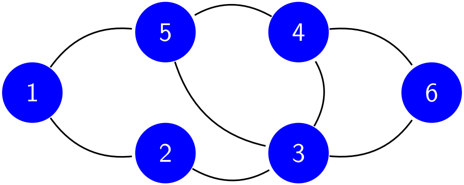
1.022 Introduction to Network Models

Amir Ajorlou

Laboratory for Information and Decision Systems
Institute for Data, Systems, and Society
Massachusetts Institute of Technology

Lecture 2

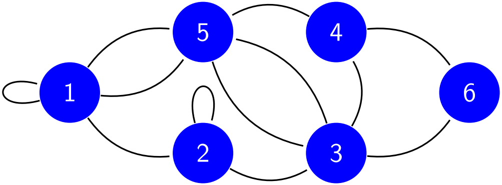
Graphs



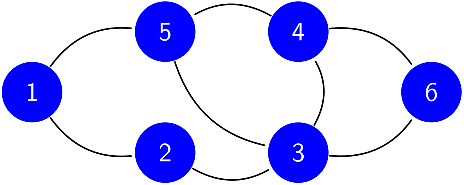
- Graph $G(V, E)$ $\Rightarrow$ A set V of vertices or nodes
  $\Rightarrow$ Connected by a set E of edges or links
  $\Rightarrow$ Elements of E are unordered pairs (u, v), $u, v \in V$
- In figure $\Rightarrow$ Vertices are $V = \{1, 2, 3, 4, 5, 6\}$
  $\Rightarrow$ Edges $E = \{(1, 2), (1, 5), (2, 3), (3, 4), \ldots$
  $(3, 5), (3, 6), (4, 5), (4, 6)\}$

Simple and multi-graphs

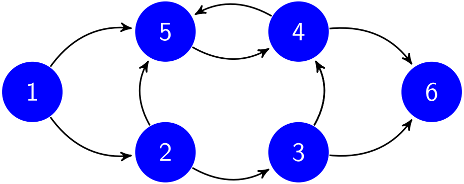
- In general, graphs may have self-loops and multi-edges
  - ⇒ A graph with either is called a multi-graph



- Mostly work with simple graphs, with no self-loops or multi-edges



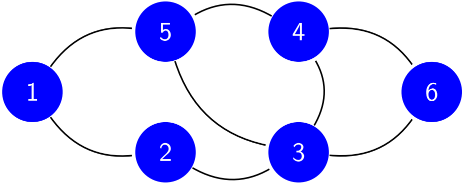
Directed graphs



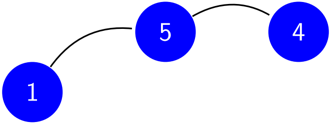
- In directed graphs, elements of E are ordered pairs (u, v), $u, v \in V$
  - $\Rightarrow$ Means (u, v) distinct from (v, u)
- Directed graphs often called digraphs
  - $\Rightarrow$ By convention (u, v) points to v
  - $\Rightarrow$ If both $\{(u, v), (v, u)\} \subseteq E$, the edges are said to be mutual
- Ex: who-calls-whom phone networks, Twitter follower networks

Subgraphs

- Consider a given graph $G(V, E)$



- **Def:** Graph $G'(V', E')$ is an induced subgraph of G if $V' \subseteq V$ and $E' \subseteq E$ is the collection of edges in G among that subset of vertices
- Ex: Graph induced by $V' = \{1, 4, 5\}$



Weighted graphs

- Oftentimes one labels edges with numerical values

  $\Rightarrow$ Such graphs are called weighted graphs

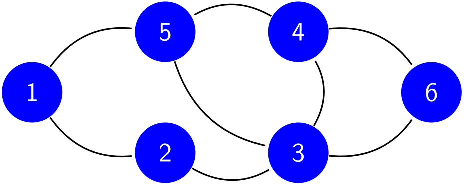
- Typical network representations:

Network	**Graph representation**
WWW	Directed multi-graph (with loops), unweighted
Facebook friendships	Undirected, unweighted
Citation network	Directed, unweighted, acyclic
Collaboration network	Undirected, unweighted
Mobile phone calls	Directed, weighted
Protein interaction	Undirected multi-graph (with loops), unweighted
⋮	⋮

- Note that multi-edges are often encoded as edge weights (counts)

Adjacency

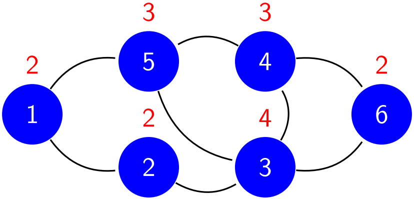
- Useful to develop a language to discuss the connectivity of a graph
- A simple and local notion is that of adjacency
  - $\Rightarrow$ Vertices $u, v \in V$ are said adjacent if joined by an edge in E
  - $\Rightarrow$ Edges $e_1, e_2 \in E$ are adjacent if they share an endpoint in V



- In figure $\Rightarrow$ Vertices 1 and 5 are adjacent; 2 and 4 are not
  $\Rightarrow$ Edge $(1, 2)$ is adjacent to $(1, 5)$, but not to $(4, 6)$

Degree

- An edge (u, v) is incident with the vertices u and v
- **Def:** The degree d_v of vertex v is its number of incident edges



- In figure $\Rightarrow$ Vertex degrees shown in red, e.g., $d_1 = 2$ and $d_5 = 3$
- High-degree vertices likely influential, central, prominent.
- The neighborhood $\mathcal{N}_i$ of a node i is the set of all its adjacent nodes

  $\Rightarrow \mathcal{N}_5 = \{1, 3, 4\}$ $\Rightarrow$ In general, $|\mathcal{N}_i| = d_i$

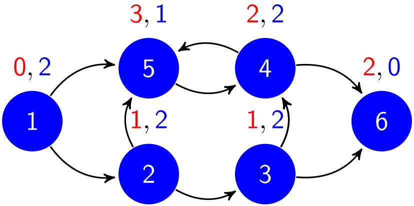
Properties and observations about degrees

- Degree values range from 0 to $|V| - 1$
- The sum of the degree sequence is twice the size of the graph

$$\sum_{v=1}^{|V|} d_v = 2|E|$$

$\Rightarrow$ The number of vertices with odd degree is even

- In digraphs, we have vertex in-degree d_v^{in} and out-degree d_v^{out}



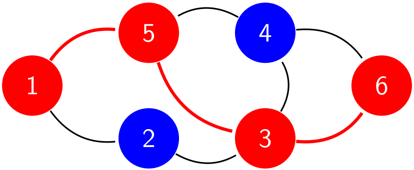
- In figure $\Rightarrow$ Vertex in-degrees shown in red, out-degrees in blue
  $\Rightarrow$ For example, $d_1^{in} = 0, d_1^{out} = 2$ and $d_5^{in} = 3, d_5^{out} = 1$

Movement in a graph

- A path of length l from v_0 to v_l is a consecutive sequence of **distinct** vertices

$$\{v_0, v_1, \ldots, v_{l-1}, v_l\}, \text{ where } v_i \text{ and } v_{i+1} \text{ are adjacent}$$

- A Walk: vertices do not have to be distinct.



- A closed walk ($v_0 = v_l$) is called a circuit

  $\Rightarrow$ A closed path is a cycle

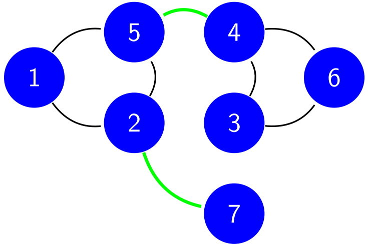
- All these notions generalize naturally to directed graphs

Distances in a graph

- Length of a path $\Rightarrow$ is the sum of the weights of traversed edges
- The distance between two nodes i and j is the length of the shortest path linking i and j
  - $\Rightarrow$ In the absence of such a path, the distance is ∞
  - $\Rightarrow$ The diameter of the graph is the value of the largest distance
- There exist efficient algorithms to compute distances in graphs
  - $\Rightarrow$ Dijkstra, Floyd-Warshall, Johnson, ...

Connectivity

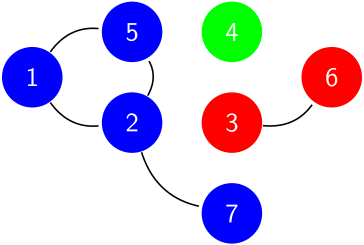
- Vertex v is reachable from u if there exists a $u - v$ path
- **Def:** Graph is connected if every vertex is reachable from every other



- If bridge edges are removed, the graph becomes disconnected

Connected components

- **Def:** A component is a maximally connected subgraph
  $\Rightarrow$ Maximal means adding a vertex will ruin connectivity

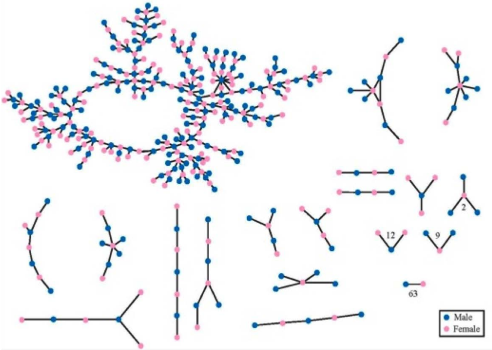


- In figure $\Rightarrow$ Components are $\{1, 2, 5, 7\}$, $\{3, 6\}$ and $\{4\}$
  $\Rightarrow$ Subgraph $\{3, 4, 6\}$ not connected, $\{1, 2, 5\}$ not maximal

- Disconnected graphs have 2 or more components
  $\Rightarrow$ Largest component often called giant component

Giant connected components

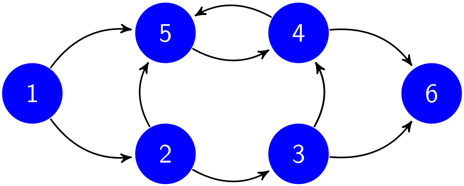
- Large real-world networks typically exhibit one giant component
- Ex: romantic relationships in a US high school [Bearman et al'04]



Bearman, Peter S., James Moody, and Katherine Stovel. "Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks."

- Q: Why do we expect to find a single giant component?
- A: It only takes one edge to merge two giant components

- Connectivity is more subtle with directed graphs. Two notions
- Digraph is strongly connected if for every pair $u, v \in V$, u is reachable from v (via a directed path) and vice versa
- Digraph is weakly connected if connected after disregarding edge directions, i.e., the underlying undirected graph is connected

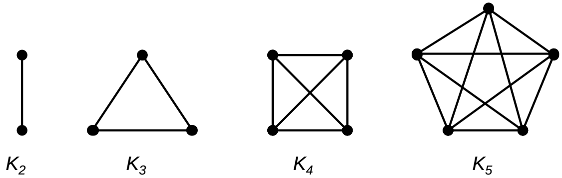


- Above graph is weakly connected but not strongly connected

  $\Rightarrow$ Strong connectivity implies weak connectivity

Complete graphs and cliques

- A complete graph K_n of order n has all possible edges

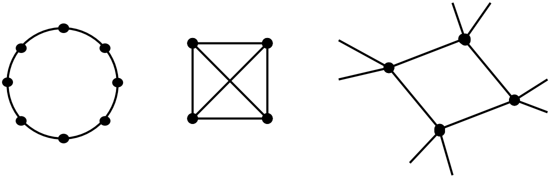


- Q: How many edges does K_n have?
- A: Number of edges in K_n = Number of vertex pairs $= \binom{n}{2} = \frac{n(n-1)}{2}$
- Of interest in network analysis are cliques, i.e., complete subgraphs

  $\Rightarrow$ Extreme notions of cohesive subgroups, communities

Regular graphs

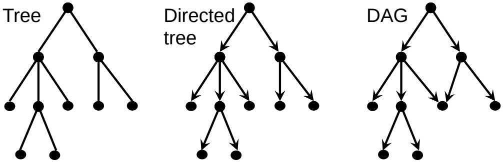
- A *d*-regular graph has vertices with equal degree d



- Naturally, the complete graph K_n is $(n-1)$-regular
  - $\Rightarrow$ Cycles are 2-regular (sub) graphs
- Regular graphs arise frequently in e.g.,
  - Physics and chemistry in the study of crystal structures
  - Geo-spatial settings as pixel adjacency models in image processing
  - Opinion formation, information cycles as regular subgraphs

Trees and directed acyclic graphs

- A tree is a connected acyclic graph
  - ⇒ A collection of trees is denominated a forest
- Ex: river network, information cascades in Twitter, citation network



- A directed tree is a digraph whose underlying undirected graph is a tree
  - ⇒ Rooted if paths from one vertex to all others
- Vertex terminology: parent, children, ancestor, descendant, leaf
- Underlying graph of a directed acyclic graph (DAG) need not be a tree