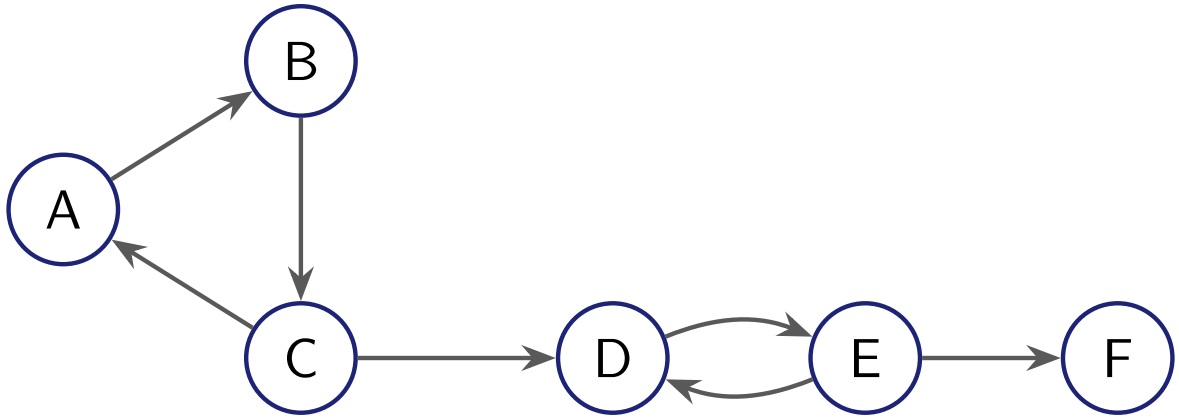
Directed Graphs

Sehyun Kwon

Mathematical Data Science
Hanyang University ERICA

Links have a direction

A webpage can link to another page without receiving a link back.



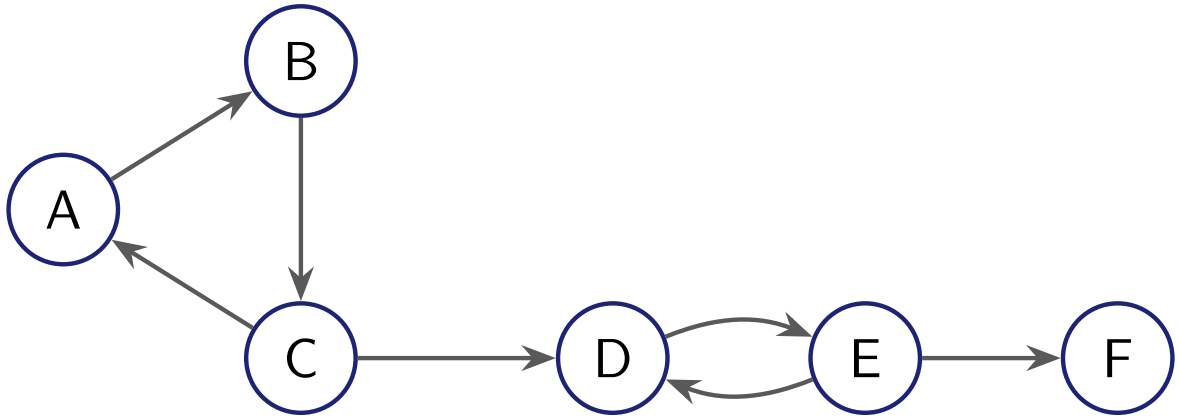
The arrow $C \to D$ means that a link on page C points to page D. It does not mean $D \to C$.

We will distinguish being connected when arrows are ignored from being able to travel along them.

Reading: Easley and Kleinberg, *Networks, Crowds, and Markets*, Sections 13.3–13.4.

Directed edges

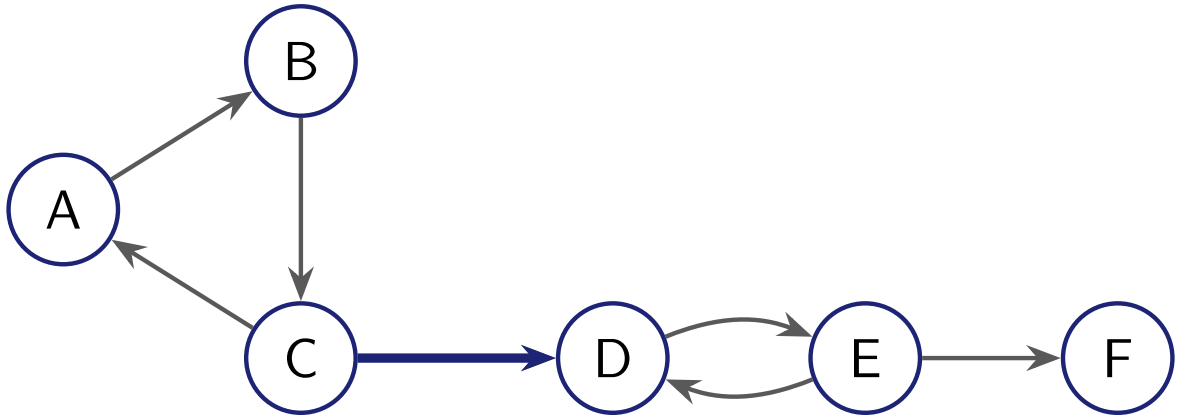
A **directed graph** is $G = (V, E)$, where each edge is an *ordered* pair (u, v). We also write $u \to v$.



$$V = \{A, B, C, D, E, F\},$$
$$E = \{(A, B), (B, C), (C, A), (C, D), (D, E), (E, D), (E, F)\}.$$

Two opposite arrows, such as $D \to E$ and $E \to D$, are two edges.

Incoming and outgoing neighbors



For a vertex v, define

$$N^-(v) = \{u : u \to v\}, \qquad N^+(v) = \{w : v \to w\}.$$

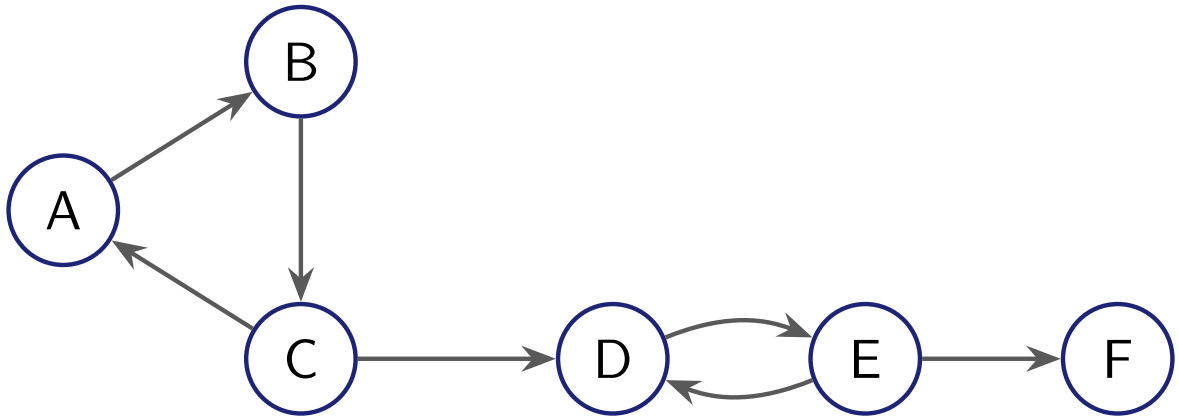
These are its **incoming** and **outgoing** neighbors.

$$N^-(D) = \{C, E\}, \qquad N^+(D) = \{E\}.$$

For PageRank, a node *sends* score along outgoing links and *receives* score along incoming links.

In-degree and out-degree

$$d^-(v) = |N^-(v)|, \qquad d^+(v) = |N^+(v)|.$$

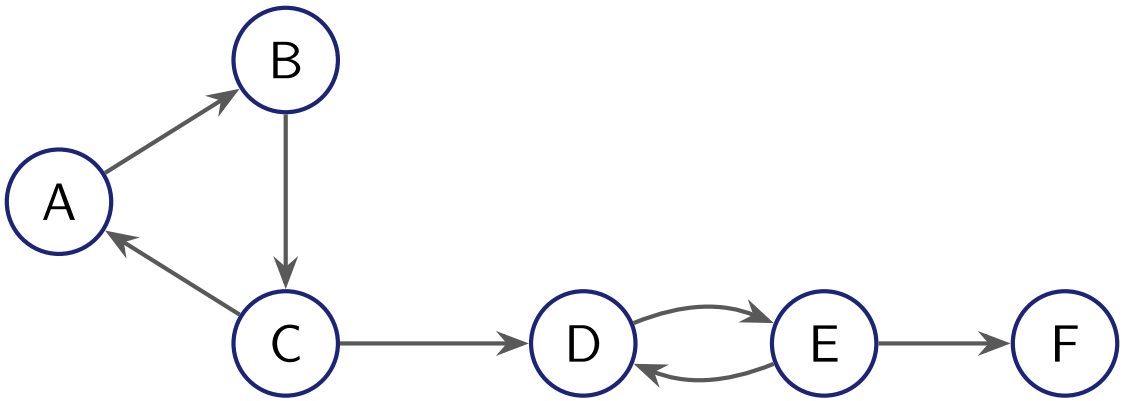


v	A	B	C	D	E	F
$d^-(v)$	1	1	1	2	1	1
$d^+(v)$	1	1	2	1	2	0

Each edge has one tail and one head. Therefore

$$\sum_{v \in V} d^+(v) = \sum_{v \in V} d^-(v) = |E| = 7.$$

Outgoing adjacency lists



Page	Links to
A	B
B	C
C	A, D
D	E
E	D, F
F	none

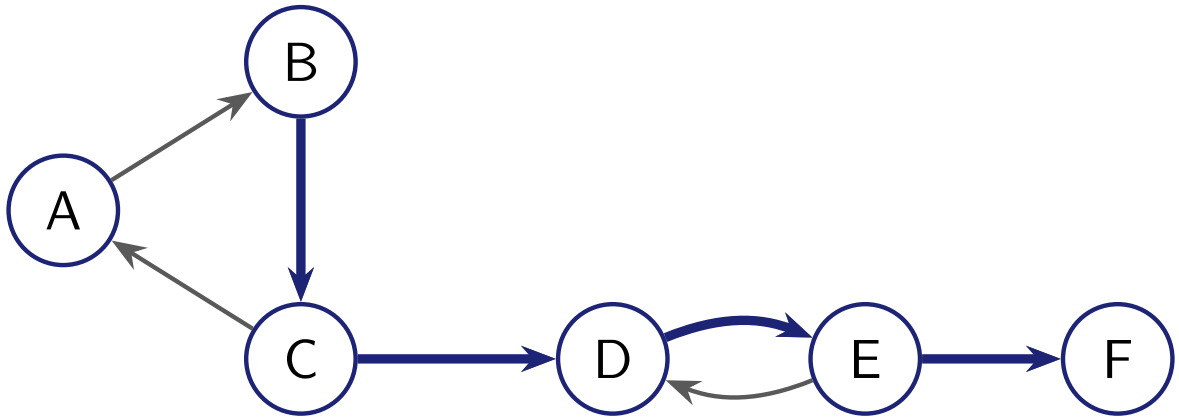
An empty list means that the vertex has no outgoing links.

Unlike an undirected adjacency list, these lists need not be symmetric.

This will also be the input format for our PageRank implementation.

Directed paths and reachability

A **directed path** must follow every arrow in its direction. Vertex v is **reachable** from u if such a path exists.

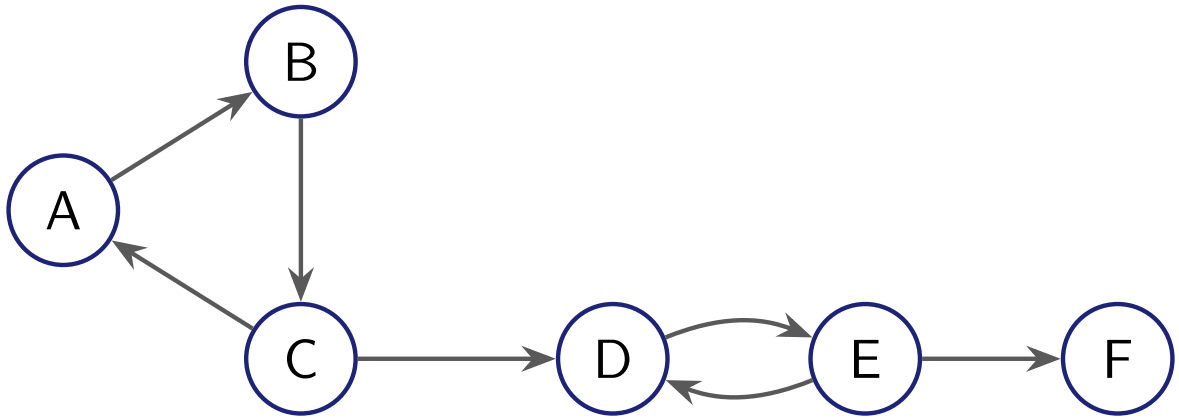


$$B \to C \to D \to E \to F$$

F is reachable from B, but B is not reachable from F. We allow a length-zero path, so a vertex is reachable from itself.

Reachability is not adjacency

Problem. Which vertices are reachable from D? Can a user starting at D eventually visit A?

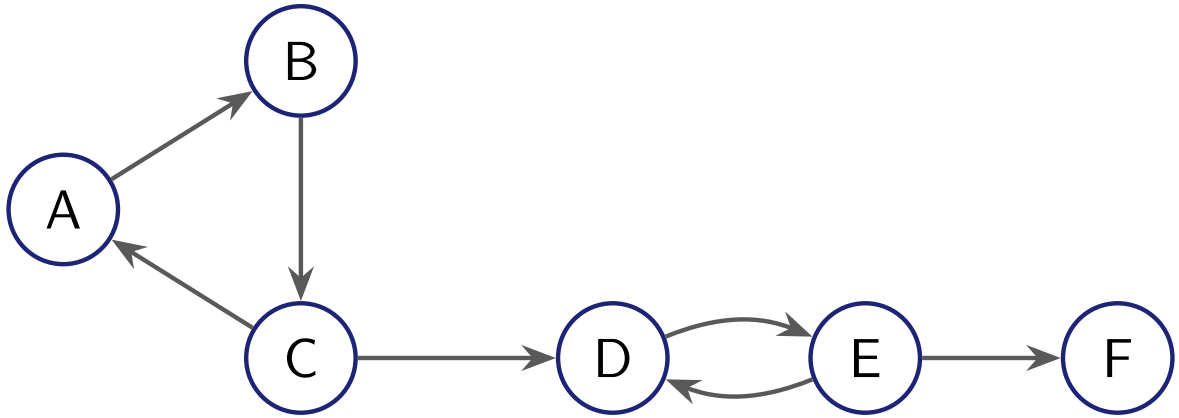


Solution. From D we can visit E, and from E we can visit D or F. Thus the reachable set is $\{D, E, F\}$.

There is no link from this set back to $\{A, B, C\}$. The arrow $C \to D$ cannot be followed backwards.

Weak connectivity

A directed graph is **weakly connected** if ignoring all arrow directions produces a connected undirected graph.

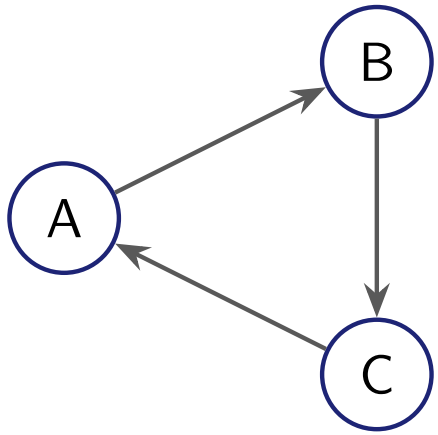


Our graph is weakly connected.

This does not imply that every page is reachable from every other page. Weak connectivity describes the underlying connections, not navigation.

Strong connectivity

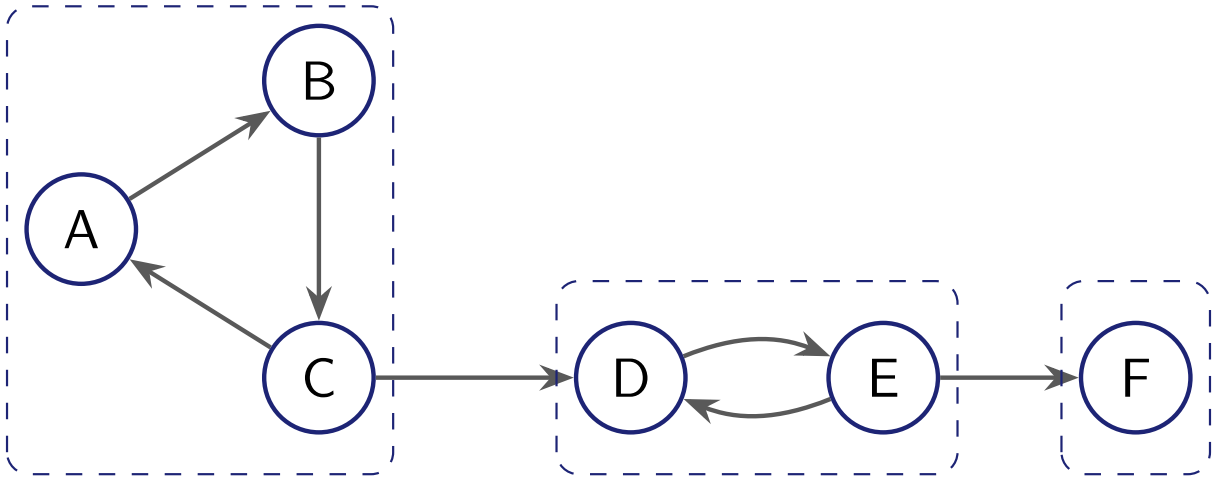
A directed graph is **strongly connected** if every vertex can reach every other vertex by directed paths.



Here, A reaches C via B, and C reaches B via A. The graph is strongly connected even though not every pair has a direct link.

Strongly connected components

A **strongly connected component** (SCC) is a maximal set of vertices that can all reach one another.

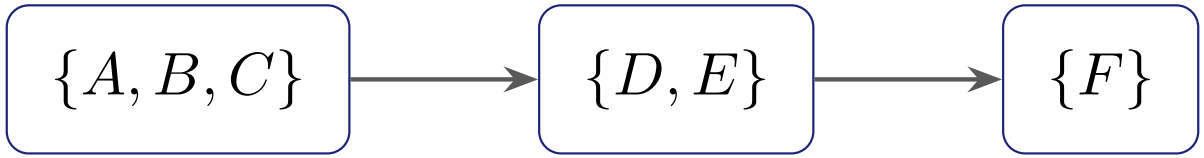


$$\{A, B, C\}, \qquad \{D, E\}, \qquad \{F\}.$$

The first two groups cannot be merged: there is no path from D back to A. A single vertex can also be an SCC.

A simpler map of the graph

Replace each SCC by one node and keep arrows between different SCCs.



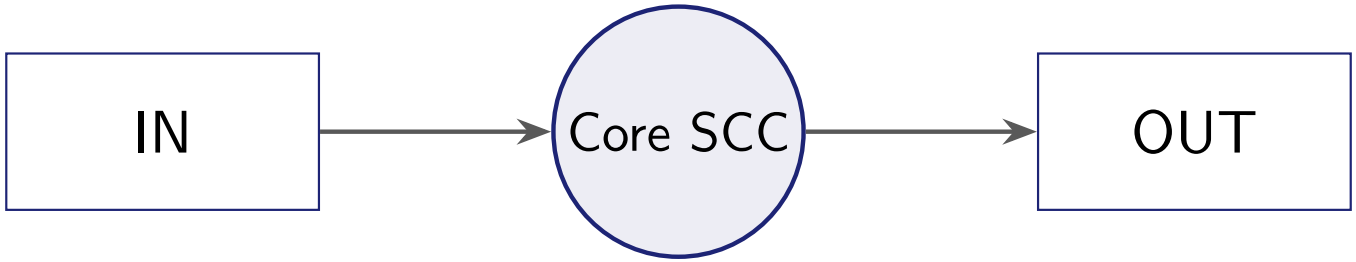
This component graph has no directed cycle.

Why? A cycle of components would let their vertices reach one another in both directions. They would belong to one larger SCC instead.

This separates movement *inside* a group from movement *between* groups.

The bow-tie view

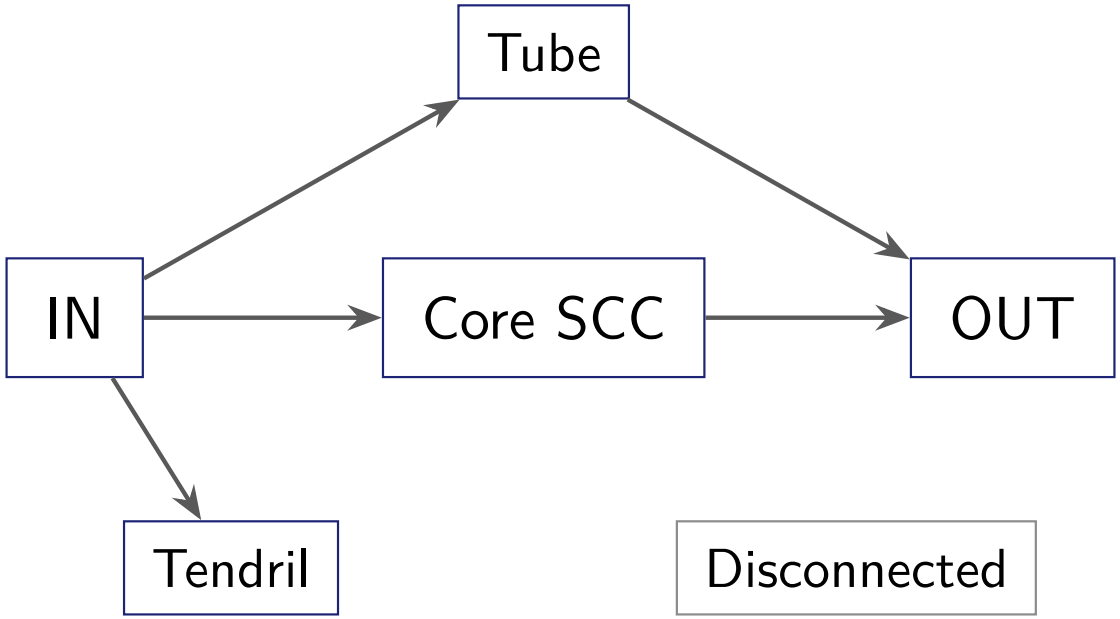
Large directed networks can be studied around a central SCC.



- **IN:** can reach the core, but cannot be reached from it.
- **Core SCC:** its vertices can reach one another.
- **OUT:** can be reached from the core, but cannot reach it.

Conceptual schematic following Easley and Kleinberg, Section 13.4; not a measurement of the current Web.

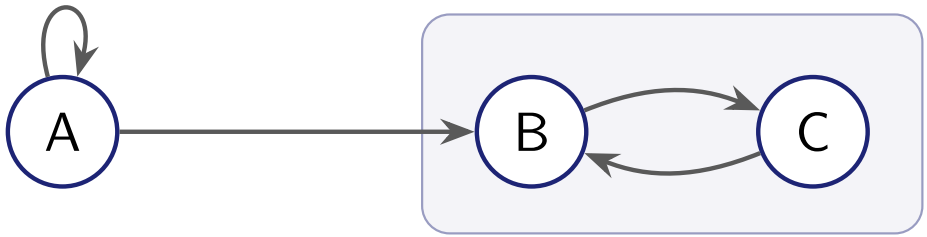
Not every vertex belongs to the core



A **tube** connects IN to OUT without passing through the core. A **tendril** may branch out from IN without reaching the core or OUT.

Other vertices may be disconnected even after directions are ignored. The point is the pattern of possible routes, not the physical layout.

A region with no exit



no link leaves $\{B, C\}$

Once a user enters $\{B, C\}$, following links cannot take them back to A.

This is a **closed** group: no outgoing edge leaves the group. A closed group can contain several vertices, not just a dead-end page.

If importance flows along links, could too much of it accumulate here? This will matter when we build PageRank.

From navigation to ranking

A graph tells us who links to whom. We would also like to compare the importance of its vertices.

- In-degree counts how many pages point to a page.
- Reachability tells us which pages can lead to it.
- Strongly connected components reveal regions where navigation can circulate.

PageRank combines the link structure with a rule for passing score from one page to another.

The next question is not just *how many* pages link to a page, but *which* pages do so.