

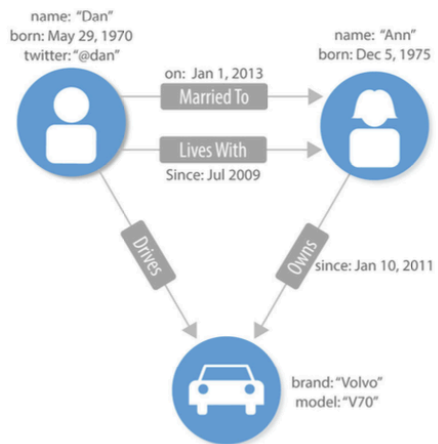
Graph Data Models

Wednesday, February 19, 2025 9:29 AM

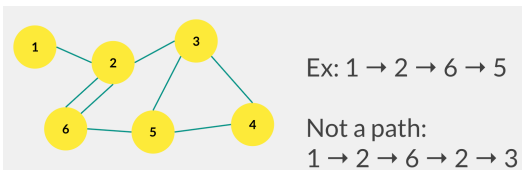
- Data model based on the graph data structure
- Composed of nodes and edges
 - edges connect nodes
 - each is uniquely identified
 - each can contain properties (e.g. name, occupation, etc)
 - supports queries based on graph-oriented operations
 - traversals
 - shortest path
 - lots of others

Labeled Property Graph

- Composed of a set of node (vertex) objects and relationship (edge) objects
- Labels are used to mark a node as part of a group
- Properties are attributes (think KV pairs) and can exist on nodes and relationships
- Nodes with no associated relationships are OK.
Edges not connected to nodes are not permitted.

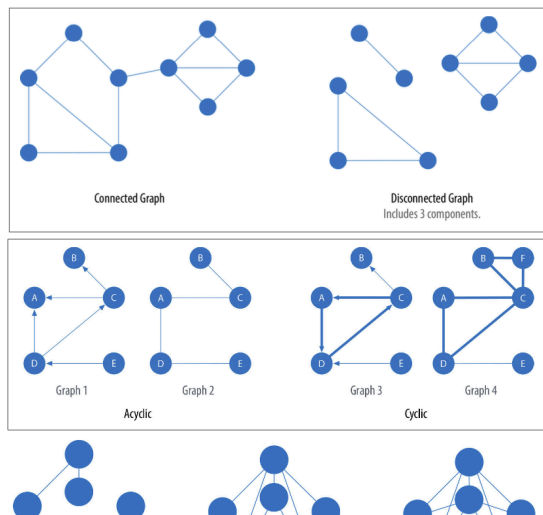


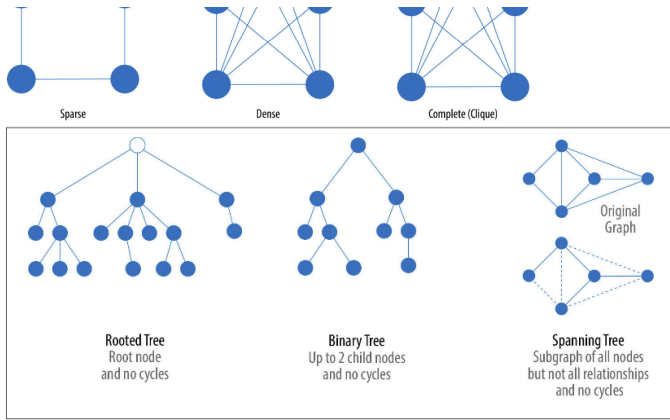
A **path** is an ordered sequence of nodes connected by edges in which no nodes or edges are repeated.



Flavors of Graphs

- **Connected (vs. Disconnected)** there is a path between any two nodes in the graph
- **Weighted (vs. Unweighted)** – edge has a weight property (important for some algorithms)
- **Directed (vs. Undirected)** – relationships (edges) define a start and end node
- **Acyclic (vs. Cyclic)** – Graph contains no cycles

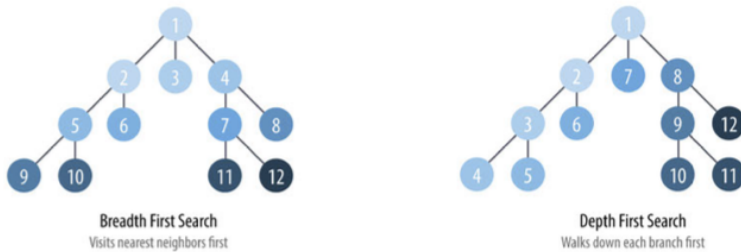




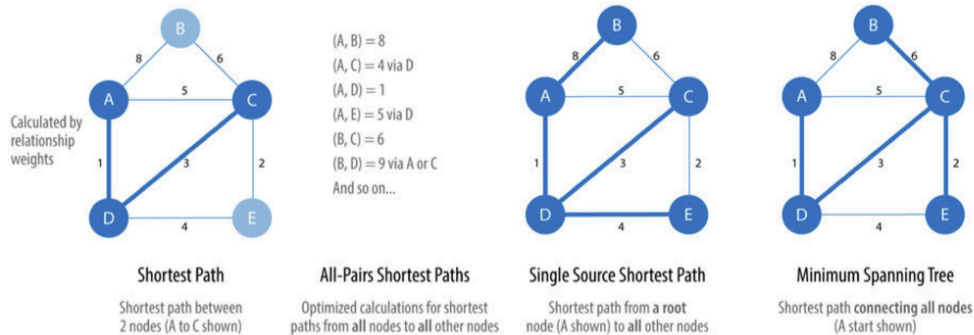
Pathfinding

- finding the shortest path between two nodes, if one exists, is probably the most common operation
- "shortest" means fewest edges or lowest weight
- Average Shortest Path* can be used to monitor efficiency and resiliency of networks.
 - Minimum spanning tree, cycle detection, max/min flow...* are other types of pathfinding

BFS vs DFS

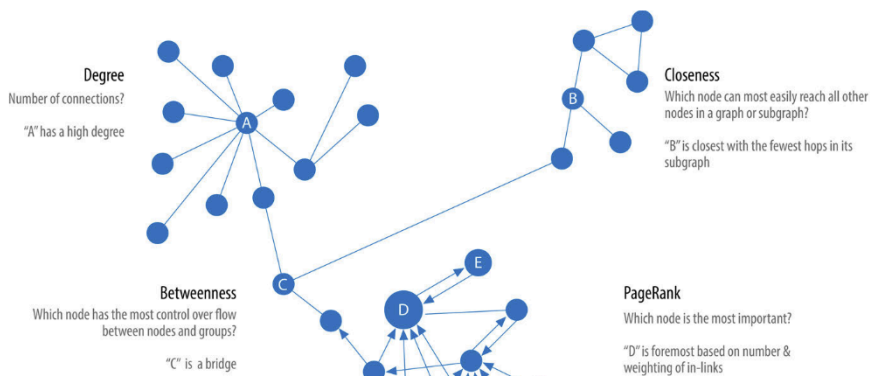


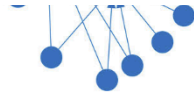
Shortest Path



- Centrality**
 - determining which nodes are "more important" in a network compared to other nodes
 - EX: Social Network Influencers?
- Community Detection**
 - evaluate clustering or partitioning of nodes of a graph and tendency to strengthen or break apart

Centrality





"E" is next, due to the influence of D's link

Some Famous Graph Algos

- **Dijkstra's Algorithm** - single-source shortest path algo for positively weighted graphs
- **A* Algorithm** - Similar to Dijkstra's with added feature of using a heuristic to guide traversal
- **PageRank** - measures the importance of each node within a graph based on the number of incoming relationships and the importance of the nodes from those incoming relationships

Neo4j

- A Graph Database System that supports both transactional and analytical processing of graph-based data
- Relatively new class of no-sql DBs
- Considered schema optional (one can be imposed)
- Supports various types of indexing
- ACID compliant
- Supports distributed computing
- Similar: Microsoft CosmosDB, Amazon Neptune