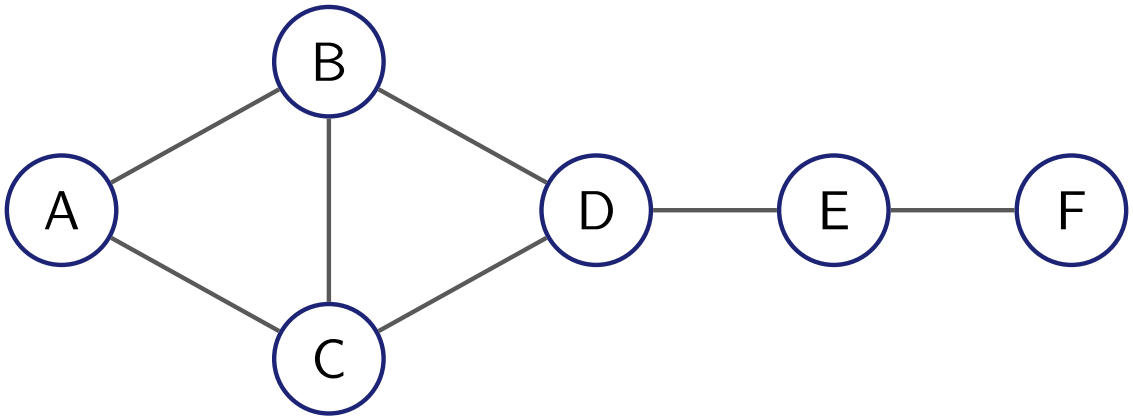
Graphs

Sehyun Kwon

Mathematical Data Science
Hanyang University ERICA

From relationships to graphs

Which computers can communicate? Which students can reach one another through mutual friends? Which pages link to a website?



A **graph** records objects and the connections between them.

We will use this language to understand how PageRank assigns importance to pages in a network.

Reading: Easley and Kleinberg, *Networks, Crowds, and Markets*, Sections 2.1–2.3 (selected topics).

Choosing the model

Before drawing a graph, decide what its nodes and edges mean.

Network	Nodes	An edge means
Friendship	Students	They are friends
Collaboration	Researchers	They coauthored a paper
Web	Pages	One links to the other

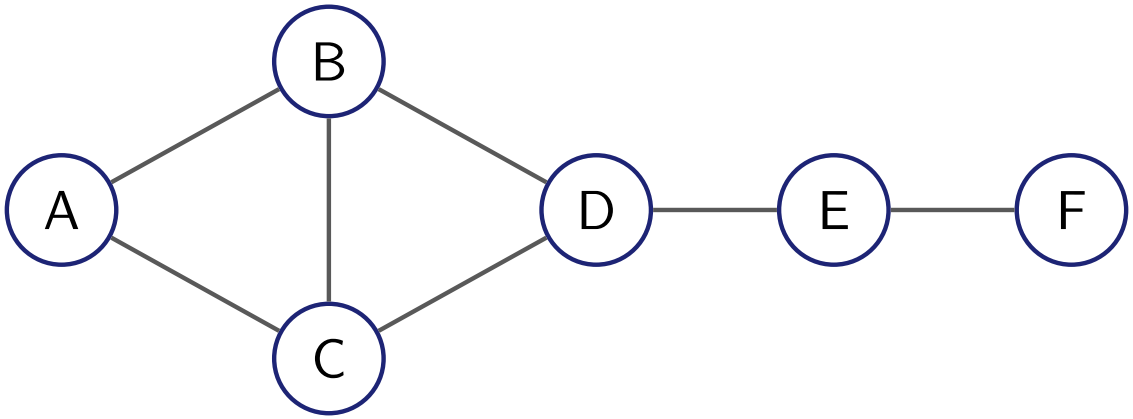
Friendship is modeled here as mutual. A hyperlink need not be. The choice between undirected and directed edges matters.

We first study **undirected graphs**.

Vertices and edges

An undirected graph is a pair $G = (V, E)$.

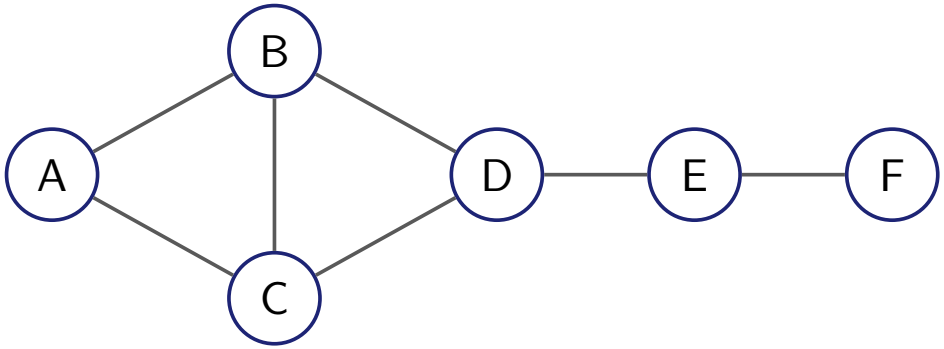
- V is the set of **vertices**, also called **nodes**.
- E is the set of **edges**: unordered pairs of vertices.

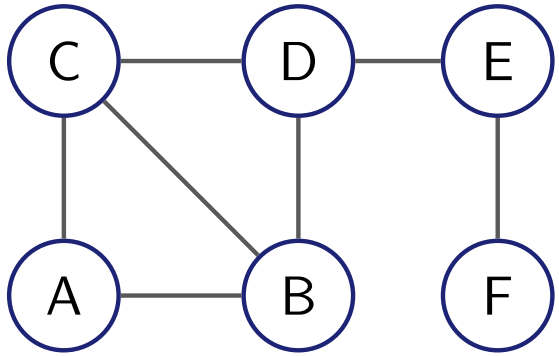


$$V = \{A, B, C, D, E, F\}, \qquad |V| = 6, \quad |E| = 7.$$

For example, $\{A, B\} \in E$, but $\{A, D\} \notin E$. Here we use finite simple graphs: no self-loops or repeated edges.

A graph is not its drawing



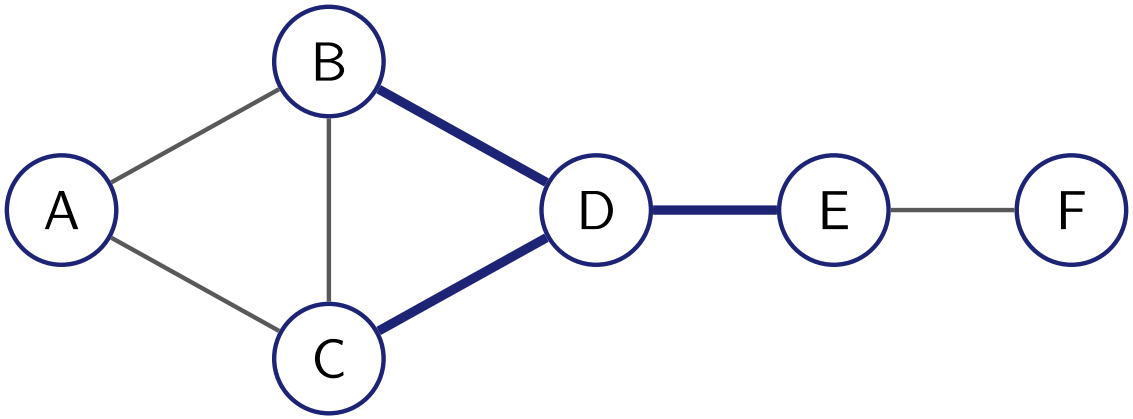


These drawings have exactly the same vertices and edges. Moving a vertex on the page does not change the graph.

Line lengths and geometric positions have no meaning unless the model explicitly assigns them one.

Neighbors and degree

Two vertices are **adjacent** if an edge joins them. The **neighbors** of v form the set $N(v)$.

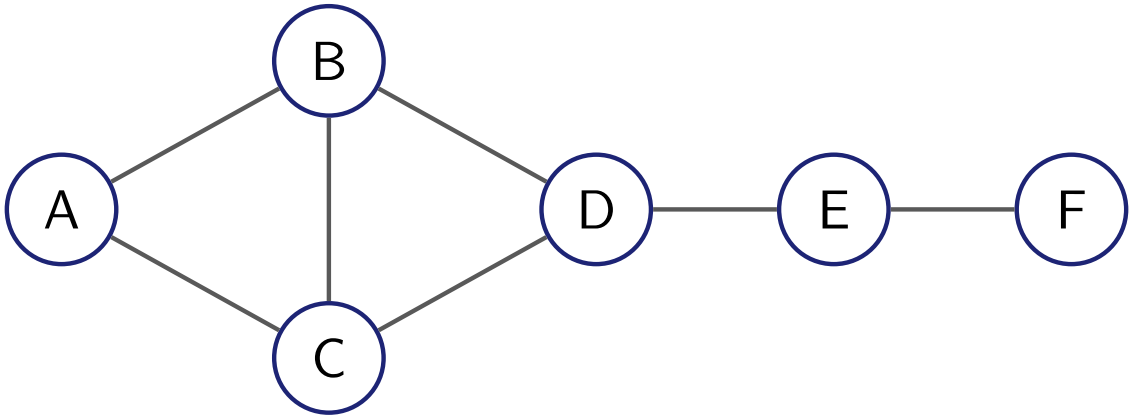


The **degree** of v is the number of its neighbors:

$$\deg(v) = |N(v)|.$$

For vertex D, $N(D) = \{B, C, E\}$, so $\deg(D) = 3$. For vertex A, $N(A) = \{B, C\}$, so $\deg(A) = 2$.

Counting degrees



v	A	B	C	D	E	F
$\deg(v)$	2	3	3	3	2	1

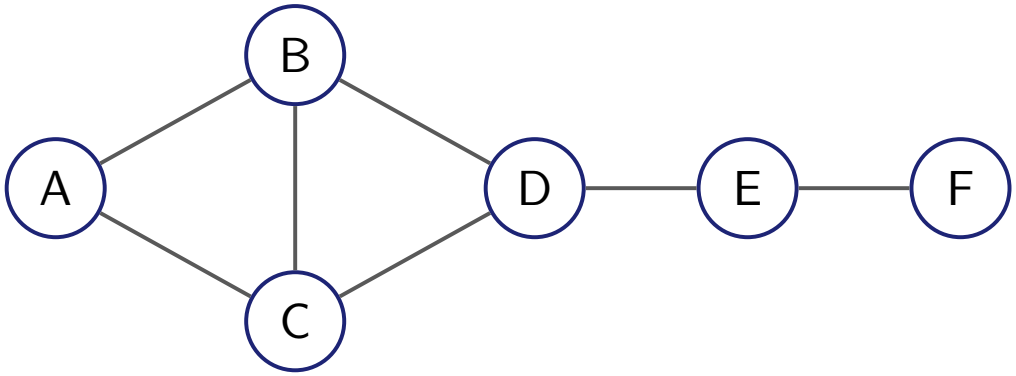
Each edge contributes one to the degree of each endpoint. Thus

$$\boxed{\sum_{v \in V} \deg(v) = 2|E|.}$$

Here, $2 + 3 + 3 + 3 + 2 + 1 = 14 = 2 \cdot 7$. This is a useful check when constructing a graph from data.

Storing a graph

An **adjacency list** records the neighbors of each vertex.



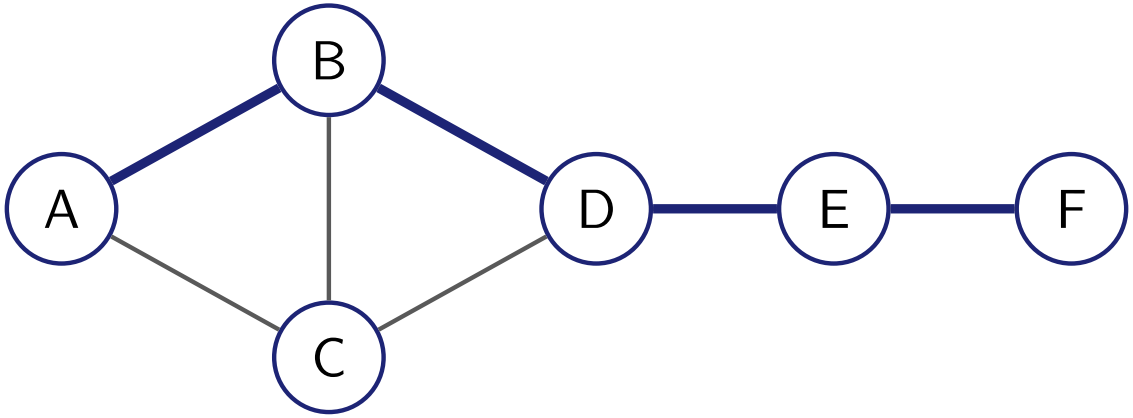
A	B, C
B	A, C, D
C	A, B, D
D	B, C, E
E	D, F
F	E

Because the graph is undirected, B appears in A's list and A appears in B's list.

Later, PageRank will use lists of *outgoing* neighbors. No matrix representation is needed.

Walks and paths

A **walk** is a sequence of vertices in which consecutive vertices are adjacent. Its length is its number of edges.



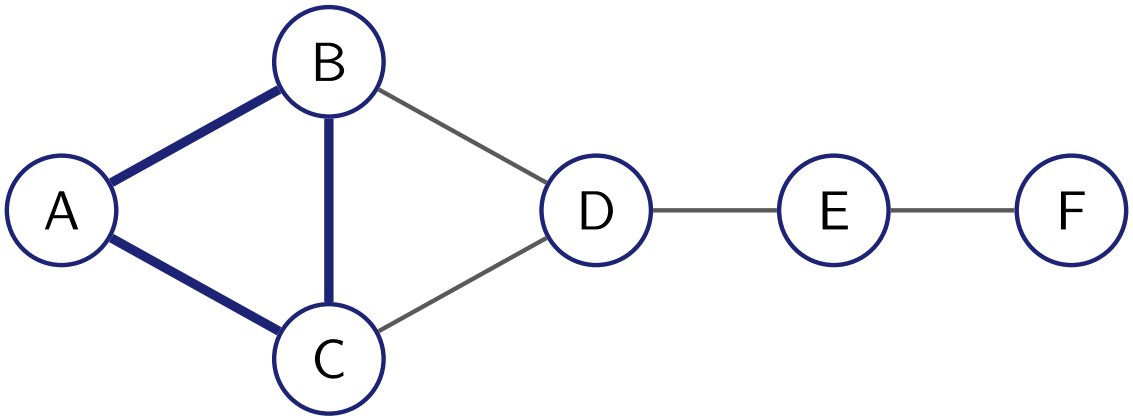
A, B, D, E, F has length 4.

We call a walk with no repeated vertices a **path**. The highlighted walk is a path; A, B, C, A, B is not.

Terminology varies: some texts call a walk a path and use "simple path" when vertices do not repeat.

Cycles

A **cycle** returns to its starting vertex without repeating any other vertex. In a simple undirected graph, it has at least three edges.

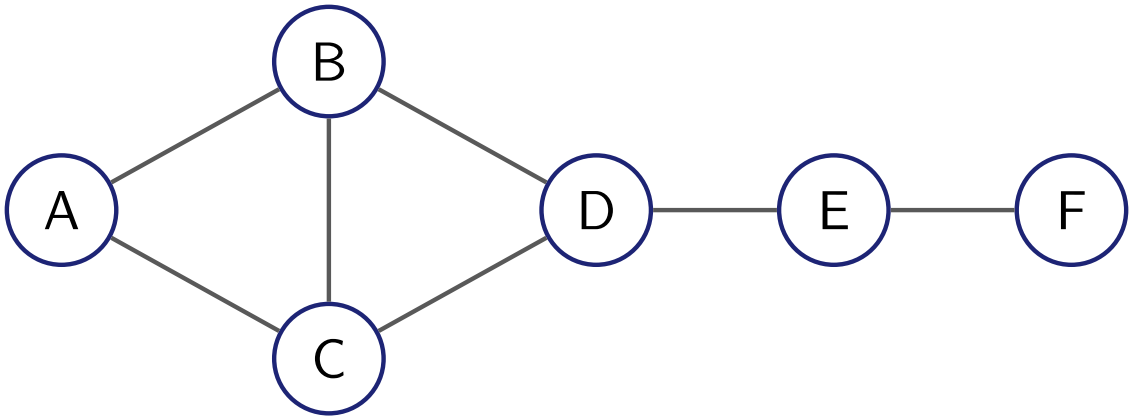


A, B, C, A is a cycle of length 3.

A cycle provides an alternative route: even without edge $\{A, B\}$, we can still go from A to B through C.

Connected graphs

An undirected graph is **connected** if a path exists between every pair of vertices.

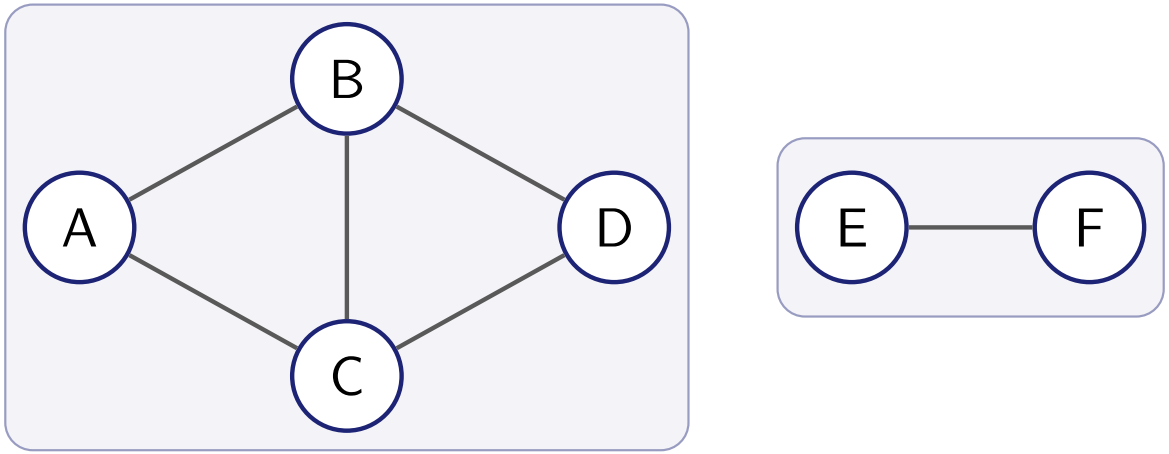


This graph is connected. For example, A, B, D, E, F connects A to F.

Connected does *not* mean that every pair is adjacent. There is no edge $\{A, F\}$, but a path connects them.

Connected components

Remove the edge $\{D, E\}$.

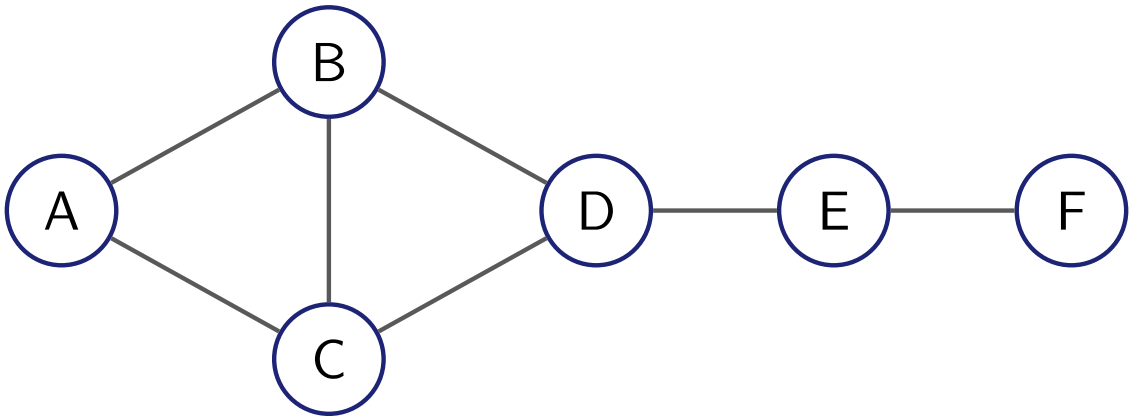


A **connected component** is a maximal set of vertices connected to one another by paths.

The components are $\{A, B, C, D\}$ and $\{E, F\}$. "Maximal" means that no vertex outside the set can be added while keeping it connected.

Which connection is critical?

Problem. In our original network, would removing $\{A, B\}$ disconnect the graph? What about $\{D, E\}$?

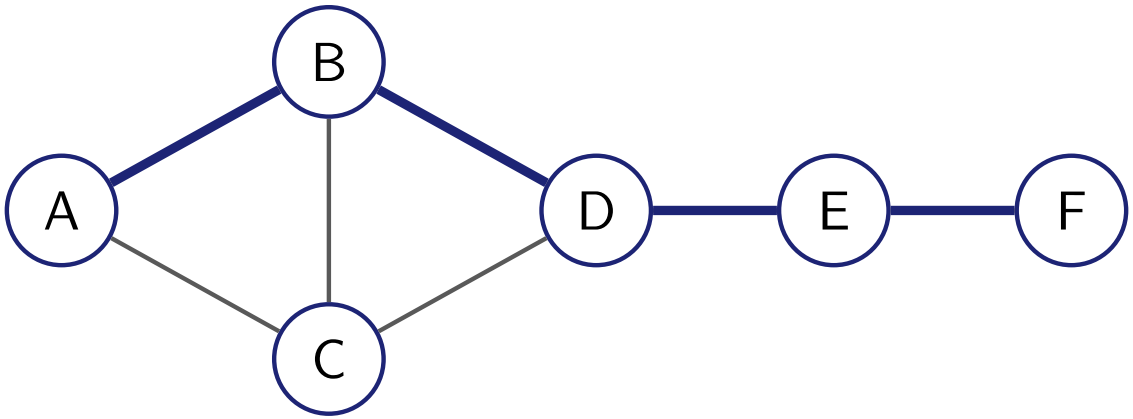


Solution. Removing $\{A, B\}$ leaves an alternative route A, C, B. The graph remains connected.

Removing $\{D, E\}$ separates $\{A, B, C, D\}$ from $\{E, F\}$. An edge whose removal increases the number of components is a **bridge**.

Distance

The **distance** $d(u, v)$ is the length of a shortest path from u to v. We set $d(u, v) = \infty$ if no such path exists.

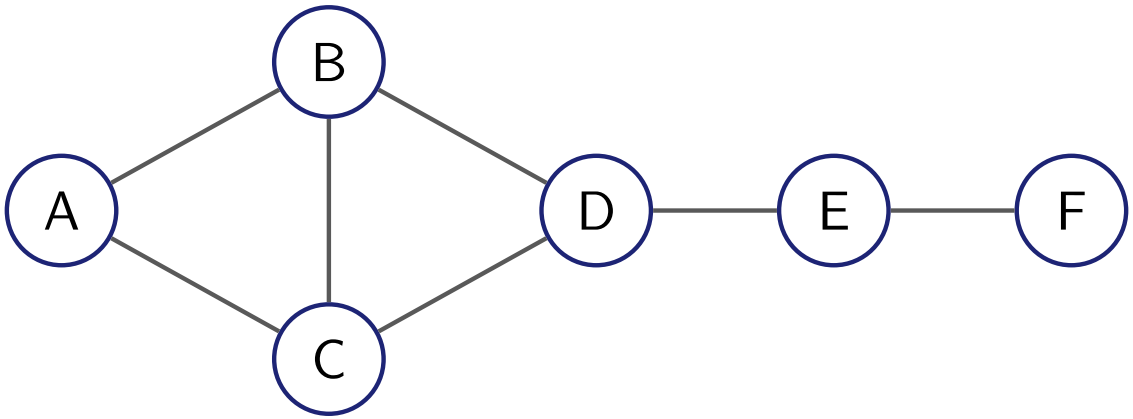


Here, $d(A, D) = 2$ and $d(A, F) = 4$.

Every route from A to F must reach D, then use D, E, F. Reaching D takes at least two edges, so four edges are necessary.

Finding reachable vertices layer by layer

Starting at A, first find its neighbors, then their unvisited neighbors.



Distance from A	0	1	2	3	4
Vertices	A	B, C	D	E	F

This is the idea of **breadth-first search** (BFS). A vertex first discovered in layer j has distance j from A.

Remember visited vertices so that a cycle does not make the search repeat forever.

What a graph lets us ask

Our model separates three different questions:

- **Local:** Who is directly connected to a vertex?
  Neighbors and degree.
- **Global:** Can we get from one vertex to another?
  Paths, components, and distance.
- **Structural:** What changes if a connection disappears?
  Alternative routes and bridges.

For the Web, connections have a direction. A link from A to B does not promise a way back. We next add arrows to the model.