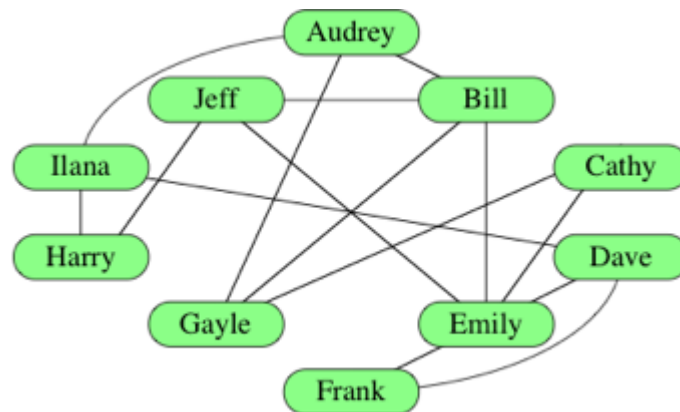# Describing graphs

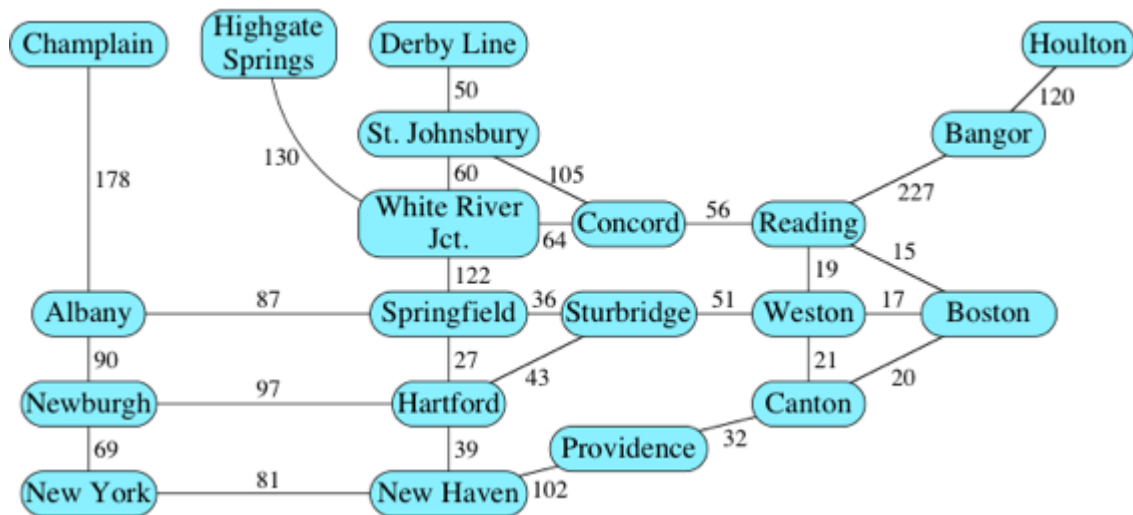Here's one way to represent a social network:

A line between the names of two people means that they know each other. If there's no line between two names, then the people do not know each other. The relationship "know each other" goes both ways; for example, because Audrey knows Gayle, that means Gayle knows Audrey.

This social network is a **graph**. The names are the **vertices** of the graph. (If you're talking about just one of the vertices, it's a **vertex**.) Each line is an **edge**, connecting two vertices. We denote an edge connecting vertices **u** and **v** by the pair **(u,v)**. Because the "know each other" relationship goes both ways, this graph is **undirected**. An undirected edge **(u,v)** is the same as **(v,u)**. Later, we'll see **directed graphs**, in which relationships between vertices don't necessarily go both ways. In an undirected graph, an edge between two vertices, such as the edge between Audrey and Gayle, is **incident** on the two vertices, and we say that the vertices connected by an edge are **adjacent** or **neighbors**. The number of edges incident on a vertex is the **degree** of the vertex.

Audrey and Frank do not know each other. Suppose that Frank wanted to be introduced to Audrey. How could he get an introduction? Well, he knows Emily, who knows Bill, who knows Audrey. We say that there is a **path** of three edges between Frank and Audrey. In fact, that is the most direct way for Frank

edges between Frank and Audrey. In fact, that is the most direct way for Frank to meet Audrey; there is no path between them with fewer than three edges.

We call a path between two vertices with the fewest edges a **shortest path**. We've highlighted that particular shortest path below:

When a path goes from a particular vertex back to itself, that's a **cycle**. The social network contains many cycles; one of them goes from Audrey to Bill to Emily to Jeff to Harry to Ilana and back to Audrey. There's a shorter cycle containing Audrey, shown below: Audrey to Bill to Gayle and back to Audrey. What other cycles can you find?

Sometimes we put numeric values on the edges. For example, in the social network, we might use values to indicate how well two people know each other. To bring in another example, let's represent a road map as a graph. Assuming that there are no one-way streets, a road map is also an undirected graph, with cities as vertices, roads as edges, and the values on edges indicating the distance of each road. For example, here's a road map, not to scale, of some of the interstate highways in the northeastern U.S., with distances next to edges:

The general term we use for a number that we put on an edge is its **weight**, and a graph whose edges have is a **weighted graph**. In the case of a road map, if you want to find the shortest route between two locations, you're looking for a path between two vertices with the minimum sum of edge weights over all paths between the two vertices. As with unweighted graphs, we call such a path a shortest path. For example, the **shortest path** in this graph from New York to Concord goes from New York to New Haven to Hartford to Sturbridge to Weston to Reading to Concord, totaling 289 miles.

The relationship between vertices does not always go both ways. In a road map, for example, there could be one-way streets. Or here's a graph showing the order in which a goalie in ice hockey could get dressed:
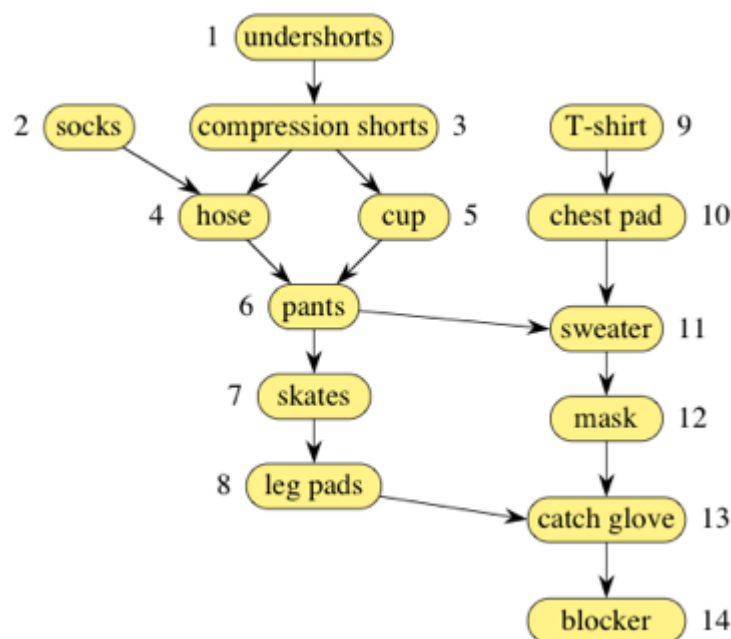
Now edges, shown with arrows, are **directed**, and we have a **directed graph**.

Here, the directions show which pieces of equipment must be put on before other pieces. For example, the edge from chest pad to sweater indicates that

the chest pad must be put on before the sweater. The numbers next to the vertices show one of the many possible orders in which to put on the equipment, so that undershorts go on first, then socks, then compression shorts, and so on, with the blocker going on last. You might have noticed that this particular directed graph has no cycles; we call such a graph a **directed acyclic graph**, or **DAG**. Of course, we can have weighted directed graphs, such as road maps with one-way streets and road distances.

We use different terminology with directed edges. We say that a directed edge **leaves** one vertex and **enters** another. For example, one directed edge leaves the vertex for chest pad and enters the vertex for sweater. If a directed edge leaves vertex **u** and enters vertex **v**, we denote it by **(u,v)**, and the order of the vertices in the pair matters. The number of edges leaving a vertex is its **out-degree**, and the number of edges entering is the **in-degree**.

As you might imagine, graphs—both directed and undirected—have many applications for modeling relationships in the real world.

Graph sizes

When we work with graphs, it's helpful to be able to talk about the set of vertices and the set of edges. We usually denote the vertex set by **V** and the edge set by **E**. When we represent a graph or run an algorithm on a graph, we often want to use the sizes of the vertex and edge sets in asymptotic notation. For example, suppose that we want to talk about a running time that is linear in the number of vertices. Strictly speaking, we should say that it's $\Theta(|V|)$, using the notation $|\cdot|$ to denote the size of a set. But using this set-size notation in asymptotic notation is cumbersome, and so we adopt the convention that in asymptotic notation, and only in asymptotic notation, we drop the set-size notation with the understanding that we're talking about set sizes. So, instead of writing $\Theta(|V|)$, we write just $\Theta(V)$. Similarly, instead of $\Theta(\lg|E|)$, we write $\Theta(\lg E)$.