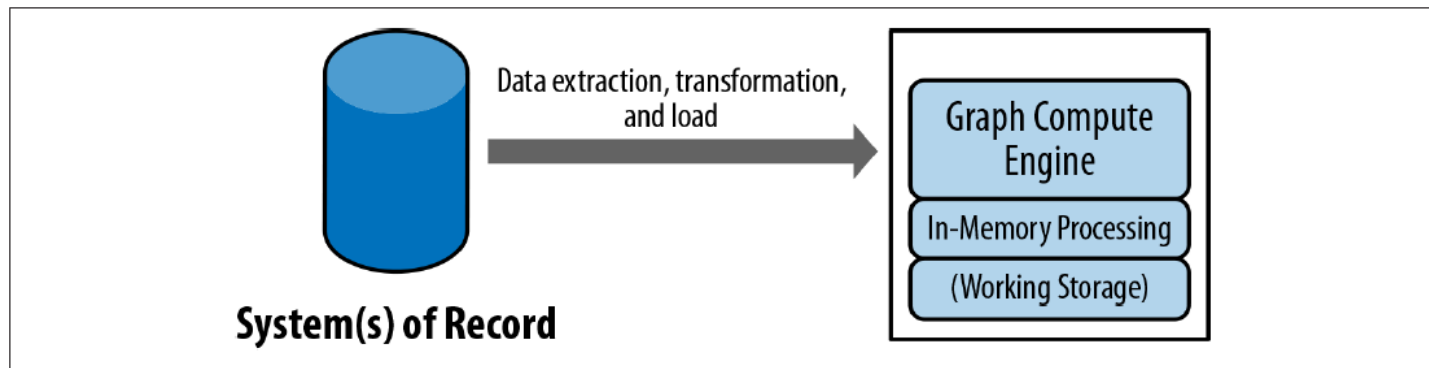


What should we do if we
want to summarize and
clarify many complex
relationships in our lives?

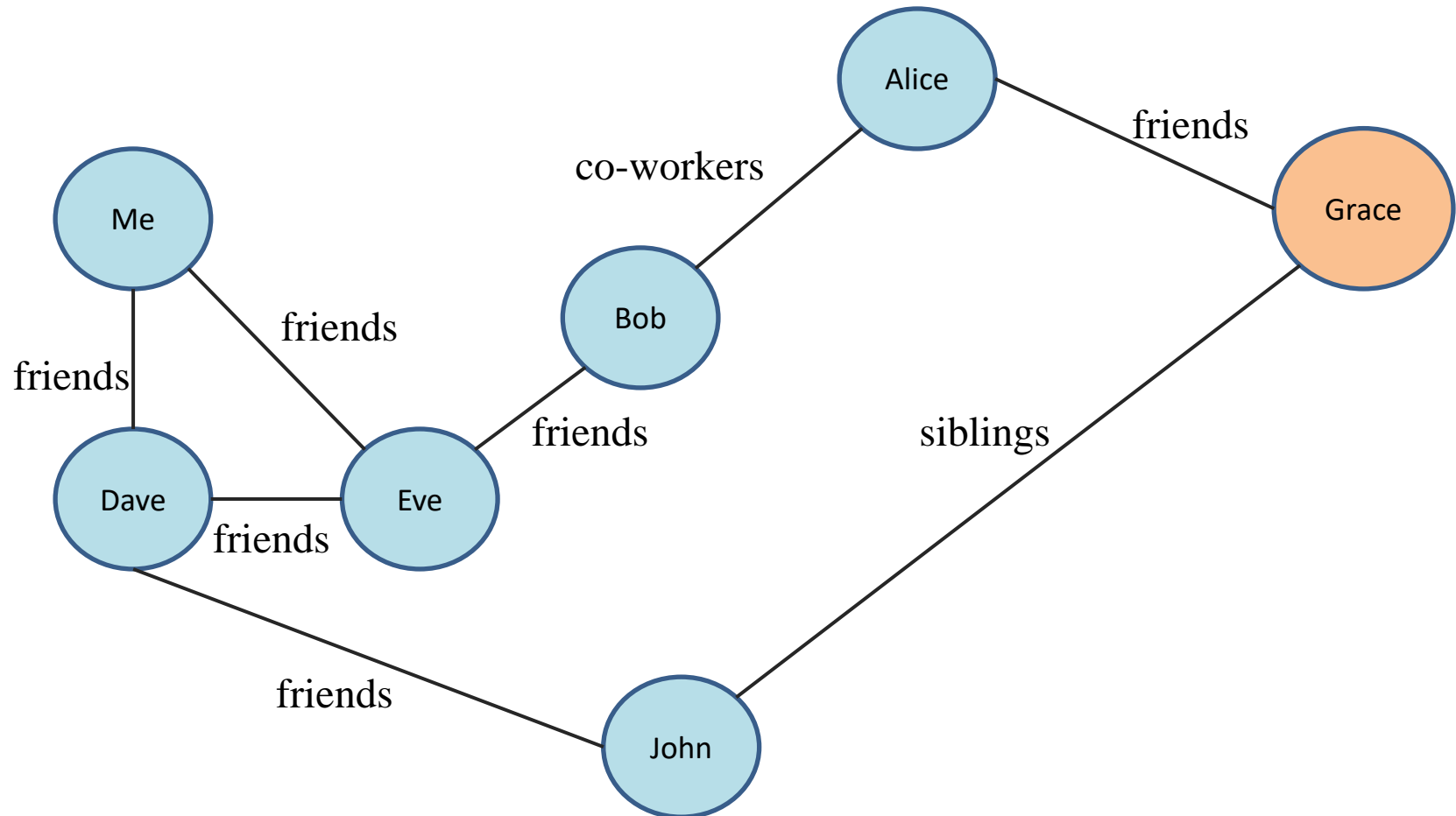
How to introduce graph
into multimodal machine
learning?

- A graph database management system features a CRUD (Create, Read, Update, Delete) interface.
- They are built for use with OLTP (online transactional processing) systems.
- Graph databases have two key properties:
 - The **underlying storage** (*native graph storage vs serialized storage*).
 - The **processing engine** (most Graph DBs feature *index-free adjacency*, meaning nodes point to each other in the underlying database).

- A graph compute engine enables **global graph computational algorithms** to be run against large datasets.
- They are optimised for processing information in batches, similarly to OLAP (online analytical processing).



Graph structure



Why Graph Databases?

1 - Performance

- Graph databases have excellent performance on queries involving highly-connected data.
- The execution time for each query is proportional only to the size of the **part of the graph** traversed to satisfy the query, rather than the size entire graph.

Depth	RDBMS execution time(s)	Neo4j execution time(s)	Records returned
2	0.016	0.01	~2500
3	30.267	0.168	~110,000
4	1543.505	1.359	~600,000
5	Unfinished	2.132	~800,000

Above: an experiment finding friends of friends in an RDBMS vs in Neo4j.

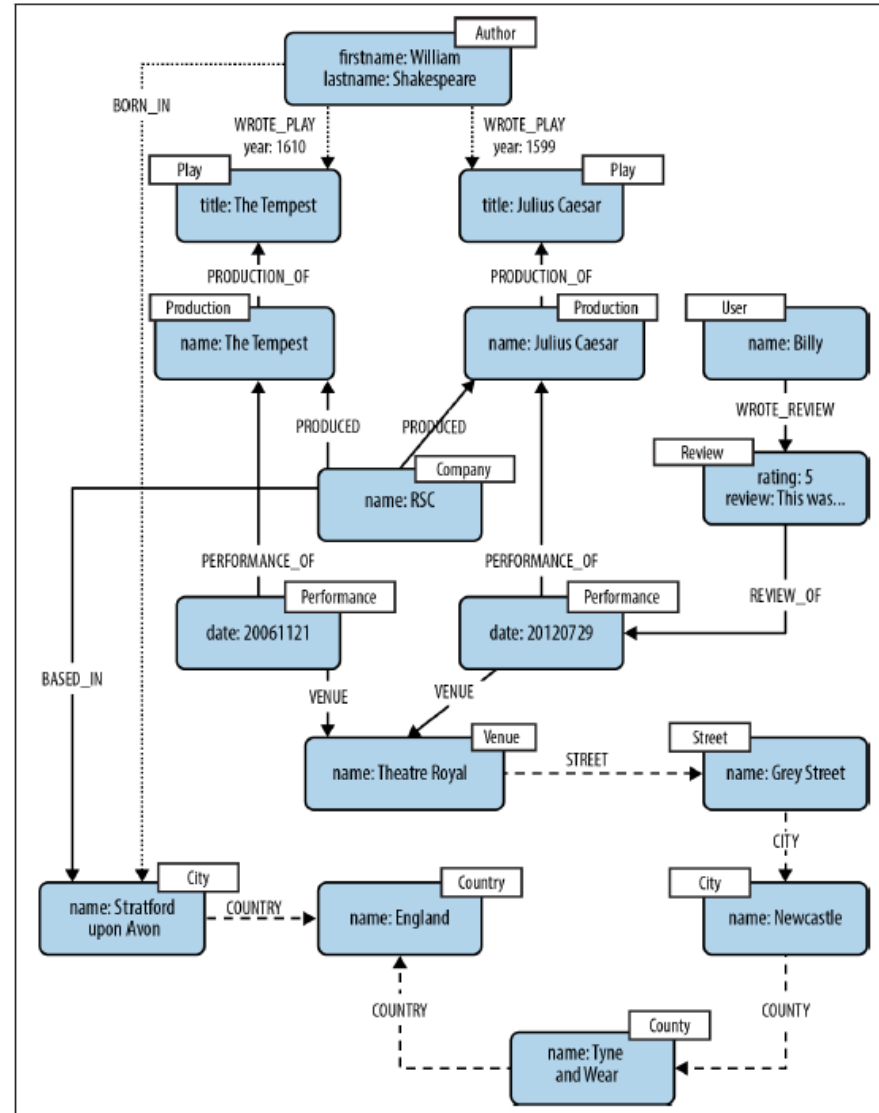
2 - Flexibility

- We often need to modify an existing database, such as capturing a new type of relationship between two entities, or adding a new property.
- Graphs are **naturally additive**, meaning we can add new kinds of relationships, new nodes, labels and subgraphs without affecting existing queries.
- This means we don't have to model our domain ahead of time, and can update our graph ad hoc.

Why Graph Databases?

2 - Flexibility

Graph databases are excellent at combining data from across multiple domains:

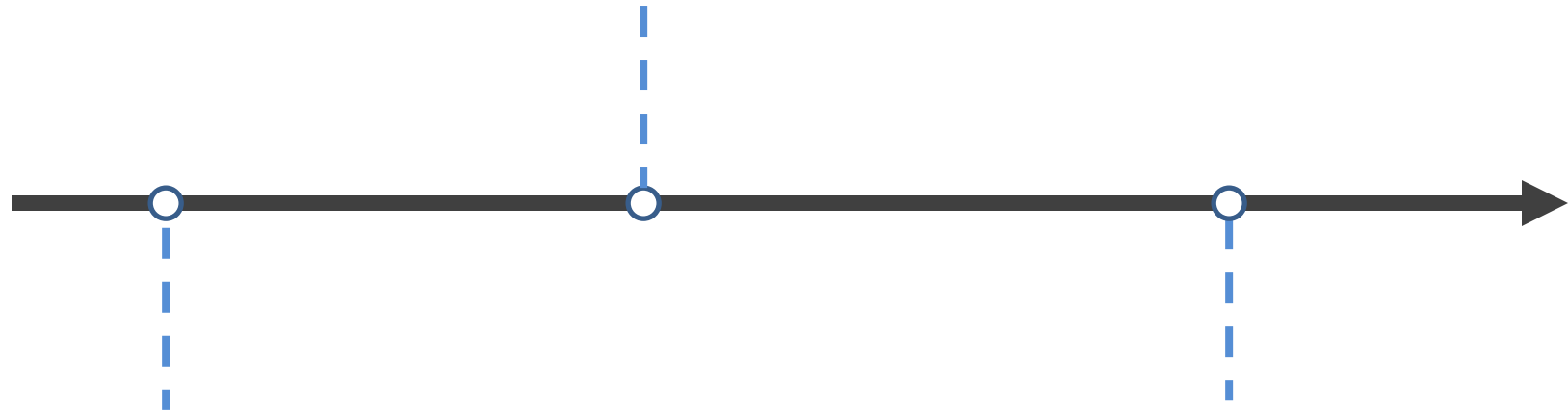


3 - Agility

- Graph databases are **schema-free**, rapidly speeding up development.
- Graph databases **do not adhere to the ACID principle** (Atomicity, Consistency, Isolation, Durability) of relational databases.

Second Questions of Today's Lecture

How to manage these
graphs composed of
large datasets?

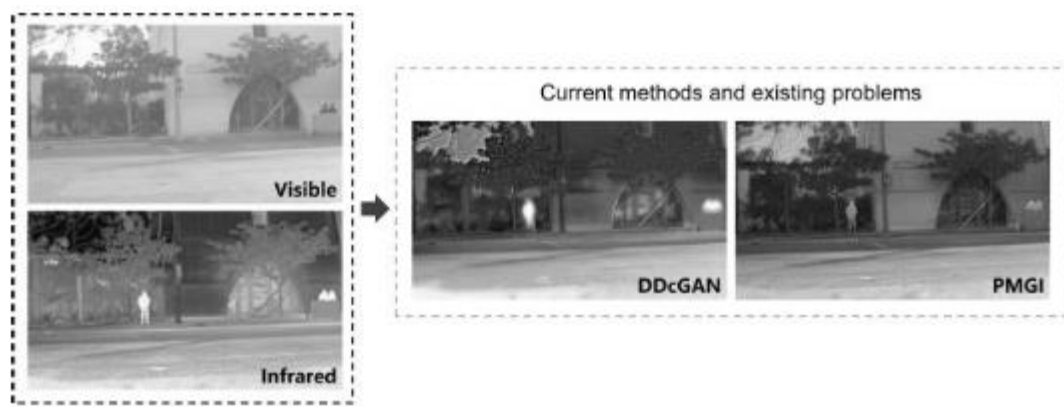


What should we do if we
want to summarize and
clarify many complex
relationships in our lives?

How to introduce graph
into multimodal machine
learning?



Multimodal learning tasks



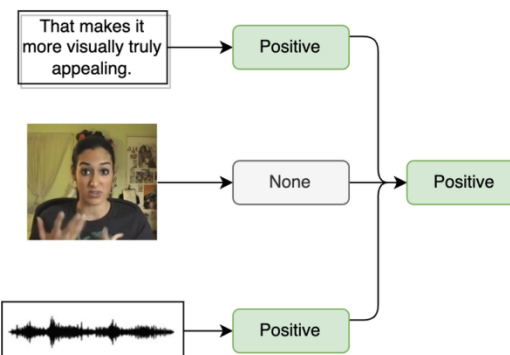
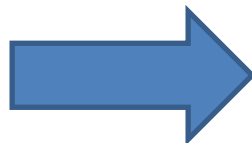
"I am happy with this water bottle."



"This is a bad investment."



"I am going to walk today."



The problems of Multimodal learning



This part of my
life is called "Happiness"



- Different schema
- Unclear relationship semantics



I am so happy



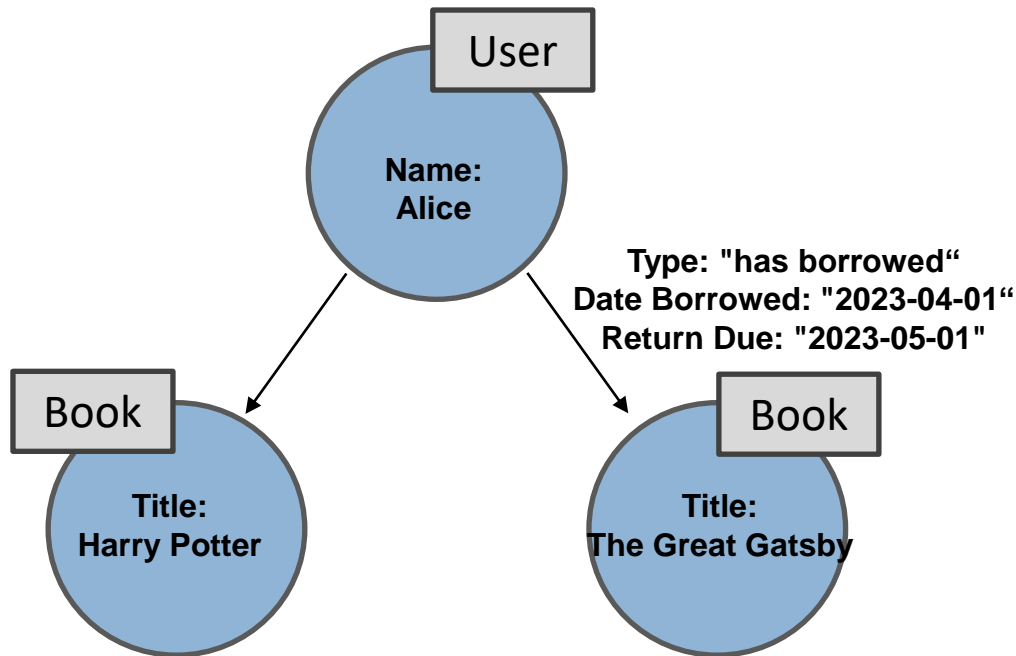
I am so happy

The problems of Multimodal learning

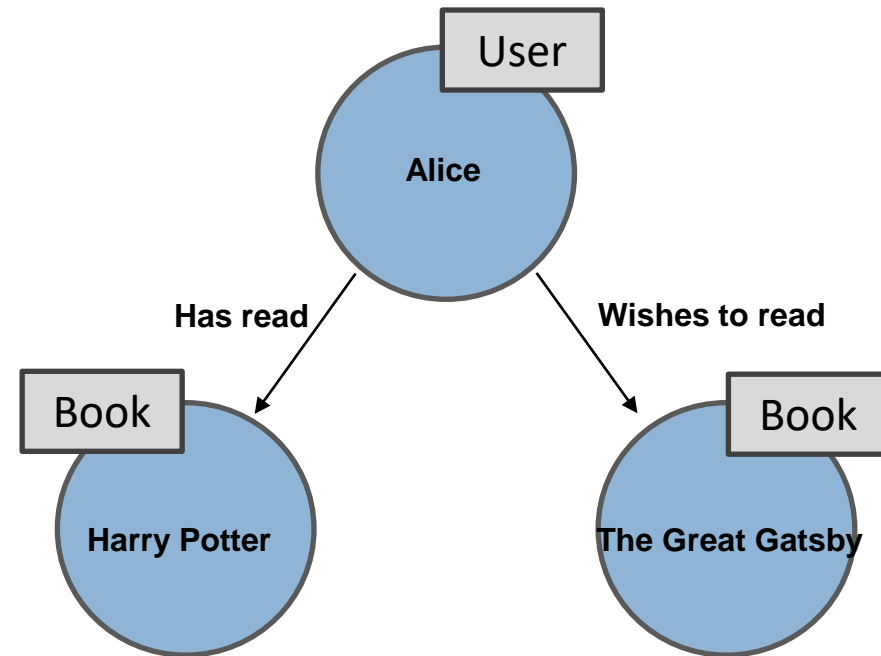
- Multimodal datasets
- Different schema
- Unclear relationship semantics



Knowledge graph

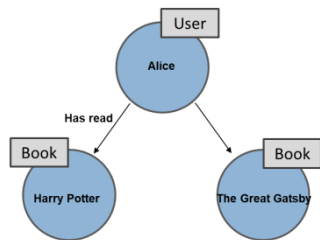


Property graph

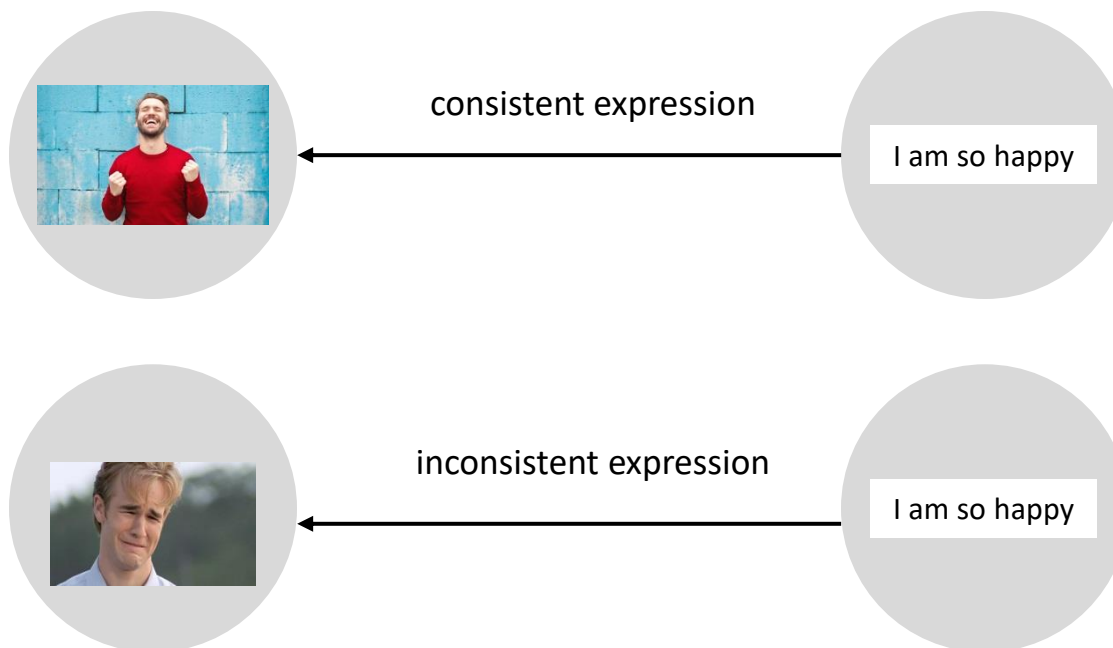


Knowledge graph

Multimodal learning with graph



Knowledge graph



- A graph is data structure that depicts key relationships and patterns within complex datasets.
- A graph database is a specialized data management system optimized for storing, querying, and analyzing interconnected data represented as graphs.
- If you want to analyze multimodal datasets that contains complex relationship semantics between different modalities, knowledge graph may be a good data representation method.