11 Simple Graphs

Simple graphs model relationships that are *symmetric*, meaning that the relationship is mutual. Examples of such mutual relationships are being married, speaking the same language, not speaking the same language, occurring during overlapping time intervals, or being connected by a conducting wire. They come up in all sorts of applications, including scheduling, constraint satisfaction, computer graphics, and communications, but we'll start with an application designed to get your attention: we are going to make a professional inquiry into sexual behavior. Specifically, we'll look at some data about who, on average, has more opposite-gender partners: men or women.

Sexual demographics have been the subject of many studies. In one of the largest, researchers from the University of Chicago interviewed a random sample of 2500 people over several years to try to get an answer to this question. Their study, published in 1994 and entitled *The Social Organization of Sexuality*, found that men have on average 74% more opposite-gender partners than women.

Other studies have found that the disparity is even larger. In particular, ABC News claimed that the average man has 20 partners over his lifetime, and the average woman has 6, for a percentage disparity of 233%. The ABC News study, aired on Primetime Live in 2004, purported to be one of the most scientific ever done, with only a 2.5% margin of error. It was called "American Sex Survey: A peek between the sheets"—raising some questions about the seriousness of their reporting.

Yet again in August, 2007, the New York Times reported on a study by the National Center for Health Statistics of the U.S. government showing that men had seven partners while women had four. So, whose numbers do you think are more accurate: the University of Chicago, ABC News, or the National Center?

Don't answer—this is a trick question designed to trip you up. Using a little graph theory, we'll explain why none of these findings can be anywhere near the truth.

11.1 Vertex Adjacency and Degrees

Simple graphs are defined as digraphs in which edges are *undirected*—they connect two vertices without pointing in either direction between the vertices. So instead of a directed edge $\langle v \to w \rangle$ which starts at vertex v and ends at vertex w, a simple graph only has an undirected edge, $\langle v\text{—}w \rangle$, that connects v and w.

Definition 11.1.1. A *simple graph*, G, consists of a nonempty set, $V(G)$, called the *vertices* of G, and a set $E(G)$ called the *edges* of G. An element of $V(G)$ is called a *vertex*. A vertex is also called a *node*; the words "vertex" and "node" are used interchangeably. An element of $E(G)$ is an *undirected edge* or simply an "edge." An undirected edge has two vertices $u \neq v$ called its *endpoints*. Such an edge can be represented by the two element set $\{u, v\}$. The notation $\langle u\text{—}v \rangle$ denotes this edge.

Both $\langle u\text{—}v \rangle$ and $\langle v\text{—}u \rangle$ define the same undirected edge, whose endpoints are u and v.

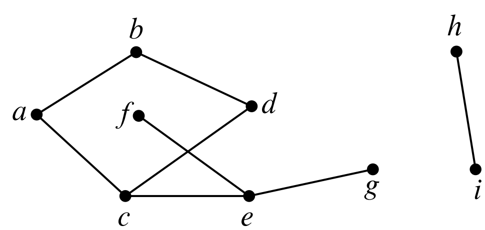

Figure 11.1 An example of a graph with 9 nodes and 8 edges.

For example, let H be the graph pictured in Figure 11.1. The vertices of H correspond to the nine dots in Figure 11.1, that is,

$$V(H) = \{a, b, c, d, e, f, g, h, i\}.$$

The edges correspond to the eight lines, that is,

$$E(H) = \{\, \langle a\text{—}b \rangle, \langle a\text{—}c \rangle, \langle b\text{—}d \rangle, \langle c\text{—}d \rangle, \langle c\text{—}e \rangle, \langle e\text{—}f \rangle, \langle e\text{—}g \rangle, \langle h\text{—}i \rangle \,\}.$$

Mathematically, that's all there is to the graph H.

Definition 11.1.2. Two vertices in a simple graph are said to be *adjacent* iff they are the endpoints of the same edge, and an edge is said to be *incident* to each of its endpoints. The number of edges incident to a vertex v is called the *degree* of the vertex and is denoted by $\deg(v)$. Equivalently, the degree of a vertex is the number of vertices adjacent to it.

For example, for the graph H of Figure 11.1, vertex a is adjacent to vertex b, and b is adjacent to d. The edge $\langle a\text{—}c \rangle$ is incident to its endpoints a and c. Vertex h has degree 1, d has degree 2, and $\deg(e) = 3$. It is possible for a vertex to have degree 0, in which case it is not adjacent to any other vertices. A simple graph, G, does not need to have any edges at all. $|E(G)|$ could be zero, implying that the degree of every vertex would also be zero. But a simple graph must have at least one vertex—$|V(G)|$ is required to be at least one.

An edge whose endpoints are the same is called a *self-loop*. Self-loops aren't allowed in simple graphs.[1] In a more general class of graphs called *multigraphs*, there can be more than one edge with the same two endpoints, but this doesn't happen in simple graphs, because every edge is uniquely determined by its two endpoints. Sometimes graphs with no vertices, with self-loops, or with more than one edge between the same two vertices are convenient to have, but we don't need them, and sticking with simple graphs is simpler.

For the rest of this chapter we'll use "graphs" as an abbreviation for "simple graphs."

A synonym for "vertices" is "*nodes*," and we'll use these words interchangeably. Simple graphs are sometimes called *networks*, edges are sometimes called *arcs*. We mention this as a "heads up" in case you look at other graph theory literature; we won't use these words.

11.2 Sexual Demographics in America

Let's model the question of heterosexual partners in graph theoretic terms. To do this, we'll let G be the graph whose vertices, V, are all the people in America. Then we split V into two separate subsets: M, which contains all the males, and F, which contains all the females.[2] We'll put an edge between a male and a female iff they have been sexual partners. This graph is pictured in Figure 11.2 with males on the left and females on the right.

Actually, this is a pretty hard graph to figure out, let alone draw. The graph is *enormous*: the US population is about 300 million, so $|V| \approx 300M$. Of these, approximately 50.8% are female and 49.2% are male, so $|M| \approx 147.6M$, and $|F| \approx 152.4M$. And we don't even have trustworthy estimates of how many edges there are, let alone exactly which couples are adjacent. But it turns out that we don't need to know any of this—we just need to figure out the relationship between the average number of partners per male and partners per female. To do this, we note that every edge has exactly one endpoint at an M vertex (remember, we're only considering male-female relationships); so the sum of the degrees of the M vertices equals the number of edges. For the same reason, the sum of the degrees

[1] You might try to represent a self-loop going between a vertex v and itself as $\{v, v\}$, but this equals $\{v\}$. It wouldn't be an edge, which is defined to be a set of *two* vertices.

[2] For simplicity, we'll ignore the possibility of someone being *both* a man and a woman, or neither.

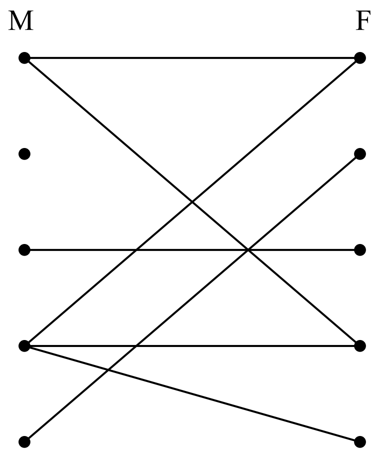

Figure 11.2 The sex partners graph.

of the F vertices equals the number of edges. So these sums are equal:

$$\sum_{x \in M} \deg(x) = \sum_{y \in F} \deg(y).$$

Now suppose we divide both sides of this equation by the product of the sizes of the two sets, $|M| \cdot |F|$:

$$\left(\frac{\sum_{x \in M} \deg(x)}{|M|}\right) \cdot \frac{1}{|F|} = \left(\frac{\sum_{y \in F} \deg(y)}{|F|}\right) \cdot \frac{1}{|M|}$$

The terms above in parentheses are the *average degree of an M vertex* and the *average degree of an F* vertex. So we know:

$$\text{Avg. deg in } M = \frac{|F|}{|M|} \cdot \text{Avg. deg in } F \tag{11.1}$$

In other words, we've proved that the average number of female partners of males in the population compared to the average number of males per female is *determined solely by the relative number of males and females in the population*.

Now the Census Bureau reports that there are slightly more females than males in America; in particular $|F|/|M|$ is about 1.035. So we know that males have on average 3.5% more opposite-gender partners than females, and that this tells us nothing about any sex's promiscuity or selectivity. Rather, it just has to do with the relative number of males and females. Collectively, males and females have the same number of opposite gender partners, since it takes one of each set for every partnership, but there are fewer males, so they have a higher ratio. This means that the University of Chicago, ABC, and the Federal government studies are way off. After a huge effort, they gave a totally wrong answer.

There's no definite explanation for why such surveys are consistently wrong. One hypothesis is that males exaggerate their number of partners—or maybe females downplay theirs—but these explanations are speculative. Interestingly, the principal author of the National Center for Health Statistics study reported that she knew the results had to be wrong, but that was the data collected, and her job was to report it.

The same underlying issue has led to serious misinterpretations of other survey data. For example, a couple of years ago, the Boston Globe ran a story on a survey of the study habits of students on Boston area campuses. Their survey showed that on average, minority students tended to study with non-minority students more than the other way around. They went on at great length to explain why this "remarkable phenomenon" might be true. But it's not remarkable at all. Using our graph theory formulation, we can see that all it says is that there are fewer minority students than non-minority students, which is, of course, what "minority" means.

11.2.1 Handshaking Lemma

The previous argument hinged on the connection between a sum of degrees and the number of edges. There is a simple connection between these in any graph:

Lemma 11.2.1. *The sum of the degrees of the vertices in a graph equals twice the number of edges.*

Proof. Every edge contributes two to the sum of the degrees, one for each of its endpoints. ■

We refer to Lemma 11.2.1 as the *Handshaking Lemma*: if we total up the number of people each person at a party shakes hands with, the total will be twice the number of handshakes that occurred.

11.3 Some Common Graphs

Some graphs come up so frequently that they have names. A *complete graph* K_n has n vertices and an edge between every two vertices, for a total of $n(n-1)/2$ edges. For example, K_5 is shown in Figure 11.3.

The *empty graph* has no edges at all. For example, the empty graph with 5 nodes is shown in Figure 11.4.

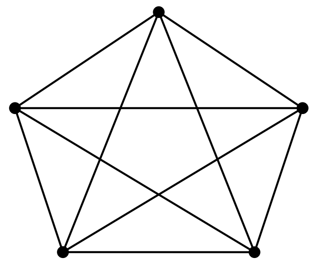

Figure 11.3 K_5: the complete graph on 5 nodes.

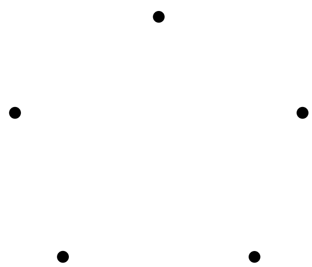

Figure 11.4 An empty graph with 5 nodes.

An n-node graph containing $n-1$ edges in sequence is known as a *line graph* L_n. More formally, L_n has

$$V(L_n) = \{v_1, v_2, \ldots, v_n\}$$

and

$$E(L_n) = \{\, \langle v_1 — v_2 \rangle, \langle v_2 — v_3 \rangle, \ldots, \langle v_{n-1} — v_n \rangle \,\}$$

For example, L_5 is pictured in Figure 11.5.

There is also a one-way infinite line graph L_∞ which can be defined by letting the nonnegative integers $\mathbb{N}$ be the vertices with edges $\langle k—(k+1) \rangle$ for all $k \in \mathbb{N}$.

If we add the edge $\langle v_n — v_1 \rangle$ to the line graph L_n, we get a graph called a *length-n cycle* C_n. Figure 11.6 shows a picture of length-5 cycle.

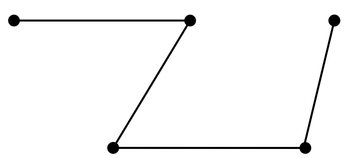

Figure 11.5 L_5: a 5-node line graph.

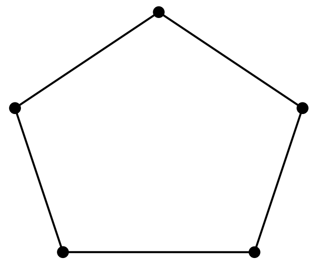

Figure 11.6 C_5: a 5-node cycle graph.

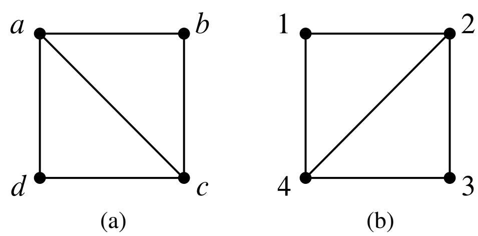

Figure 11.7 Two Isomorphic graphs.

11.4 Isomorphism

Two graphs that look different might actually be the same in a formal sense. For example, the two graphs in Figure 11.7 are both 4-vertex, 5-edge graphs and you get graph (b) by a 90° clockwise rotation of graph (a).

Strictly speaking, these graphs are different mathematical objects, but this difference doesn't reflect the fact that the two graphs can be described by the same picture—except for the labels on the vertices. This idea of having the same picture "up to relabeling" can be captured neatly by adapting Definition 9.7.1 of isomorphism of digraphs to handle simple graphs. An isomorphism between two graphs is an edge-preserving bijection between their sets of vertices:

Definition 11.4.1. An isomorphism between graphs G and H is a bijection $f : V(G) \to V(H)$ such that

$$\langle u—v \rangle \in E(G) \quad \text{iff} \quad \langle f(u)—f(v) \rangle \in E(H)$$

for all $u, v \in V(G)$. Two graphs are isomorphic when there is an isomorphism between them.

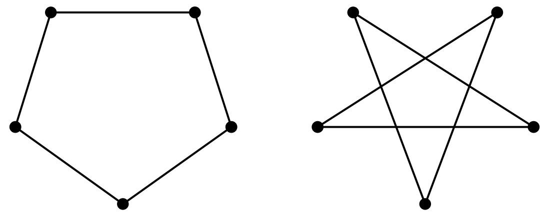

Figure 11.8 Isomorphic C_5 graphs.

Here is an isomorphism, f, between the two graphs in Figure 11.7:

$$f(a) ::= 2 \qquad f(b) ::= 3$$
$$f(c) ::= 4 \qquad f(d) ::= 1.$$

You can check that there is an edge between two vertices in the graph on the left if and only if there is an edge between the two corresponding vertices in the graph on the right.

Two isomorphic graphs may be drawn very differently. For example, Figure 11.8 shows two different ways of drawing C_5.

Notice that if f is an isomorphism between G and H, then f^{-1} is an isomorphism between H and G. Isomorphism is also transitive because the composition of isomorphisms is an isomorphism. In fact, isomorphism is an equivalence relation.

Isomorphism preserves the connection properties of a graph, abstracting out what the vertices are called, what they are made out of, or where they appear in a drawing of the graph. More precisely, a property of a graph is said to be *preserved under isomorphism* if whenever G has that property, every graph isomorphic to G also has that property. For example, since an isomorphism is a bijection between sets of vertices, isomorphic graphs must have the same number of vertices. What's more, if f is a graph isomorphism that maps a vertex, v, of one graph to the vertex, $f(v)$, of an isomorphic graph, then by definition of isomorphism, every vertex adjacent to v in the first graph will be mapped by f to a vertex adjacent to $f(v)$ in the isomorphic graph. Thus, v and $f(v)$ will have the same degree. If one graph has a vertex of degree 4 and another does not, then they can't be isomorphic. In fact, they can't be isomorphic if the number of degree 4 vertices in each of the graphs is not the same.

Looking for preserved properties can make it easy to determine that two graphs are not isomorphic, or to guide the search for an isomorphism when there is one. It's generally easy in practice to decide whether two graphs are isomorphic. However, no one has yet found a procedure for determining whether two graphs are isomorphic that is *guaranteed* to run in polynomial time on all pairs of graphs.[3]

Having such a procedure would be useful. For example, it would make it easy to search for a particular molecule in a database given the molecular bonds. On the other hand, knowing there is no such efficient procedure would also be valuable: secure protocols for encryption and remote authentication can be built on the hypothesis that graph isomorphism is computationally exhausting.

The definitions of bijection and isomorphism apply to infinite graphs as well as finite graphs, as do most of the results in the rest of this chapter. But graph theory focuses mostly on finite graphs, and we will too. *In the rest of this chapter we'll assume graphs are finite.*

We've actually been taking isomorphism for granted ever since we wrote "K_n has n vertices..." at the beginning of Section 11.3.

Graph theory is all about properties preserved by isomorphism.

11.5 Bipartite Graphs & Matchings

There were two kinds of vertices in the "Sex in America" graph, males and females, and edges only went between the two kinds. Graphs like this come up so frequently that they have earned a special name: *bipartite graphs*.

Definition 11.5.1. A *bipartite graph* is a graph whose vertices can be partitioned[4] into two sets, $L(G)$ and $R(G)$, such that every edge has one endpoint in $L(G)$ and the other endpoint in $R(G)$.

So every bipartite graph looks something like the graph in Figure 11.2.

11.5.1 The Bipartite Matching Problem

The bipartite matching problem is related to the sex-in-America problem that we just studied; only now, the goal is to get everyone happily married. As you might imagine, this is not possible for a variety of reasons, not the least of which is the fact that there are more women in America than men. So, it is simply not possible to marry every woman to a man so that every man is married at most once.

But what about getting a mate for every man so that every woman is married at most once? Is it possible to do this so that each man is paired with a woman that

[3] A procedure runs in *polynomial time* when it needs an amount of time of at most $p(n)$, where n is the total number of vertices and $p()$ is a fixed polynomial.

[4] Partitioning a set means cutting it up into *nonempty* pieces. In this case, it means that $L(G)$ and $R(G)$ are nonempty, $L(G) \cup R(G) = V(G)$, and $L(G) \cap R(G) = \emptyset$.