Connected Components

A connected component of an undirected graph is a maximal set of nodes such that each pair of nodes is connected by a path. Connected components form a partition of the set of graph vertices, meaning that connected components are non-empty, they are pairwise disjoints, and the union of connected components forms the set of all vertices. Equivalently, we can say that the relation that associates two nodes if and only if they belong to the same connected component is an equivalence relation, that is it is reflexive, symmetric, and transitive.

Basic Concepts of Connected Components

Before we can understand what a connected component is, we first have to understand two adjacent concepts: graphs and connectivity.

Graphs

A graph is a set of nodes (also called vertices) connected by edges. It's a foundational structure for representing relationships—between users, devices, events, or any entities that interact. Following convention, we typically denote the number of nodes by n and the number of edges by m.

Graphs can be undirected, where edges imply mutual relationships (like friendships), or directed, where edges have direction (like who follows whom on a social platform).

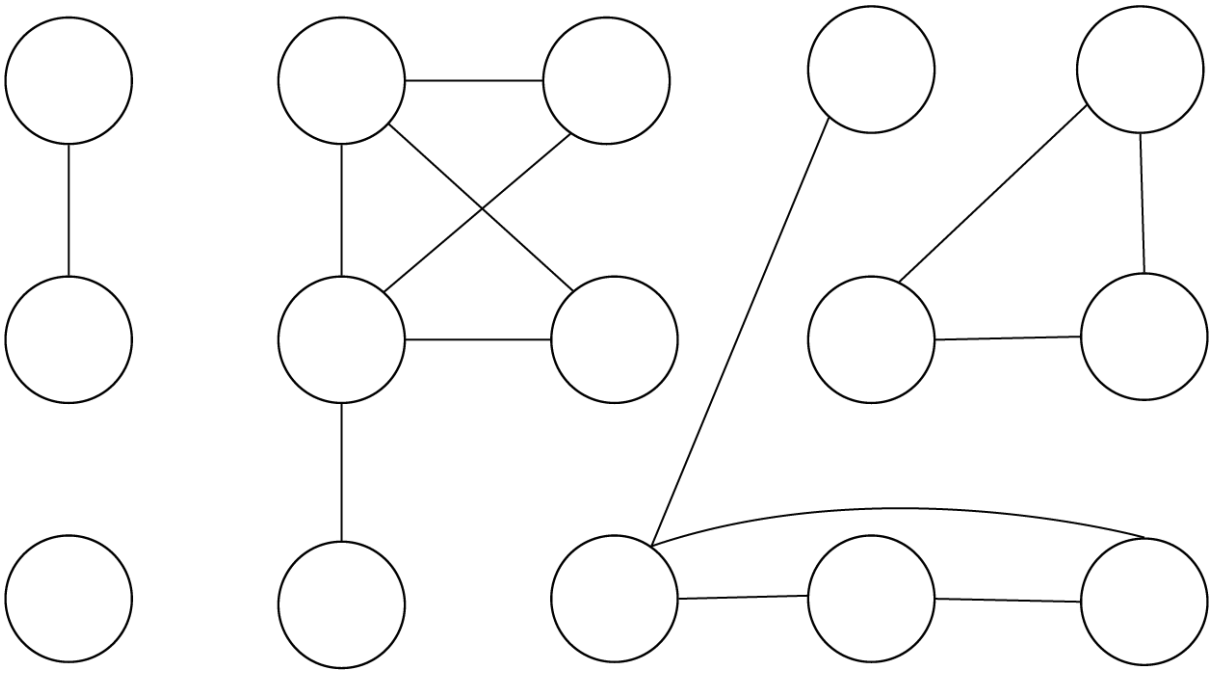
In the property graph model, which is widely used in practice, both nodes and edges can carry labels and attributes. For example, a node might represent a user with attributes like name and role, and an edge might represent an access relationship with an associated timestamp. These properties make graphs more expressive and useful in domains like fraud detection, cybersecurity, and infrastructure modeling.

Connectivity

In an undirected graph, connectivity means there is a path between two nodes. If you can get from node A to node B by following a sequence of edges, then A and B are connected.

What are Connected Components?

A connected component is a maximal group of nodes where every node is reachable from every other node in the same group. Maximal means that you cannot add any other nodes from the graph and still keep the subgraph connected. Equivalently, there are no edges from the component to nodes outside it. A graph can have one or many connected components.



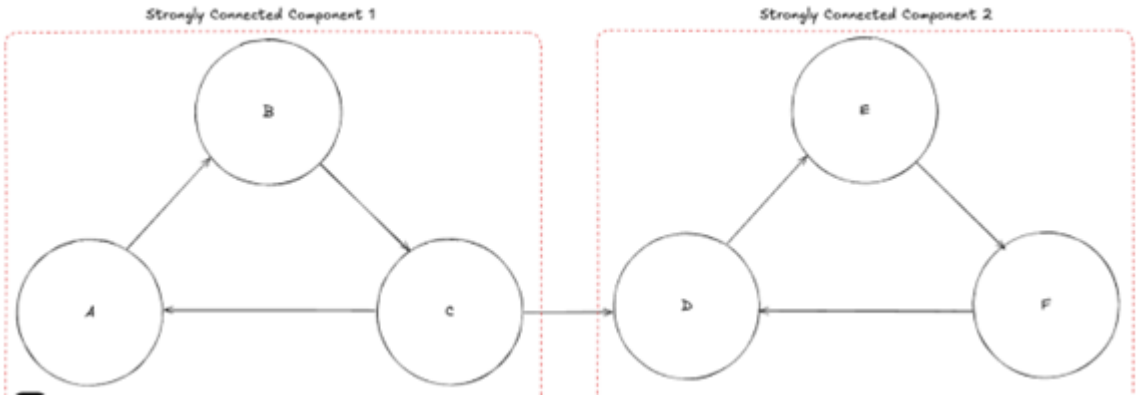
For a more formal treatment, connectivity can also be defined in terms of vertex cuts. A vertex cut (or separating set) is a set of nodes whose removal increases the number of connected components in the graph. A graph is said to be k-connected if it has more than k vertices and every vertex cut contains at least k nodes. The connectivity of a graph is the largest integer k for which the graph is k-connected.

This formal definition is useful when reasoning about fault tolerance or structural robustness—for example, in network design or critical infrastructure analysis. But in this post, we focus on computing basic connected components: the simpler notion of which nodes are reachable from one another in the graph.

Types of Connected Components

Strongly connected components (SCC)

A strongly connected component is a maximal set of vertices in a directed graph where every vertex can reach every other vertex in both directions. While trivial SCCs are just isolated nodes, non-trivial SCCs help to capture tight cyclical structure like feedback loops and mutually dependent services.



Weakly connected components (WCC)

A weakly connected component is a maximal set of vertices that is connected if you ignore edge directions. WCCs provide a coarse partition of a directed graph and are often used as a first pass to bound subsequent work. Conceptually, they are the connected components of the graph you get by replacing each directed edge with an undirected one.

