CHAPTER
ONE

An Introduction to Combinatorics

As we hope you will sense right from the beginning, we believe that combinatorial mathematics is one of the most fascinating and captivating subjects on the planet. Combinatorics is *very* concrete and has a wide range of applications, but it also has an intellectually appealing theoretical side. Our goal is to give you a taste of both. In order to begin, we want to develop, through a series of examples, a feeling for what types of problems combinatorics addresses.

1.1 Introduction

There are three principal themes to our course:

Discrete Structures Graphs, digraphs, networks, designs, posets, strings, patterns, distributions, coverings, and partitions.

Enumeration Permutations, combinations, inclusion/exclusion, generating functions, recurrence relations, and Pólya counting.

Algorithms and Optimization Sorting, spanning trees, shortest paths, eulerian circuits, hamiltonian cycles, graph coloring, planarity testing, network flows, bipartite matchings, and chain partitions.

To illustrate the accessible, concrete nature of combinatorics and to motivate topics that we will study, this preliminary chapter provides a first look at combinatorial problems, choosing examples from enumeration, graph theory, number theory, and optimization. The discussion is very informal—but this should serve to explain why we have to be more precise at later stages. We ask lots of questions, but at this stage, you'll only be able to answer a few. Later, you'll be able to answer many more . . . but as promised earlier, most likely you'll never be able to answer them all. And if we're wrong in making that statement, then you're certain to become *very* famous. Also, you'll get an A++ in the course and maybe even a Ph.D. too.

Our discussion will also introduce you to Alice and Bob, who are almost always on opposite sides of any issue. So let's begin.

1.2 Enumeration

The roots of combinatorics lie in counting discrete (as opposed to continuous, as you studied in calculus) objects. The classical problems of this type are partition problems and counting in light of symmetries. We present examples of each of these types of problems below.

Alice has three children named Dawn, Keesha and Seth.

1. Alice has ten one dollar bills and decides to give the full amount to her children. How many ways can she do this? For example, one way she might distribute the funds is to give Dawn and Keesha four dollars each with Seth receiving the balance—two dollars. Another way is to give the entire amount to Keesha, an option that probably won't make Dawn and Seth very happy. Note that hidden within this question is the assumption that Alice need only decide the *amount* each of the three children is to receive.

2. The amounts of money distributed to the three children form a sequence which if written in non-increasing order has the form: a_1, a_2, a_3 with $a_1 \geq a_2 \geq a_3$ and $a_1 + a_2 + a_3 = 10$. How many such sequences are there?

3. Suppose Alice decides to give each child at least one dollar. How does this change the answers to the first two questions?

4. Now suppose that Alice has ten books, in fact the top 10 books from the New York Times best-seller list, and decides to give them to her children. How many ways can she do this? Again, we note that there is a hidden assumption—the ten books are all different.

5. Suppose the ten books are labeled $B_1, B_2, \ldots, B_{10}$. The sets of books given to the three children are pairwise disjoint and their union is $\{B_1, B_2, \ldots, B_{10}\}$. How many different sets of the form $\{S_1, S_2, S_3\}$ where S_1, S_2 and S_3 are pairwise disjoint and $S_1 \cup S_2 \cup S_3 = \{B_1, B_2, \ldots, B_{10}\}$?

6. Suppose Alice decides to give each child at least one book. How does this change the answers to the preceding two questions?

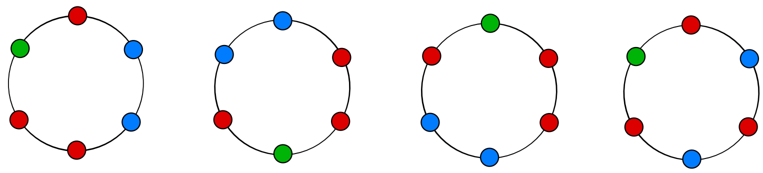

FIGURE 1.1: NECKLACES MADE WITH THREE COLORS

7. How would we possibly answer these kinds of questions if ten was really ten thousand (OK, we're not talking about children any more!) and three was three thousand?

A circular necklace with a total of six beads will be assembled using beads of three different colors. In Figure 1.1, we show four such necklaces—however, note that the first three are actually the *same* necklace. Each has three red beads, two blues and one green. On the other hand, the fourth necklace has the same number of beads of each color but it is a *different* necklace.

1. How many different necklaces of six beads can be formed using three reds, two blues and one green?
2. How many different necklaces of six beads can be formed using red, blue and green beads (not all colors have to be used)?
3. How many different necklaces of six beads can be formed using red, blue and green beads if all three colors have to be used?
4. How would we possibly answer these questions for necklaces of six thousand beads made with beads from three thousand different colors?

1.3 Combinatorics and Graph Theory

A *graph* G consists of a *vertex* set V and a collection E of 2-element subsets of V. Elements of E are called edges. In our course, we will (almost always) use the convention that $V = \{1, 2, 3, \ldots, n\}$ for some positive integer n. With this convention, graphs can be described *precisely* with a text file:

1. The first line of the file contains a single integer n, the number of vertices in the graph.

```
graph1.txt
9
6 2
1 5
1 7
6 8
9 1
4 3
5 7
1 3
5 9
7 9
```

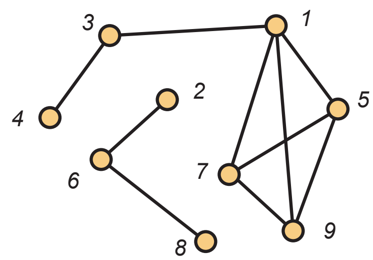

FIGURE 1.2: A GRAPH DEFINED BY DATA

2. Each of the remaining lines of the file contains a pair of distinct integers and specifies an edge of the graph.

We illustrate this convention in Figure 1.2 with a text file and the diagram for the graph G it defines.

Much of the notation and terminology for graphs is quite natural. See if you can make sense out of the following statements which apply to the graph G defined above:

1. G has 9 vertices and 10 edges.
2. $\{2, 6\}$ is an edge.
3. Vertices 5 and 9 are adjacent.
4. $\{5, 4\}$ is not an edge.
5. Vertices 3 and 7 are not adjacent.
6. $P = (4, 3, 1, 7, 9, 5)$ is a path of length 5 from vertex 4 to vertex 5.
7. $C = (5, 9, 7, 1)$ is cycle of length 4.
8. G is disconnected and has two components. One of the components has vertex set $\{2, 6, 8\}$.
9. $\{1, 5, 7\}$ is a triangle.
10. $\{1, 7, 5, 9\}$ is a clique of size 4.
11. $\{4, 2, 8, 5\}$ is an independent set of size 4.

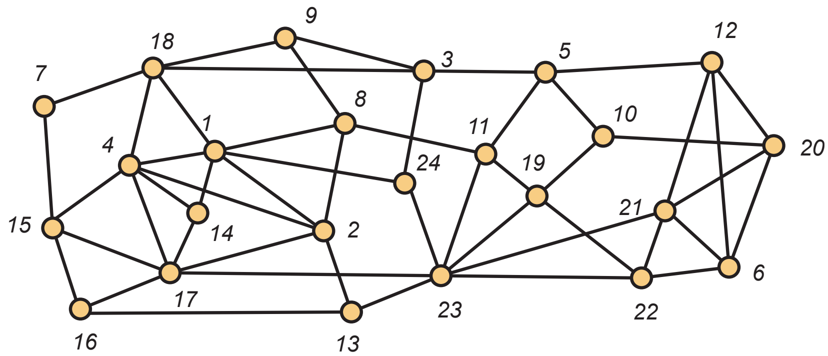

FIGURE 1.3: A CONNECTED GRAPH

Equipped only with this little bit of background material, we are already able to pose a number of interesting and challenging problems.

Example 1.1. Consider the graph G shown in Figure 1.3.

1. What is the largest k for which G has a path of length k?
2. What is the largest k for which G has a cycle of length k?
3. What is the largest k for which G has a clique of size k?
4. What is the largest k for which G has an independent set of size k?
5. What is the shortest path from vertex 7 to vertex 6?

Suppose we gave the class a text data file for a graph on 1500 vertices and asked whether the graph contains a cycle of length at least 500. Alice says yes and Bob says no. How do we decide who is right?

We will frequently study problems in which graphs arise in a very natural manner. Here's an example.

Example 1.2. In Figure 1.4, we show the location of some radio stations in the plane, together with a scale indicating a distance of 200 miles. Radio stations that are closer than 200 miles apart must broadcast on different frequencies to avoid interference.

We've shown that 6 different frequencies are enough. Can you do better?

Can you find 4 stations each of which is within 200 miles of the other 3? Can you find 8 stations each of is more than 200 miles away from the other 7? Is there a natural way to define a graph associated with this problem?

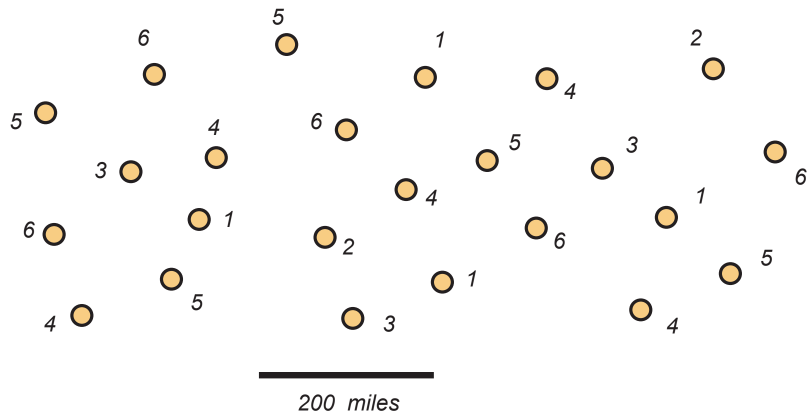

FIGURE 1.4: RADIO STATIONS

1.4 Combinatorics and Number Theory

Broadly, number theory concerns itself with the properties of the positive integers. G.H. Hardy was a brilliant British mathematician who lived through both World Wars and conducted a large deal of number-theoretic research. He was also a pacifist who was happy that, from his perspective, his research was not "useful". He wrote in his 1940 essay *A Mathematician's Apology* "[n]o one has yet discovered any warlike purpose to be served by the theory of numbers or relativity, and it seems very unlikely that anyone will do so for many years."[1] Little did he know, the purest mathematical ideas of number theory would soon become indespensible for the cryptographic techniques that kept communications secure. Our subject here is not number theory, but we will see a few times where combinatorial techniques are of use in number theory.

Example 1.3. Form a sequence of positive integers using the following rules. Start with a positive integer $n > 1$. If n is odd, then the next number is $3n + 1$. If n is even, then the next number is $n/2$. Halt if you ever reach 1. For example, if we start with 28, the sequence is

$$28, 14, 7, 22, 11, 34, 17, 52, 26, 13, 40, 20, 10, 5, 16, 8, 4, 2, 1.$$

Now suppose you start with 19. Then the first few terms are

$$19, 58, 29, 88, 44, 22.$$

But now we note that the integer 22 appears in the first sequence, so the two sequences will agree from this point on.

[1]G.H. Hardy, *A Mathematician's Apology*, Cambridge University Press, p. 140. (1993 printing)

Pick a number somewhere between 100 and 200 and write down the sequence you get. Regardless of your choice, you will eventually halt with a 1. However, is there some positive integer n (possibly quite large) so that if you start from n, you will never reach 1?

Example 1.4. Students in middle school are taught to add fractions by finding least common multiples. For example, the least common multiple of 15 and 12 is 60, so:

$$\frac{2}{15} + \frac{7}{12} = \frac{8}{60} + \frac{35}{60} = \frac{43}{60}.$$

How hard is it to find the least common multiple of two integers? It's really easy if you can factor them into primes. For example, consider the problem of finding the least common multiple of 351785000 and 316752027900 if you just happen to know that

$$\begin{aligned} 351785000 =& 2^3 \times 5^4 \times 7 \times 19 \times 23^2 \quad \text{and} \\ 316752027900 =& 2^2 \times 3 \times 5^2 \times 7^3 \times 11 \times 23^4. \end{aligned}$$

Then the least common multiple is

$$300914426505000 = 2^3 \times 3 \times 5^4 \times 7^3 \times 11 \times 19 \times 23^4.$$

So to find the least common multiple of two numbers, we just have to factor them into primes. That doesn't sound too hard. For starters, can you factor 1961? OK, how about 1348433? Now for a real challenge. Suppose you are told that the integer

$$c = 5220070641387698449504000148751379227274095462521$$

is the product of two primes a and b. Can you find them?

What if factoring is hard? Can you find the least common multiple of two relatively large integers, say each with about 500 digits, by another method? How should middle school students be taught to add fractions?

As an aside, we note that most calculators can't add or multiply two 20 digits numbers, much less two numbers with more than 500 digits. But it is relatively straightforward to write a computer program that will do the job for us. Also, there are some powerful mathematical software tools available. Two very well known examples are *Maple*® and *Mathematica*®. For example, if you open up a *Maple* workspace and enter the command:

```
ifactor(300914426505000);
```

then about as fast as you hit the carriage return, you will get the prime factorization shown above.

Now here's how we made up the challenge problem. First, we found a site on the web that lists large primes and found these two values:

$$a = 45095080578985454453 \quad \text{and}$$
$$b = 115756986668303657898962467957.$$

We then used *Maple* to multiply them together using the following command:

$$45095080578985454453 * 115756986668303657898962467957;$$

Almost instantly, *Maple* reported the value for c given above.

Out of curiosity, we then asked *Maple* to factor c. It took almost 12 minutes on a reasonably good computer (a dual Opteron).

Questions arising in number theory can also have an enumerative flair, as the following example shows.

Example 1.5. In Table 1.1, we show the integer partitions of 8. There are 22 partitions

8 distinct parts	7+1 distinct parts, odd parts	6+2 distinct parts
6+1+1	5+3 distinct parts, odd parts	5+2+1 distinct parts
5+1+1+1 odd parts	4+4	4+3+1 distinct parts
4+2+2	4+2+1+1	4+1+1+1+1
3+3+2	3+3+1+1 odd parts	3+2+2+1
3+2+1+1+1	3+1+1+1+1+1 odd parts	2+2+2+2
2+2+2+1+1	2+2+1+1+1+1	2+1+1+1+1+1+1
	1+1+1+1+1+1+1+1 odd parts	

TABLE 1.1: THE PARTITIONS OF 8, NOTING THOSE INTO DISTINCT PARTS AND THOSE INTO ODD PARTS.

altogether, and as noted, exactly 6 of them are partitions of 8 into odd parts. Also, exactly 6 of them are partitions of 8 into distinct parts.

What would be your reaction if we asked you to find the number of integer partitions of 25892? Do you think that the number of partitions of 25892 into odd parts equals the number of partitions of 25892 into distinct parts? Is there a way to answer this question *without* actually calculating the number of partitions of each type?

1.5 Combinatorics and Geometry

There are many problems in geometry that are innately combinatorial or for which combinatorial techniques shed light on the problem.

Example 1.6. In Figure 1.5, we show a family of 4 lines in the plane. Each pair of lines intersects and no point in the plane belongs to more than two lines. These lines determine 11 regions.

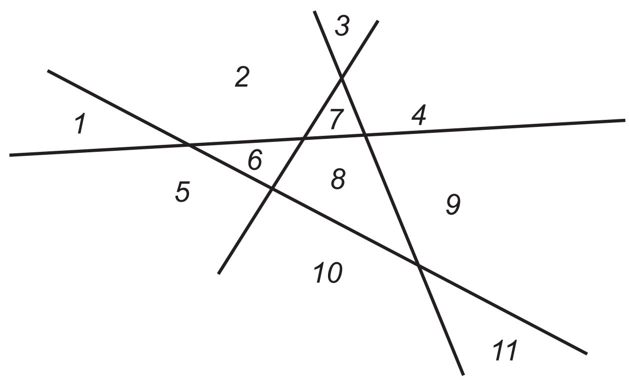

FIGURE 1.5: LINES AND REGIONS

Under these same restrictions, how many regions would a family of 8947 lines determine? Can different arrangements of lines determine different numbers of regions?

Example 1.7. Bob says he has found a set of 882 points in the plane that determine exactly 752 lines. His nemesis Alice again reports that Bob is out to lunch. Why?

Example 1.8. There are many different ways to draw a graph in the plane. Some may have crossing edges while others may not. But sometimes, crossing edges will appear in any diagram. Consider the graph G shown in Figure 1.6. Can you redraw G without crossing edges?

Suppose Alice and Bob were given a homework problem asking whether a particular graph on 2843952 vertices and 9748032 edges could be drawn without edge crossings. Alice just looked at the number of vertices and the number of edges and said that the answer is "no." This time, it is Bob who is the skeptic and demands proof. How can Alice defend her negative answer?

1.6 Combinatorics and Optimization

You likely have already been introduce to optimization problems, as calculus students around the world are familiar with the plight of farmers trying to fence the largest area of land given a certain amount of fence or people needing to cross rivers downstream

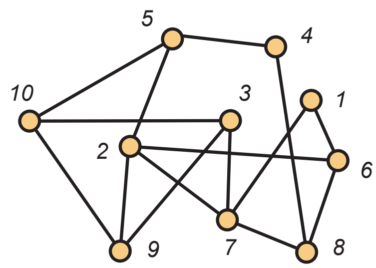

FIGURE 1.6: A GRAPH WITH CROSSING EDGES

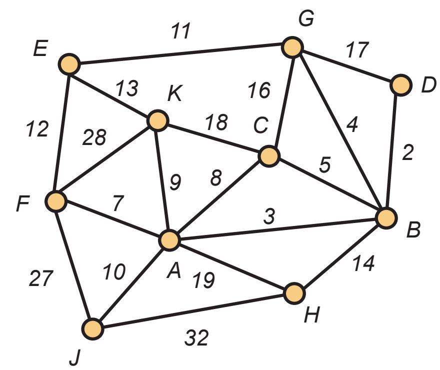

FIGURE 1.7: A LABELED GRAPH WITH WEIGHTED EDGES

from their current location who must decide where they should cross based on the speed at which they can run and swim. However, these problems are inherently continuous. In theory, you can cross the river at any point you want, even if it were irrational. (OK, so not exactly irrational, but a good decimal approximation.) In this course, we will examine a few optimization problems that are not continuous, as only integer values for the variables will make sense. It turns out that many of these problems are very hard to solve in general.

Example 1.9. In Figure 1.7, we use letters for the labels on the vertices to help distinguish visually from the integer weights on the edges.

Suppose the vertices are cities, the edges are highways and the weights on the edges represent *distance*.

Q_1: What is the shortest path from vertex E to vertex B?

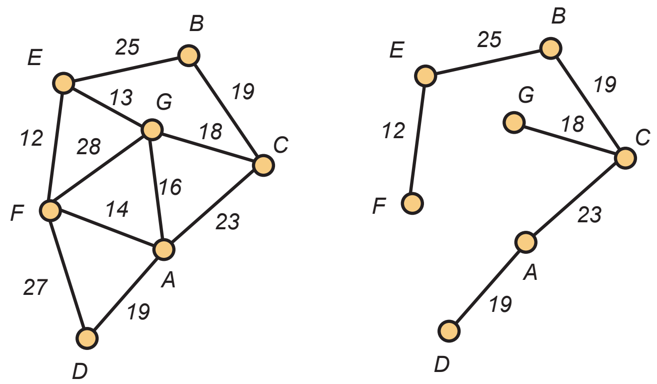

FIGURE 1.8: A WEIGHTED GRAPH AND SPANNING TREE

Suppose Alice is a salesperson whose home base is city A.

Q_2: In what order should Alice visit the other cities so that she goes through each of them at least once and returns home at the end—while keeping the total distance traveled to a minimum? Can Alice accomplish such a tour visiting each city *exactly* once?

Bob is a highway inspection engineer and must traverse every highway each month. Bob's homebase is City E.

Q_3: In what order should Bob traverse the highways to minimize the total distance traveled? Can Bob make such a tour traveling along each highway exactly once?

Example 1.10. Now suppose that the vertices are locations of branch banks in Atlanta and that the weights on an edge represents the cost, in millions of dollars, of building a high capacity data link between the branch banks at it two end points. In this model, if there is no edge between two branch banks, it means that the cost of building a data link between this particular pair is prohibitively high (here we are tempted to say the cost is infinite, but the authors don't admit to knowing the meaning of this word).

Our challenge is to decide which data links should be constructed to form a network in which any branch bank can communicate with any other branch. We assume that data can flow in either direction on a link, should it be built, and that data can be relayed through any number of data links. So to allow full communication, we should contruct a *spanning tree* in this network. In Figure 1.8, we show a graph G on the left and one of its many *spanning trees* on the right.

The weight of the spanning tree is the sum of the weights on the edges. In our model, this represents the costs, again in millions of dollars, of building the data links associated with the edges in the spanning tree. For the spanning tree shown in Figure 1.8, this total is

$$12 + 25 + 19 + 18 + 23 + 19 = 116.$$

Of all spanning trees, the bank would naturally like to find one having minimum weight.

How many spanning trees does this graph have? For a large graph, say one with 2875 vertices, does it make sense to find all spanning trees and simply take the one with minimum cost? In particular, for a positive integer n, how many trees have vertex set $\{1, 2, 3, \ldots, n\}$?

1.7 Sudoku Puzzles

Here's an example which has more substance than you might think at first glance. It involves Sudoku puzzles, which have become immensely popular in recent years.

Example 1.11. A Sudoku puzzle is a 9×9 array of cells that when completed have the integers $1, 2, \ldots, 9$ appearing exactly once in each row and each column. Also (and this is what makes the puzzles so fascinating), the numbers 1, 2, $3, \ldots, 9$ appear once in each of the nine 3×3 subquares identified by the darkened borders. To be considered a legitimate Sudoku puzzle, there should be a *unique* solution. In Figure 1.9, we show two Sudoku puzzles. The one on the left is fairly easy, and the one on the right is far more challenging.

		7				8	2	
	9				1			
	4		9	7				
					5	4		6
		3				7		
5		6	7					
				8	4		5	
			6				1	
	2	4				6		

	8	1	3		2	6		
6		9	5		1		2	
2	3							
5		2		3		7	8	9
4	6	3		8		2		1
							6	2
	2		7		9	5		3
		6	8		3	9	4	

FIGURE 1.9: SUDOKU PUZZLES

There are many sources of Sudoku puzzles, and software that generates Sudoku puzzles and then allows you to play them with an attractive GUI is available with any distribution of Linux (not at all advisable to play them during class!). Also, you can find Sudkoku puzzles on the web at:

http://www.websudoku.com

On this site, the “Evil” ones are just that.

How does Alice make up good Sudoku puzzles, ones that are difficult for Bob to solve? How could Bob use a computer to solve puzzles that Alice has constructed? What makes some Sudoku puzzles easy and some of them hard?

1.8 Closing Comments

Hopefully, these examples have piqued your interest in combinatorics, and you are ready to study matters in greater depth. Let’s start!