

Graph Theoretic Characterization of Metadata Structures

**Frank Olken,
Bruce Bargmeyer,
Kevin D. Keck**

Lawrence Berkeley National Laboratory
Computational Research Division

presented to ISO JTC1 SC32 WG2 Mtg.
Washington, DC
Nov. 8, 2004

Our Goal

- Create a taxonomy of metadata structures:
 - Terminologies, Taxonomies, partonomies, ontologies, ...
- Taxonomy will be framed in graph theoretic terms:
 - What kind of graph would be used to represent a given metadata structure

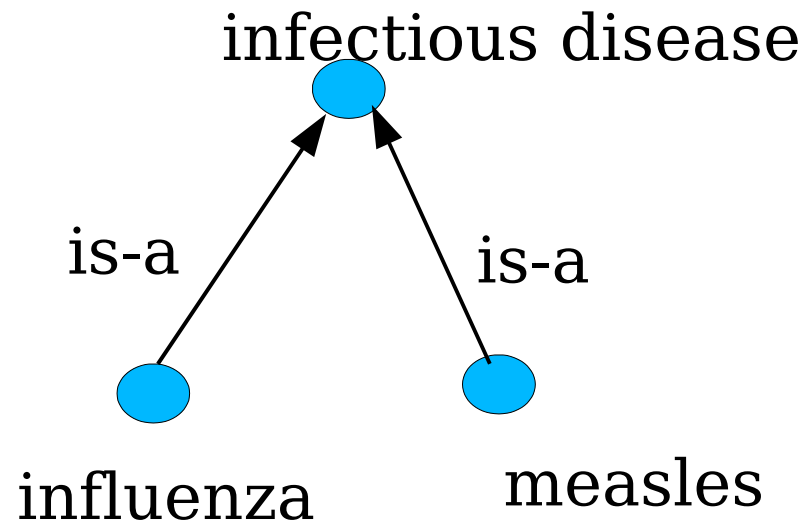
Why use a graph theoretic characterization?

- Readily comprehensible characterization of metadata structures
- Graph structure has implications for:
 - Integrity Constraint Enforcement
 - Data structures
 - Query languages
 - Combining metadata sets
 - Algorithms for query processing

Definition of a graph

- Graph = vertex (node) set + edge set
- Nodes, edges may be labeled
- Edge set = binary relation over nodes
 - cf. NIAM
- Labeled edge set
 - RDF triples (subject, predicate, object)
 - predicate = edge label
- Typically edges are directed

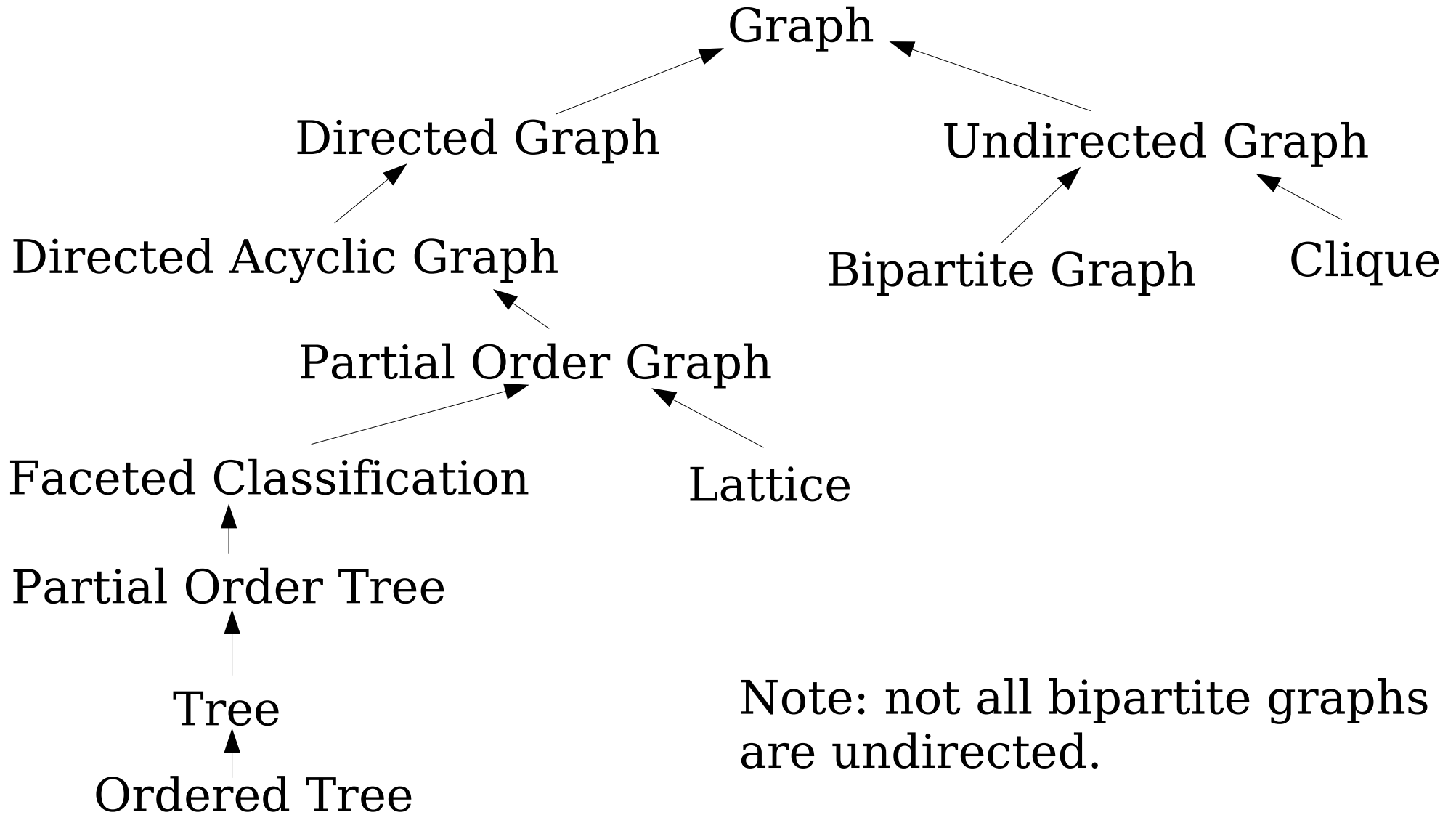
Example of a graph



Types of Metadata Graph Structures

- Trees
- Partially Ordered Trees
- Ordered Trees
- Faceted Classifications
- Directed Acyclic Graphs
- Partially Ordered Graphs
- Lattices
- Bipartite Graphs
- Directed Graphs
- Cliques
- Compound Graphs

Graph Taxonomy

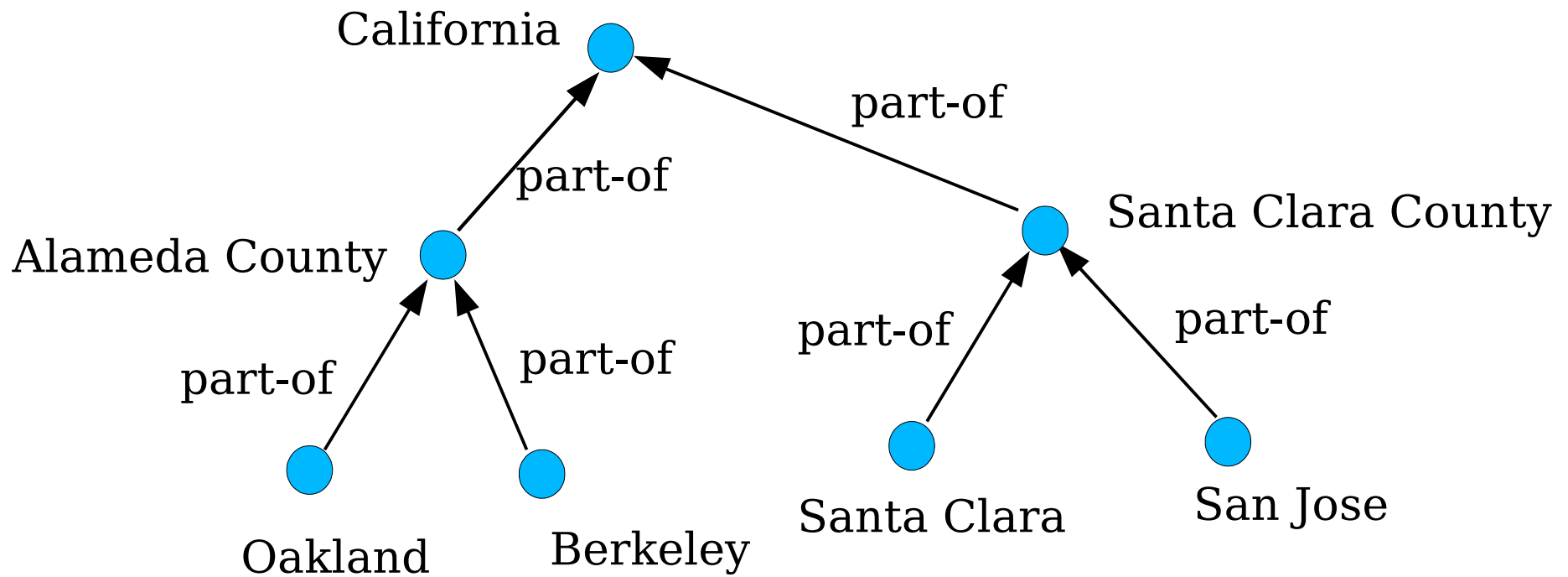


Note: not all bipartite graphs are undirected.

Trees

- In metadata settings trees are almost invariably directed
 - edges indicate direction
- In metadata settings trees are usually partial orders
 - Transitivity is implied (see next slide)
 - Not true for some trees with mixed edge types.
 - Not always true for all paronomies

Example: Tree



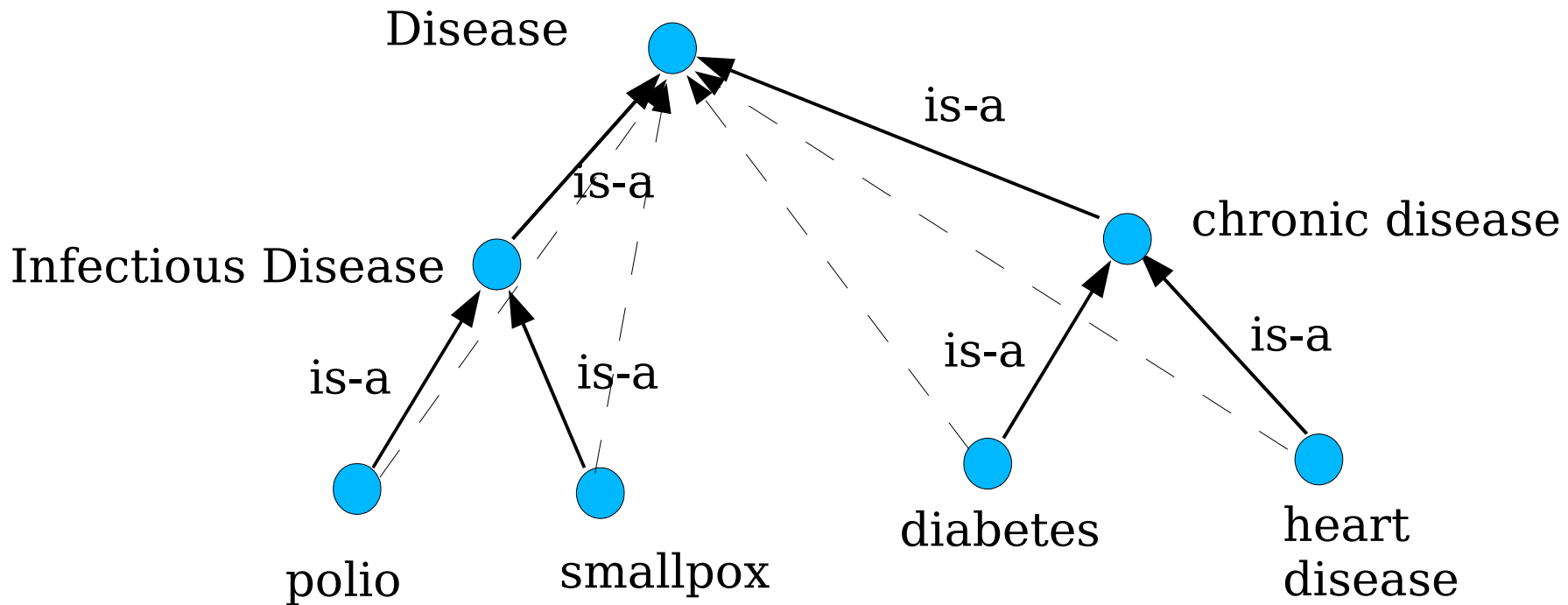
Trees - cont.

- Uniform vs. non-uniform height subtrees
- Uniform height subtrees
 - fixed number of levels
 - common in dimensions of multi-dimensional data models
- Non-uniform height subtrees
 - common terminologies

Partially Ordered Trees

- A conventional directed tree
- Plus, assumption of transitivity
- Usually only show immediate ancestors (transitive reduction)
- Edges of transitive closure are implied
- Classic Example:
 - Simple Taxonomy, “is-a” relationship

Example: Partial Order Tree

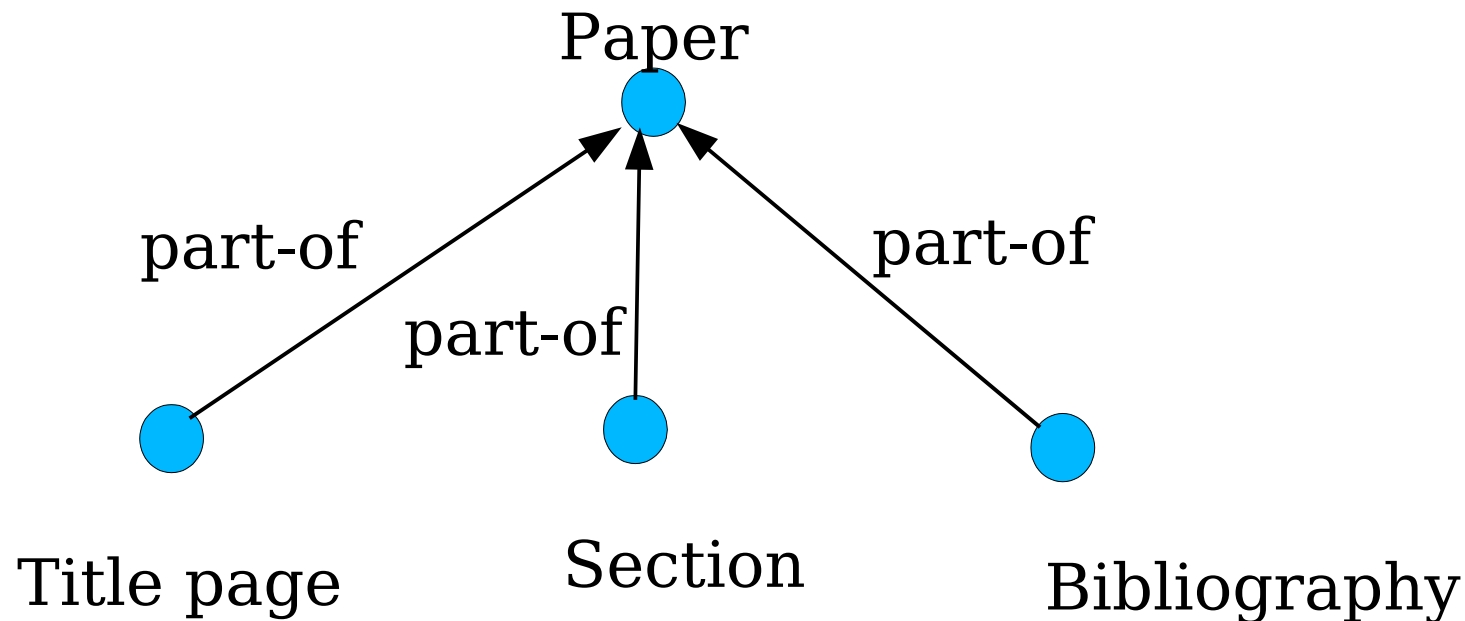


— — — ► Signifies inferred is-a relationship

Ordered Trees

- Order here refers to order among sibling nodes (not related to partial order discussed elsewhere)
- XML documents are ordered trees
 - Ordering of “sub-elements” is to support classic linear encoding of documents

Example: Ordered Tree



Note: implicit ordering relation among parts of paper.

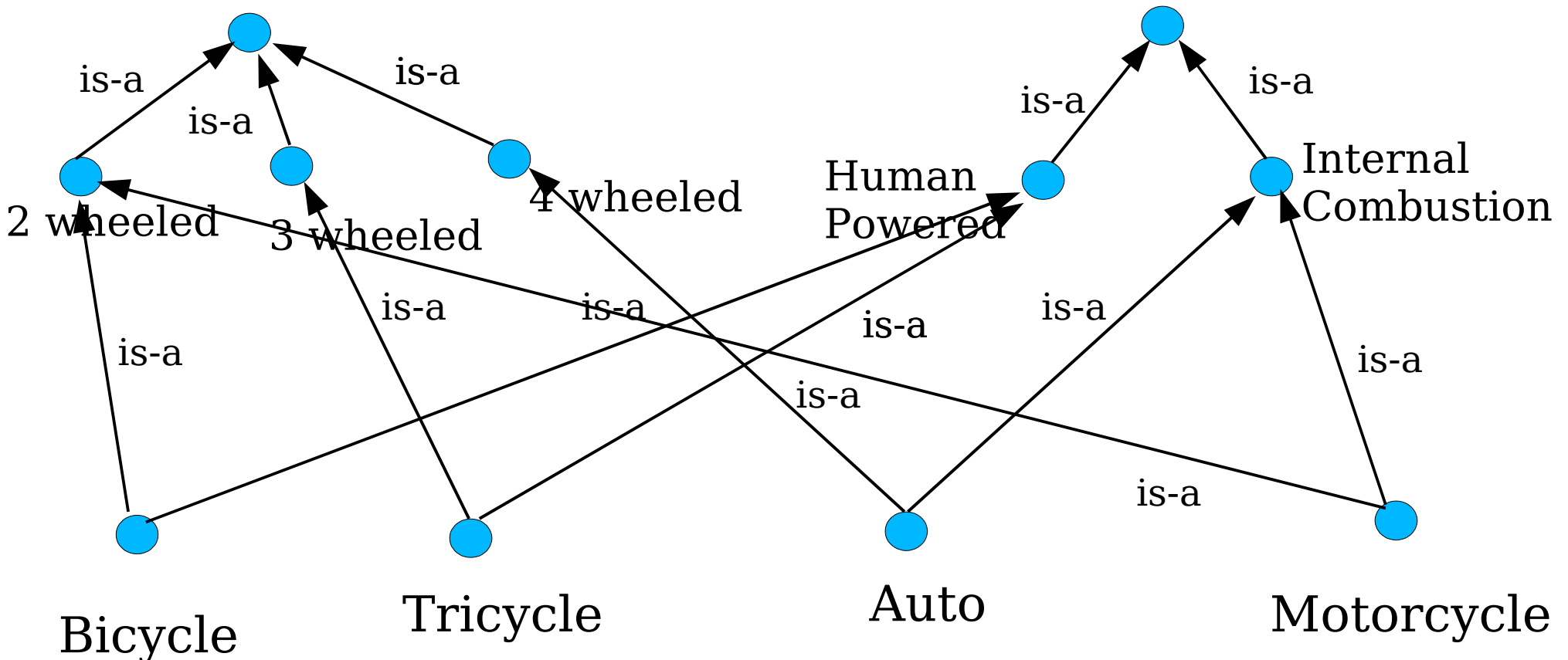
Faceted Classification

- Classification scheme has multiple facets
- Each facet = partial order tree
- Categories = conjunction of facet values (often written as [facet1, facet2, facet3])
- Faceted classification = a simplified partial order graph
- Introduced by Ranganathan in 19th century, as Colon Classification scheme
- Faceted classification can be described with Description Logic, e.g., OWL-DL

Example: Faceted Classification

Wheeled Vehicle Facet

Vehicle Propulsion Facet



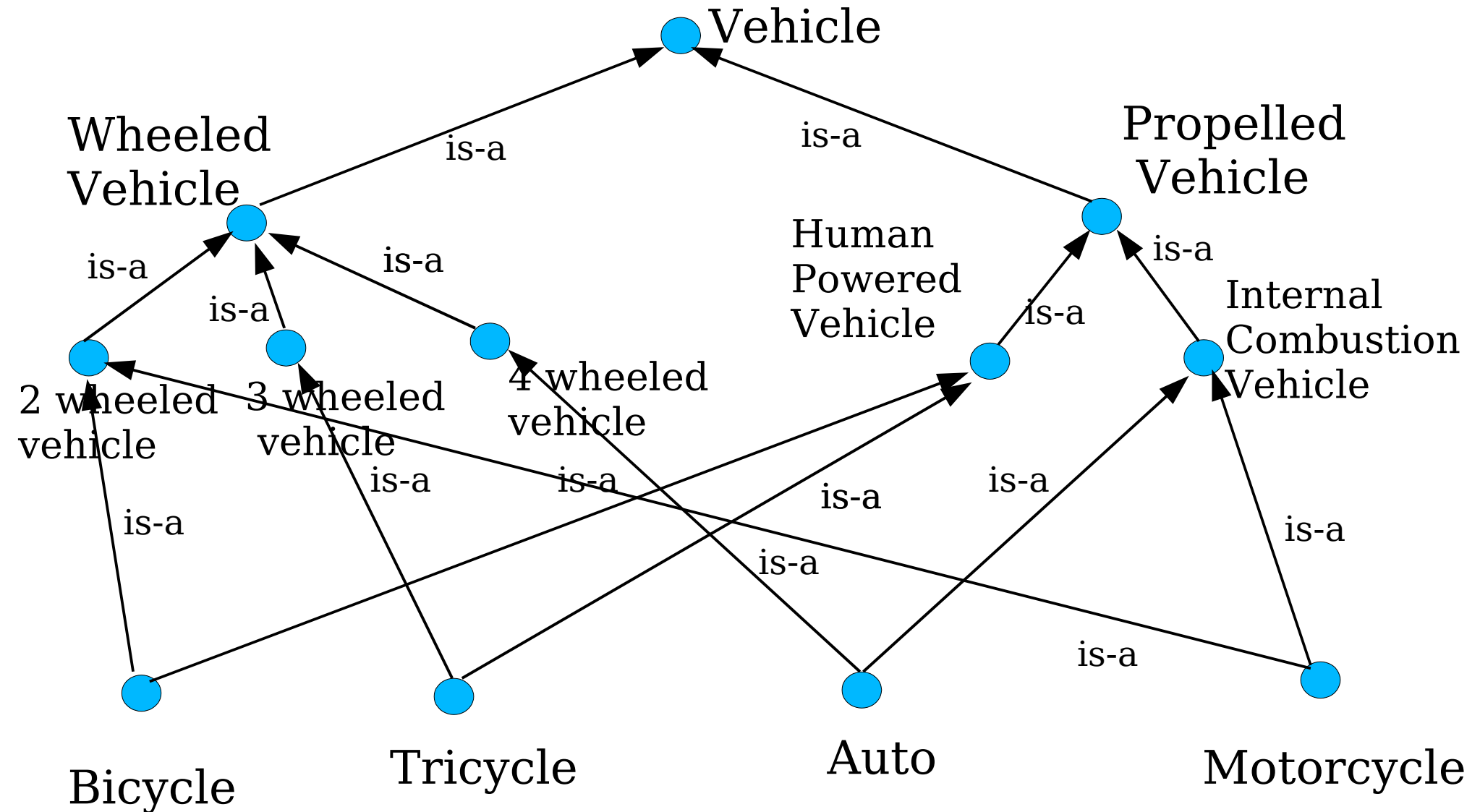
Faceted Classifications and Multi-dimensional Data Model

- MDM – a.k.a. OLAP data model
 - Online Analytical Processing data model
 - Star / Snowflake schemas
- Fact Tables
 - fact = function over Cartesian product of dimensions
 - dimensions = facets
 - geographic region, product category, year, ...

Directed Acyclic Graphs

- Graph:
 - Directed edges
 - No cycles
 - No assumptions about transitivity (e.g., mixed edge types, some paronomies)
 - Nodes may have multiple parents
- Examples:
 - Paronomies (“part-of”) - transitivity is not always true

Example: Directed Acyclic Graph

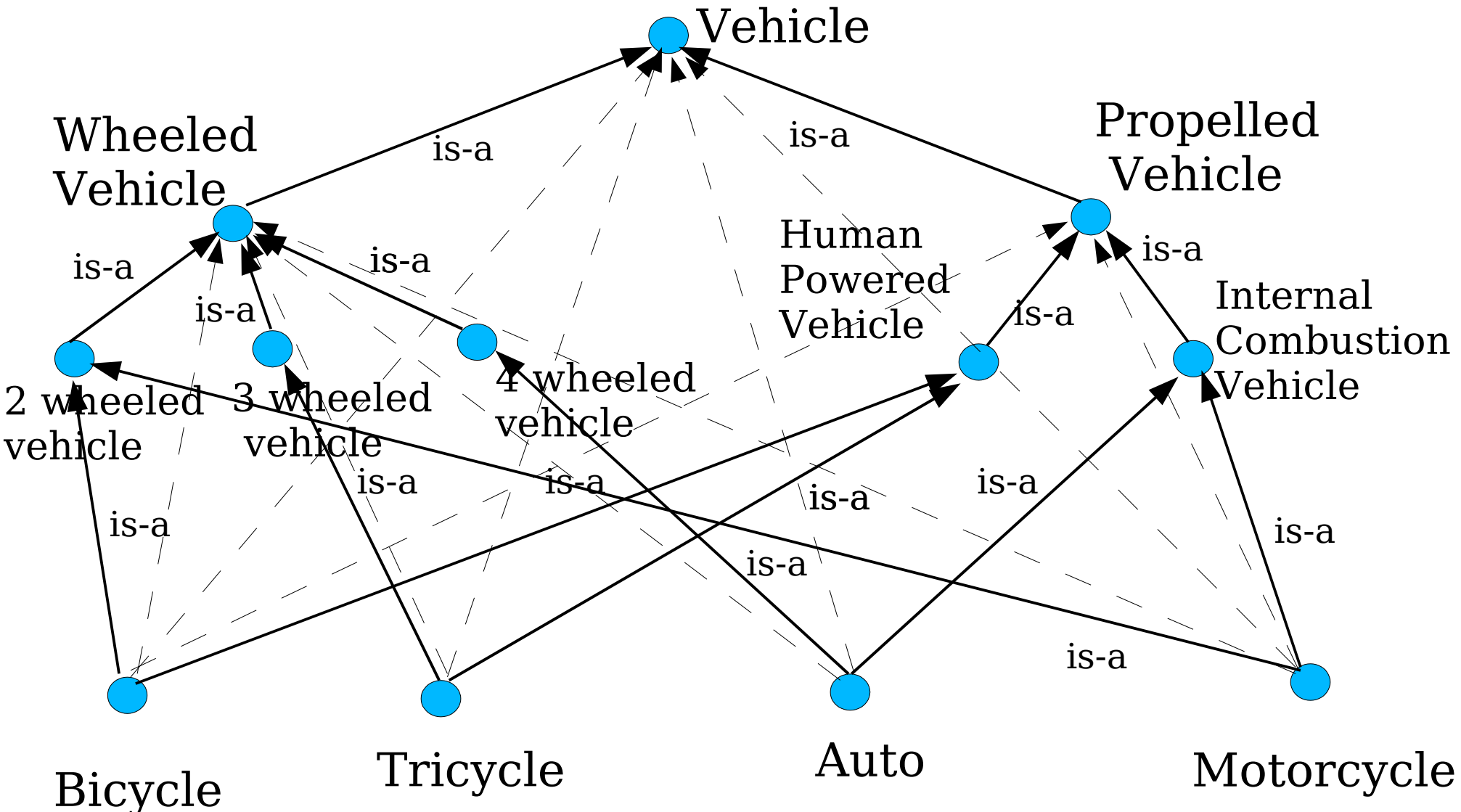


Observe: multiple inheritance.

Partial Order Graphs

- Directed acyclic graphs + inferred transitivity
- Nodes may have multiple parents
- Most taxonomies drawn as transitive reduction, transitive closure edges are implied.
- Examples:
 - all taxonomies
 - most partonomies
 - multiple inheritance
- POGs can be described in Description Logic, e.g., OWL-DL

Example: Partial Order Graph



— — — — — ▶ Dashed line = inferred is-a (transitive closure)

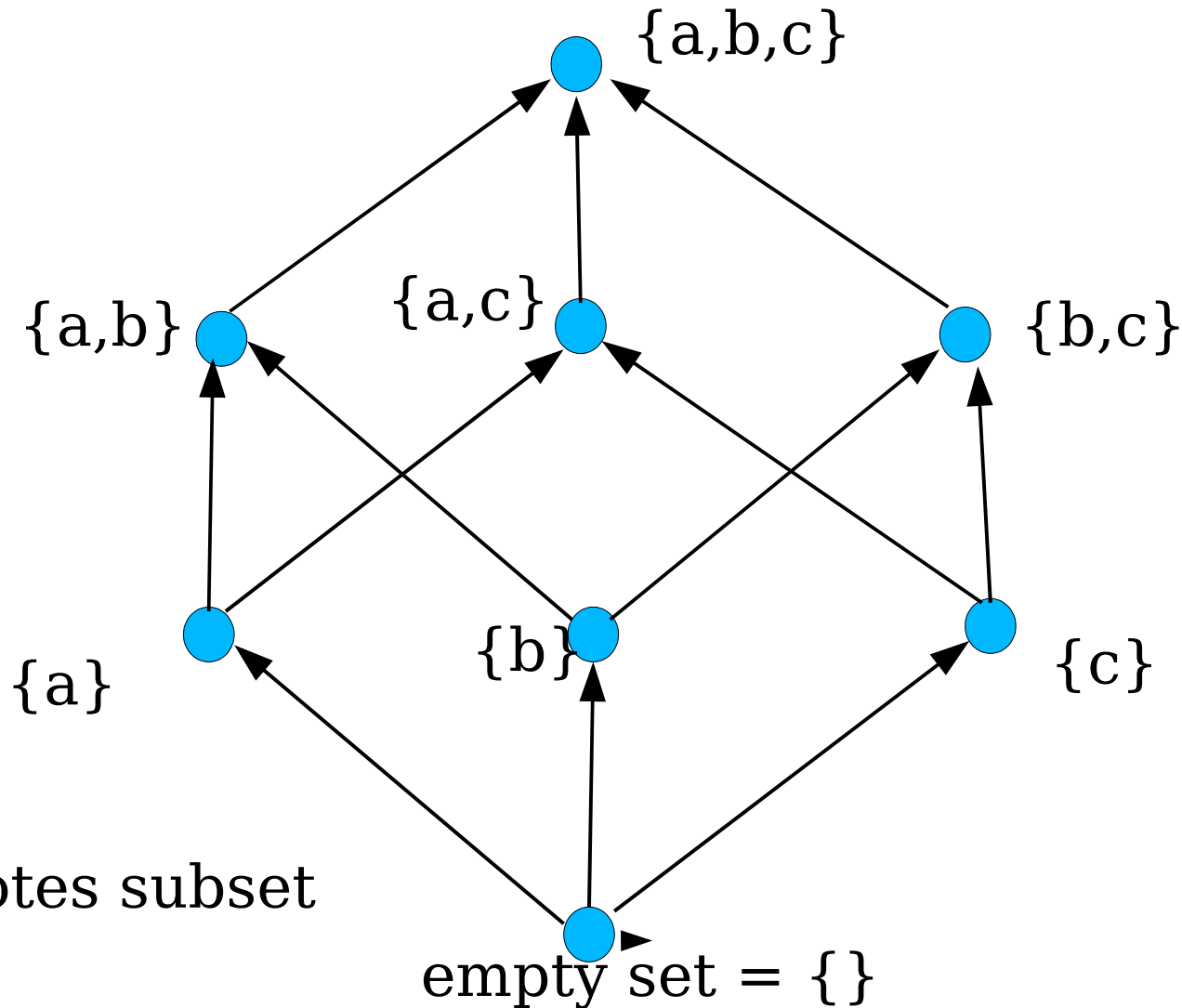
Directed Graph

- Generalization of DAG (directed acyclic graph)
- Cycles are allowed
- Arises when many edge types allowed
- Example: UMLS

Lattices

- A partial order
- For every pair of elements A and B
 - There exists a least upper bound
 - There exists a greatest lower bound
- Example:
 - The power set (all possible subsets) of a finite set
 - $LUB(A,B)$ = union of two sets A, B
 - $GLB(A,B)$ = intersect of two sets A,B

Example Lattice: Powerset of 3 element set



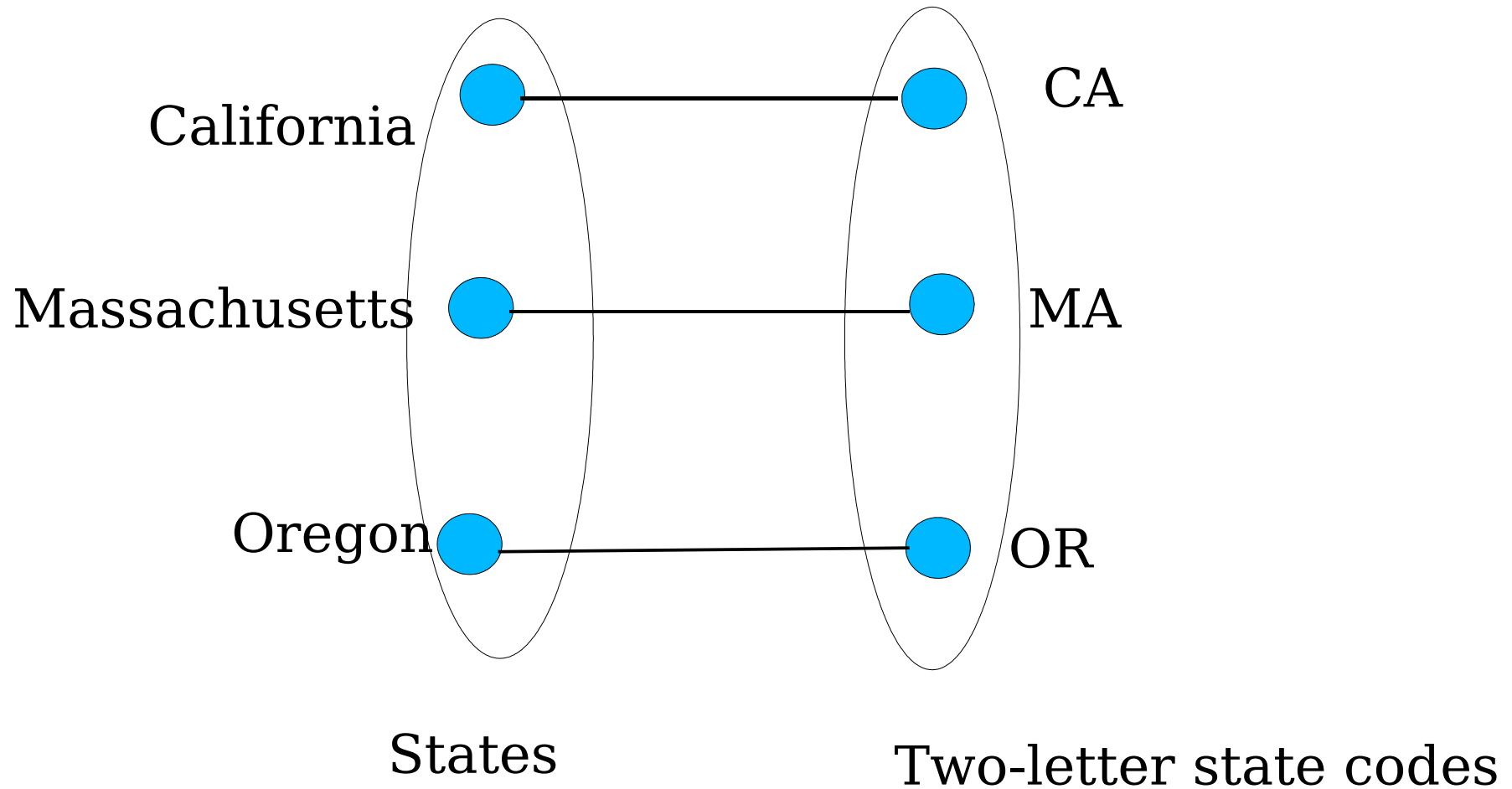
Lattices - Applications

- Formal Concept Analysis
 - synthesizing taxonomies
- Machine Learning
 - concept learning

Bipartite Graphs

- Vertices = two disjoint sets, V and W
- All edges connect one vertex from V and one vertex from W
- Examples:
 - mappings among value representations
 - mappings among schemas
 - (entity/attribute, relationship) nodes in Conceptual Graphs

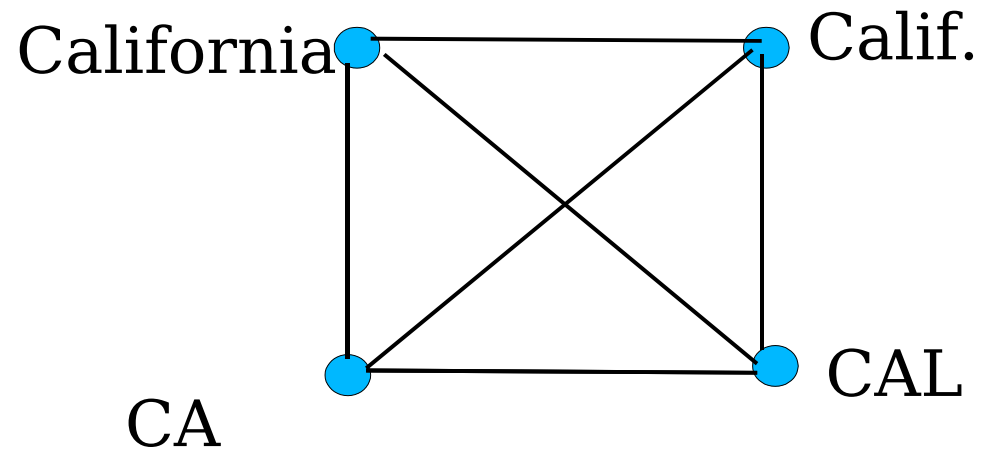
Example Bipartite Graph



Clique

- Clique = complete graph (or subgraph)
 - all possible edges are present
- Used to represent equivalence classes
- Typically, on undirected graphs

Example of Clique

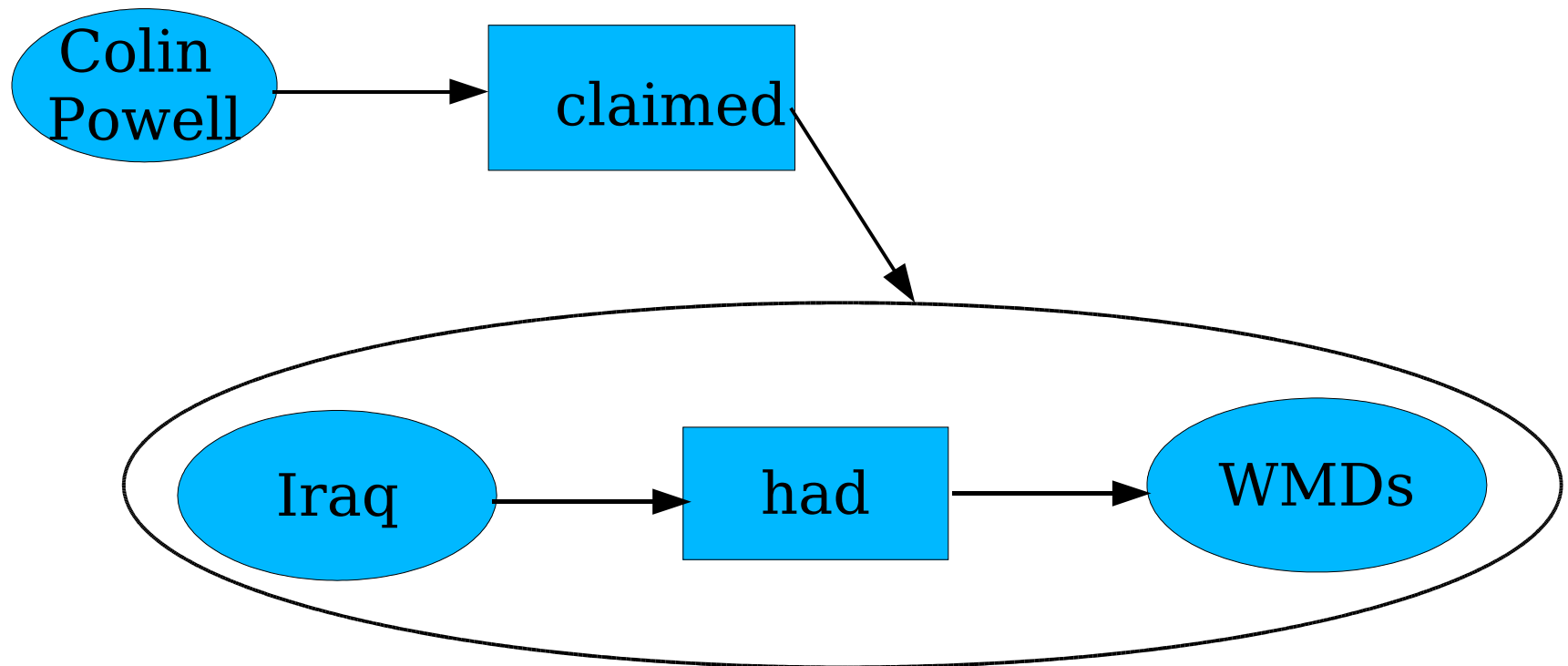


Here edges denote synonymy.

Compound Graphs

- Edges can point to/from subgraphs, not just simple nodes
- Used in conceptual graphs
- CG is isomorphic to First Order Logic
- Could be used to specify contexts for subgraphs

Example Compound Graph



Conclusions

- We can characterize metadata structure in terms of graph structures
- Partial Order Graphs are the most common structure:
 - used for taxonomies, partonomies
 - support multiple inheritance, faceted classification
 - implicit inclusion of inferred transitive closure edges

Additional Resources

- Faceted classification
 - <http://www.kmconnection.com/DOC100100.htm>
- Lattices
 - http://www.rci.rutgers.edu/~cfs/472_html/Learn/SimpleLattice_472.html

Contact Information

- Frank Olken
 - olken@lbl.gov, 510-486-5891
 - <http://www.lbl.gov/~olken>
- XMDR Project
 - <http://hpcrd.lbl.gov/SDM/XMDR/>
- Bruce Bargmeyer
 - bebargmeyer@lbl.gov
- Kevin D. Keck
 - <mailto:kdkeck@lbl.gov>