Basics of Graphs

1 Definition of graphs and digraphs

Graphs are important discrete structures that are studied in discrete mathematics and computer science. We start with basic definitions and later give many examples.

Definition 1. *An* undirected graph*, or simply a* graph*, consists of points and lines between the points. Points are called* vertices *and lines are called* edges*. If there is an edge in a graph between two vertices then we say that the edge connects these vertices. No edge in a graph connects a vertex to itself. We also do not allow more than one edge between any two vertices.*

Thus, every graph G consists of the set V of vertices and the set E of undirected edges. Therefore, G can be written as the pair (V, E). We can represent graphs pictorially. Consider, for example, the graph presented in Figure 1.

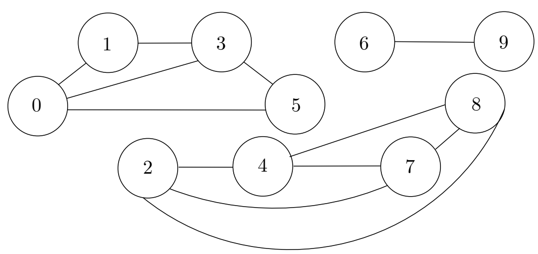

Fig. 1. An example of a graph

The set V of vertices of this graph is:

$$V = \{0, 1, \ldots, 9\}.$$

To denote edges of this graph we use set-theoretic notations. If an edge connects two vertices, say u and v, then we represent it by $\{u, v\}$ or $\{v, u\}$. So, the edge can be identified with the unordered pair consisting of u and v. So, the set E of edges of the graph in Figure 1 can be written as

$$E = \{\{0,1\}, \{0,3\}, \{1,3\}, \{3,5\}, \{0,5\}, \{2,4\}, \{2,7\}, \{2,8\}, \{4,7\}, \{7,8\}, \{4,8\}, \{6,9\}\}$$

Note E does not contain, for instance, $\{3,1\}$ since it already contains $\{1,3\}$ representing the same edge. Putting $\{3,1\}$ into the edge set E is redundant.

A technical but important concept for graphs is the notion of degree. Let v be a vertex in a graph. Then the *degree* of v, written $deg(v)$, is the number of edges connected to v. For example, for graph G in Figure 1, $deg(4) = deg(7) = deg(0) = 3$ and $deg(5) = 2$.

Edges in undirected graphs have no directions. We can put directions on the edges, and obtained directed graphs. Formally, a *directed graph* (or *digraph* for short) consists of points and directed lines connecting the points. Points are usually called *vertices* and the directed lines are called *edges*. From this definition we see that in order to describe a directed graph, one needs to specify (1) its vertices and (2) its edges. As for graphs, the collection of all vertices of the digraph is denoted by the upper case letter V, and the collection of all edges by the letter E. As undirected graphs, digraphs can be represented pictorially. Consider, as an example, the digraph in Figure 2. The digraph in Figure 2 has 10 vertices represented as

$$V = \{0,\ 1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 7,\ 8,\ 9\}.$$

The edges in this digraph can be represented as *ordered pairs* of vertices. For example, the edge from 2 to 4 is represented as the ordered pair $(2,4)$. There is no edge from 4 to 2, and hence the pair $(4,2)$ is not included in our description of the edge. We write the collection E of edges as

$$E = \{(0,1),\ (0,2), (0,3), (2,2),\ (2,4), (3,4), (3,5),\ (4,0), (5,6), (5,8), \\ (6,8), (7,9), (8,6), (8,7), (9,8)\}.$$

Here we stress the difference in the notation $\{u,v\}$ and (u,v). The former represents unordered pair and the latter ordered pair. So, (u,v) is not the same as (v,u).

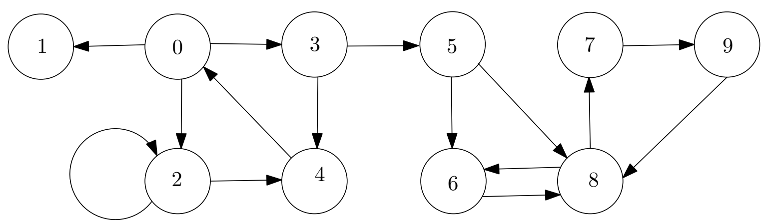

Fig. 2. An example of a directed graph

Thus, we can translate a picture of a digraph into a written description, which enumerates all its vertices V and all its edges E.

Given a set V of vertices and its edges E, we can also draw the corresponding digraph. Consider the digraph whose set V of vertices is $V = \{a,b,c,d,e,f\}$, and whose set E of edges is $E = \{(a,b),(a,c),(a,d),(b,c),(b,d),(c,a),(c,f),(d,c),(e,a),(e,b),(f,d)\}$, A pictorial presentation of this digraph is in Figure 3.

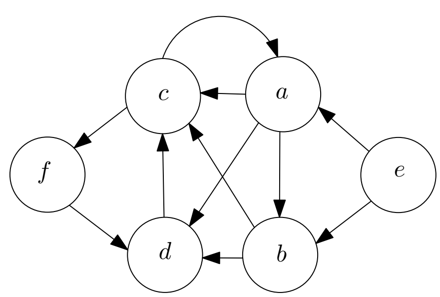

Fig. 3. An example of a directed graph

2 Examples: digraphs

Most of our examples below have pictorial presentations. In these pictures, we represent vertices as circles and we do not name the vertices.

Example 1 (Directed cycles). A pictorial presentation of the first five directed cycles is in Figure 4. The first directed cycle C_1 is a trivial cycle. The second C_2 is a directed cycle of length 2. The third C_3 is of length 3, etc. Formally the directed cycle C_n of length n is defined as follows. The vertices of C_n are 0, 1, $\ldots$, $n-1$. The edges are $(0,1)$, $(1,2)$, $\ldots$, $(n-2,n-1)$, $(n-1,0)$.

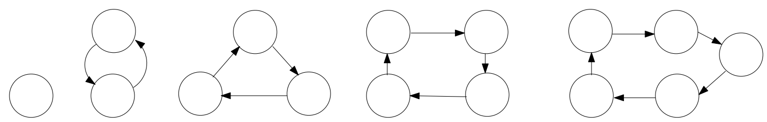

Fig. 4. Examples of directed cycles

Example 2 (Directed wheels). A pictorial presentation of the first three directed wheels is in Figure 5. The first directed wheel has four vertices. The second has five vertices, etc.

Formally the directed wheel W_n, where $n > 2$, has $n+1$ vertices and is defined as follows. The vertices of W_n are 0, 1, $\ldots$, $n-1$, and n. The edges are $(1,2)$, $(2,3)$, $\ldots$, $(n-1,n)$, $(n,1)$, and $(1,0)$, $(2,0)$, $\ldots$, $(n,0)$.

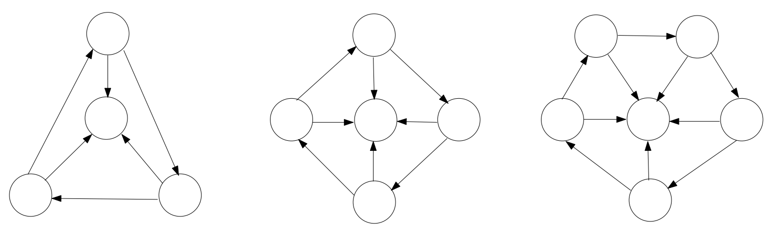

Fig. 5. Examples of directed wheels

Example 3 (Directed grids). A pictorial presentation of three two dimensional directed grids is in Figure 6. The first directed grid has four vertices. The second has eight vertices, etc.

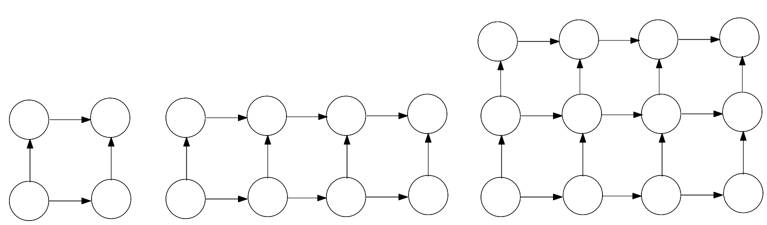

Fig. 6. Examples of directed grids

Each directed grid is defined by two parameters n and m, where n indicates the length and m indicates the height of the grid. Therefore, we write directed grids as $Grid_{n,m}$. For example, the length of the third grid in Figure 6 is 3 and the height is 2, hence the notation for that grid is $Grid_{3,2}$. Formally, the directed grid $Grid_{n,m}$ is described as follows. Its vertices are ordered pairs of the form (i, j), where $i = 0, \ldots, n$ and $j = 0, \ldots, m$. Hence the directed grid $Gird_{n,m}$ has $(n+1) \cdot (m+1)$ vertices. Its edges are of the form $((i, j), (i, j+1))$ if $j < m$ and $i \leq n$, as well as edges of the form $((i, j), (i+1, j))$ if $i < n$ and $j \leq m$.

Example 4 (Directed cubes). A pictorial presentation of the first three directed cubes is in Figure 7. The first directed cube $Cube_1$ has two vertices. The second $Cube_2$ has four vertices, etc.

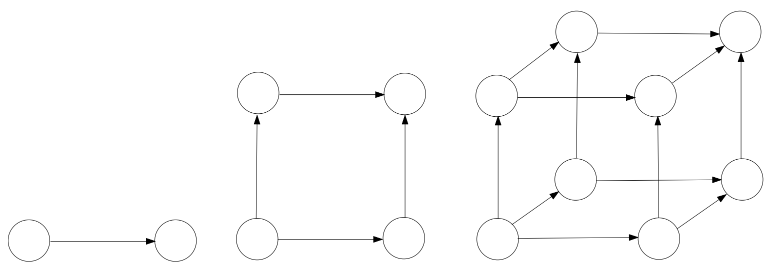

Fig. 7. Examples of directed cubes

To give an intuition for a general description of $Cube_n$, we give a formal descriptions of the directed cube $Cube_3$. The directed cube $Cube_3$ can be viewed as a three-dimensional cube. Its vertices are the points P with coordinates (x, y, z) such that x, y and z are either 0 or 1. Thus, there are 8 vertices. To describe the edges of $Cube_3$, we need to recall how we add points in the 3-dimensional space. Given two points (x_1, y_1, z_1) and (x_2, y_2, z_2) in the space, the addition of these points produces the point whose coordinates are $(x_1 + x_2, y_1 + y_2, z_1 + z_2)$. We write this:

$$(x_1, y_1, z_1) + (x_2, y_2, z_2) = (x_1 + x_2, y_1 + y_2, z_1 + z_2).$$

For instance, the addition of $(1, 2, -3)$ and $(2, -4, 7)$ produces the point $(3, -2, 4)$.

Now consider the following points that we call the *base points* in the 3-dimensional space:

$$B_1 = (1, 0, 0),\ B_2 = (0, 1, 0),\ \text{and } B_3 = (0, 0, 1).$$

Using these base points we can describe the edges of the directed cube $Cube_3$ as follows. There is an edge from a vertex (x_1, y_1, z_1) of the cube to another vertex (x_2, y_2, z_2) if and only if $(x_2, y_2, z_2) = (x_1, y_1, z_1) + B$ for some base point B. For instance, there are edges from the vertices $(0, 1, 1)$, $(1, 0, 1)$ and $(1, 1, 0)$ to the vertex $(1, 1, 1)$.

Here is a general description of the directed cube $Cube_n$. The vertices are points $(x_1, \ldots, x_n)$ in n-dimensional space such that each of the coordinates $x_1, \ldots, x_n$ is either 0 or 1. There are 2^n such points. Consider the following n base points:

$$B_1 = (1, 0, \ldots, 0),\ B_2 = (0, 1, 0, \ldots, 0),\ \ldots,\ \text{and } B_n = (0, \ldots, 0, 1).$$

Using these base points, we can describe the edges of $Cube_n$ as follows. There is an edge from a vertex $(x_1, x_2, \ldots, x_n)$ of the directed cube to another vertex $(y_1, y_2, \ldots, y_n)$ if and only if

$(y_1, y_2, \ldots, y_n) = (x_1, x_2, \ldots, x_n) + B$ for some base point B. For instance, there are edges to the vertex $(1, 1, \ldots, 1)$ from all of the following n vertices:

$$(0, 1, 1, 1, \ldots, 1),\ (1, 0, 1, 1, \ldots, 1),\ (1, 1, 0, 1, \ldots, 1),\ \ldots,\ (1, \ldots, 1, 0).$$

3 Examples: graphs

In this section, we present several examples of graphs. Recall that by graph we mean undirected graphs. Some examples are just undirected version of the digraphs from the previous section:

Example 5 (Complete graphs). We depict five complete graphs in Figure 8. The first graph is a

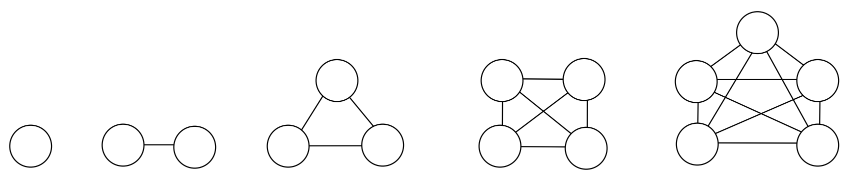

Fig. 8. Examples of complete graphs

trivial one, the second is just a two-vertex graph with an edge between the vertices, the third is the cycle of length 3, the fourth is the graph with four vertices that are all connected with edges, etc. A complete graph with n vertices is denoted by K_n. Formally, K_n is defined as follows. The vertices of K_n are 0, 1, …, $n-1$. The edges are of the form $\{i, j\}$ where $i = 0, \ldots, n-1$, $j = 0, \ldots, n-1$, and $i \neq j$.

Example 6 (Cycles). A pictorial presentation of the first four cycles is in Figure 9. The first cycle is of length 2, the second cycle is of length 3, etc. Formally, the vertices of C_n are 0, 1, …, $n-1$. The edges are $\{0, 1\}$, $\{1, 2\}$, …, $\{n-2, n-1\}$, $\{n-1, 0\}$.

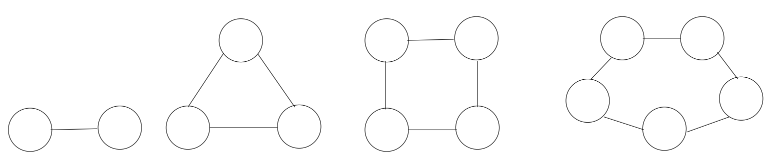

Fig. 9. Examples of cycles

Example 7 (Wheels). A pictorial presentation of the first three wheels is in Figure 10. The first

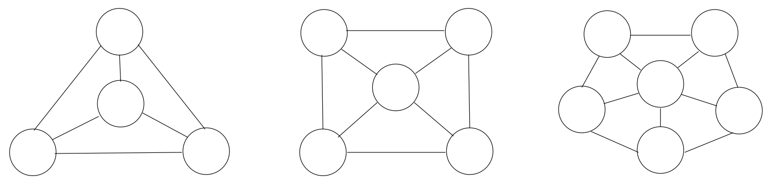

Fig. 10. Examples of wheels

wheel has four vertices. The second has five, and the third has 6 vertices.

Formally, a wheel W_n has $n+1$ vertices and is defined as follows. The vertices of W_n are 0, 1, …, $n-1$, and n. The edges are $\{1,2\}$, $\{2,3\}$, …, $\{n-1,n\}$,$\{n,1\}$, and $\{1,0\}$, $\{2,0\}$, …, $\{n,0\}$.

Example 8 (Grids). Each grid has two parameters n and m. The first parameter indicates the length of the grid, and the second parameter indicates the height of the grid. Therefore, we write directed grids as $Grid_{n,m}$ A pictorial presentation of grid $Grid_{7,2}$ is in Figure 11.

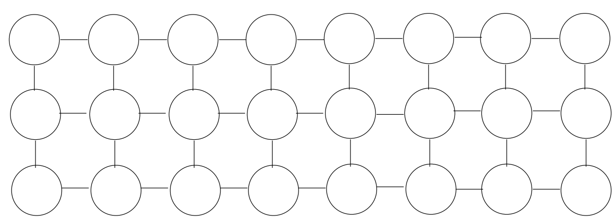

Fig. 11. Grid $G_{7,2}$

Formally, vertices of $Grid_{n,m}$ are of the form (i,j), where $i = 0,\ldots,n$ and $j = 0,\ldots,m$. It has edges of the form $\{(i,j),(i,j{+}1)\}$ if $j < m$ and $i \leq n$, as well as edges of the form $\{(i,j),(i{+}1,j)\}$ if $i < n$ and $j \leq m$.

Example 9. The following graphs are called *complete bipartite graphs* denoted by $K_{n,m}$. The vertex set of these graphs consist of $-m,-m+1,\ldots,-1,1,2,\ldots,n$. So the graph $K_{n,m}$ has exactly $n+m$ vertices. The edges are of the form $\{i,j\}$, where $-m \leq i \leq -1$ and $1 \leq j \leq n$. Thus there are exactly $n \cdot m$ edges of the graph $K_{n,m}$.

4 Examples: models of networks

Example 10 (Transportation networks). Consider a map of routes served by an airline company. This map forms a graph: the vertices are airports, and there is an edge from u to v if there is a nonstop flight from u to v. This a directed graph description of the network. In practice, if there is a flight from u to v then there is a flight from v to u. So, we can ignore the directions on edges and treat the network as a graph.

Example 11 (Communication networks). A collection of computers connected via a communication network can be modelled as a graph. One simple way to model the network as a graph is to declare that vertices are computers, and an edge is put from vertex u to vertex v if there is a physical link from u to v. Again, as in the example above, if we assume that the physical link between u and v treats u and v equally then we have undirected graph model of the network.

Example 12 (Large communication networks). For large communication networks, such as an internet, we can create a model of the network which differs from the example above. Vertices are internet service provides. We put an edge between two such vertices u and v if there is an agreement between them to exchange data (say under some protocol that governs global internet routing). This graph is undirected. We can view this as a "virtual" representation of the network since the edges indicate agreements. Another interesting point is that this graph can viewed as an example of a large scale structure of the internet.

Example 13 (Information networks). The world wide web is an example of an information network. The directed graph model of this network consists of vertices representing the web pages, and there is an edge from u to v if there is a hyperlink that starts at u and ends at v. It is important to note that this graph is directed. For instance, many webpages link to popular websites (such as New York times or CNN), and usually the links in opposite direction do not happen. The structure of all these links can be used by algorithms to infer various types of information.

Example 14 (Social networks). A social network is roughly a group of people that interact. These could be students of a school, members of a tennis club, employees of a company, residents of a city. These define graphs whose vertices are people, and edges joining u and v in case u and v are friends. One can put other types of edges that represent various relationships between the people. For instance, an undirected edge $\{u, v\}$ between u and v can represent that u and v are relatives, or u and v have a financial relationship, or u and v have a romantic relationship, or u and v previously worked in the same company. The directed edge (u, v) could mean that u seeks advice from v, or that u sent an email to v.

The study of social networks is popular nowadays. Networks are used by sociologists, computer scientists, mathematicians to investigate the dynamics of interaction between people. Important issues focus on identifying the "influential" people in a company or organisation, detecting various groups in the networks, predicting the strength of the friendships, modelling trust relationships in a financial setting, etc.

Example 15 (Biological networks). Biology provides plethora of examples of networks that can be modelled as graphs. Vertices in these graphs are biological entities and edges are the interactions between them. Here are some examples of biological networks. For each of these, one can identify vertices and edges between them: (a) protein-to-protein interaction networks; in these networks vertices are proteins and edges are interactions between them. (b) gene regulatory networks; (c) metabolic networks, (d) neuronal networks.

Example 16 (Dependency networks). A good examples of a dependency network is a list of tasks linked by dependency relationship. Such a network can be presented as a directed graph. Another example of a dependency network is the list of courses provided by, say, The University of Auckland. The vertices are the courses, and there an edge from u to v if course u is a prerequisite for course v. Another example is a large software system. It consists of many functions, methods, test predicates, and calls. These represent the vertices of the network, and an edge is put from vertex u to vertex v if execution of u invokes v.

5 Connectivity in graphs

From now on we mostly consider underected graphs unless we specify otherwise. Given a graph, an important operation on graph is to walk through its vertices using the edges that connect the vertices. Such walks are called paths.

Formally, a *path* in a graph $G = (V, E)$ is a sequence v_0, v_1, …, v_n of vertices such that $\{v_0, v_1\}$, $\{v_1, v_2\}$, …, $\{v_{n-1}, v_n\}$ are all edges. The number n is called the length of this path. If all the vertices in the path v_0, v_1, …, v_n are pairwise distinct then we call the path a *simple path*. Clearly, if v_0, v_1, …, v_n is a path then so is v_n, v_{n-1}, …, v_1, v_0. This is not always true for digraphs. A*cycle* is a path v_0, v_1, …, v_n such that $n > 1$, the first n vertices are all pairwise distinct, and $v_0 = v_n$.

Two vertices u and v in graph G are *connected* if there is a path from u to v. The graph G is *connected* if all pairs of vertices u, v of the graph are connected. For instance, for graph G in Figure 1 all the vertices 2, 4, 7 and 8 are connected.

Exercise 1. Which vertices of the graph in Figure 1 are connected with each other and which vertices are not?

Exercise 2. Consider the undirected graph model of Facebook. Reason about if the graph is connected or not.

Connectedness in graphs has the following properties in a given graph G.

1. Every vertex v of the graph is connected to itself.
2. If a vertex v is connected to a vertex u then u is connected to v.
3. If a vertex v is connected to u and u is connected to w then v is connected to w.

Exercise 3. Prove all the three properties above.

The *component* of vertex v, written $C(v)$, is the collection of *all* vertices u that are connected to v. For example, graph G in Figure 1 has three components

$$C(0) = \{0, 1, 3, 5\}, \; C(2) = \{2, 4, 7, 8\}, \text{ and } C(6) = \{6, 9\}.$$

Note that a connected graph has exactly one component; and the component is the graph itself.

Exercise 4. Let $C(u)$ and $C(v)$ be two components of graph G. Explain that either $C(u) = C(v)$ (that is, these two components coincide) or the components $C(u)$ and $C(v)$ have no vertices in common (written $C(u) \cap C(v) = \emptyset$).

Now we list two important computational problems in the study of graphs:

Problem 1: Given a graph G, its vertices u and v, is there a path from u to v?

Problem 2: Given a graph G, is the graph G connected?

Apart from knowing the existence of a path between some vertices u and v, we may also want to know whether there is a shortest path. This leads us to define the notion of a distance from u to v, written $d(u, v)$. The distance between u and v is the minimum length among the lengths of all paths from u to v. This definition makes sense if u and v are connected. If they are not connected, we can simply declare that the distance $d(u, v) = \infty$. The term distance comes from imagining G as a transportation network; if we plan to go from station u to station v, we would like to do that with as little number of stops as possible.

Exercise 5. Prove that if there is a path from u to v then there is a simple path from u to v (that is, a path in which no vertex is repeated).

6 Trees

Let $G = (V, E)$ be a graph. Recall that a *cycle* in the graph is a path $v_0, \ldots, v_n$ of length greater than 2 such that no vertices in this path are repeated, apart from the start vertex v_0 being equal to the last one v_n. The length of this cycle is n.

Definition 2. *The graph $G = (V, E)$ is a* **tree** *if G is connected and contains no cycles.*

For a tree G, we call its vertices *nodes*. Anytime, when we refer to nodes, this underlines the fact that we are considering trees. Figure 12 presents two examples of trees.

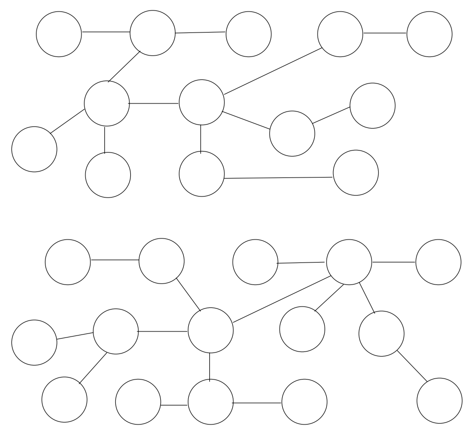

Fig. 12. Two examples of trees

The definition of trees does not specify a root node. The definition also excludes directions on the edges of the graph. For example, there is no concept of parent and children nodes. However, this definition can *informally* be explained as follows by "rooting" the tree.

Select a vertex $r \in V$ in the graph. Pull the vertex r up and let the rest of the vertices dangle down from r. Visually, the vertex r has become the root of the tree and the direction moves down from the root to the leaves. The important part in this visualization is that it is based on the hypothesis that the graph G has no cycles and that G is connected. This visual explanation allows us to orient the edges of the tree. For each v, declare the parent of v to be the node u that directly precedes v on its path from r; thus, v has become a child of u. Generally, we say that w is a descendant of v (or v is an ancestor of w) if v appears in the path from the root r to node w. A node w is a leaf if it has no children. By selecting another node r' as a root, we can build a tree quite different from the one where r was the root.

Exercise 6. Explain why a tree with n nodes has exactly $n - 1$ edges.

Exercise 7. Let G be a graph with exactly n nodes. Consider the following three statements about G:

1. G is connected.
2. G has no cycle.
3. G has exactly $n - 1$ edges.

Explain why any two of these statements implies the third.