CSIT882 Data Management Systems

Graph Data Models

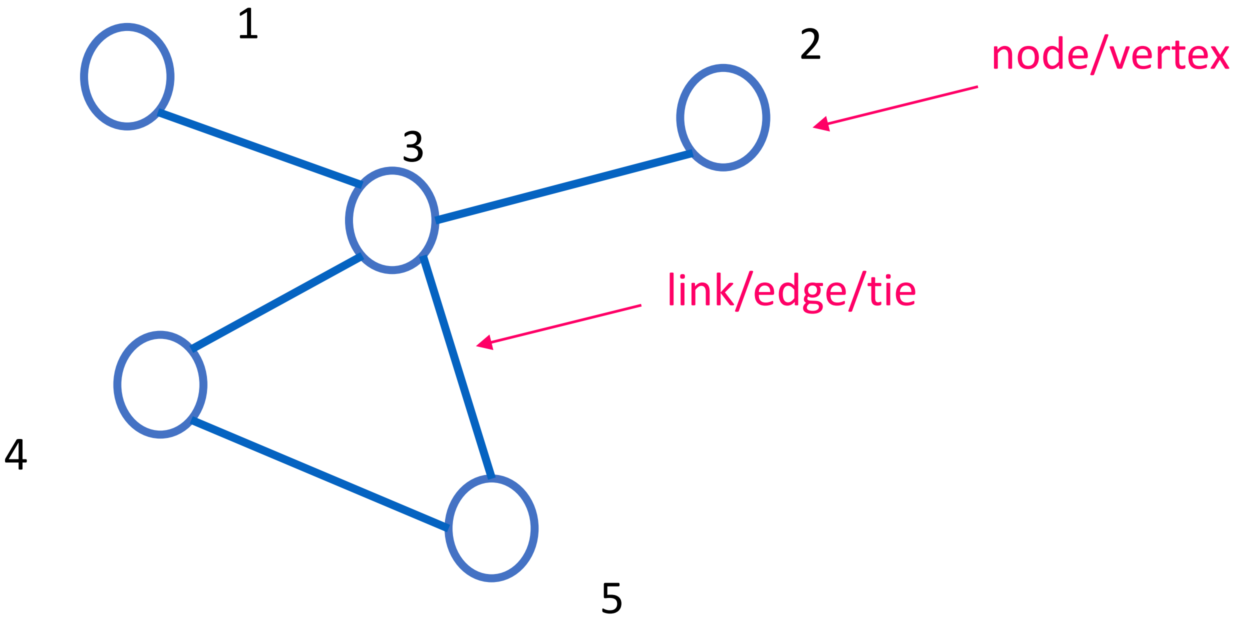
Subject Coordinator: Dr Chen Chen

Slides created by Dr Janusz R. Getta

School of Computing and Information Technology - University of Wollongong

INTRODUCTION – What are Networks

- **Networks** are collections of **nodes (points)** connected by Edges **(lines)**



“Network” ≡ “Graph”

INTRODUCTION – Why we study graphs

- **Networks** are everywhere, and we live in a **highly-connected world**.
- e.g., easy to reach a place, to reach a person or to find information
- In many applications, we need to analyze **in the context of networks**, not just individuals.
  - Social networks
  - Information networks
  - Road networks
  - Protein-Protein-interaction networks
  - …

Social Networks – Network of thrones

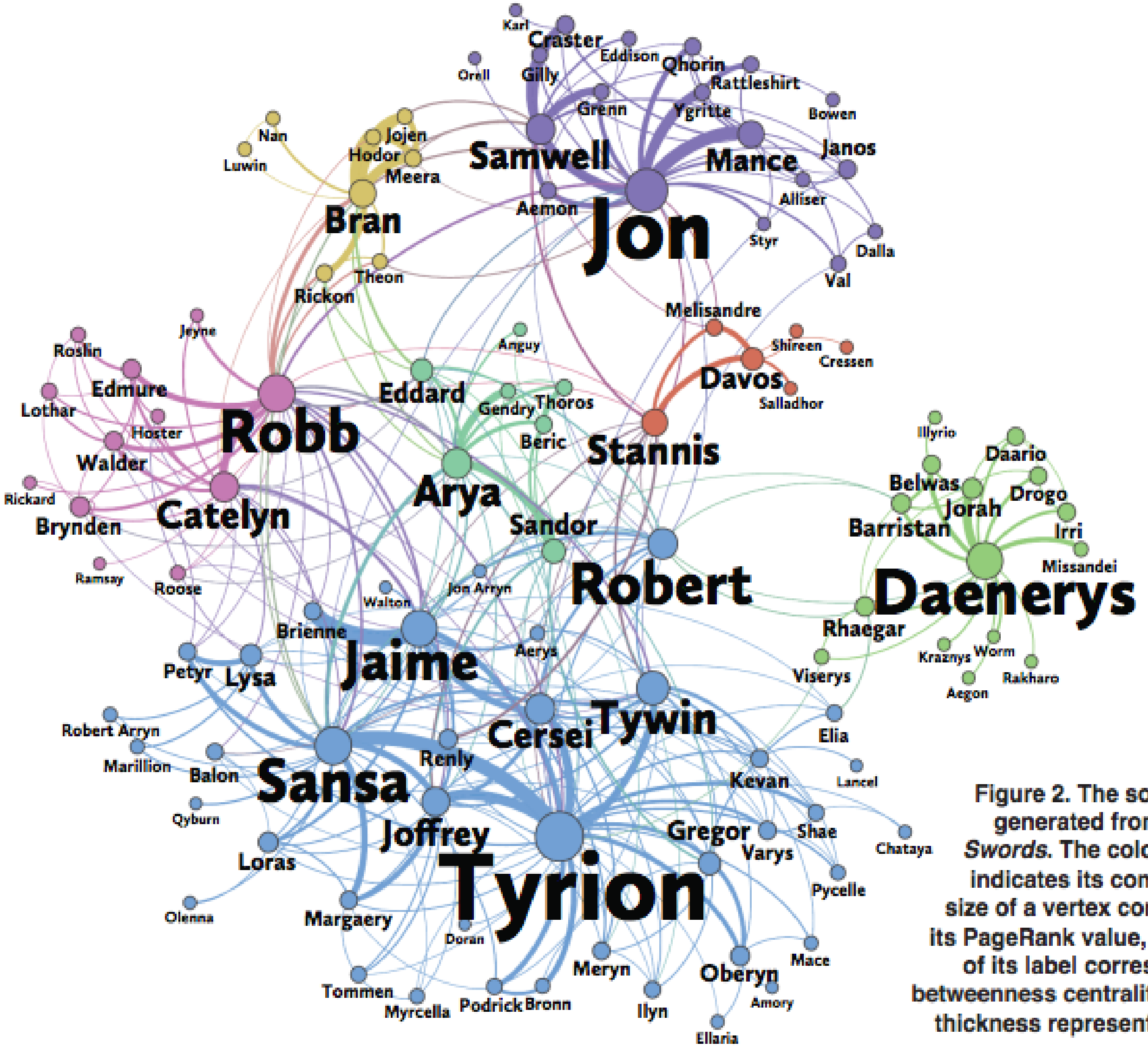


Figure 2. The social network generated from *A Storm of Swords*. The color of a vertex indicates its community. The size of a vertex corresponds to its PageRank value, and the size of its label corresponds to its betweenness centrality. An edge's thickness represents its weight.

http://www.maa.org/sites/default/files/pdf/Mathhorizons/NetworkofThrones%20%281%29.pdf

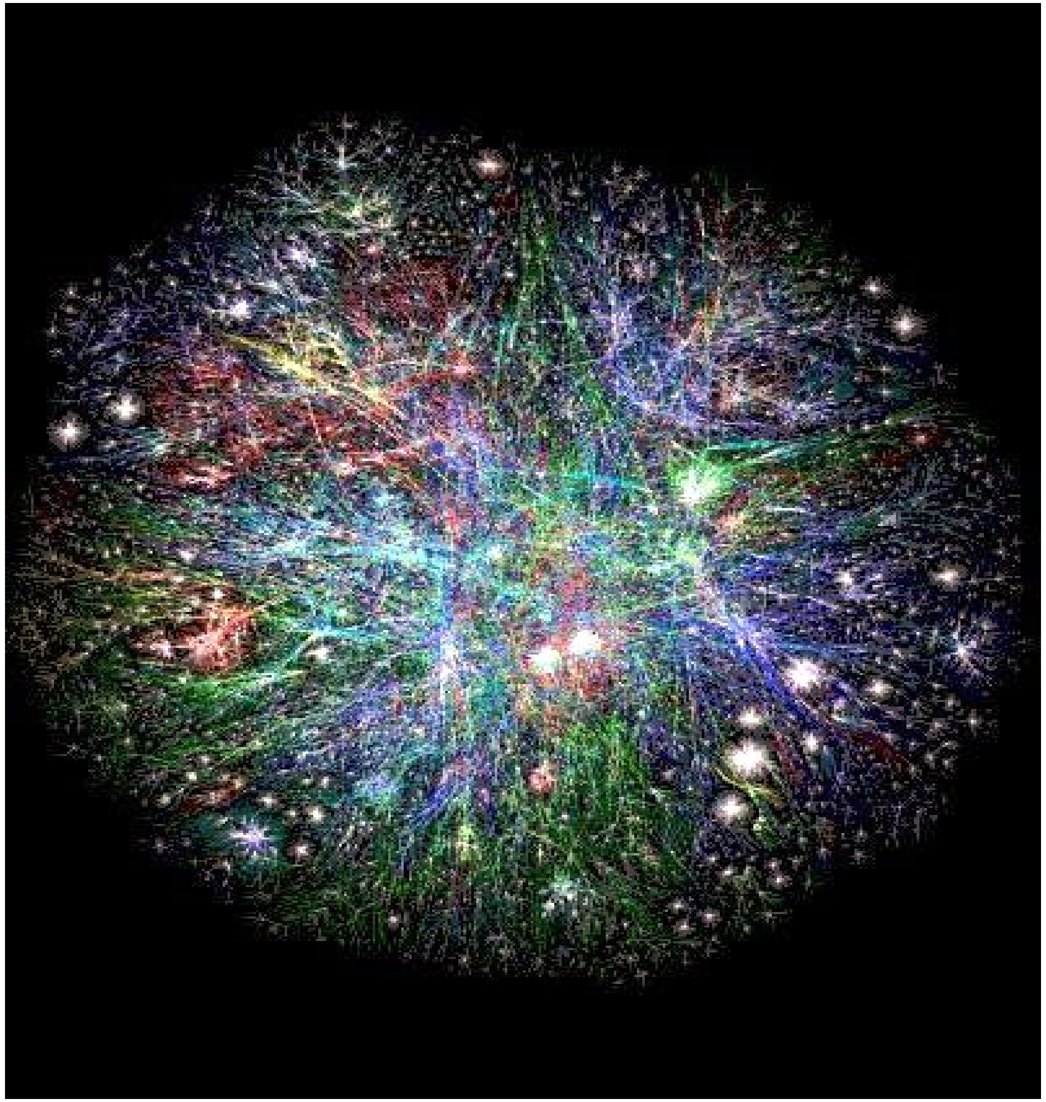
Social Networks – Facebook (location based visualization in 2011)

Monthly active users: around **1 billion** in **2012** and **2.32 billion** (2017) **-** now around **2.5 billion (2020)**

facebook
December 2010

*Facebook social graph, [Backstrom-Boldi-Rosa-Ugander-Vigna, **2011**]*

Information Networks – WWW

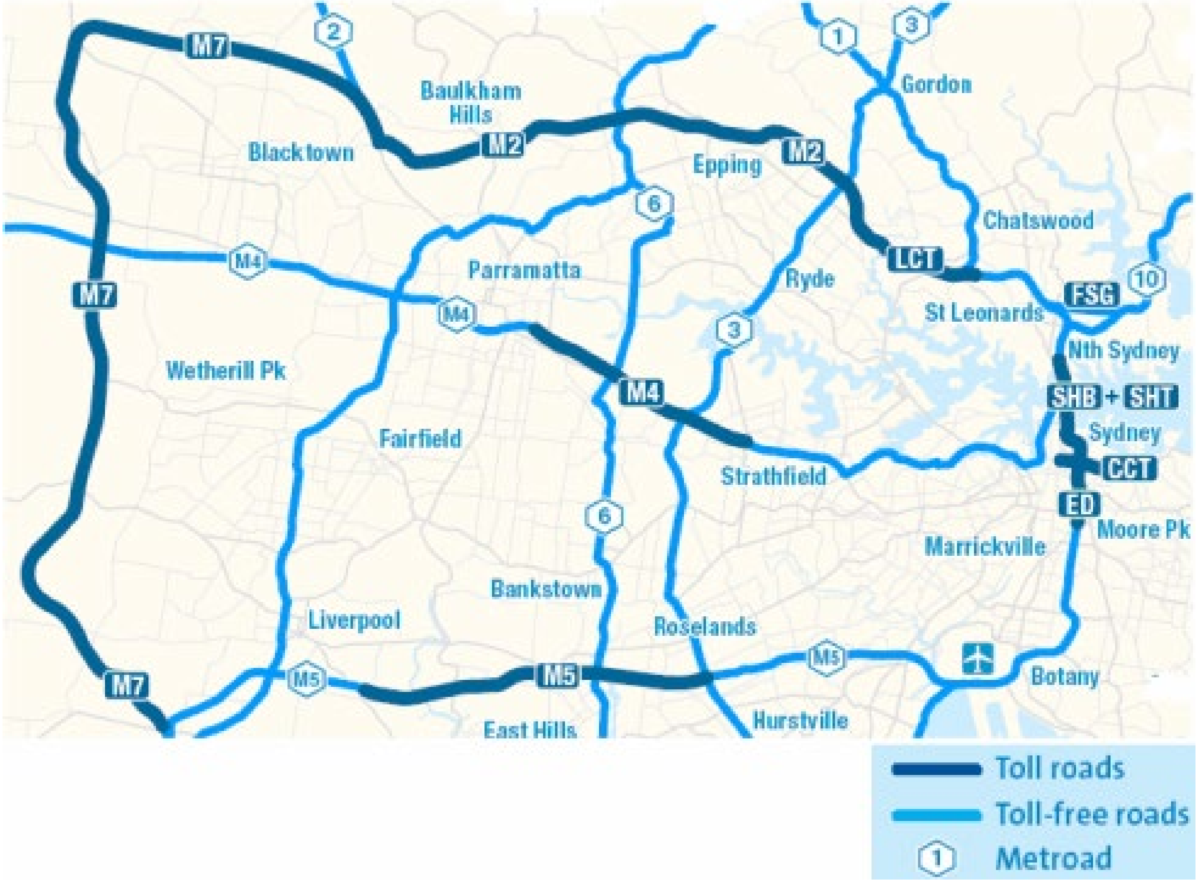


http://www.vlib.us/web/worldwideweb3d.html

3D Map of the World Wide Web

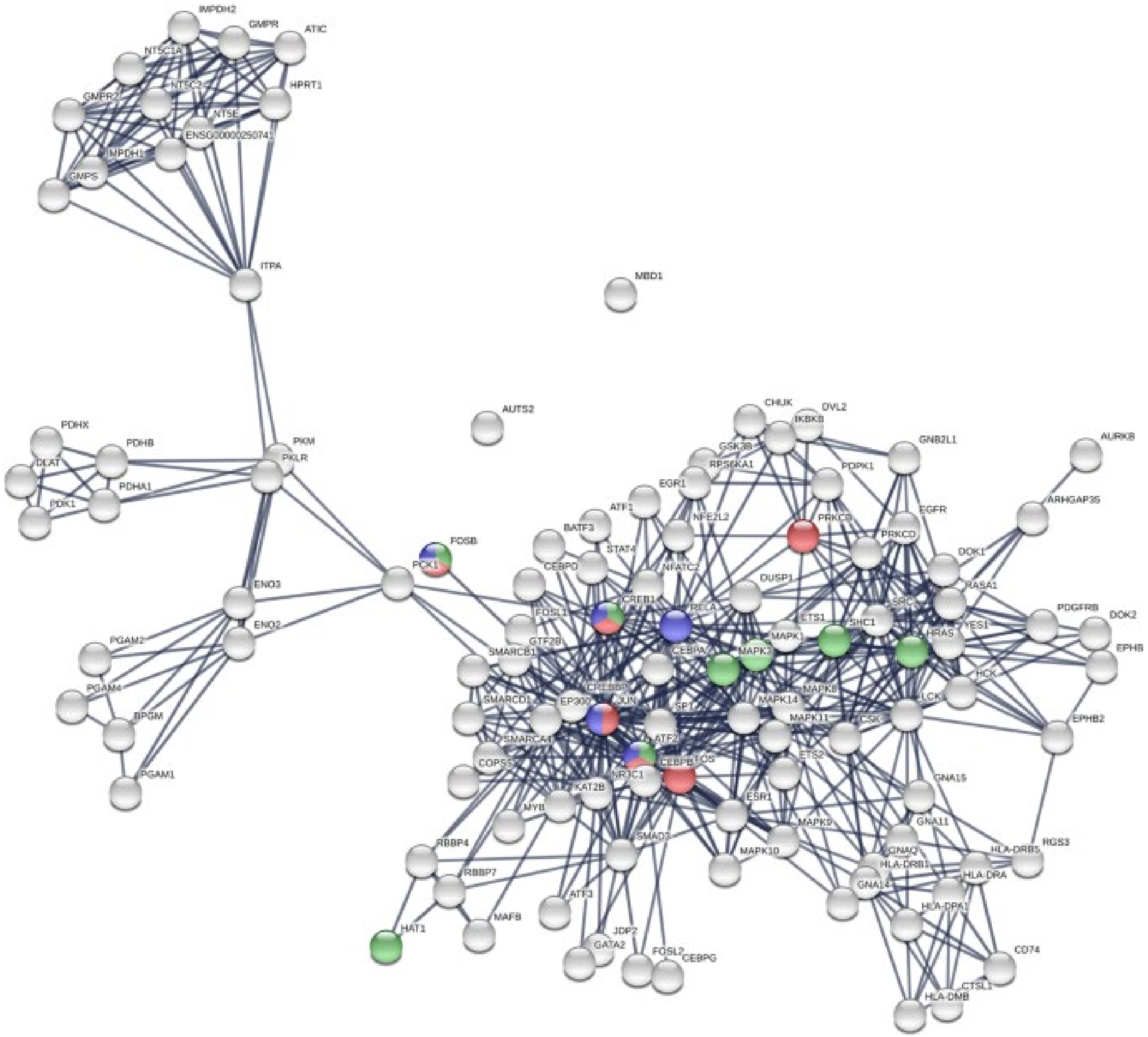
This illustrates in 3-D the actual domains and connections of the world wide web. **Colors** have been added to represent .edu, .gov, .com, etc. domains.

Other Networks



Sydney Road Network

http://www.brt.cl/opinion-pieces-i-tolled-you-so



Protein interaction network

https://www.nature.com/articles/s41598-019-41552-z

Graph Data Models

Outline

- Property Graph Model
- Hypergraph Data Model
- Nested Graph Data Model

Property Graph Data Model

A graph G consists of a set of nodes (N) and a set of edges (E), G = (N, E)

In an undirected graph an edge is represented by a set of 2 nodes $\{n_i, n_j\}$

In a directed graph an edge is represented by an ordered tuple of 2 nodes (n_i, n_j)

Multi-relational graph is a graph where types are introduced for nodes and edges

For example a multi-relational graph may have the names like STUDENT or PROJECT for node type and a name like WORKS-ON for edge type

A node is labelled with a node label and an edge is labelled with an edge label

A specification of type also includes attributes for nodes and edges

A pair (attribute,value) describes property of nodes and edges

Each node and each edge has a unique identifier

In HTML view press 'p' to see the lecture notes
TOP

Property Graph Data Model

A property graph is a labelled and attributed directed multigraph with identifiers

A property graph G is defined as $G = (N, E, L_n, L_e, ID)$ where

- N is a set of nodes
- E is a set of edges
- L_n is a set of node labels where a label has a set of node type definitions attached; a node type definition is a pair (type, set of attribute definitions) where an attribute definition is a pair (attribute name, attribute domain)
- L_e is a set of edge labels where a label has a set of attribute type definitions and restrictions for the types of source and target nodes attached;
- edge type definition is a quadruple (type, set of attribute definitions, type, type) where an attribute definition is a pair (attribute name, attribute domain)
- ID is a set of node and edge identifiers

In HTML view press 'p' to see the lecture notes

TOP

Property Graph Data Model

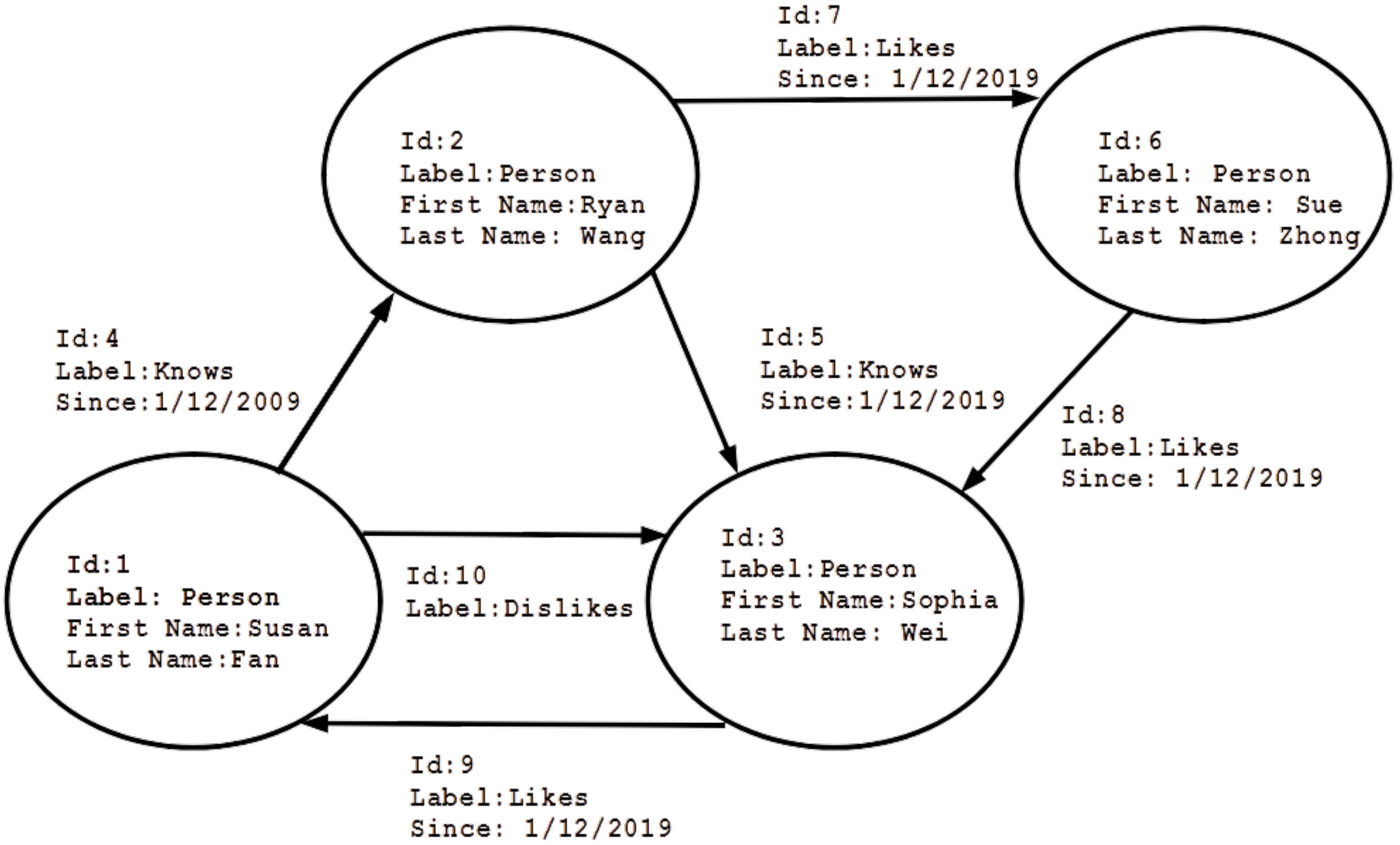
For example, students enrolling subjects taught by lecturers can be described by a property graph UNI = (N, E, L_n, L_e, ID) where

- N = {n_1, n_2, n_3}
- E = {e_1}
- L_n = {Student, Lecturer} where node type definitions are
  (Student, $A_{Student}$), $A_{Student}$ = {(name, string), (age, int)}
  (Lecturer, $A_{Lecturer}$), $A_{Lecturer}$ ={(name,string),(expert, string)}
- L_e ={Likes, Teaches}
  (Likes, A_{Likes}, Student, Student U Lecturer}, A_{Likes} ={(level,int)}
  (Teaches,$A_{Teaches}$, Lecturer, Student}, $A_{Teaches}$ ={(since,date)}
- ID = {1,2,3,4}
- n_1 = {1, Student, {name: Alice, age: 23}}
- n_2 = {2, Student,{name: Bob, age: 25}}
- n_3 = {3, Lecturer,{name: James, expert: databases}}
- e_1 = {4, Teaches,{since: 01-01-2023}, 3, 2}

In HTML view press 'p' to see the lecture notes

Property Graph Data Model

Visualisation of a sample instance of property graph



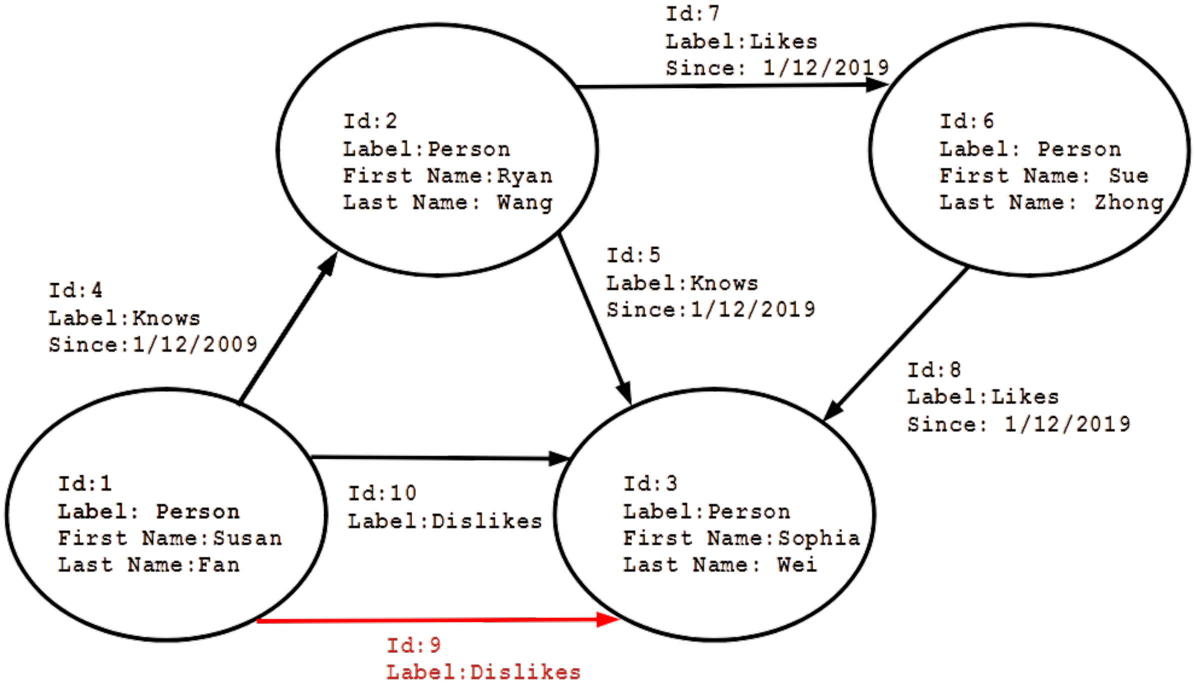
In HTML view press 'p' to see the lecture notes

TOP

Property Graph Data Model

Property graph must satisfy a number of constraints

- Edges of certain type are allowed between the nodes of certain types, e.g. Teaches can occur only between the pairs of nodes (Lecturer, Student)
- Two nodes cannot be linked by two edges of the same type



In HTML view press 'p' to see the lecture notes

Graph Data Models

Outline

- Property Graph Data Model
- Hypergraph Data Model
- Nested Graph Data Model

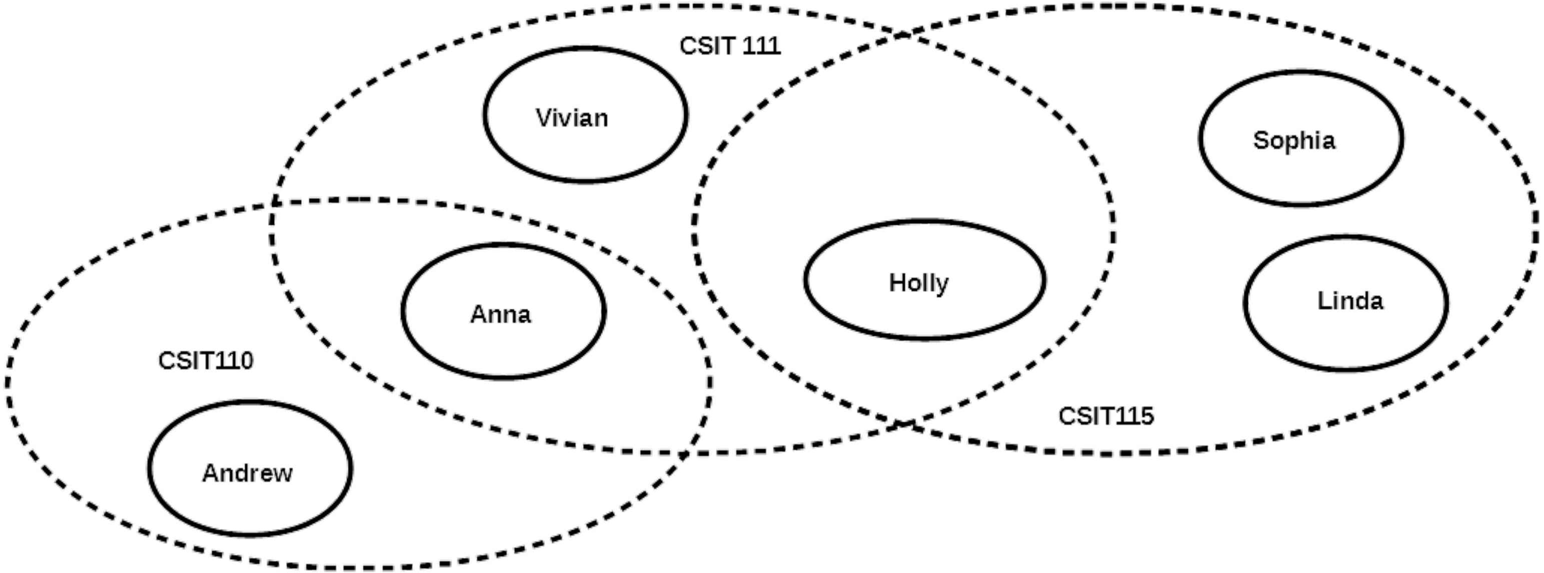
Hypergraph Graph Data Model

A hypergraph is a graph with hyperedges

An undirected hyperedge is a set of nodes

A hypergraph G consists of a set of nodes (N) and a set of hyperedges (E), G= (N, E)

In an undirected hypergraph an edge is represented by a set of k nodes $\{n_{i1}, n_{i2}, \dots , n_{ik}\}$



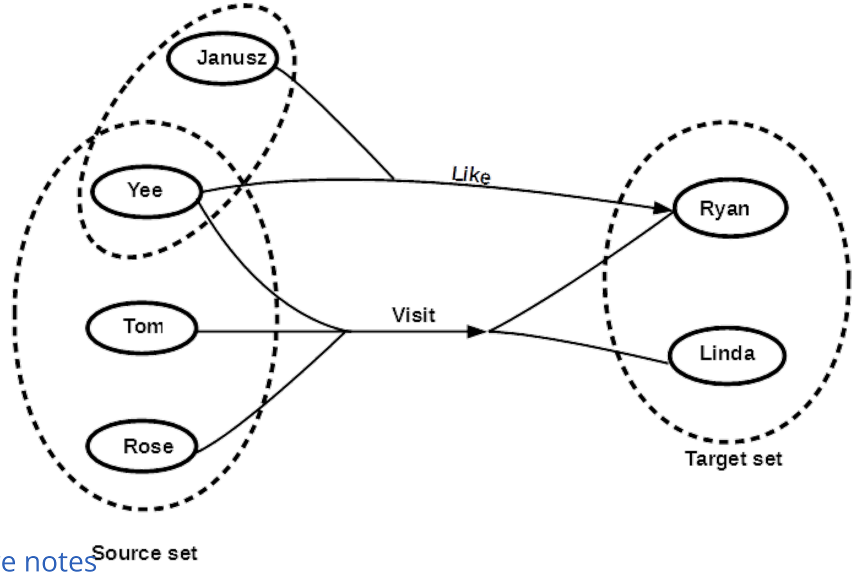
In HTML view press 'p' to see the lecture notes

Hypergraph Graph Data Model

A Hypergraph Data Model can be used for data modeling where a database contains information about the groups of the real world objects

For example, a group of students taking French or Spanish classes, or a group of vehicles being cars, trucks, and trains

In directed hypergraph an edge is represented by an ordered tuple of 2 sets of nodes $(\{n_{i1}, n_{i2}, \ldots , n_{ik}\},\{n_{j1}, n_{j2}, \ldots , n_{jm}\})$



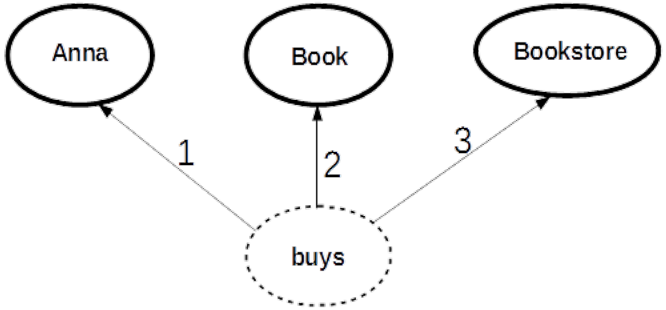
Hypergraph Graph Data Model

Multirelational hypergraph is a hypergraph where the types are introduced for nodes and edges

A Directed Hypergraph Data Model can be used for data modeling when a database contains information about the associations between entire groups of the real world objects

For example, a group pf people visiting another group of people or a group of researchers working on hardware computing devices and passing the results of research to a group working on software for the same computing devices

An oriented hyperedge is an arbitrary sequence of nodes, i.e. it is a tuple $(n_{i1}, \ldots, n_{ik})$



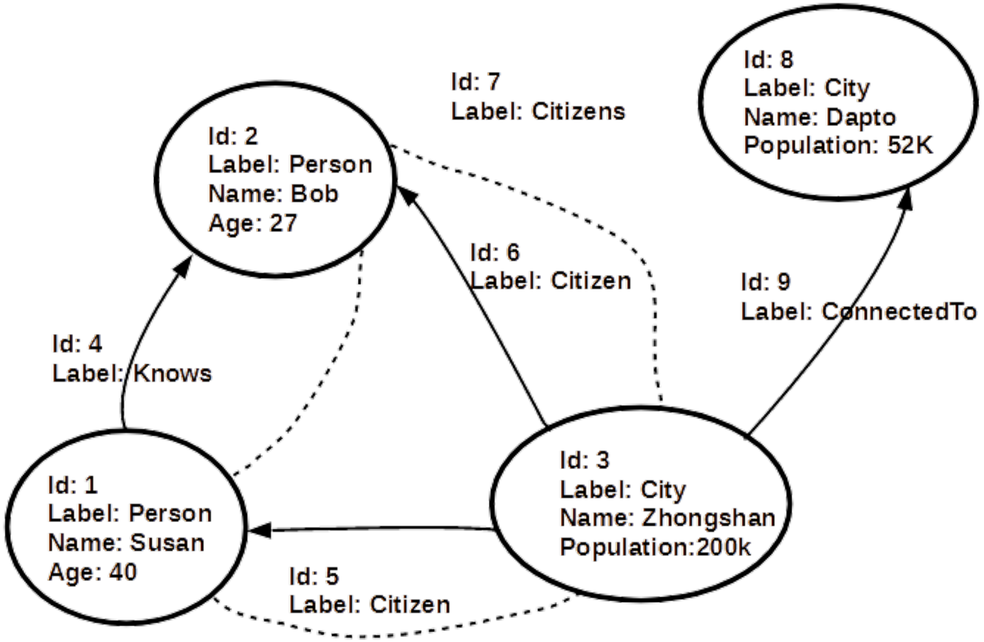
In HTML view press 'p' to see the lecture notes

TOP

Hypergraph Graph Data Model

A generalized hyperedge groups together a set of nodes an edges

A generalized hyperedge is a set $e = \{a_i,\ldots a_j\}$ of an arbitrary number of elements where the elements a_i and a_j are either nodes or egdes, $a_i, a_j \in V \cup E$



Graph Data Models

Outline

- Property Graph Data Model
- Hypergraph Data Model
- Nested Graph Data Model

Nested Graph Data Model

A nested graph consists of hypernodes that can be used to represent complex objects

Hypernodes can encapsulate entire subgraphs that themselves can be nested

A definition of a hypernode can be recursive and because of that the depth of nestings can be unlimited

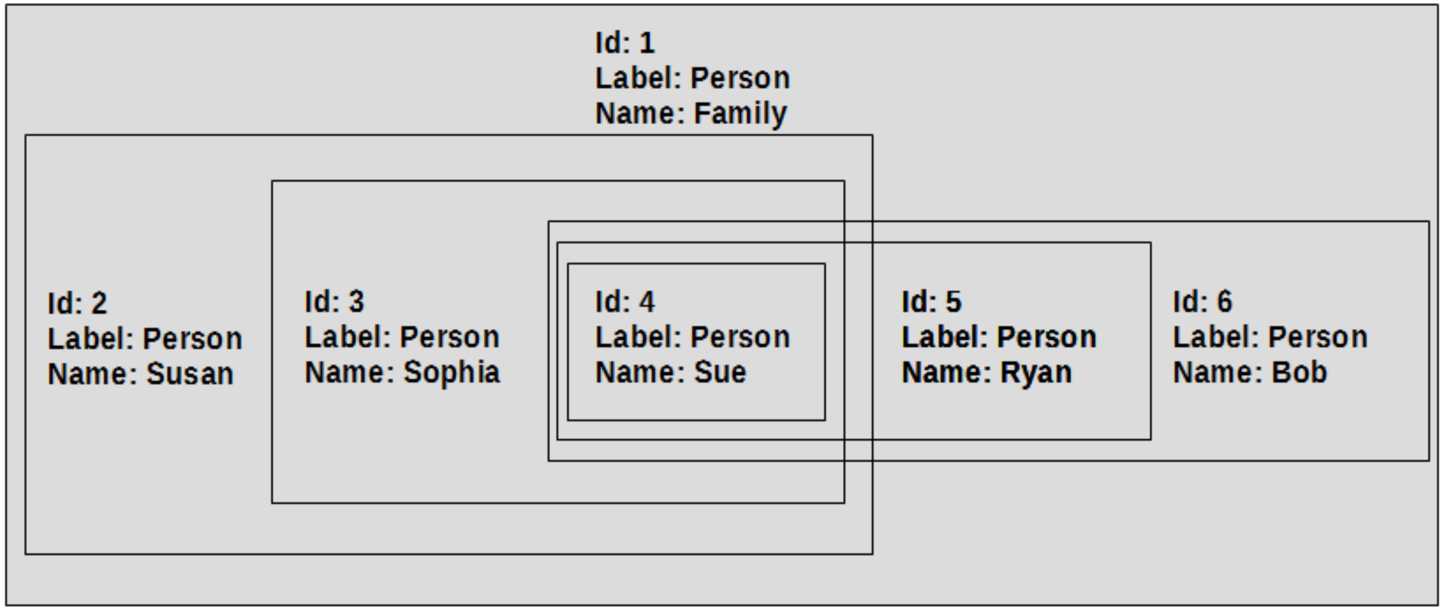
A nested graph consists of the following components:

- a set P of primitive nodes that contains keys and values of (key,value) pairs,
- a set I of identfiers
- A nested graph is defined by choosing and identifier G in I and then assigning a set of hypernodes and binary edges G =(N,E) such that each n ∈ N is either a primitive node or another identifier G' (N is included or equal to P ∪ I); a complex hypernode G' ∈ N is itself a nested graph
- a set E included in N × N is defined as a binary edge between two hypernodes restricted in such a way that an edge e ∈ E can map a key to a value or to an identifier

In HTML view press 'p' to see the lecture notes

Nested Graph Data Model

A sample nested graph



References

Wiese L., Advanced Data Management, 2015, Chapter 4 Graph Databases