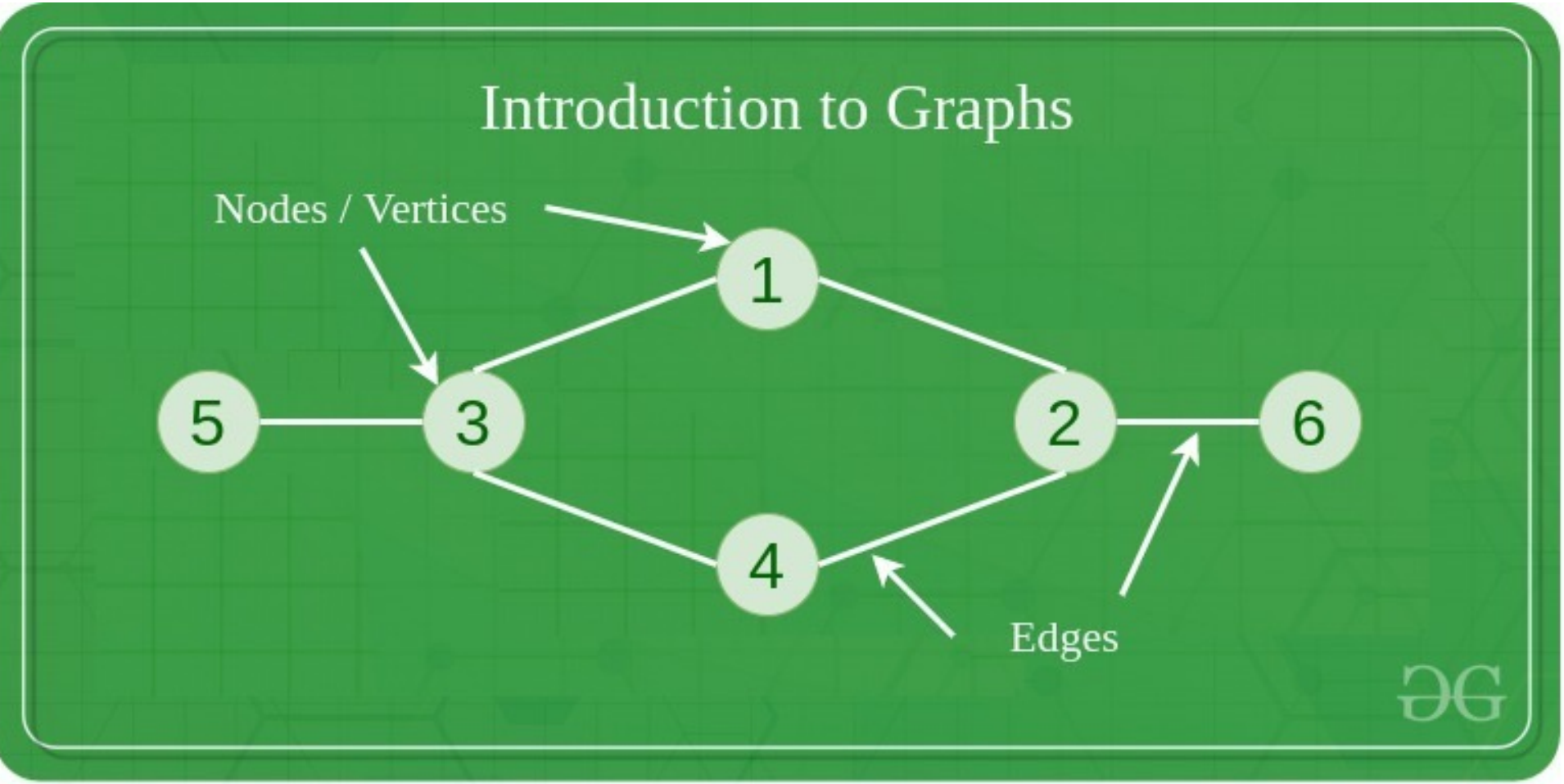
INTRODUCTION TO KNOWLEDGE GRAPHS

UNIVERSIDAD DE
MURCIA

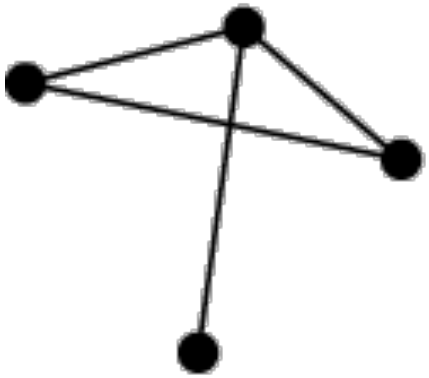
Structure

UNIVERSIDAD DE MURCIA

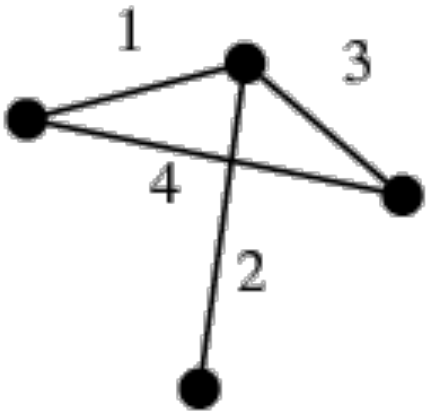


Labels

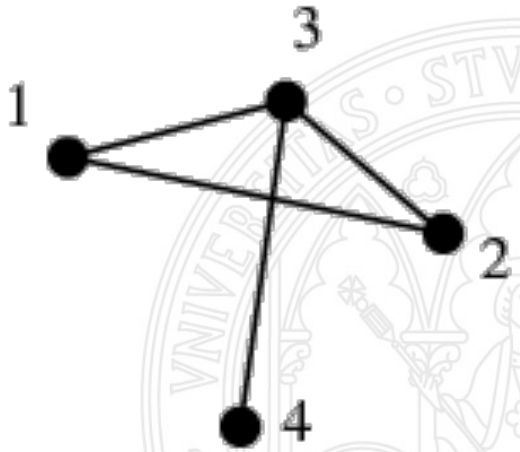
UNIVERSIDAD DE MURCIA



unlabeled graph

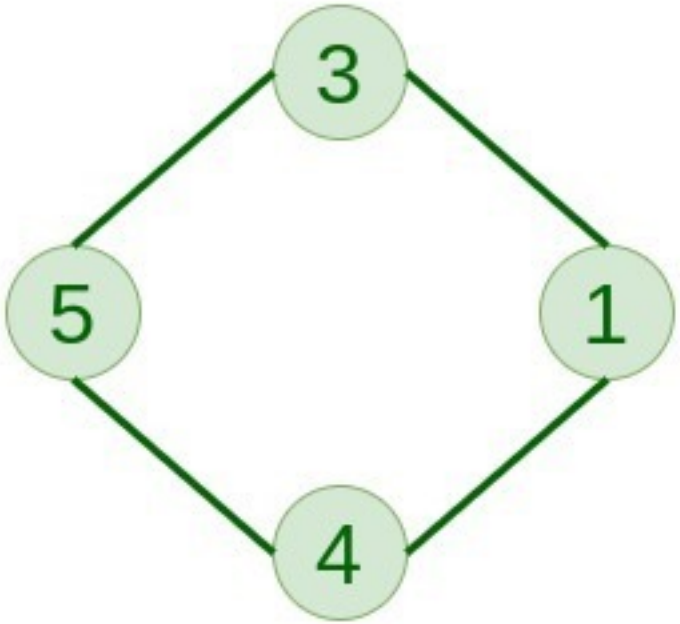


edge-labeled graph

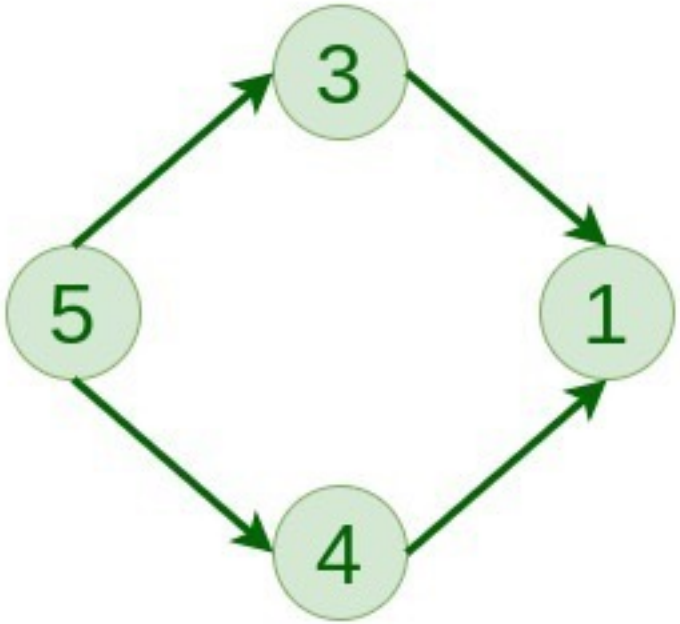


vertex-labeled graph

Directionality

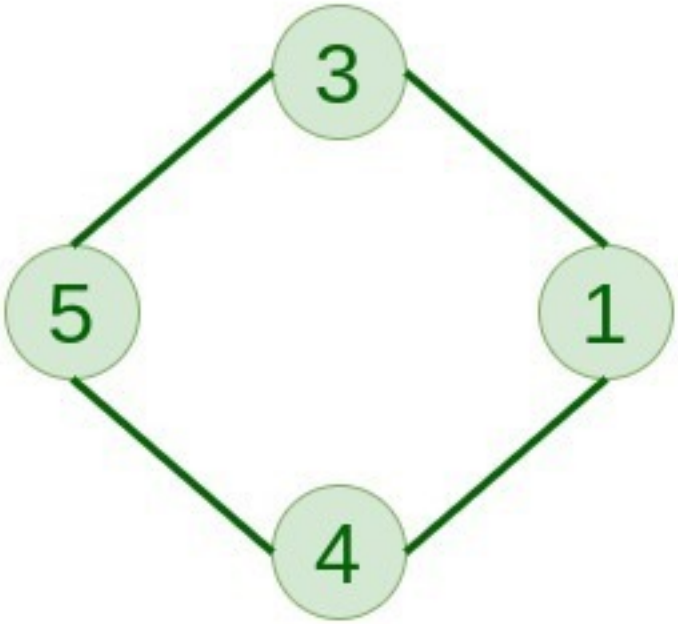


Undirected Graph

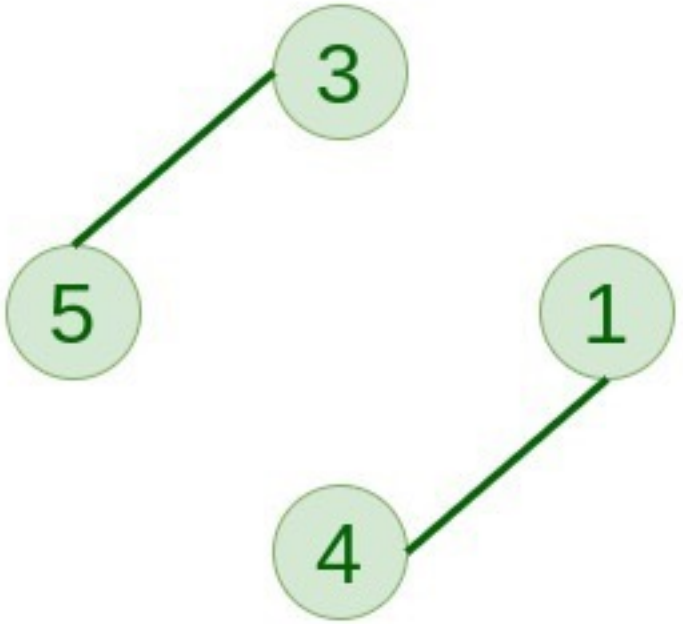


Directed Graph

Topology

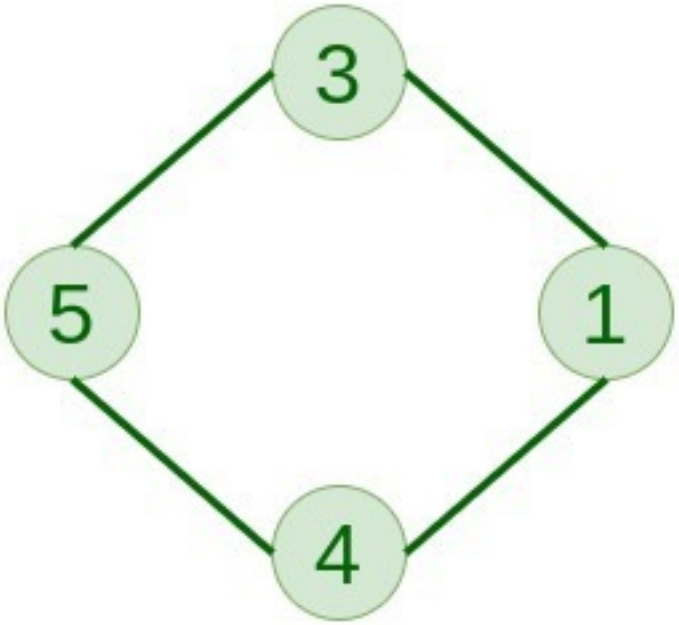


Connected Graph

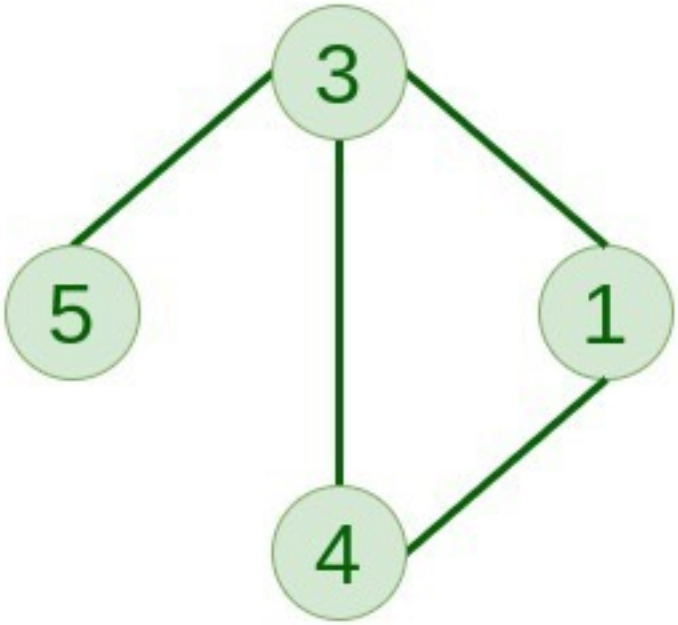


Disconnected Graph

Cycles

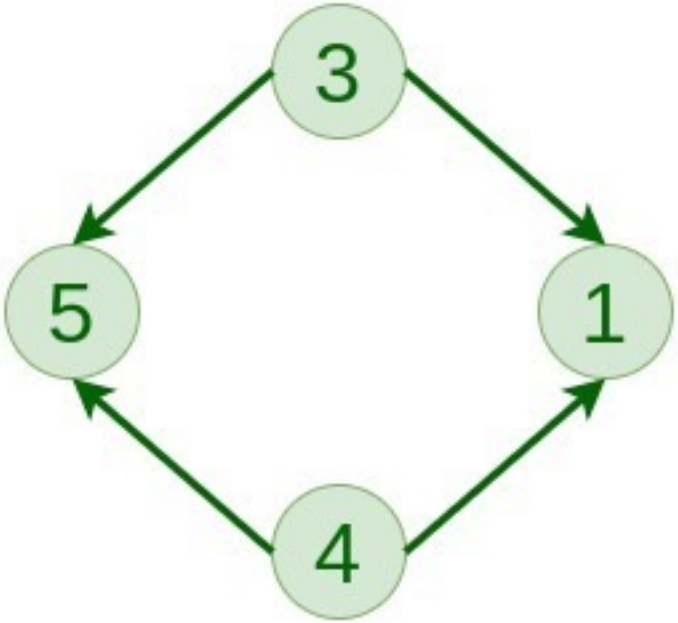


Cycle Graph

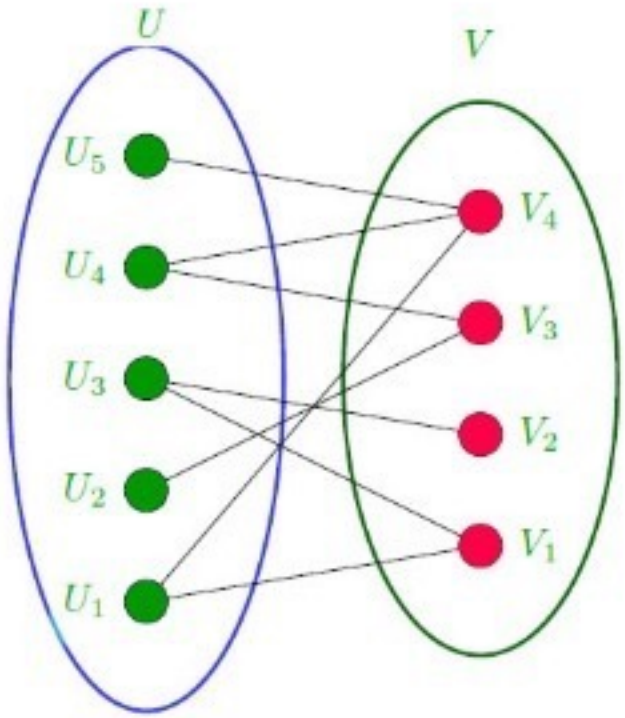


Cyclic Graph

Cycles



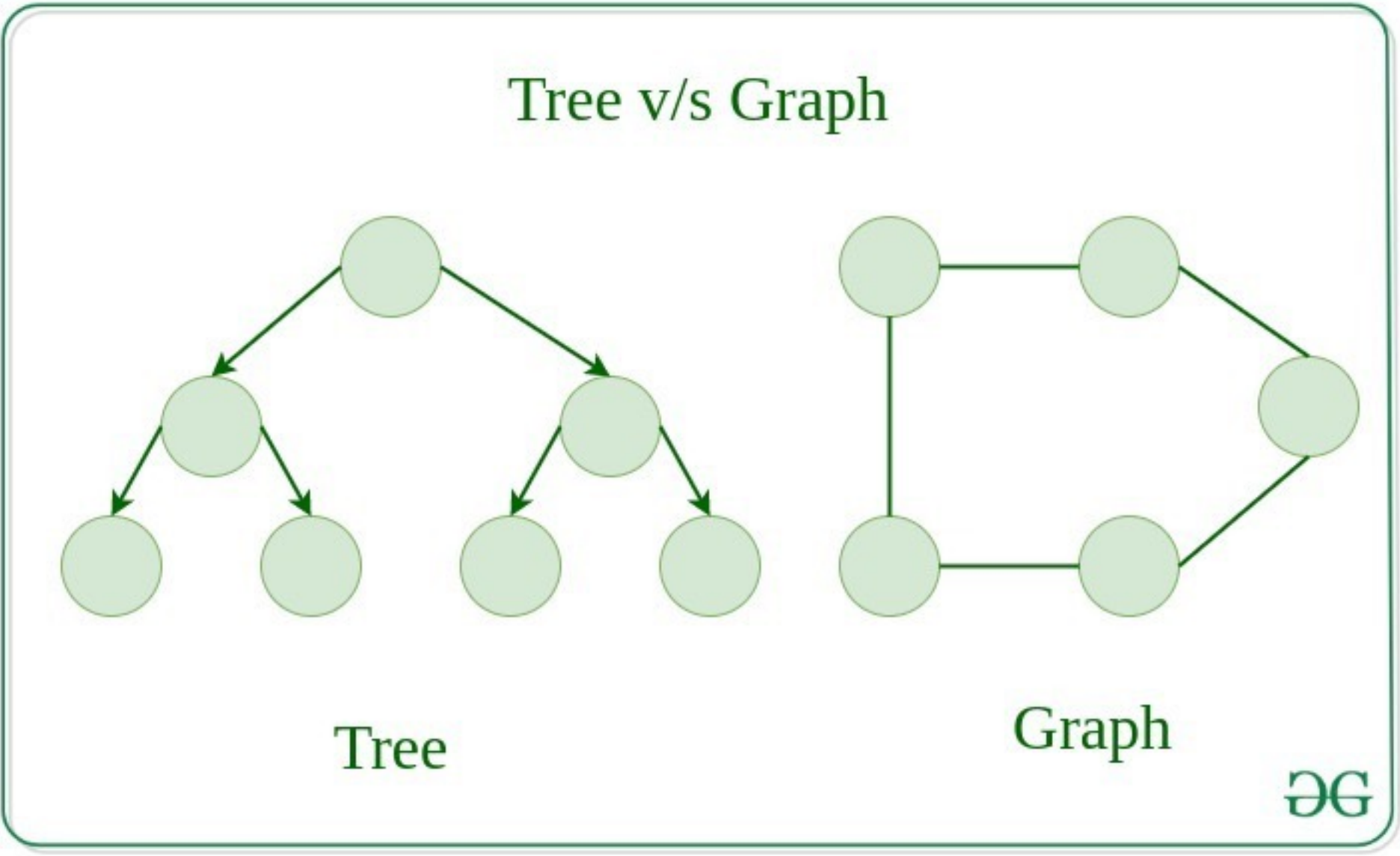
Directed Acyclic Graph



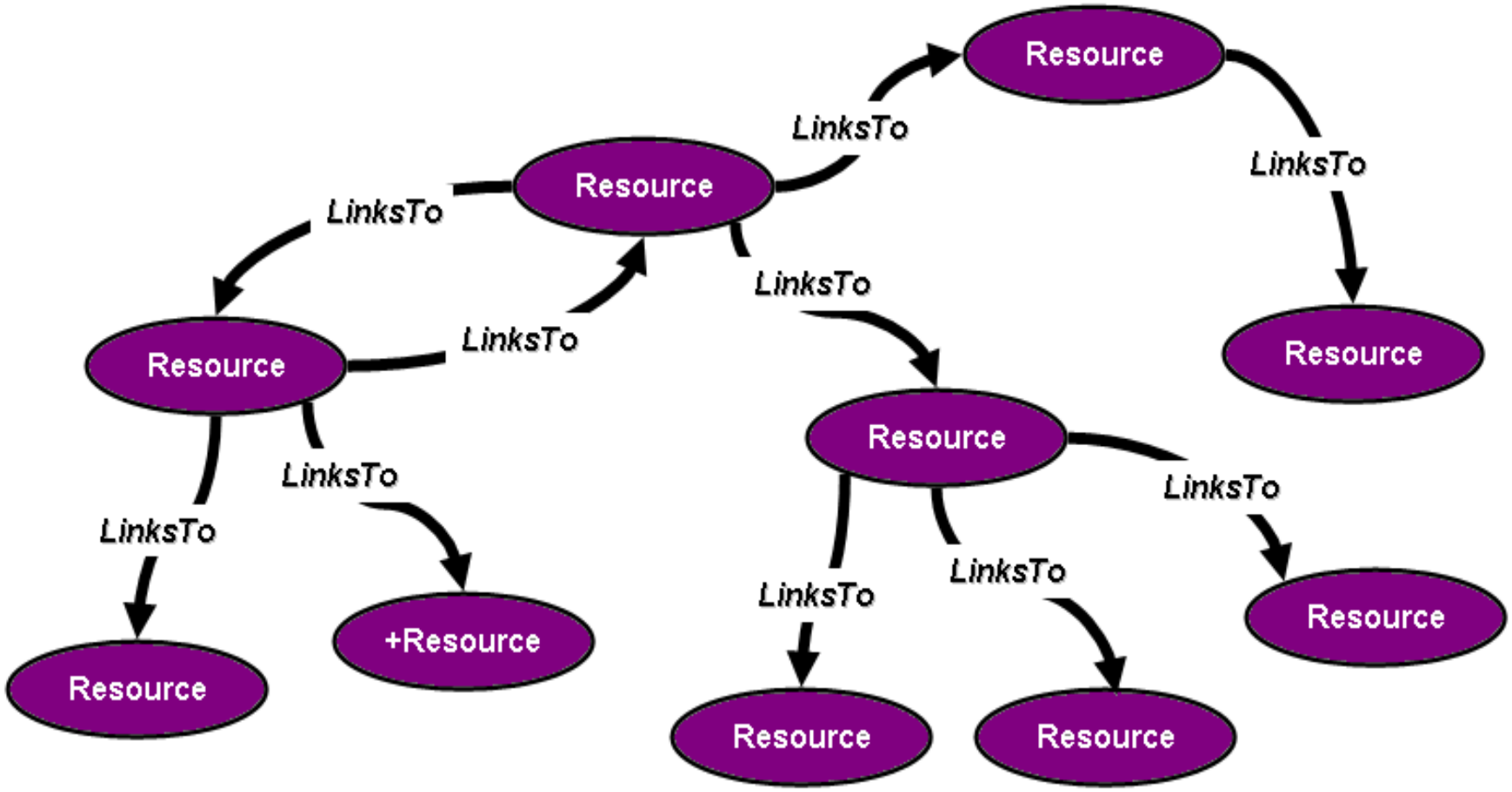
Bipartite Graph

Data structure

UNIVERSIDAD DE MURCIA

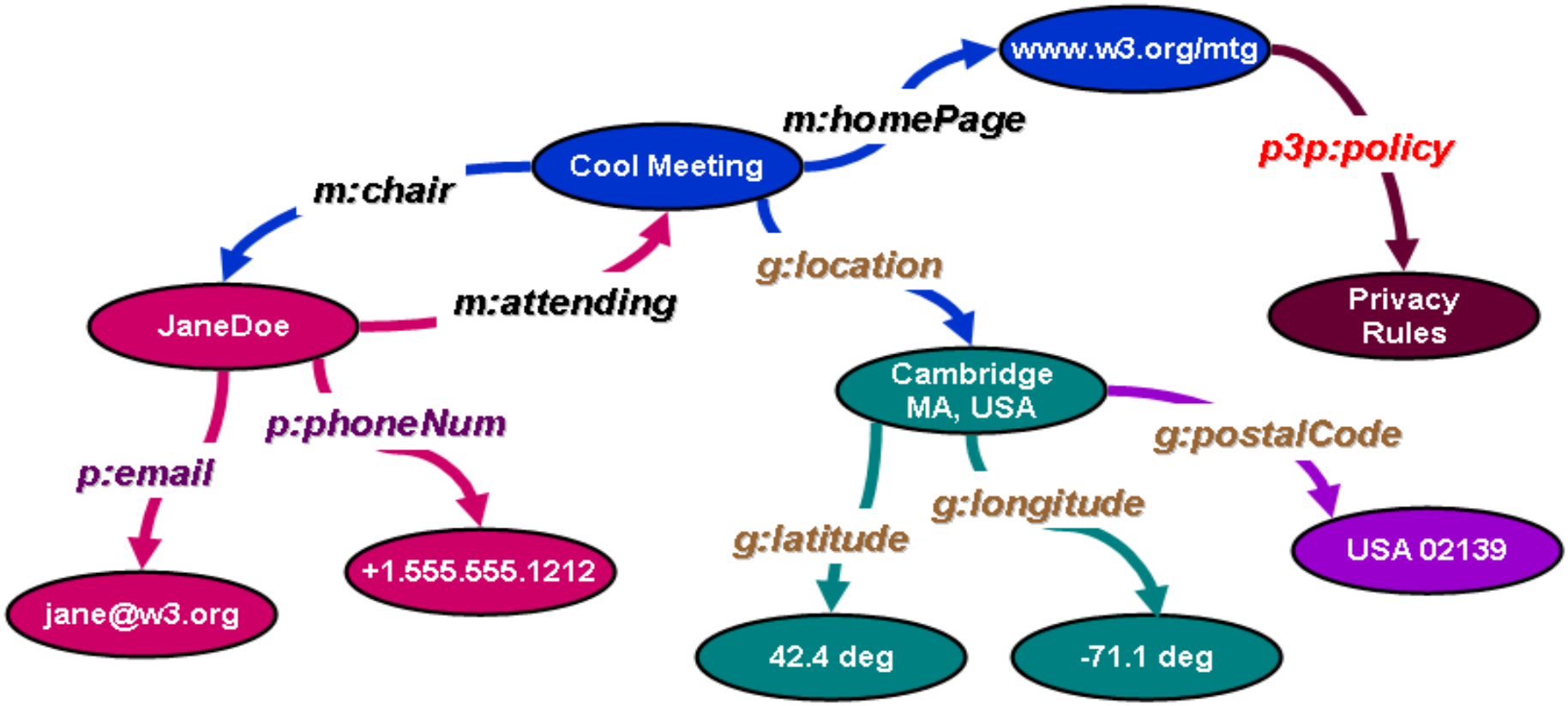


The web as a graph



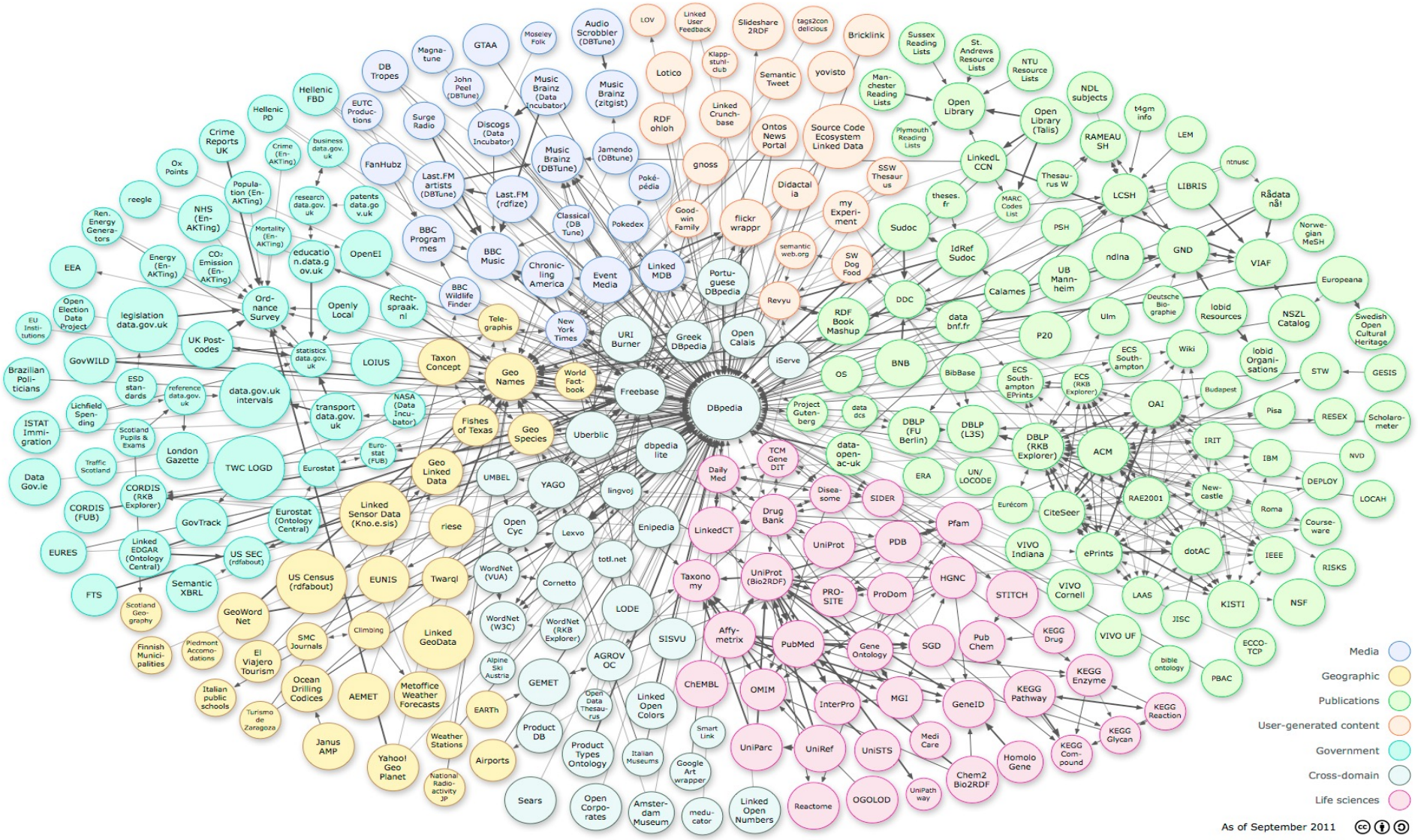
The semantic web as a graph

UNIVERSIDAD DE MURCIA



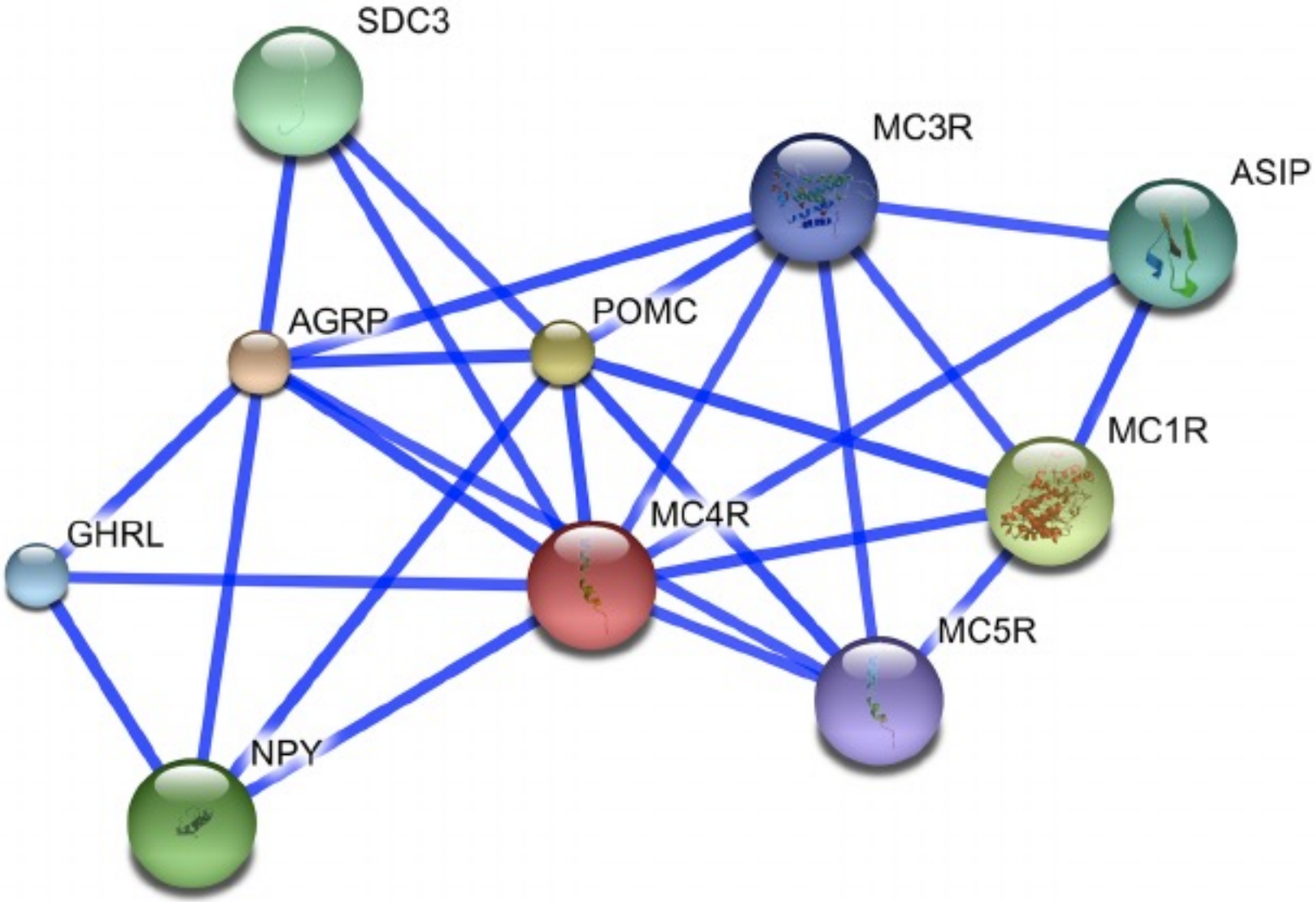
Some graphs

UNIVERSIDAD DE MURCIA

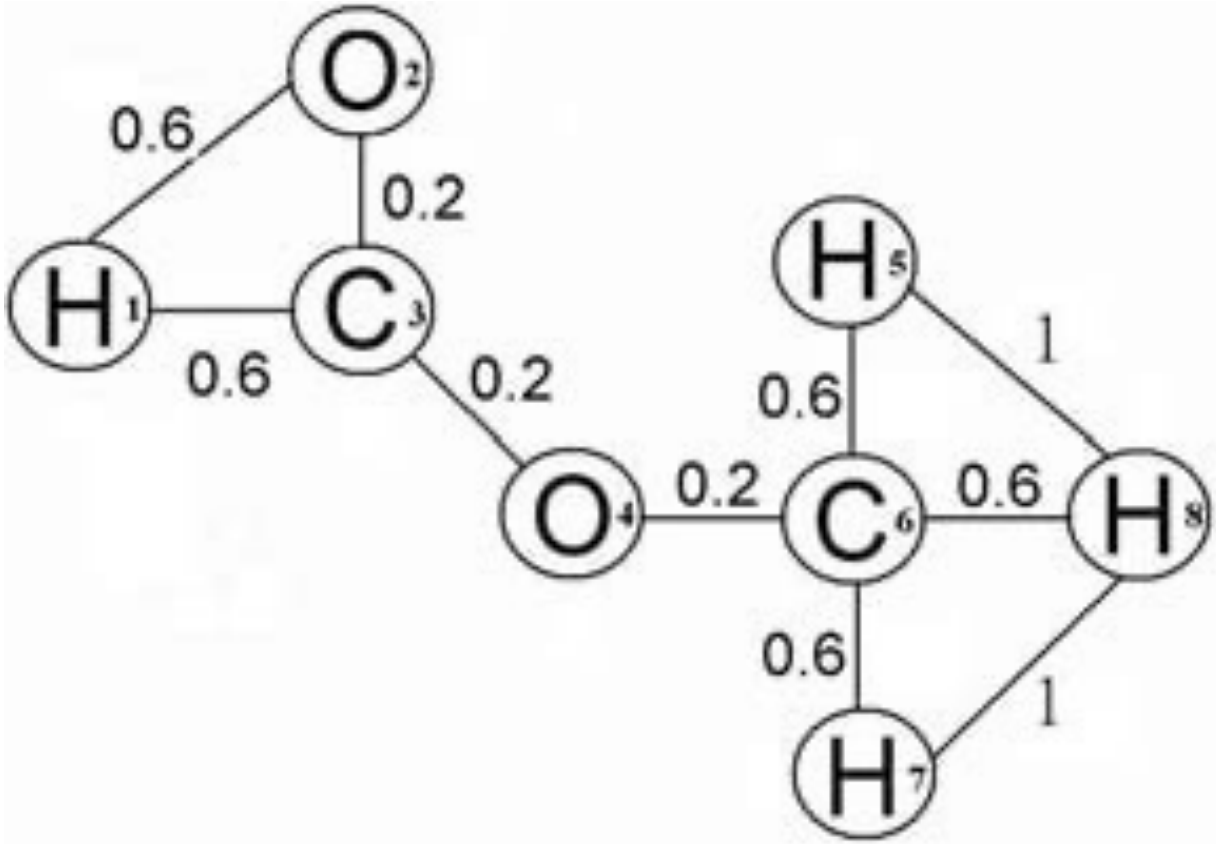


Protein interactions as graphs

UNIVERSIDAD DE MURCIA

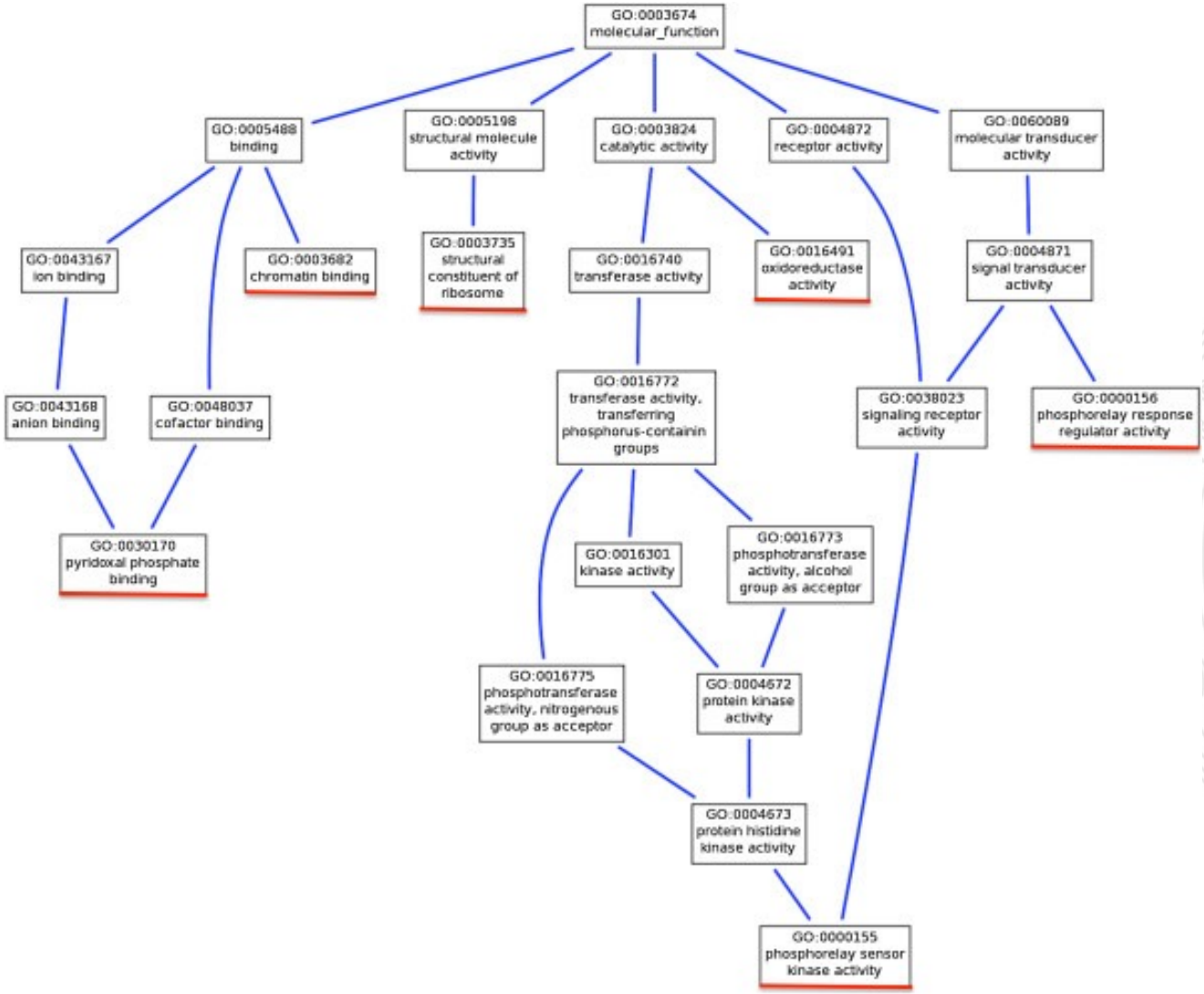


Some graphs



Is this a graph or a tree?

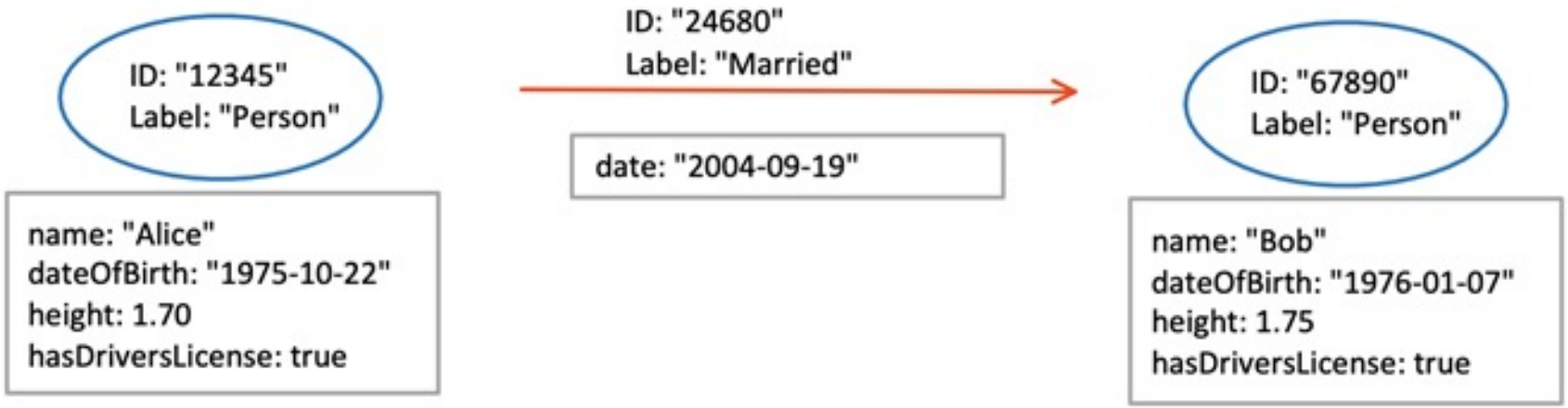
UNIVERSIDAD DE MURCIA



Label Property Graphs (LPG)

UNIVERSIDAD DE MURCIA

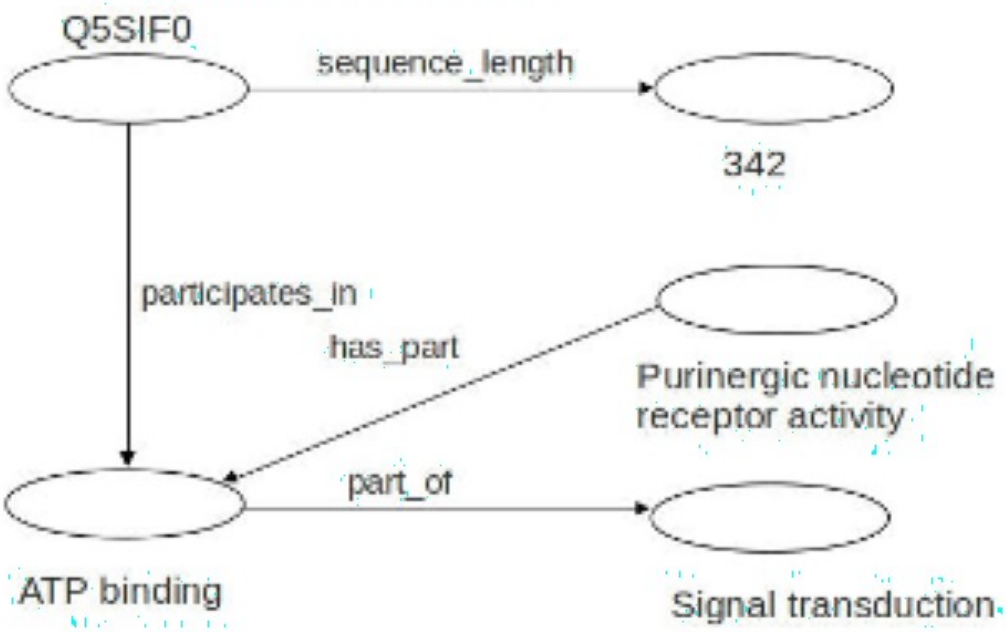
- Vertices:
  - Nodes with identifier + set of key-value pairs
- Edges:
  - Relations with identifier + type + set of key-value pairs



Semantic graphs (RDF)

UNIVERSIDAD DE MURCIA

- **Statements or RDF triples (subject, predicate, object):** combination of:
  - The resource described **(Subject)**
  - The property or relation between the subject and the object **(Predicate)**
  - Value of the property **(Object),** which is a resource or a literal
  - Each resource or property has an identifier (URI)



Types of graph databases

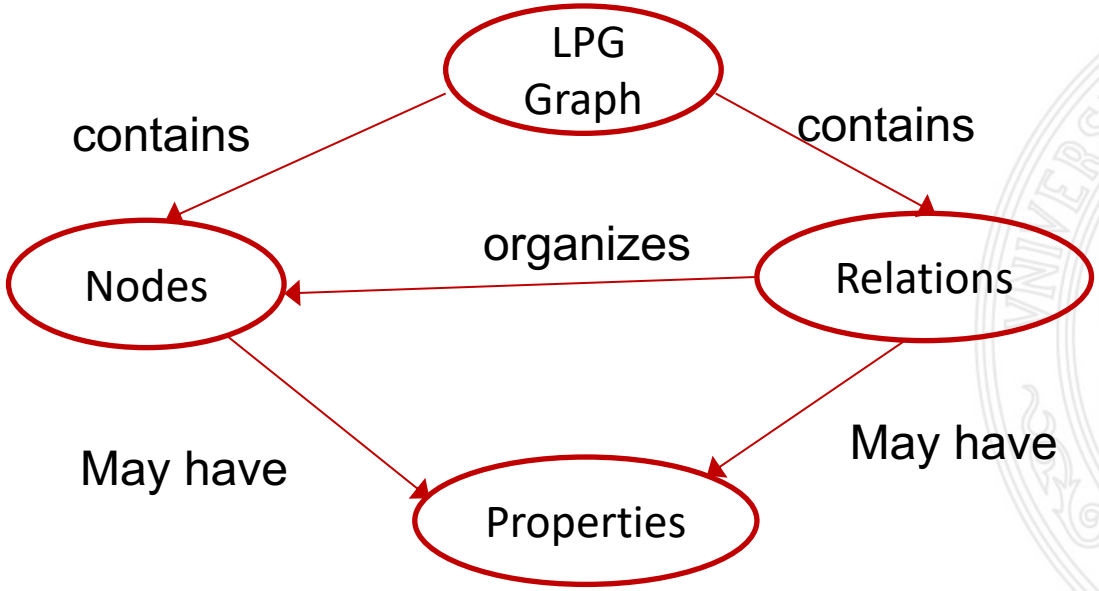
UNIVERSIDAD DE MURCIA

- There are different types of graph databases.
- The difference is in how graph are modeled what can be associated with a node, with an edge, how each element in the graph is interpreted
- Types
  - Labeled Property Graph (LPG)
  - Resource Description Framework Graph (RDF Graph)

Labeled property graph

UNIVERSIDAD DE MURCIA

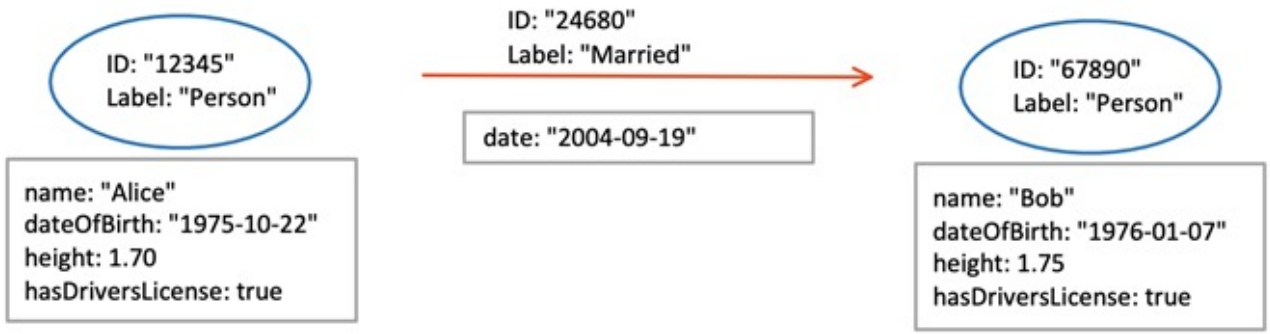
- Nodes, edges, properties and labels



Labeled property graph

UNIVERSIDAD DE MURCIA

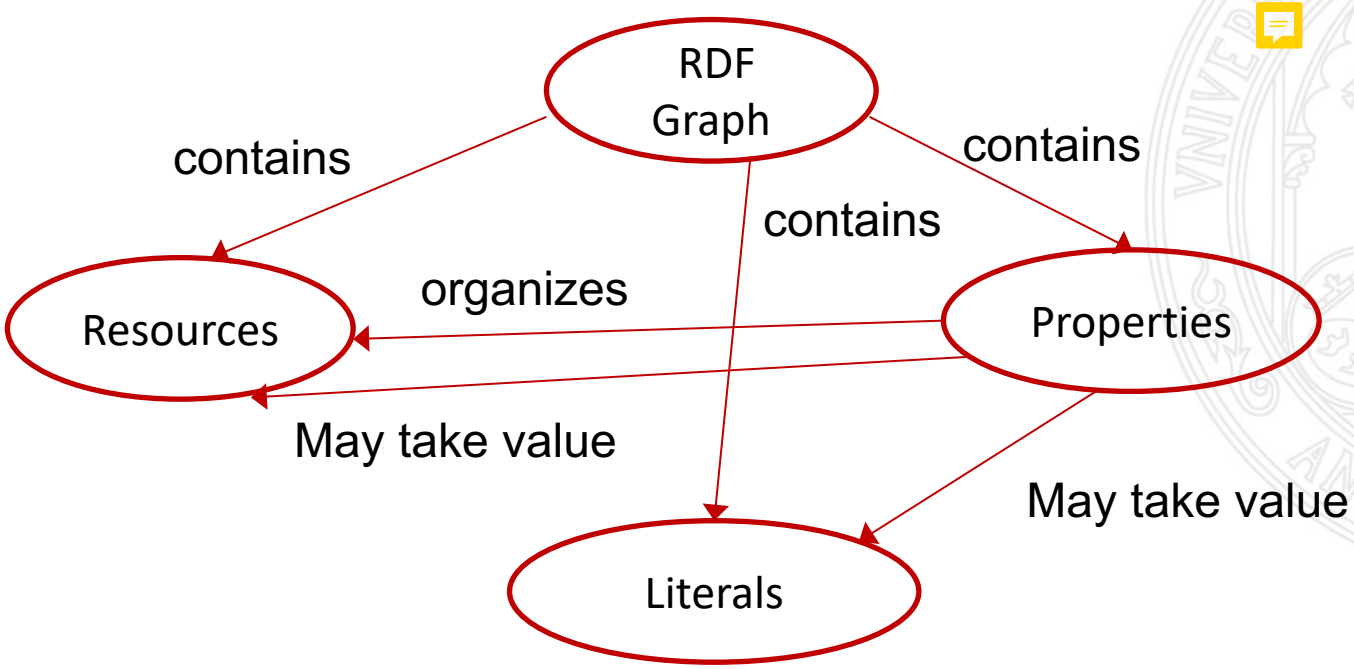
- Node (Vertex): identifier, label (type), set of properties <attribute, value>
- Relations (Edges): identifier, label (type), set of properties <attribute, value>



RDF graph

UNIVERSIDAD DE MURCIA

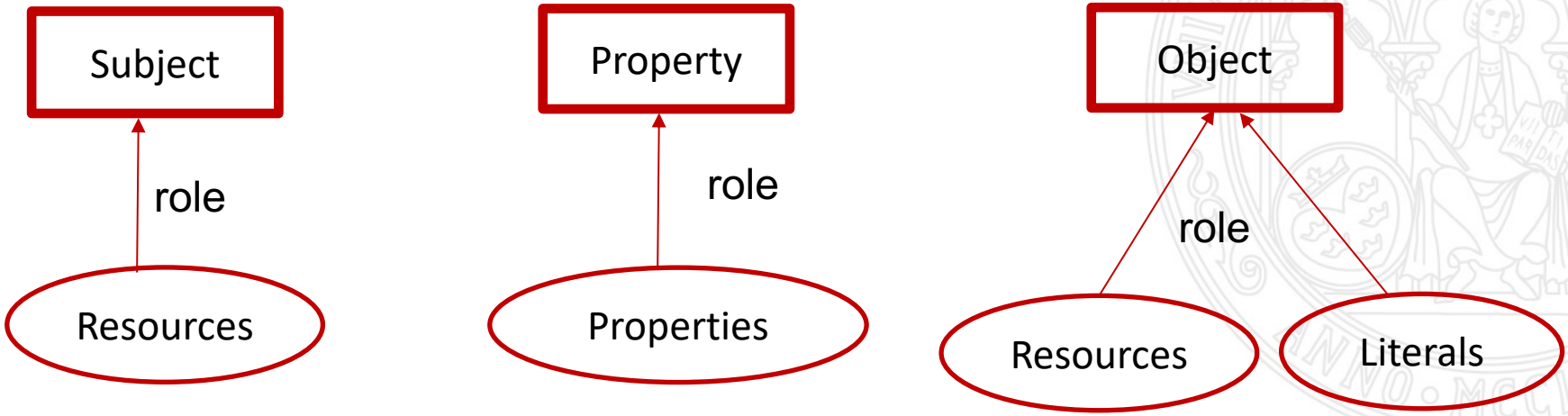
- RDF is a W3C recommendation for data exchange
- Resources, properties and literals (value)



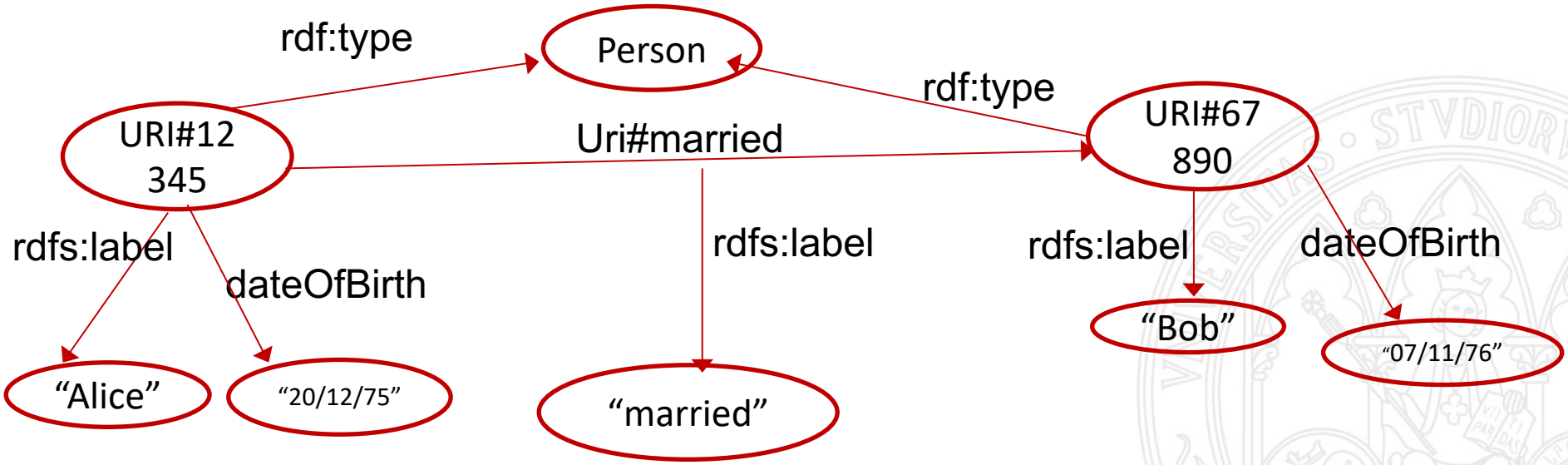
RDF graph

UNIVERSIDAD DE MURCIA

- Formal model based on RDF **statements**: triples <Subject, Property, Object>
- Resources and properties have universal identifiers (URI/IRI)
- RDF provides a set of standardized properties and resources to facilitate interoperability across graphs and data machine understanding
- Natural integration with ontologies, which proivide the terms for concepts and properties



Example of RDF graph



Properties do not have attributes, only annotations (metadata)

Comparison LPG-RDF

UNIVERSIDAD DE MURCIA

Aspect	LPG	RDF
Data model	Nodes, edges and properties	Triples <S,P,O>
Schema definition	Schema-less, schema-flexible	Not oriented to be schema-les. Schema based on RDFS and OWL.
Query language	Database-dependent	SPARQL, W3C recommendation
Formats	Database-dependent	Formats based on standards: XML, JSON, Turtle
Data integration	Do it yourself!	Formalism oriented to data exchange and interoperability
Federated queries	No native support	Native support for distributed queries
Scalability	Database-dependent	Designed for billions of data relations
Expressiveness	Allows attributes for properties	Explicit meaning of data
Graph processing	Facilitates graph analysis, paths finding (Property Path Discovery)	No native support

Stardog

- https://www.stardog.com/platform
- Oriented to RDF knowledge graphs
- Query language: SPARQL
- Cloud support and deployment
- API, multiple programming connectors

Amazon Neptune

UNIVERSIDAD DE MURCIA

- https://aws.amazon.com/es/neptune/
- LPG and RDF graphs
- Query languages: Gremlin (Apache TinkerPop) and Cypher (Neo4J) for LPG, SPARQL for RDF
- AWS Cloud support and deployment
- API for the use of the database

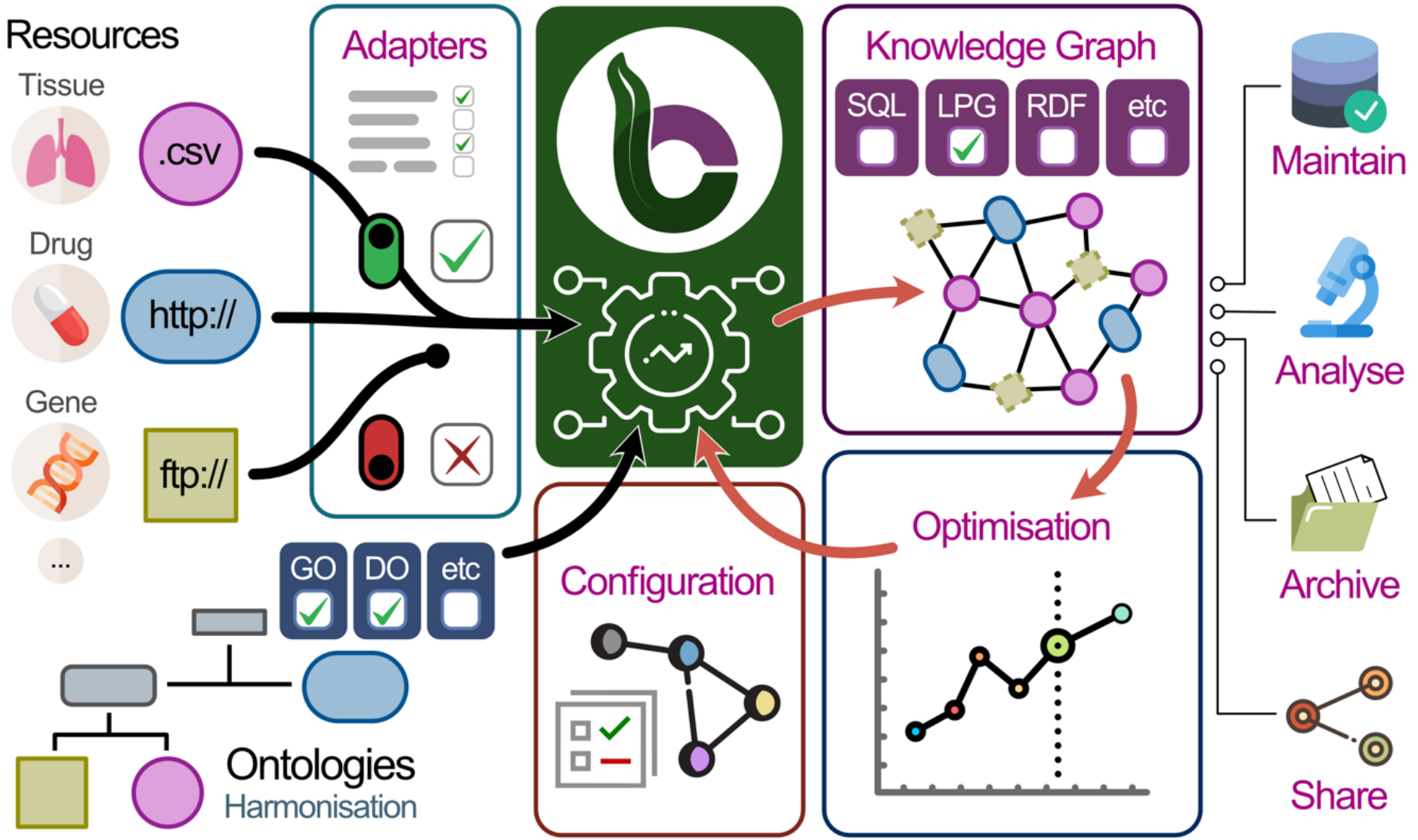
Neo4J

UNIVERSIDAD DE MURCIA

- https://neo4j.com
- LPG graphs, with plugins to import RDF graphs
- Query language: Cypher
- Free, Enterprise and cloud versions(AuraDB)
- API and connectors for big data processing tools: Kafka, Spark, etc.

BioCypher: https://biocypher.org

UNIVERSIDAD DE MURCIA



Questions, comments…

UNIVERSIDAD DE MURCIA