1
Graph Theory

"Begin at the beginning," the King said, gravely, "and go on till you come to the end; then stop."

— Lewis Carroll, *Alice in Wonderland*

The Pregolya River passes through a city once known as Königsberg. In the 1700s seven bridges were situated across this river in a manner similar to what you see in Figure 1.1. The city's residents enjoyed strolling on these bridges, but, as hard as they tried, no resident of the city was ever able to walk a route that crossed each of these bridges exactly once. The Swiss mathematician Leonhard Euler learned of this frustrating phenomenon, and in 1736 he wrote an article [98] about it. His work on the "Königsberg Bridge Problem" is considered by many to be the beginning of the field of graph theory.

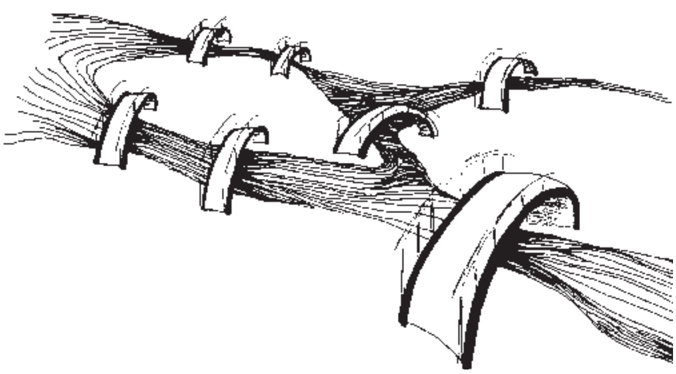


FIGURE 1.1. The bridges in Königsberg.

J.M. Harris et al., *Combinatorics and Graph Theory*, DOI: 10.1007/978-0-387-79711-3_1,

At first, the usefulness of Euler's ideas and of "graph theory" itself was found only in solving puzzles and in analyzing games and other recreations. In the mid 1800s, however, people began to realize that graphs could be used to model many things that were of interest in society. For instance, the "Four Color Map Conjecture," introduced by DeMorgan in 1852, was a famous problem that was seemingly unrelated to graph theory. The conjecture stated that four is the maximum number of colors required to color any map where bordering regions are colored differently. This conjecture can easily be phrased in terms of graph theory, and many researchers used this approach during the dozen decades that the problem remained unsolved.

The field of graph theory began to blossom in the twentieth century as more and more modeling possibilities were recognized — and the growth continues. It is interesting to note that as specific applications have increased in number and in scope, the theory itself has developed beautifully as well.

In Chapter 1 we investigate some of the major concepts and applications of graph theory. Keep your eyes open for the Königsberg Bridge Problem and the Four Color Problem, for we will encounter them along the way.

1.1 Introductory Concepts

A definition is the enclosing a wilderness of idea within a wall of words.

— Samuel Butler, *Higgledy-Piggledy*

1.1.1 Graphs and Their Relatives

A *graph* consists of two finite sets, V and E. Each element of V is called a *vertex* (plural *vertices*). The elements of E, called *edges*, are unordered pairs of vertices. For instance, the set V might be $\{a, b, c, d, e, f, g, h\}$, and E might be $\{\{a, d\}$, $\{a, e\}$, $\{b, c\}$, $\{b, e\}$, $\{b, g\}$, $\{c, f\}$, $\{d, f\}$, $\{d, g\}$, $\{g, h\}\}$. Together, V and E are a graph G.

Graphs have natural visual representations. Look at the diagram in Figure 1.2. Notice that each element of V is represented by a small circle and that each element of E is represented by a line drawn between the corresponding two elements of V.

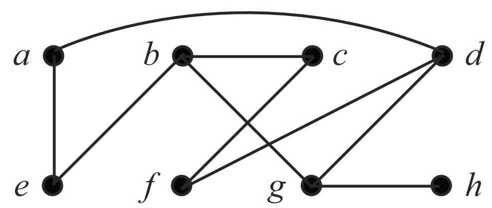


FIGURE 1.2. A visual representation of the graph G.

As a matter of fact, we can just as easily define a graph to be a diagram consisting of small circles, called vertices, and curves, called edges, where each curve connects two of the circles together. When we speak of a graph in this chapter, we will almost always refer to such a diagram.

We can obtain similar structures by altering our definition in various ways. Here are some examples.

1. By replacing our set E with a set of *ordered* pairs of vertices, we obtain a *directed graph*, or *digraph* (Figure 1.3). Each edge of a digraph has a specific orientation.

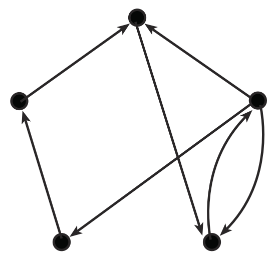


FIGURE 1.3. A digraph.

2. If we allow repeated elements in our set of edges, technically replacing our set E with a multiset, we obtain a *multigraph* (Figure 1.4).

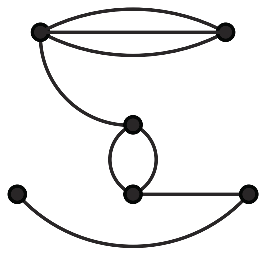


FIGURE 1.4. A multigraph.

3. By allowing edges to connect a vertex to itself ("loops"), we obtain a *pseudograph* (Figure 1.5).

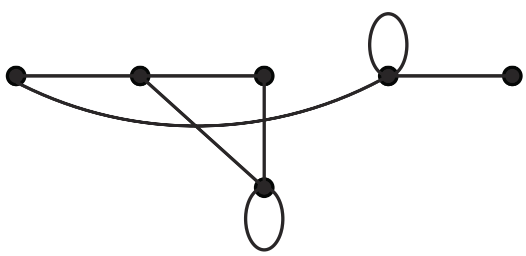


FIGURE 1.5. A pseudograph.

4. Allowing our edges to be arbitrary subsets of vertices (rather than just pairs) gives us *hypergraphs* (Figure 1.6).

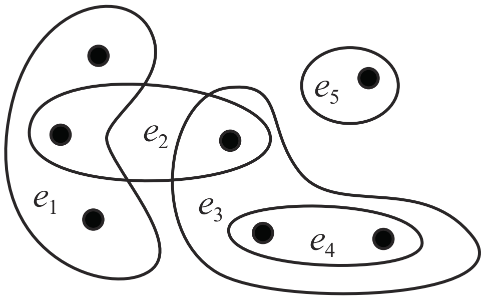


FIGURE 1.6. A hypergraph with 7 vertices and 5 edges.

5. By allowing V or E to be an infinite set, we obtain *infinite graphs*. Infinite graphs are studied in Chapter 3.

In this chapter we will focus on finite, simple graphs: those without loops or multiple edges.

Exercises

1. Ten people are seated around a circular table. Each person shakes hands with everyone at the table except the person sitting directly across the table. Draw a graph that models this situation.

2. Six fraternity brothers (Adam, Bert, Chuck, Doug, Ernie, and Filthy Frank) need to pair off as roommates for the upcoming school year. Each person has compiled a list of the people with whom he would be willing to share a room.

   Adam's list: Doug
   Bert's list: Adam, Ernie
   Chuck's list: Doug, Ernie
   Doug's list: Chuck
   Ernie's list: Ernie
   Frank's list: Adam, Bert

   Draw a digraph that models this situation.

3. There are twelve women's basketball teams in the Atlantic Coast Conference: Boston College (B), Clemson (C), Duke (D), Florida State (F), Georgia Tech (G), Miami (I), NC State (S), Univ. of Maryland (M), Univ. of North Carolina (N), Univ. of Virginia (V), Virginia Tech (T), and Wake Forest Univ. (W). At a certain point in midseason,

   B has played I, T*, W

   C has played D*, G

D has played C*, S, W

F has played N*, V

G has played C, M

I has played B, M, T

S has played D, V*

M has played G, I, N

N has played F*, M, W

V has played F, S*

T has played B*, I

W has played B, D, N

The asterisk(*) indicates that these teams have played each other twice. Draw a multigraph that models this situation.

4. Can you explain why no resident of Königsberg was ever able to walk a route that crossed each bridge exactly once? (We will encounter this question again in Section 1.4.1.)

1.1.2 The Basics

Your first discipline is your vocabulary;

— Robert Frost

In this section we will introduce a number of basic graph theory terms and concepts. Study them carefully and pay special attention to the examples that are provided. Our work together in the sections that follow will be enriched by a solid understanding of these ideas.

The Very Basics

The vertex set of a graph G is denoted by $V(G)$, and the edge set is denoted by $E(G)$. We may refer to these sets simply as V and E if the context makes the particular graph clear. For notational convenience, instead of representing an edge as $\{u, v\}$, we denote this simply by uv. The *order* of a graph G is the cardinality of its vertex set, and the *size* of a graph is the cardinality of its edge set.

Given two vertices u and v, if $uv \in E$, then u and v are said to be *adjacent*. In this case, u and v are said to be the *end vertices* of the edge uv. If $uv \notin E$, then u and v are *nonadjacent*. Furthermore, if an edge e has a vertex v as an end vertex, we say that v is *incident* with e.

The *neighborhood* (or *open neighborhood*) of a vertex v, denoted by $N(v)$, is the set of vertices adjacent to v:

$$N(v) = \{x \in V \mid vx \in E\}.$$

The *closed neighborhood* of a vertex v, denoted by $N[v]$, is simply the set $\{v\} \cup N(v)$. Given a set S of vertices, we define the neighborhood of S, denoted by $N(S)$, to be the union of the neighborhoods of the vertices in S. Similarly, the closed neighborhood of S, denoted $N[S]$, is defined to be $S \cup N(S)$.

The *degree* of v, denoted by $\deg(v)$, is the number of edges incident with v. In simple graphs, this is the same as the cardinality of the (open) neighborhood of v. The *maximum degree* of a graph G, denoted by $\Delta(G)$, is defined to be

$$\Delta(G) = \max\{\deg(v) \mid v \in V(G)\}.$$

Similarly, the *minimum degree* of a graph G, denoted by $\delta(G)$, is defined to be

$$\delta(G) = \min\{\deg(v) \mid v \in V(G)\}.$$

The *degree sequence* of a graph of order n is the n-term sequence (usually written in descending order) of the vertex degrees.

Let's use the graph G in Figure 1.2 to illustrate some of these concepts: G has order 8 and size 9; vertices a and e are adjacent while vertices a and b are nonadjacent; $N(d) = \{a, f, g\}$, $N[d] = \{a, d, f, g\}$; $\Delta(G) = 3$, $\delta(G) = 1$; and the degree sequence is 3, 3, 3, 2, 2, 2, 2, 1.

The following theorem is often referred to as the First Theorem of Graph Theory.

Theorem 1.1. *In a graph G, the sum of the degrees of the vertices is equal to twice the number of edges. Consequently, the number of vertices with odd degree is even.*

Proof. Let $S = \sum_{v \in V} \deg(v)$. Notice that in counting S, we count each edge exactly twice. Thus, $S = 2|E|$ (the sum of the degrees is twice the number of edges). Since S is even, it must be that the number of vertices with odd degree is even. □

Perambulation and Connectivity

A *walk* in a graph is a sequence of (not necessarily distinct) vertices $v_1, v_2, \ldots, v_k$ such that $v_i v_{i+1} \in E$ for $i = 1, 2, \ldots, k-1$. Such a walk is sometimes called a v_1–v_k *walk*, and v_1 and v_k are the *end vertices* of the walk. If the vertices in a walk are distinct, then the walk is called a *path*. If the edges in a walk are distinct, then the walk is called a *trail*. In this way, every path is a trail, but not every trail is a path. Got it?

A *closed path*, or *cycle*, is a path $v_1, \ldots, v_k$ (where $k \geq 3$) together with the edge $v_k v_1$. Similarly, a trail that begins and ends at the same vertex is called a *closed trail*, or *circuit*. The *length* of a walk (or path, or trail, or cycle, or circuit) is its number of edges, counting repetitions.

Once again, let's illustrate these definitions with an example. In the graph of Figure 1.7, a, c, f, c, b, d is a walk of length 5. The sequence b, a, c, b, d represents a trail of length 4, and the sequence d, g, b, a, c, f, e represents a path of length 6.

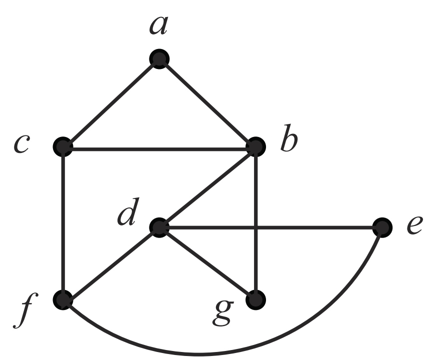


FIGURE 1.7.

Also, g, d, b, c, a, b, g is a circuit, while e, d, b, a, c, f, e is a cycle. In general, it is possible for a walk, trail, or path to have length 0, but the least possible length of a circuit or cycle is 3.

The following theorem is often referred to as the Second Theorem in this book.

Theorem 1.2. *In a graph G with vertices u and v, every u–v walk contains a u–v path.*

Proof. Let W be a u–v walk in G. We prove this theorem by induction on the length of W. If W is of length 1 or 2, then it is easy to see that W must be a path. For the induction hypothesis, suppose the result is true for all walks of length less than k, and suppose W has length k. Say that W is

$$u = w_0, w_1, w_2, \ldots, w_{k-1}, w_k = v$$

where the vertices are not necessarily distinct. If the vertices are in fact distinct, then W itself is the desired u–v path. If not, then let j be the smallest integer such that $w_j = w_r$ for some $r > j$. Let W_1 be the walk

$$u = w_0, \ldots, w_j, w_{r+1}, \ldots, w_k = v.$$

This walk has length strictly less than k, and therefore the induction hypothesis implies that W_1 contains a u–v path. This means that W contains a u–v path, and the proof is complete. □

We now introduce two different operations on graphs: *vertex deletion* and *edge deletion*. Given a graph G and a vertex $v \in V(G)$, we let $G - v$ denote the graph obtained by removing v and all edges incident with v from G. If S is a set of vertices, we let $G - S$ denote the graph obtained by removing each vertex of S and all associated incident edges. If e is an edge of G, then $G - e$ is the graph obtained by removing only the edge e (its end vertices stay). If T is a set of edges, then $G - T$ is the graph obtained by deleting each edge of T from G. Figure 1.8 gives examples of these operations.

A graph is *connected* if every pair of vertices can be joined by a path. Informally, if one can pick up an entire graph by grabbing just one vertex, then the

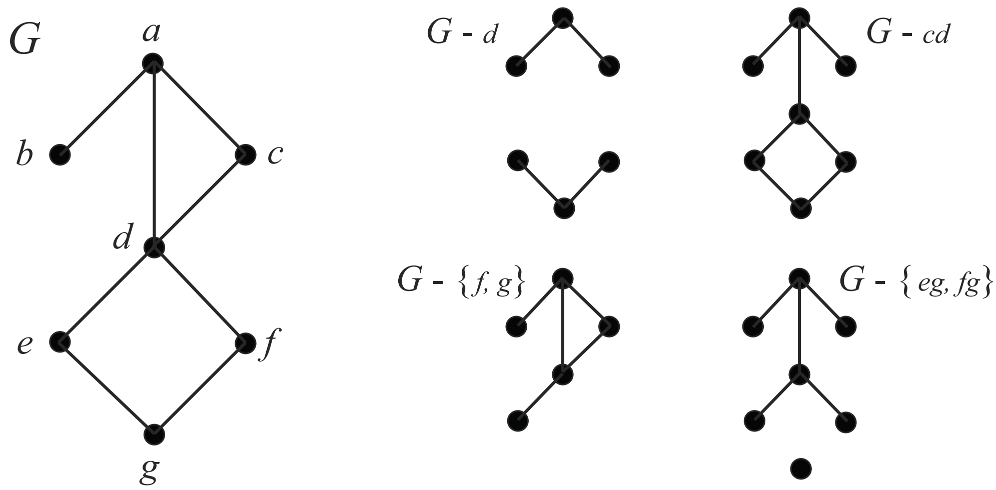


FIGURE 1.8. Deletion operations.

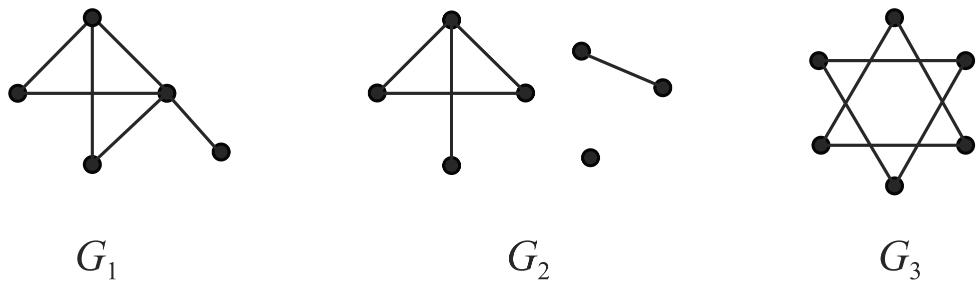


FIGURE 1.9. Connected and disconnected graphs.

graph is connected. In Figure 1.9, G_1 is connected, and both G_2 and G_3 are not connected (or *disconnected*). Each maximal connected piece of a graph is called a *connected component*. In Figure 1.9, G_1 has one component, G_2 has three components, and G_3 has two components.

If the deletion of a vertex v from G causes the number of components to increase, then v is called a *cut vertex*. In the graph G of Figure 1.8, vertex d is a cut vertex and vertex c is not. Similarly, an edge e in G is said to be a *bridge* if the graph $G - e$ has more components than G. In Figure 1.8, the edge ab is the only bridge.

A proper subset S of vertices of a graph G is called a *vertex cut set* (or simply, a *cut set*) if the graph $G - S$ is disconnected. A graph is said to be *complete* if every vertex is adjacent to every other vertex. Consequently, if a graph contains at least one nonadjacent pair of vertices, then that graph is not complete. Complete graphs do not have any cut sets, since $G - S$ is connected for all proper subsets S of the vertex set. Every non-complete graph has a cut set, though, and this leads us to another definition. For a graph G which is not complete, the *connectivity* of G, denoted $\kappa(G)$, is the minimum size of a cut set of G. If G is a connected, non-complete graph of order n, then $1 \leq \kappa(G) \leq n - 2$. If G is disconnected, then $\kappa(G) = 0$. If G is complete of order n, then we say that $\kappa(G) = n - 1$.

Further, for a positive integer k, we say that a graph is *k-connected* if $k \le \kappa(G)$. You will note here that "1-connected" simply means "connected."

Here are several facts that follow from these definitions. You will get to prove a couple of them in the exercises.

i. A graph is connected if and only if $\kappa(G) \ge 1$.

ii. $\kappa(G) \ge 2$ if and only if G is connected and has no cut vertices.

iii. Every 2-connected graph contains at least one cycle.

iv. For every graph G, $\kappa(G) \le \delta(G)$.

Exercises

1. If G is a graph of order n, what is the maximum number of edges in G?

2. Prove that for any graph G of order at least 2, the degree sequence has at least one pair of repeated entries.

3. Consider the graph shown in Figure 1.10.

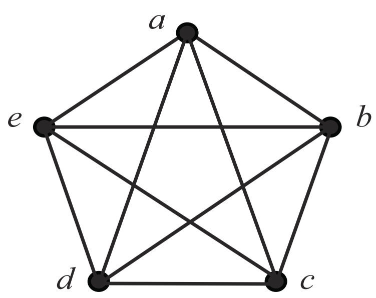


FIGURE 1.10.

   (a) How many different paths have c as an end vertex?
   (b) How many different paths avoid vertex c altogether?
   (c) What is the maximum length of a circuit in this graph? Give an example of such a circuit.
   (d) What is the maximum length of a circuit that does not include vertex c? Give an example of such a circuit.

4. Is it true that a finite graph having exactly two vertices of odd degree must contain a path from one to the other? Give a proof or a counterexample.

5. Let G be a graph where $\delta(G) \ge k$.

   (a) Prove that G has a path of length at least k.
   (b) If $k \ge 2$, prove that G has a cycle of length at least $k + 1$.

6. Prove that every closed odd walk in a graph contains an odd cycle.

7. Draw a connected graph having at most 10 vertices that has at least one cycle of each length from 5 through 9, but has no cycles of any other length.

8. Let P_1 and P_2 be two paths of maximum length in a connected graph G. Prove that P_1 and P_2 have a common vertex.

9. Let G be a graph of order n that is not connected. What is the maximum size of G?

10. Let G be a graph of order n and size strictly less than $n - 1$. Prove that G is not connected.

11. Prove that an edge e is a bridge of G if and only if e lies on no cycle of G.

12. Prove or disprove each of the following statements.

    (a) If G has no bridges, then G has exactly one cycle.
    (b) If G has no cut vertices, then G has no bridges.
    (c) If G has no bridges, then G has no cut vertices.

13. Prove or disprove: If every vertex of a connected graph G lies on at least one cycle, then G is 2-connected.

14. Prove that every 2-connected graph contains at least one cycle.

15. Prove that for every graph G,

    (a) $\kappa(G) \leq \delta(G)$;
    (b) if $\delta(G) \geq n - 2$, then $\kappa(G) = \delta(G)$.

16. Let G be a graph of order n.

    (a) If $\delta(G) \geq \frac{n-1}{2}$, then prove that G is connected.
    (b) If $\delta(G) \geq \frac{n-2}{2}$, then show that G need not be connected.

1.1.3 Special Types of Graphs

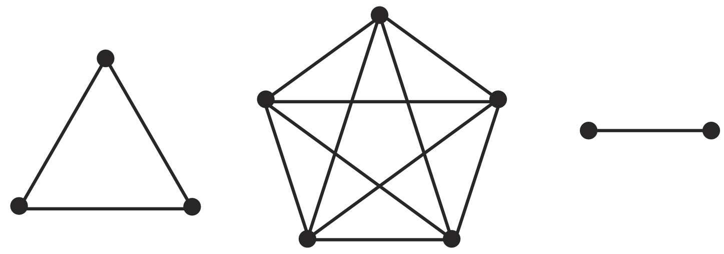
until we meet again . . .

— from *An Irish Blessing*

In this section we describe several types of graphs. We will run into many of them later in the chapter.

1. Complete Graphs

   We introduced complete graphs in the previous section. A complete graph of order n is denoted by K_n, and there are several examples in Figure 1.11.

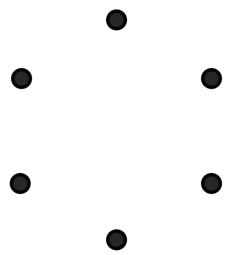


K_3 K_5 K_2

FIGURE 1.11. Examples of complete graphs.

2. Empty Graphs

   The *empty graph* on n vertices, denoted by E_n, is the graph of order n where E is the empty set (Figure 1.12).

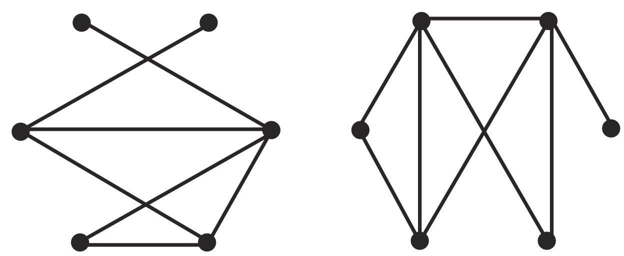


E_6

FIGURE 1.12. An empty graph.

3. Complements

   Given a graph G, the *complement* of G, denoted by $\overline{G}$, is the graph whose vertex set is the same as that of G, and whose edge set consists of all the edges that are *not* present in G (Figure 1.13).



G $\overline{G}$

FIGURE 1.13. A graph and its complement.

4. Regular Graphs

   A graph G is *regular* if every vertex has the same degree. G is said to be *regular of degree* r (or r-*regular*) if $\deg(v) = r$ for all vertices v in G. Complete graphs of order n are regular of degree $n - 1$, and empty graphs are regular of degree 0. Two further examples are shown in Figure 1.14.

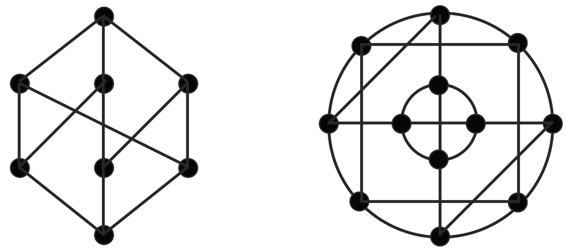


FIGURE 1.14. Examples of regular graphs.

5. Cycles

   The graph C_n is simply a cycle on n vertices (Figure 1.15).

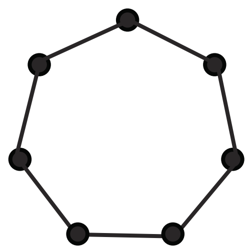


FIGURE 1.15. The graph C_7.

6. Paths

   The graph P_n is simply a path on n vertices (Figure 1.16).

FIGURE 1.16. The graph P_6.

7. Subgraphs

   A graph H is a *subgraph* of a graph G if $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$. In this case we write $H \subseteq G$, and we say that G contains H. In a graph where the vertices and edges are unlabeled, we say that $H \subseteq G$ if the vertices *could* be labeled in such a way that $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$. In Figure 1.17, H_1 and H_2 are both subgraphs of G, but H_3 is not.

8. Induced Subgraphs

   Given a graph G and a subset S of the vertex set, the *subgraph of G induced by S*, denoted $\langle S \rangle$, is the subgraph with vertex set S and with edge set $\{uv \mid u, v \in S \text{ and } uv \in E(G)\}$. So, $\langle S \rangle$ contains all vertices of S and all edges of G whose end vertices are *both* in S. A graph and two of its induced subgraphs are shown in Figure 1.18.

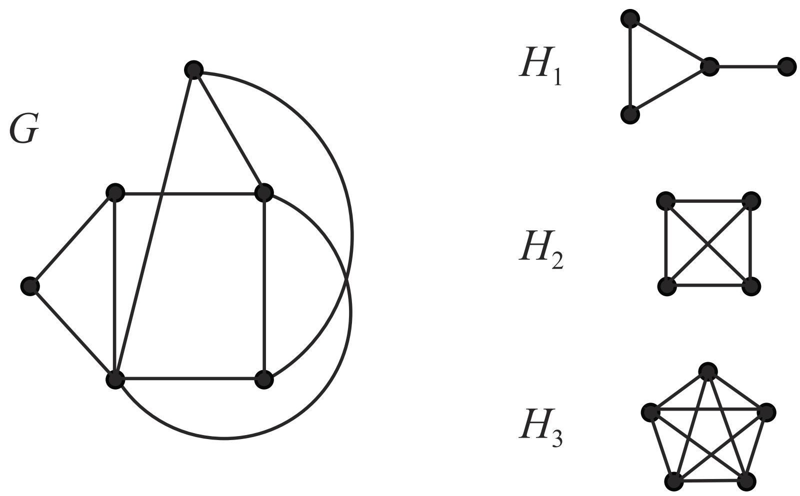


FIGURE 1.17. H_1 and H_2 are subgraphs of G.

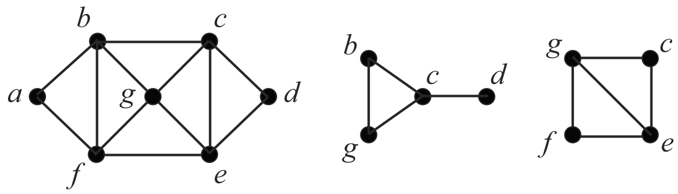


FIGURE 1.18. A graph and two of its induced subgraphs.

9. Bipartite Graphs

   A graph G is *bipartite* if its vertex set can be partitioned into two sets X and Y in such a way that every edge of G has one end vertex in X and the other in Y. In this case, X and Y are called the *partite sets*. The first two graphs in Figure 1.19 are bipartite. Since it is not possible to partition the vertices of the third graph into two such sets, the third graph is not bipartite.

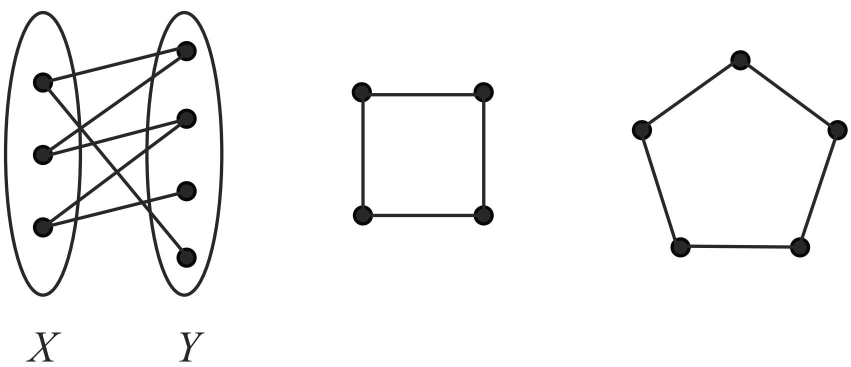


FIGURE 1.19. Two bipartite graphs and one non-bipartite graph.

A bipartite graph with partite sets X and Y is called a *complete bipartite graph* if its edge set is of the form $E = \{xy \mid x \in X, y \in Y\}$ (that is, if every possible connection of a vertex of X with a vertex of Y is present in the graph). Such a graph is denoted by $K_{|X|,|Y|}$. See Figure 1.20.

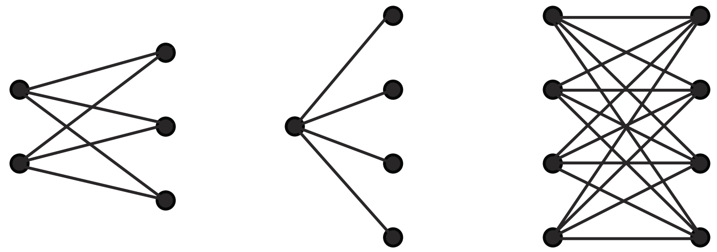


FIGURE 1.20. A few complete bipartite graphs.

The next theorem gives an interesting characterization of bipartite graphs.

Theorem 1.3. *A graph with at least two vertices is bipartite if and only if it contains no odd cycles.*

Proof. Let G be a bipartite graph with partite sets X and Y. Let C be a cycle of G and say that C is $v_1, v_2, \ldots, v_k, v_1$. Assume without loss of generality that $v_1 \in X$. The nature of bipartite graphs implies then that $v_i \in X$ for all odd i, and $v_i \in Y$ for all even i. Since v_k is adjacent to v_1, it must be that k is even; and hence C is an even cycle.

For the reverse direction of the theorem, let G be a graph of order at least two such that G contains no odd cycles. Without loss of generality, we can assume that G is connected, for if not, we could treat each of its connected components separately. Let v be a vertex of G, and define the set X to be

$$X = \{x \in V(G) \mid \text{the shortest path from } x \text{ to } v \text{ has even length}\},$$

and let $Y = V(G) \setminus X$.

Now let x and x' be vertices of X, and suppose that x and x' are adjacent. If $x = v$, then the shortest path from v to x' has length one. But this implies that $x' \in Y$, a contradiction. So, it must be that $x \neq v$, and by a similar argument, $x' \neq v$. Let P_1 be a path from v to x of shortest length (a shortest v–x path) and let P_2 be a shortest v–x' path. Say that P_1 is $v = v_0, v_1, \ldots, v_{2k} = x$ and that P_2 is $v = w_0, w_1, \ldots, w_{2t} = x'$. The paths P_1 and P_2 certainly have v in common. Let v' be a vertex on both paths such that the v'–x path, call it P_1', and the v'–x' path, call it P_2', have only the vertex v' in common. Essentially, v' is the "last" vertex common to P_1 and P_2. It must be that P_1' and P_2' are shortest v'–x and v'–x' paths, respectively, and it must be that $v' = v_i = w_i$ for some i. But since x and x' are adjacent, $v_i, v_{i+1}, \ldots, v_{2k}, w_{2t}, w_{2t-1}, \ldots, w_i$ is a cycle of length $(2k - i) + (2t - i) + 1$, which is odd, and that is a contradiction.

Thus, no two vertices in X are adjacent to each other, and a similar argument shows that no two vertices in Y are adjacent to each other. Therefore, G is bipartite with partite sets X and Y. $\square$

We conclude this section with a discussion of what it means for two graphs to be the same. Look closely at the graphs in Figure 1.21 and convince yourself that one could be re-drawn to look just like the other. Even though these graphs

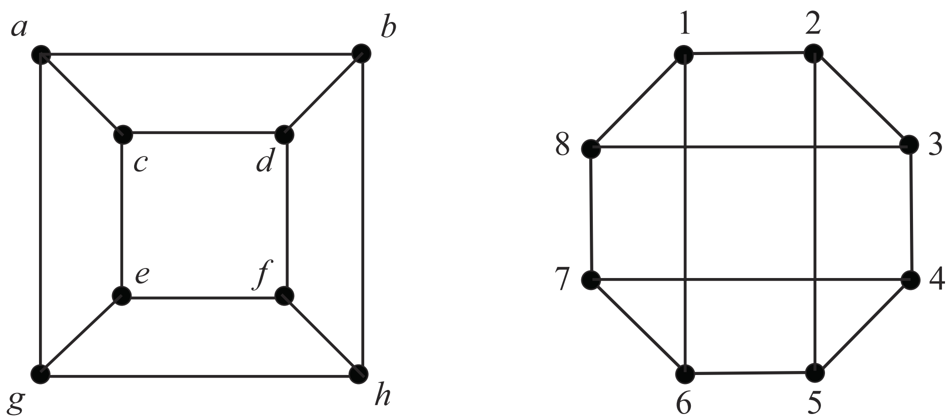


FIGURE 1.21. Are these graphs the same?

have different vertex sets and are drawn differently, it is still quite natural to think of these graphs as being the same. The idea of isomorphism formalizes this phenomenon.

Graphs G and H are said to be *isomorphic* to one another (or simply, isomorphic) if there exists a one-to-one correspondence $f : V(G) \to V(H)$ such that for each pair x,y of vertices of G, $xy \in E(G)$ if and only if $f(x)f(y) \in E(H)$. In other words, G and H are isomorphic if there exists a mapping from one vertex set to another that preserves adjacencies. The mapping itself is called an *isomorphism*. In our example, such an isomorphism could be described as follows:

$$\{(a,1),(b,2),(c,8),(d,3),(e,7),(f,4),(g,6),(h,5)\}.$$

When two graphs G and H are isomorphic, it is not uncommon to simply say that $G = H$ or that "G is H." As you will see, we will make use of this convention quite often in the sections that follow.

Several facts about isomorphic graphs are immediate. First, if G and H are isomorphic, then $|V(G)| = |V(H)|$ and $|E(G)| = |E(H)|$. The converse of this statement is not true, though, and you can see that in the graphs of Figure 1.22. The vertex and edge counts are the same, but the two graphs are clearly not iso-

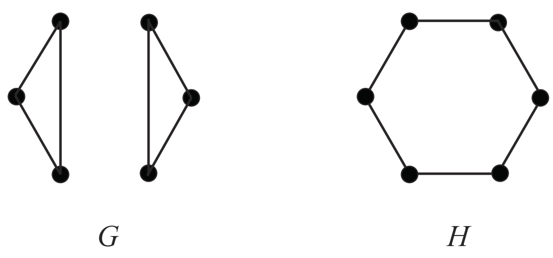


FIGURE 1.22.

morphic.

A second necessary fact is that if G and H are isomorphic then the degree sequences must be identical. Again, the graphs in Figure 1.22 show that the converse of this statement is not true. A third fact, and one that you will prove in Exercise 8, is that if graphs G and H are isomorphic, then their complements $\overline{G}$ and $\overline{H}$ must also be isomorphic.

In general, determining whether two graphs are isomorphic is a difficult problem. While the question is simple for small graphs and for pairs where the vertex counts, edge counts, or degree sequences differ, the general problem is often tricky to solve. A common strategy, and one you might find helpful in Exercises 9 and 10, is to compare subgraphs, complements, or the degrees of adjacent pairs of vertices.

Exercises

1. For $n \geq 1$, prove that K_n has $n(n-1)/2$ edges.

2. If K_{r_1,r_2} is regular, prove that $r_1 = r_2$.

3. Determine whether K_4 is a subgraph of $K_{4,4}$. If yes, then exhibit it. If no, then explain why not.

4. Determine whether P_4 is an induced subgraph of $K_{4,4}$. If yes, then exhibit it. If no, then explain why not.

5. List all of the unlabeled connected subgraphs of C_{34}.

6. The concept of complete bipartite graphs can be generalized to define the *complete multipartite graph* $K_{r_1,r_2,\ldots,r_k}$. This graph consists of k sets of vertices A_1, A_2, $\ldots$, A_k, with $|A_i| = r_i$ for each i, where all possible "interset edges" are present and no "intraset edges" are present. Find expressions for the order and size of $K_{r_1,r_2,\ldots,r_k}$.

7. The *line graph* $L(G)$ of a graph G is defined in the following way: the vertices of $L(G)$ are the edges of G, $V(L(G)) = E(G)$, and two vertices in $L(G)$ are adjacent if and only if the corresponding edges in G share a vertex.

   (a) Let G be the graph shown in Figure 1.23. Find $L(G)$.

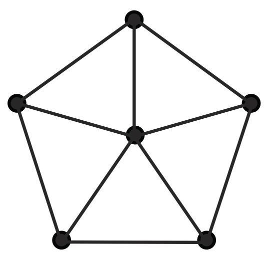


FIGURE 1.23.

(b) Find the complement of $L(K_5)$.

(c) Suppose G has n vertices, labeled $v_1, \ldots v_n$, and the degree of vertex v_i is r_i. Let m denote the size of G, so $r_1 + r_2 + \cdots + r_n = 2m$. Find formulas for the order and size of $L(G)$ in terms of n, m, and the r_i.

8. Prove that if graphs G and H are isomorphic, then their complements $\overline{G}$ and $\overline{H}$ are also isomorphic.

9. Prove that the two graphs in Figure 1.24 are not isomorphic.

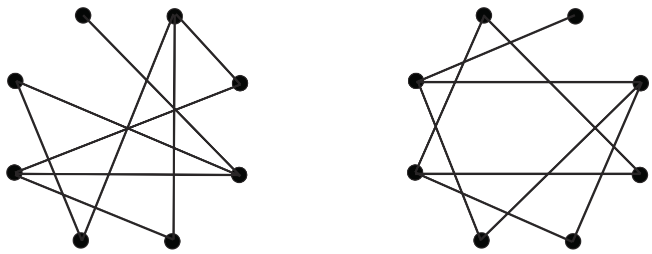


FIGURE 1.24.

10. Two of the graphs in Figure 1.25 are isomorphic.

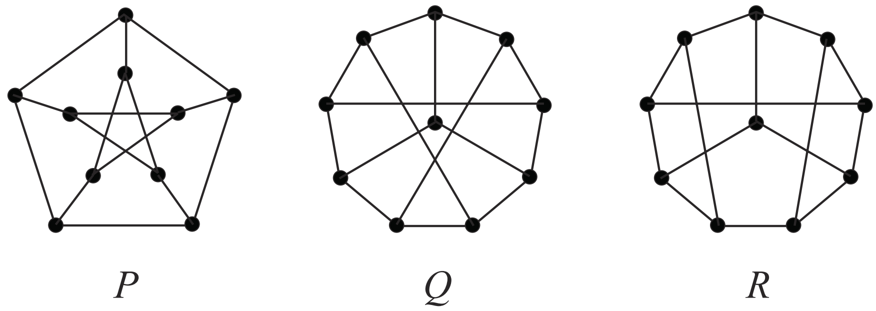


FIGURE 1.25.

(a) For the pair that is isomorphic, give an appropriate one-to-one correspondence.

(b) Prove that the remaining graph is not isomporhic to the other two.

1.2 Distance in Graphs

'Tis distance lends enchantment to the view . . .
— Thomas Campbell, *The Pleasures of Hope*

How far is it from one vertex to another? In this section we define distance in graphs, and we consider several properties, interpretations, and applications. Distance functions, called metrics, are used in many different areas of mathematics, and they have three defining properties. If M is a metric, then

i. $M(x, y) \geq 0$ for all x, y, and $M(x, y) = 0$ if and only if $x = y$;

ii. $M(x, y) = M(y, x)$ for all x, y;

iii. $M(x, y) \leq M(x, z) + M(z, y)$ for all x, y, z.

As you encounter the distance concept in the graph sense, verify for yourself that the function is in fact a metric.

1.2.1 Definitions and a Few Properties

I prefer the term 'eccentric.'

— Brenda Bates, *Urban Legend*

Distance in graphs is defined in a natural way: in a connected graph G, the *distance* from vertex u to vertex v is the length (number of edges) of a shortest u–v path in G. We denote this distance by $d(u, v)$, and in situations where clarity of context is important, we may write $d_G(u, v)$. In Figure 1.26, $d(b, k) = 4$ and $d(c, m) = 6$.

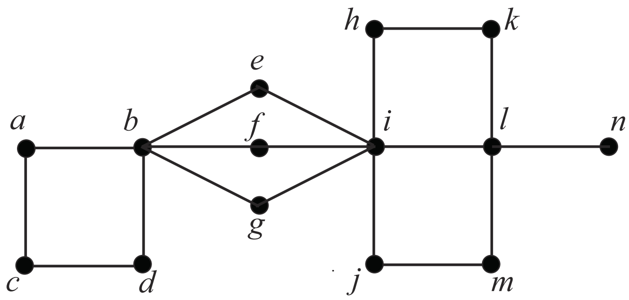


FIGURE 1.26.

For a given vertex v of a connected graph, the *eccentricity* of v, denoted $\mathrm{ecc}(v)$, is defined to be the greatest distance from v to any other vertex. That is,

$$\mathrm{ecc}(v) = \max_{x \in V(G)} \{d(v, x)\}.$$

In Figure 1.26, $\mathrm{ecc}(a) = 5$ since the farthest vertices from a (namely k, m, n) are at a distance of 5 from a.

Of the vertices in this graph, vertices c, k, m and n have the greatest eccentricity (6), and vertices e, f and g have the smallest eccentricity (3). These values and types of vertices are given special names. In a connected graph G, the *radius* of G, denoted $\mathrm{rad}(G)$, is the value of the smallest eccentricity. Similarly, the *diameter* of G, denoted $\mathrm{diam}(G)$, is the value of the greatest eccentricity. The *center* of the graph G is the set of vertices, v, such that $\mathrm{ecc}(v) = \mathrm{rad}(G)$. The *periphery* of G is the set of vertices, u, such that $\mathrm{ecc}(u) = \mathrm{diam}(G)$. In Figure 1.26, the radius is 3, the diameter is 6, and the center and periphery of the graph are, respectively, $\{e, f, g\}$ and $\{c, k, m, n\}$.

Surely these terms sound familiar to you. On a disk, the farthest one can travel from one point to another is the disk's diameter. Points on the rim of a disk are on the periphery. The distance from the center of the disk to any other point on the disk is at most the radius. The terms for graphs have similar meanings.

Do not be misled by this similarity, however. You may have noticed that the diameter of our graph G is twice the radius of G. While this does seem to be a natural relationship, such is not the case for all graphs. Take a quick look at a cycle or a complete graph. For either of these graphs, the radius and diameter are equal!

The following theorem describes the proper relationship between the radii and diameters of graphs. While not as natural, tight, or "circle-like" as you might hope, this relationship does have the advantage of being correct.

Theorem 1.4. *For any connected graph G,* $\operatorname{rad}(G) \le \operatorname{diam}(G) \le 2\operatorname{rad}(G)$.

Proof. By definition, $\operatorname{rad}(G) \le \operatorname{diam}(G)$, so we just need to prove the second inequality. Let u and v be vertices in G such that $d(u,v) = \operatorname{diam}(G)$. Further, let c be a vertex in the center of G. Then,

$$\operatorname{diam}(G) = d(u,v) \le d(u,c) + d(c,v) \le 2\operatorname{ecc}(c) = 2\operatorname{rad}(G). \qquad \Box$$

The definitions in this section can also be extended to graphs that are not connected. In the context of a single connected component of a disconnected graph, these terms have their normal meanings. If two vertices are in different components, however, we say that the distance between them is infinity.

We conclude this section with two interesting results. Choose your favorite graph. It can be large or small, dense with edges or sparse. Choose anything you like, as long as it is your favorite. Now, wouldn't it be neat if there existed a graph in which your favorite graph was the "center" of attention? The next theorem (credited to Hedetneimi in [44]) makes your wish come true.

Theorem 1.5. *Every graph is (isomorphic to) the center of some graph.*

Proof. Let G be a graph (your favorite!). We now construct a new graph H (see Figure 1.27) by adding four vertices (w, x, y, z) to G, along with the following edges:

$$\{wx, yz\} \cup \{xa \mid a \in V(G)\} \cup \{yb \mid b \in V(G)\}.$$

Now, $\operatorname{ecc}(w) = \operatorname{ecc}(z) = 4$, $\operatorname{ecc}(y) = \operatorname{ecc}(x) = 3$, and for any vertex $v \in V(G)$,

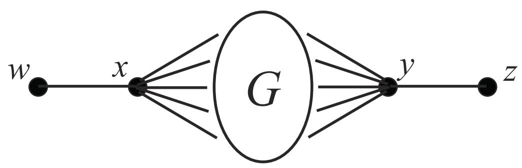


FIGURE 1.27. G is the center.

$\operatorname{ecc}(v) = 2$. Therefore, G is the center of H. $\Box$

Suppose you don't like being the center of attention. Maybe you would rather your favorite graph avoid the spotlight and stay on the periphery. The next theorem (due to Bielak and Sysło, [25]) tells us when that can happen.

Theorem 1.6. *A graph G is (isomorphic to) the periphery of some graph if and only if either every vertex has eccentricity* 1*, or no vertex has eccentricity* 1.

Proof. Suppose that every vertex of G has eccentricity 1. Not only does this mean that G is complete, it also means that every vertex of G is in the periphery. G is the periphery of itself!

On the other hand, suppose that no vertex of G has eccentricity 1. This means that for every vertex u of G, there is some vertex v of G such that $uv \notin E(G)$. Now, let H be a new graph, constructed by adding a single vertex, w, to G, together with the edges $\{wx \mid x \in V(G)\}$. In the graph H, the eccentricity of w is 1 (w is adjacent to everything). Further, for any vertex $x \in V(G)$, the eccentricity of x in H is 2 (no vertex of G is adjacent to everything else in G, and everything in G is adjacent to w). Thus, the periphery of H is precisely the vertices of G.

For the reverse direction, let us suppose that G has some vertices of eccentricity 1 and some vertices of eccentricity greater than 1. Suppose also (in anticipation of a contradiction) that G forms the periphery of some graph, say H. Since the eccentricities of the vertices in G are not all the same, it must be that $V(G)$ is a proper subset of $V(H)$. This means that H is not the periphery of itself and that $\mathrm{diam}(H) \geq 2$. Now, let v be a vertex of G whose eccentricity in G is 1 (v is therefore adjacent to all vertices of G). Since $v \in V(G)$ and since G is the periphery of H, there exists a vertex w in H such that $d(v, w) = \mathrm{diam}(H) \geq 2$. The vertex w, then, is also a peripheral vertex (see Exercise 4) and therefore must be in G. This contradicts the fact that v is adjacent to everything in G. □

Exercises

1. Find the radius, diameter and center of the graph shown in Figure 1.28.

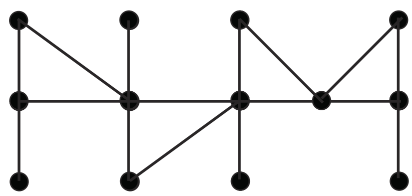


FIGURE 1.28.

2. Find the radius and diameter of each of the following graphs: P_{2k}, P_{2k+1}, C_{2k}, C_{2k+1}, K_n, $K_{m,n}$.

3. For each graph in Exercise 2, find the number of vertices in the center.

4. If x is in the periphery of G and $d(x, y) = \mathrm{ecc}(x)$, then prove that y is in the periphery of G.

5. If u and v are adjacent vertices in a graph, prove that their eccentricities differ by at most one.

6. A graph G is called *self-centered* if $C(G) = V(G)$. Prove that every complete bipartite graph, every cycle, and every complete graph is self-centered.

7. Given a connected graph G and a positive integer k, the *kth power of G*, denoted G^k, is the graph with $V(G^k) = V(G)$ and where vertices u and v are adjacent in G^k if and only if $d_G(u, v) \leq k$.

   (a) Draw the 2nd and 3rd powers of P_8 and C_{10}.

   (b) For a graph G of order n, what is $G^{\mathrm{diam}(G)}$?

8. (a) Find a graph of order 7 that has radius 3 and diameter 6.

   (b) Find a graph of order 7 that has radius 3 and diameter 5.

   (c) Find a graph of order 7 that has radius 3 and diameter 4.

   (d) Suppose r and d are positive integers and $r \leq d \leq 2r$. Describe a graph that has radius r and diameter d.

9. Suppose that u and v are vertices in a graph G, $\mathrm{ecc}(u) = m$, $\mathrm{ecc}(v) = n$, and $m < n$. Prove that $d(u, v) \geq n - m$. Then draw a graph G_1 where $d(u, v) = n - m$, and another graph G_2 where $d(u, v) > n - m$. In each case, label the vertices u and v, and give the values of m and n.

10. Let G be a connected graph with at least one cycle. Prove that G has at least one cycle whose length is less than or equal to $2\,\mathrm{diam}(G) + 1$.

11. (a) Prove that if G is connected and $\mathrm{diam}(G) \geq 3$, then $\overline{G}$ is connected.

    (b) Prove that if $\mathrm{diam}(G) \geq 3$, then $\mathrm{diam}(\overline{G}) \leq 3$.

    (c) Prove that if G is regular and $\mathrm{diam}(G) = 3$, then $\mathrm{diam}(\overline{G}) = 2$.

1.2.2 Graphs and Matrices

Unfortunately no one can be told what the Matrix is. You have to see it for yourself.

— Morpheus, *The Matrix*

What do matrices have to do with graphs? This is a natural question — nothing we have seen so far has suggested any possible relationship between these two types of mathematical objects. That is about to change!

As we have seen, a graph is a very visual object. To this point, we have determined distances by looking at the diagram, pointing with our fingers, and counting edges. This sort of analysis works fairly well for small graphs, but it quickly breaks down as the graphs of interest get larger. Analysis of large graphs often requires computer assistance.

Computers cannot just look and point at graphs like we can. Instead, they understand graphs via matrix representations. One such representation is an adjacency matrix. Let G be a graph with vertices $v_1, v_2, \ldots, v_n$. The *adjacency matrix* of G is the $n \times n$ matrix A whose (i, j) entry, denoted by $[A]_{i,j}$, is defined by

$$[A]_{i,j} = \begin{cases} 1 & \text{if } v_i \text{ and } v_j \text{ are adjacent,} \\ 0 & \text{otherwise.} \end{cases}$$

The graph in Figure 1.29 has six vertices. Its adjacency matrix, A, is

$$A = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}.$$

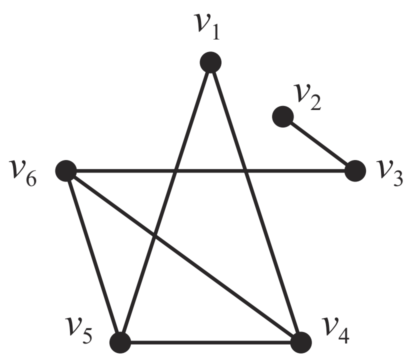


FIGURE 1.29.

Note that for simple graphs (where there are no loops) adjacency matrices have all zeros on the main diagonal. You can also see from the definition that these matrices are symmetric.[1]

A single graph can have multiple adjacency matrices — different orderings of the vertices will produce different matrices. If you think that these matrices ought to be related in some way, then you are correct! In fact, if A and B are two different adjacency matrices of the same graph G, then there must exist a permutation of the vertices such that when the permuation is applied to the corresponding rows and columns of A, you get B.

This fact can be used in reverse to determine if two graphs are isomorphic, and the permutation mentioned here serves as an appropriate bijection: Given two graphs G_1 and G_2 with respective adjacency matrices A_1 and A_2, if one can apply

[1] Can you think of a context in which adjacency matrices might not be symmetric? Direct your attention to Figure 1.3 for a hint.

a permutation to the rows and columns of A_1 and produce A_2, then G_1 and G_2 are isomorphic.

Let's take a closer look at the previous example. The fact that the $(1,6)$ entry is 0 indicates that v_1 and v_6 are not adjacent. Consider now the $(1,6)$ entry of the matrix A^2. This entry is just the dot product of row one of A with column six of A:

$$\begin{aligned}[A^2]_{1,6} &= (0,0,0,1,1,0)\cdot(0,0,1,1,1,0)\\ &= (0\cdot 0)+(0\cdot 0)+(0\cdot 1)+(1\cdot 1)+(1\cdot 1)+(0\cdot 0)\\ &= 2.\end{aligned}$$

Think about what makes this dot product nonzero. It is the fact that there was at least one place (and here there were two places) where a 1 in row one corresponded with a 1 in column six. In our case, the 1 in the fourth position of row one (representing the edge v_1v_4) matched up with the 1 in the fourth position of column six (representing the edge v_4v_6). The same thing occurred in the fifth position of the row and column (where the edges represented were v_1v_5 and v_5v_6).

Can you see what is happening here? The entry in position $(1,6)$ of A^2 is equal to the number of two-edge walks from v_1 to v_6 in G. As the next theorem shows us, this is not a coincidence.

Theorem 1.7. *Let G be a graph with vertices labeled v_1, v_2, $\ldots$, v_n, and let A be its corresponding adjacency matrix. For any positive integer k, the (i,j) entry of A^k is equal to the number of walks from v_i to v_j that use exactly k edges.*

Proof. We prove this by induction on k. For $k=1$, the result is true since $[A]_{i,j}=1$ exactly when there is a one-edge walk between v_i and v_j.

Now suppose that for every i and j, the (i,j) entry of A^{k-1} is the number of walks from v_i to v_j that use exactly $k-1$ edges. For each k-edge walk from v_i to v_j, there exists an h such that the walk can be thought of as a $(k-1)$-edge walk from v_i to v_h, combined with an edge from v_h to v_j. The total number of these k-edge walks, then, is

$$\sum_{v_h\in N(v_j)} \text{(number of } (k-1)\text{-edge walks from } v_i \text{ to } v_h).$$

By the induction hypothesis, we can rewrite this sum as

$$\sum_{v_h\in N(v_j)} [A^{k-1}]_{i,h} = \sum_{h=1}^{n}[A^{k-1}]_{i,h}[A]_{h,j} = [A^k]_{i,j}\ ,$$

and this proves the result. □

This theorem has a straightforward corollary regarding distance between vertices.

Corollary 1.8. *Let G be a graph with vertices labeled v_1, v_2, $\ldots$, v_n, and let A be its corresponding adjacency matrix. If $d(v_i,v_j)=x$, then $[A^k]_{i,j}=0$ for $1\le k<x$.*

Let's see if we can relate these matrices back to earlier distance concepts. Given a graph G of order n with adjacency matrix A, and given a positive integer k, define the matrix sum S_k to be

$$S_k = I + A + A^2 + \cdots + A^k,$$

where I is the $n \times n$ identity matrix. Since the entries of I and A are ones and zeros, the entries of S_k (for any k) are nonnegative integers. This implies that for every pair i, j, we have $[S_k]_{i,j} \leq [S_{k+1}]_{i,j}$.

Theorem 1.9. *Let G be a connected graph with vertices labeled v_1, v_2, …, v_n, and let A be its corresponding adjacency matrix.*

1. *If k is the smallest positive integer such that row j of S_k contains no zeros, then $\operatorname{ecc}(v_j) = k$.*

2. *If r is the smallest positive integer such that all entries of at least one row of S_r are positive, then $\operatorname{rad}(G) = r$.*

3. *If m is the smallest positive integer such that all entries of S_m are positive, then $\operatorname{diam}(G) = m$.*

Proof. We will prove the first part of the theorem. The proofs of the other parts are left for you as exercises.[2]

Suppose that k is the smallest positive integer such that row j of S_k contains no zeros. The fact that there are no zeros on row j of S_k implies that the distance from v_j to any other vertex is at most k. If $k = 1$, the result follows immediately. For $k > 1$, the fact that there is at least one zero on row j of S_{k-1} indicates that there is at least one vertex whose distance from v_j is greater than $k - 1$. This implies that $\operatorname{ecc}(v_j) = k$. □

We can use adjacency matrices to create other types of graph-related matrices. The steps given below describe the construction of a new matrix, using the matrix sums S_k defined earlier. Carefully read through the process, and (before you read the paragraph that follows!) see if you can recognize the matrix that is produced.

Creating a New Matrix, M

Given: A connected graph of order n, with adjacency matrix A, and with S_k as defined earlier.

1. For each $i \in \{1, 2, \ldots, n\}$, let $[M]_{i,i} = 0$.

2. For each pair i, j where $i \neq j$, let $[M]_{i,j} = k$ where k is the least positive integer such that $[S_k]_{i,j} \neq 0$.

[2] You're welcome.

Can you see what the entries of M will be? For each pair i, j, the (i, j) entry of M is the distance from v_i to v_j. That is,

$$[M]_{i,j} = d(v_i, v_j).$$

The matrix M is called the *distance matrix* of the graph G.

Exercises

1. Give the adjacency matrix for each of the following graphs.

   (a) P_{2k} and P_{2k+1}, where the vertices are labeled from one end of the path to the other.

   (b) C_{2k} and C_{2k+1}, where the vertices are labeled consecutively around the cycle.

   (c) $K_{m,n}$, where the vertices in the first partite set are labeled $v_1, \dots, v_m$.

   (d) K_n, where the vertices are labeled any way you please.

2. Without computing the matrix directly, find A^3 where A is the adjacency matrix of K_4.

3. If A is the adjacency matrix for the graph G, show that the (j, j) entry of A^2 is the degree of v_j.

4. Let A be the adjacency matrix for the graph G.

   (a) Show that the number of triangles that contain v_j is $\frac{1}{2}[A^3]_{j,j}$.

   (b) The *trace* of a square matrix M, denoted $\mathrm{Tr}(M)$, is the sum of the entries on the main diagonal. Prove that the number of triangles in G is $\frac{1}{6}\,\mathrm{Tr}(A^3)$.

5. Find the $(1, 5)$ entry of A^{2009} where A is the adjacency matrix of C_{10} and where the vertices of C_{10} are labeled consecutively around the cycle.

6. (a) Prove the second statement in Theorem 1.9.

   (b) Prove the third statement in Theorem 1.9.

7. Use Theorem 1.9 to design an algorithm for determining the center of a graph G.

8. The graph G has adjacency matrix A and distance matrix D. Prove that if $A = D$, then G is complete.

9. Give the distance matrices for the graphs in Exercise 1. You should create these matrices directly — it is not necessary to use the method described in the section.

1.2.3 Graph Models and Distance

Do I know you?

— Kevin Bacon, in *Flatliners*

We have already seen that graphs can serve as models for all sorts of situations. In this section we will discuss several models in which the idea of distance is significant.

The Acquaintance Graph

"Wow, what a small world!" This familiar expression often follows the discovery of a shared acquaintance between two people. Such discoveries are enjoyable, for sure, but perhaps the frequency with which they occur ought to keep us from being as surprised as we typically are when we experience them.

We can get a better feel for this phenomenon by using a graph as a model. Define the *Acquaintance Graph*, A, to be the graph where each vertex represents a person, and an edge connects two vertices if the corresponding people know each other. The context here is flexible — one could create this graph for the people living in a certain neighborhood, or the people working in a certain office building, or the people populating a country or the planet. Since the smaller graphs are all subgraphs of the graphs for larger populations, most people think of A in the largest sense: The vertices represent the Earth's human population.[3]

An interesting question is whether or not the graph A, in the large (Earth) sense, is connected. Might there be a person or a group of people with no connection (direct *or* indirect) at all to another group of people?[4] While there is a possibility of this being the case, it is most certainly true that if A is in fact disconnected, there is one *very* large connected component.

The graph A can be illuminating with regard to the "six degrees of separation" phenomenon. Made popular (at least in part) by a 1967 experiment by social psychologist Stanley Milgram [204] and a 1990 play by John Guare [142], the "six degrees theory" asserts that given any pair of people, there is a chain of no more than six acquaintance connections joining them. Translating into graph theorese, the assertion is that $\mathrm{diam}(A) \leq 6$. It is, of course, difficult (if not impossible) to confirm this. For one, A is enormous, and the kind of computation required for confirmation is nontrivial (to say the least!) for matrices with six billion rows. Further, the matrix A is not static — vertices and edges appear all of the time.[5] Still, the graph model gives us a good way to visualize this intriguing phenomenon.

Milgram's experiment [204] was an interesting one. He randomly selected several hundred people from certain communities in the United States and sent a

[3]The graph could be made even larger by allowing the vertices to represent all people, living *or* dead. We will stick with the living people only — six billion vertices is large enough, don't you think?

[4]Wouldn't it be interesting to meet such a person? Wait — it wouldn't be interesting for long because as soon as you meet him, he is no longer disconnected!

[5]Vertices will disappear if you limit A to living people. Edges disappear when amnesia strikes.

packet to each. Inside each packet was the name and address of a single "target" person. If the recipient knew this target personally, the recipient was to mail the packet directly to him. If the recipient did not know the target personally, the recipient was to send the packet to the person he/she thought had the best chance of knowing the target personally (perhaps someone in the same state as the target, or something like that). The new recipient was to follow the same rules: Either send it directly to the target (if known personally) or send it to someone who has a good chance of knowing the target. Milgram tracked how many steps it took for the packets to reach the target. Of the packets that eventually returned, the median number of steps was 5! Wow, what a small world!

The Hollywood Graph

Is the actor Kevin Bacon the center of Hollywood? This question, first asked by a group of college students in 1993, was the beginning of what was soon to become a national craze: The Kevin Bacon Game. The object of the game is to connect actors to Bacon through appearances in movies. For example, the actress Emma Thompson can be linked to Bacon in two steps: Thompson costarred with Gary Oldman in *Harry Potter and the Prisoner of Azkaban* (among others), and Oldman appeared with Bacon in *JFK*. Since Thompson has not appeared with Bacon in a movie, two steps is the best we can do. We say that Thompson has a *Bacon number* of 2.

Can you sense the underlying graph here?[6] Let us define the *Hollywood Graph*, H, as follows: The vertices of H represent actors, and an edge exists between two vertices when the corresponding actors have appeared in a movie together. So, in H, Oldman is adjacent to both Bacon and Thompson, but Bacon and Thompson are not adjacent. Thompson has a Bacon number of 2 because the distance from her vertex to Bacon's is 2. In general, an actor's Bacon number is defined to be the distance from that actor's vertex to Bacon's vertex in H. If an actor cannot be linked to Bacon at all, then that actor's Bacon number is infinity. As was the case with the Acquaintance Graph, if H is disconnected we can focus our attention on the single connected component that makes up most of H (Bacon's component).

The ease with which Kevin Bacon can be connected to other actors might lead one to conjecture that Bacon is the unique center of Hollywood. In terms of graph theory, that conjecture would be that the center of H consists only of Bacon's vertex. Is this true? Is Bacon's vertex even *in* the center at all? Like the Acquaintance Graph, the nature of H changes frequently, and answers to questions like these are elusive. The best we can do is to look at a snapshot of the graph and answer the questions based on that particular point in time.

Let's take a look at the graph as it appeared on December 25, 2007. On that day, the Internet Movie Database [165] had records for nearly 1.3 million actors. Patrick Reynolds maintains a website [234] that tracks Bacon numbers, among other things. According to Reynolds, of the 1.3 million actors in the database on

[6] or, "Can you smell the Bacon?"

that day, 917,007 could be linked to Bacon in some way via chains of shared movie appearances. The maximum distance from Bacon to any of the actors in his component was 8 (and so Bacon's eccentricity is 8). What about eccentricities of other actors? Are there any that are less than 8? According to Reynolds, the answer is no — 8 is the smallest eccentricity, and so Kevin Bacon is in the center of H. But it is very crowded there — thousands and thousands of other actors have eccentricity 8 as well.

The Mathematical Collaboration Graph

The Hungarian Paul Erdős (1913–1996) was one of the greatest and most prolific mathematicians of the twentieth century. Erdős authored or coauthored over 1500 mathematical papers covering topics in graph theory, combinatorics, set theory, geometry, number theory, and more. He collaborated with hundreds of other mathematicians, and this collaboration forms the basis of a Bacon-like ranking system. While not as widely popular as Bacon numbers, almost all mathematicians are familiar with the concept of Erdős numbers.

Erdős himself is assigned Erdős number 0. Any mathematician who coauthored a paper with Erdős has Erdős number 1. If a person has coauthored a paper with someone who has an Erdős number of 1 (and if that person himself/herself doesn't have Erdős number 1), then that person has an Erdős number of 2. Higher Erdős numbers are assigned in a similar manner.

The underlying graph here should be clear. Define the *Mathematical Collaboration Graph*, C, to have vertices corresponding to researchers, and let an edge join two researchers if the two have coauthored a paper together. A researcher's Erdős number, then, is the distance from the corresponding vertex to the vertex of Erdős. If a researcher is not in the same connected component of C as Erdős, then that researcher has an infinite Erdős number.

As you might imagine, new vertices and edges are frequently added to C. Jerry Grossman maintains a website [140] that keeps track of Erdős numbers. At one point in 2007, there were over 500 researchers with Erdős number 1 and over 8100 with Erdős number 2. You might surmise that because Erdős died in 1996, the number of people with Erdős number 1 has stopped increasing. While this is surely to be true sometime in the near future, it hasn't happened yet. A number of papers coauthored by Erdős have been published since his death. Erdős has not been communicating with collaborators from the great beyond (at least as far as we know) — it is simply the case that his collaborators continue to publish joint research that began years ago.

Small World Networks

As we saw earlier, the Acquaintance Graph provides a way to model the famous "small world phemomenon" — the sense that humans are connected via numerous recognized and unrecognized connections. The immense size and dynamic nature of that graph make it difficult to analyze carefully and completely, and so smaller models can prove to be more useful. In order for the more manageable graphs to be helpful, though, it is important that they enjoy some fundamental small world properties.

So what makes a small world small? What properties should a graph have if it is to be a model of a small world? Let's list a few. As you read through the list below, think about your own acquaintance network and see if these properties make sense to you.

1. There should be plenty of mutual acquaintances (shared neighbors). If this were the only property, then complete graphs would surely fit the bill — lots of mutual neighbors there. A complete graph, though, is not a realistic model of acquaintances in the world.

2. The graph should be sparse in edges. In a realistic model, there should be relatively few edges compared to the number of vertices in the graph.

3. Distances between pairs of vertices should be relatively small. The *characteristic path length* of a graph G, denoted L_G, is the average distance between vertices, where the average is taken over all pairs of distinct vertices. In any graph of order n, there are $|E(K_n)|$ distinct pairs of vertices, and in Exercise 1 of Section 1.1.3, you showed that $|E(K_n)| = n(n-1)/2$. So for a graph G of order n,

$$L_G = \frac{\sum_{u,v \in V(G)} d(u,v)}{|E(K_n)|} = \frac{2}{n(n-1)} \sum_{u,v \in V(G)} d(u,v).$$

   One way of obtaining this value for a graph is to find the mean of the non-diagonal entries in the distance matrix of the graph.

4. There should be a reasonable amount of *clustering* in a small world graph. In actual acquaintance networks, there are a number of factors (geography, for instance) that create little clusters of vertices — small groups of vertices among which a larger than typical portion of edges exists. For example, there are likely to be many edges among the vertices that represent the people that live in your neighborhood.

   Given a vertex v in a graph of order n, we define its *clustering coefficient*, denoted $\mathrm{cc}(v)$, as follows (recall that $\langle N[v] \rangle$ is the subgraph *induced* by the *closed* neighborhood of v).

$$\mathrm{cc}(v) = \frac{|E(\langle N[v] \rangle)|}{\left|E\left(K_{1+\deg(v)}\right)\right|} = \frac{2\,|E(\langle N[v] \rangle)|}{(1+\deg(v))\deg(v)}.$$

   For each vertex v, this is the percentage of edges that exist among the vertices in the closed neighborhood of v. For a graph G of order n, we define the *clustering coefficient of the graph* G, denoted by $\mathrm{CC}(G)$ to be the average of the clustering coefficients of the vertices of G. That is,

$$\mathrm{CC}(G) = \frac{1}{n} \sum_{v \in V(G)} \mathrm{cc}(v).$$

Small world networks have the property that characteristic path lengths are low and clustering coefficients are high. Graphs that have these properties can be used as models in the mathematical analyses of the small world phenomenon and its associated concepts. It is interesting to note that other well known networks have exhibited small world traits — the internet, electric power grids, and even neural networks are examples — and this increases even further the applicability of graph models.

Exercises

1. Compute the characteristic path length for each of each of the following graphs: P_{2k}, P_{2k+1}, C_{2k}, C_{2k+1}, K_n, $K_{m,n}$.

2. Compute the clustering coefficient for each of each of the following graphs: P_{2k}, P_{2k+1}, C_{2k}, C_{2k+1}, K_n, $K_{m,n}$.

3. (a) In the Acquaintance Graph, try to find a path from your vertex to the vertex of the President of the United States.

   (b) Your path from the previous question may not be your shortest such path. Prove that your actual distance from the President is at most one away from the shortest such distance to be found among your classmates.

Interesting Note: There are several contexts in which Bacon numbers can be calculuated. While Bacon purists only use movie connections, others include shared appearances on television and in documentaries as well. Under these more open guidelines, the mathematician Paul Erdős actually has a Bacon number of 3! Erdős was the focus of the 1993 documentary *N is a Number* [63]. British actor Alec Guinness made a (very) brief appearance near the beginning of that film, and Guinness has a Bacon number of 2 (can you find the connections?). As far as we know, Bacon has not coauthored a research article with anyone who is connected to Erdős, and so while Erdős' Bacon number is 3, Bacon's Erdős number is infinity.

1.3 Trees

"O look at the trees!" they cried, "O look at the trees!"

— Robert Bridges, *London Snow*

In this section we will look at the trees—but not the ones that sway in the wind or catch the falling snow. We will talk about graph-theoretic trees. Before moving on, glance ahead at Figure 1.30, and try to pick out which graphs are trees.